Revista Brasileira de Computação Aplicada, July, 2019

DOI: 10.5335/rbca.v11i2.8906
Vol. 11, Nº 2, pp. 74–85
Homepage: seer.upf.br/index.php/rbca/index

ISSN 2176-6649

ORIGINAL ARTICLE

Application of artificial random numbers and Monte Carlo method in the reliability analysis of geodetic networks

Maria L. S. Bonimani[1], Vinicius F. Rofatto[1,2], Marcelo T. Matsuoka[1,2] and Ivandro Klein[3,4]

[1]Universidade Federal de Uberlândia, Instituto de Geografia, Engenharia de Agrimensura e Cartográfica, Monte Carmelo - MG - Brasil and [2]Universidade Federal do Rio Grande do Sul, Programa de Pós-Graduação em Sensoriamento Remoto, Porto Alegre - RS - Brasil and [3]Universidade Federal do Paraná, Programa de Pós-Graduação em Ciências Geodésicas, Paraná - PR - Brasil and [4]Instituto Federal de Santa Catarina, Curso de Agrimensura, Florianópolis - SC - Brasil

*malubonimani@hotmail.com; vfrofatto@gmail.com; tomiomatsuoka@gmail.com; ivandroklein@gmail.com

Received: 2018-11-28. **Revised:** 2019-06-26. **Accepted:** 2019-07-05.

Abstract

A Geodetic Network is a network of point interconnected by direction and/or distance measurements or by using Global Navigation Satellite System receivers. Such networks are essential for the most geodetic engineering projects, such as monitoring the position and deformation of man-made structures (bridges, dams, power plants, tunnels, ports, etc.), to monitor the crustal deformation of the Earth, to implement an urban and rural cadastre, and others. One of the most important criteria that a geodetic network must meet is reliability. In this context, the reliability concerns the network's ability to detect and identify outliers. Here, we apply the Monte Carlo Method (MMC) to investigate the reliability of a geodetic network. The key of the MMC is the random number generator. Results for simulated closed levelling network reveal that identifying an outlier is more difficult than detecting it. In general, considering the simulated network, the relationship between the outlier detection and identification depends on the level of significance of the outlier statistical test.

Key words: Computational Simulation; Geodetic Network; Hypothesis Testing; Monte Carlo Method; Outlier Detection; Quality Control.

Resumo

Uma rede geodésica consiste de pontos devidamente materializados no terreno, cujas coordenadas são estimadas por meio de medidas angulares e de distâncias entre os vértices, e/ou por meio de técnicas de posicionamento por Sistema Global de Navegação por Satélite. Estas redes são essenciais para os diversos ramos da Ciências e Engenharia, como por exemplo, no monitoramento de estruturas (barragens, pontes, usinas hidrelétricas, portos, túneis, portos, etc), no monitoramento da deformação da crosta terrestre, na implantação de um cadastro urbano e/ou rural georreferenciado, entre outros. Um dos critérios que uma rede geodésicas deve atender é a confiabilidade. Neste contexto, a confiabilidade pode ser entendida como a capacidade da rede em detectar e identificar outliers à um certo nível de probabilidade. Aqui, usamos o Método Monte Carlo (MMC) para investigar a confiabilidade de uma rede geodésica. O elemento chave do MMC é o gerador de números aleatórios. Os resultados de uma rede de nivelamento simulada revelam que identificar um outlier é mais difícil que detectá-lo. De modo geral, a relação entre a detecção e a identificação de um outlier depende do nível de significância do teste estatístico empregado para tratar os outliers.

Palavras-Chave: Método Monte Carlo; Outliers; Redes Geodésicas; Simulação Computacional; Teste de Hipóteses; Controle de Qualidade.

1 Introduction

The foundation of the Monte Carlo Method (MMC) was Buffon's needle problem by Georges Louis Leclerc in the eighteenth century. Later, in the nineteenth century, William Sealy Gosset, otherwise known as 'Student', Fisher's disciple, discovered the form of the '*t-distribution*' by a combination of mathematical and empirical work with random numbers, which is now known as an early application of the MMC. However, the MMC became well known in the 1940s, when Stanisław Ulam, Nicholas Metropolis, and John von Neumann worked on the atomic bomb project. That method was used to solve the problem of diffusion and absorption of neutrons, which was difficult to consider in any analytical approaches (Stigler; 2002).

Despite advances in science and technology to solve highly complex systems, one of the major obstacles to run a MMC up until the 1980s was the analysis time and computing resources (run time and memory). However, the advent of personal computers with powerful processors has rendered MMC a particularly attractive and cost-effective approach to performance analysis of complex systems. Therefore, the MMC emerged as a solution to help analysts understand how well a system performs under a given regime or a set of parameters.

The key of the MMC is the random number generator. A random number generator is an algorithm that generates a deterministic sequence of numbers, which simulates a sequence of independent and identically distributed (i.i.d.) numbers chosen uniformly between 0 and 1. It is *random* in the sense that the sequence of numbers generated passes the statistical tests for randomness. For this reason, random number generators are typically referred to as pseudo-random number generators (PRNGs). PRNGs are part of many machine learning and data mining techniques. In simulation, a PRNG is implemented as a computer algorithm in some programming language, and is made available to the user via procedure calls or icons (Altiok and Melamed; 2007). A good generator produces numbers that are not distinguishable from truly random numbers in a limited computation time. This is, in particular, true for Mersenne Twister (Matsumoto and Nishimura; 1998), a popular generator with a long period length of $2^{19937} - 1$.

In essence, the MMC replaces random variables by computer PRNGs, probabilities by relative frequencies, and expectations by arithmetic means over large sets of such numbers. A computation with one set of PRNG is a Monte Carlo experiment (Lehmann and Scheffler; 2011), also referred to as the number of Monte Carlo simulations (Altiok and Melamed; 2007; Gamerman and Lopes; 2006).

It is evident that in the last decades, the use of MMC for quality control proposals in geodesy has been increasing. Hekimoglu and Koch (1999) pioneered the idea of using MMC to geodesy for evaluating some probabilities as simple ratios from simulated experiments. Aydin (2012) used 5,000 MMC simulations to investigate the global test procedure in structure deformation analysis. Yang et al. (2013) used MMC to analyze the probability levels of data snooping. Koch (2015) investigated the non-centrality parameter of the F-distribution by using 100,000 simulated random variables. Klein et al. (2017) ran 1000 experiments to verify the performance of sequential likelihood ratio tests for multiple outliers. Rofatto et al. (2018a) used MMC for designing a geodetic network.

In this work, we seek to investigate the reliability of a geodetic network. One of the frequently used reliability measures is the Minimal Detectable Bias – *MDB*, see e.g. Teunissen (2006) and Teunissen (1998). The *MDB* is a diagnostic tool which allows analyzing the network's ability to detect outliers. However, not the *MDB*, but the Minimal Identifiable Bias (*MIB*) should be used as the proper diagnostic tool for outlier identification purposes (Imparato et al.; 2018). Unlike the *MDB*, the *MIB* is too complex and even practically impossible to obtain in a closed form. On the other hand, today we have fast and powerful computers, large data storage systems and modern software, which paves the way for the use of numerical simulation. In this sense, therefore, we propose the use of the MMC in order to analyze the reliability of a geodetic network in terms of the *MIB*.

The rest of the article is organized as follows: first, we provide a brief explanation on what an outlier is and explain the difference between outlier detection and outlier identification. Second, we present a MMC approach as a computational analysis tool of the reliability of a geodetic network. Third, a numerical example of the proposed method is given for a leveling network. Finally, the concluding remarks are summarized at the end of this article.

2 Outlier Detection and Identification

The most often quoted definition of outliers is that of Hawkins (1980): "An outlier is an observation that deviates so much from other observations as to arouse suspicions that it was generated by a different mechanism". In geodesy, the term outlier is defined based on a statistical hypothesis test for the presence of gross measurement errors in the observations (Baarda; 1968). Observations that are rejected by such an outlier test are called outliers. Therefore, an observation that is not grossly erroneous but is rejected by an outlier test can also be called outlier. In this context, outliers are most often caused by gross errors and gross errors most often cause outliers. But on the one hand outliers may rarely be the result of fully correct measurements, and on the other hand, mistakes or malfunctioning instruments may not always lead to large deviations, e.g., a small correction wrongly applied (Lehmann; 2013).

Since Hawkin's and most of the other definitions of outliers restrict themselves to samples (repeated observations), we follow the Lehmann (2013) definition: "An outlier is an observation that is so probably caused by a gross error that it is better not used or not used as it is".

In this section we provide the elements related to hypothesis testing for the detection and identification of a single outlier in linear(ised) models.

2.1 Outlier Detection and Minimal Detectable Bias - MDB

Baarda (1968) proposed a procedure based on hypothesis testing for the detection of a single outlier in linear(ized) models, which he called *data snooping*. Although *data snooping* was introduced as a testing procedure for use in geodetic networks, it is a generally applicable method (Lehmann; 2012). Baarda's *data snooping* consists of screening each individual observation for a possible outlier (Teunissen; 2006). Baarda's *w*-test statistic for his *data snooping* is given by a normalised least-squares residual. This test, which is based on a linear mean-shift model, can also be derived as a particular case of the generalised likelihood ratio test.

In principle, Baarda's *w*-test only makes a decision between the null H_0 and a single alternative hypothesis H_i. The null hypothesis, which is also called the working hypothesis, corresponds to a supposedly valid model describing the physical reality of the observations without the presence of an outlier. When it is assumed to be '*true*', this model is used to estimate the unknown parameters, typically in a least-squares approach. Thus, the null hypothesis of the standard Gauss–Markov model in linear or linearised form is given by equation (1) (Koch; 1999).

$$H_0 : E(\boldsymbol{y}) = \boldsymbol{Ax}, D(\boldsymbol{y}) = \boldsymbol{\Sigma_{yy}} \tag{1}$$

Where:

- E(.) is the expectation operator;
- $\boldsymbol{y} \in \mathbb{R}^n$ is the vector of measurements;
- $\boldsymbol{A} \in \mathbb{R}^{n\times u}$ is the Jacobian matrix (also called design matrix) of full rank u;
- $\boldsymbol{x} \in \mathbb{R}^u$ is the unknown parameter vector;
- D(.) is the dispersion operator; and
- $\boldsymbol{\Sigma_{yy}} \in \mathbb{R}^{n\times n}$ is the known positive definite co-variance matrix of the measurements.

The redundancy (or freedom degrees) of the model in (1) is r=n−u, where n is the number of measurements and u the number of parameters.

Instead of H_0, Baarda (1968) proposed a mean shift alternative hypothesis H_i, also referred to as model misspecification by Teunissen (2006), as follows:

$$H_i : E(\boldsymbol{y}) = \boldsymbol{Ax} + \boldsymbol{c}_i\boldsymbol{\nabla}_i, D(\boldsymbol{y}) = \boldsymbol{\Sigma_{yy}} \tag{2}$$

In the equation (2), $\boldsymbol{c}_i$ is a canonical unit vector, which consists exclusively of elements with values of 0 and 1, where 1 means that an outlier of magnitude $\boldsymbol{\nabla}_i$ affects an *i-th* measurement and 0 otherwise, e.g. $\boldsymbol{c}_i$= [0 0 . . . 1^i 0 0 . . . 0]. Therefore, the purpose of the data snooping procedure is to screen each individual observation for an outlier.

To verify if there are sufficient evidences to reject or not the null hypothesis, the test for binary case should be performed as (3):

$$Accept\ H_0\ if\ |w_i| \leq \sqrt{\chi^2_{\alpha_0}(r=1,0)} = \sqrt{k} \tag{3}$$

Where:

$$|w_i| = \frac{\boldsymbol{c}_i^\top \boldsymbol{\Sigma_{yy}^{-1}}\boldsymbol{\widehat{e}_0}}{\boldsymbol{c}_i^\top \boldsymbol{\Sigma_{yy}^{-1}}\boldsymbol{\Sigma_{\widehat{e}_0}}\boldsymbol{\Sigma_{yy}^{-1}}} \tag{4}$$

In the equations 3 and 4, $|w_i|$ is the Baarda's w-test statistic for the data snooping, which represents the normalised least-squares residual for each measurement; $\boldsymbol{\Sigma_{\widehat{e}_0}}$ is the co-variance matrix of the best linear unbiased estimator of $\boldsymbol{\widehat{e}_0}$ under H_0; and $\boldsymbol{\widehat{e}_0}$ is the least-squares residuals vector of H_0 which has this distribution under H_0. The critical value $\sqrt{k}$=$\sqrt{\chi^2_{\alpha_0}(r=1,0)}$ is computed from the central chi-squared distribution with $r = 1$ degree of freedom and type I error, also known as false alarm or level of significance,α_0 (note: the index '0' represents the case of a single alternative hypothesis testing). The second argument of $\sqrt{\chi^2_{\alpha_0}(r=1,0)}$ is the non-centrality parameter $\lambda_{r=1}$, that in this case is $\lambda_{r=1} = 0$.

In the case of accepting in favour of H_i, there is an outlier that causes the expectation of $|w_i|$ to become $\lambda_{r=1}$. The non-centrality parameter ($\lambda_{r=1}$) describes the discrepancy between H_0 of equation (1) and H_i of equation (4), and it is given by (5):

$$\lambda_{r=1} = \boldsymbol{c}_i^\top \boldsymbol{\Sigma_{yy}^{-1}}\boldsymbol{\Sigma_{\widehat{e}_0}}\boldsymbol{\Sigma_{yy}^{-1}}\boldsymbol{\nabla}_i^2 \tag{5}$$

Because Baarda's w-test in its essence is based on binary hypothesis testing, in which one decides between the null hypothesis H_0 of equation (1) and a unique alternative hypothesis H_i of equation (2), it may lead to type I error α_0 and type II error β_0. The probability of type I error α_0 is the probability of rejecting the null hypothesis when it is true, whereas the type II error β_0 is the probability of failing to reject the null hypothesis when it is false.

Instead of α_0 and β_0, there is the confidence level ($CL = 1 - \alpha_0$) and power of the test $\gamma_0 = 1 - \beta_0$, respectively. The first deals with the probability of accepting a true null hypothesis; the second, with the probability of correctly accepting the alternative hypothesis. The Fig. 1 shows an example of the relationship between these variables.

Note in (5) that the non-centrality parameter $\lambda_{r=1}$ requires knowledge of the outlier size $\boldsymbol{\nabla}_i$, which in practice is unknown. On the other hand, $\lambda_{r=1}$ can be computed as a function of α_0, γ_0, and for $r = 1$. In such case, the term $\boldsymbol{c}_i^\top \boldsymbol{\Sigma_{yy}^{-1}}\boldsymbol{\Sigma_{\widehat{e}_0}}\boldsymbol{\Sigma_{yy}^{-1}}\boldsymbol{\nabla}_i^2$ becomes a scalar and the solution of the quadratic equation (5) is given by (6) (Teunissen; 2006):

$$|\boldsymbol{\nabla}_i| = MDB_i = \sqrt{\frac{\lambda_{r=1}(\alpha_0,\gamma_0)}{\boldsymbol{c}_i^\top \boldsymbol{\Sigma_{yy}^{-1}}\boldsymbol{\Sigma_{\widehat{e}_0}}\boldsymbol{\Sigma_{yy}^{-1}}\boldsymbol{c}_i}} \tag{6}$$

In the equation 6, $|\boldsymbol{\nabla}_i|$ is the Minimal Detectable Bias (MDB_i), which is computed for each of the n alternative hypotheses according to equation (2). For more details about MDB see e.g. (Rofatto et al.; 2018b).

Although Baarda's w-test belongs to the class of generalised likelihood ratio tests and has the property of being a uniformly most powerful invariant (UMPI)

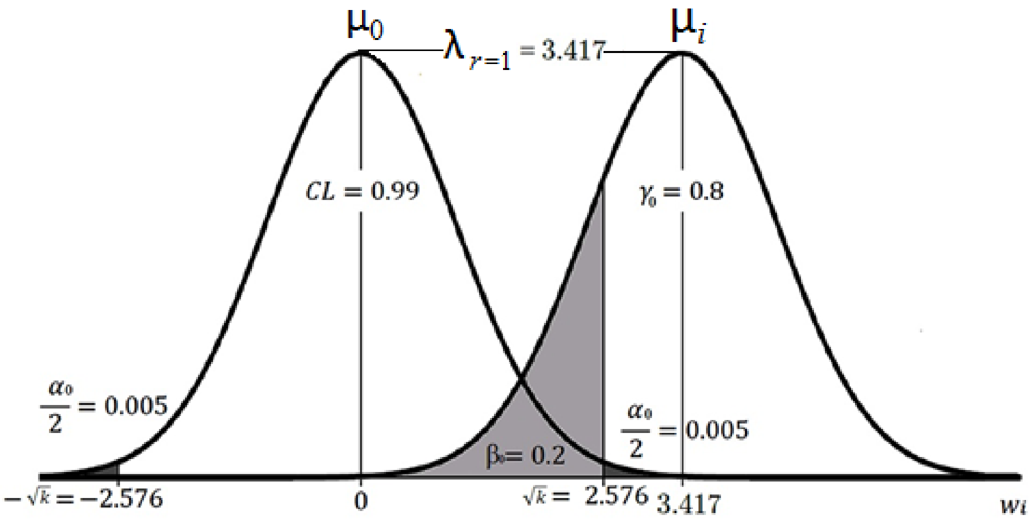

Figure 1: A non centrality parameter of $\lambda_{r=1}$ = 3.147 with α_0 = 0.01($\sqrt{k}$ = 2.576) lead to γ_0 = 0.8 (*or* β_0 = 0.2) (Adapted from Rofatto et al. (2018b)).

test when the null hypotheses is tested against a single alternative (Arnold; 1981; Teunissen; 2006), this test may not necessarily be a UMPI when more than one alternative hypothesis are considered, as is the case of the data snooping procedure (Kargoll; 2007). In the next section, we will briefly review the multiple alternative hypotheses case and the Minimal Identifiable Bias (*MIB*).

2.2 Outlier identification and Minimal Identifiable Bias

The sizes of type I and II errors are given for a single alternative hypothesis H_i of equation (2). Under this assumption, the *MDB* can be obtained as a lower bound of the outlier that can be successfully detected (Yang et al.; 2013). In practice, however, we do not have a single alternative hypothesis during the data snooping procedure, but we have multiple alternative hypotheses. Therefore, the data snooping procedure has an effect when it returns the largest absolute value among the w_i, i.e. (Teunissen; 2006):

$$w = \max |w_i|, \quad i \in \{1, \dots, n\} \tag{7}$$

The concept of multiple testing says that if H_0 is rejected, among all H_i's the one should be accepted, which would have rejected H_0 with the least α. In the case that all critical values are identical, it is most simple: H_i with the maximum test statistic should be accepted. In order to check its significance, the maximum value w should be compared with a critical value ($\sqrt{k}$) (Rofatto et al.; 2018b). In that case, the data snooping procedure is therefore given as:

$$Accept\ H_0\ if\ w \leq \sqrt{k} \tag{8}$$

Otherwise,

$$Accept\ H_i\ if\ w > \sqrt{k} \tag{9}$$

According to the inequalities (8) and (9), If none of the n w-tests gets rejected, then we accept the null hypothesis H_0.

For the test with multiples alternative hypotheses, apart from type I and type II errors, there is a third type of wrong decision when Baarda's data snooping is performed. Baarda's data snooping can also flag a non-outlying observation while the 'true' outlier remains in the dataset. We are referring to the type III error (Hawkins; 1980). The determination of the type III error (here denoted by κ_{ij}) involves a separability analysis between the alternative hypotheses (Förstner; 1983). Therefore, we are now interested in the identification of the correct alternative hypothesis. In this case, rejection of H_0 does not necessarily imply the correct identification of a particular alternative hypothesis.

Under multiple alternative hypotheses, the probabilities of type I errors in the data snooping procedure for outlier identification, when there are no outliers, are given by (10):

$$\alpha_{0i} = \int_{|w_i|>|w_j|\forall æ,\ |w_i|>\sqrt{k}} f_0' dw_1 \dots dw_n \tag{10}$$

In the equation (10), f_0' is the probability density function when the expectation of the multivariate Baarda's w-test statistics is zero (i.e. μ_n=0).

Based on the assumption that one outlier is in the ith position of the dataset, the probability of a correct identification is given by (11):

$$1 - \beta_{ii} = \int_{|w_i|>|w_j|\forall æ,\ |w_i|>\sqrt{k}} f_i' dw_1 \dots dw_n \tag{11}$$

Where f_i' is the probability density function when the expectation of the multivariate Baarda's w-test statistics is not equal to zero ($\mu_n \neq 0$).

The probability of type II error for multiple testing is given by (12):

$$\beta_{io} = \boldsymbol{P}\left[\bigcap_{i=1}^{n} |w_i| \leq \sqrt{k} \;\middle|\; H_i : true\right] \tag{12}$$

In that case, the probability of type III error is given by (13):

$$\begin{aligned}&\sum_{i=1}^{n} \boldsymbol{P}[|w_j| > |w_i| \forall_i, |w_j| > \sqrt{k}\ (i \neq j) \mid H_i : true] \\ &= \sum_{i=1}^{n} \kappa_{ij\ (i \neq j)}\end{aligned} \tag{13}$$

Testing H_0 against $H_1, H_2, H_3, \ldots, H_n$ is not a trivial task for identification purposes, because the higher the dimensionality of the alternative hypotheses, the more complicated the level probabilities associated with the data snooping procedure.

Teunissen (2018) recently introduced the concept of Minimal Identifiable Bias (*MIB*) as the smallest outlier that leads to its identification for a given correct identification rate. The detection and identification are equal in the case where we only have the one alternative hypothesis. However, under n alternative hypotheses (multiple testing), we have from equations (11), (12) and (13):

$$\beta_{ii} = \beta_{io} + \sum_{i=1}^{n}\sum_{i=1}^{n} \kappa_{ij\ (i \neq j)} \tag{14}$$

or

$$1 - \beta_{ii} = \gamma_0 - \sum_{i=1}^{n} \kappa_{ij\ (i \neq j)} \therefore \gamma_0 = 1 - \beta_{ii} + \sum_{i=1}^{n} \kappa_{ij\ (i \neq j)} \tag{15}$$

The probability of correct detection γ_0 (power of the test for a single alternative hypothesis) is the sum of the probability of correct identification $1 - \beta_{ii}$ (selecting a correct alternative hypothesis) and the probability of misidentification $\sum_{i=1}^{n} \kappa_{ij\ (i \neq j)}$ (selecting one of the n-1 other hypotheses). Thus, we have the follow inequality (Imparato et al.; 2018):

$$1 - \beta_{ii} \leq \gamma_0 \tag{16}$$

As a consequence of that inequality (16), the *MIB* will be larger than *MDB*, i.e. *MIB* $\geq$ *MDB*.

Because the acceptance region (as well as the critical region) for the multiple alternative hypotheses case is analytically intractable, the computation of *MIB* should be based on Monte Carlo integration method (MMC). In this respect, Imparato et al. (2018); Teunissen (2018) showed how to compute the *MIB*. They found that the larger the size of the outlier and/or more precisely, the estimated outlier, the higher the probability of being correctly identified. In addition, increasing the type I error (i.e. reducing the acceptance region) leads to higher probabilities of correct identification. Furthermore, increasing the number of alternative hypotheses leads to a lower probability of correct identification.

There is no difference between *MDB* and *MIB* in the case of a single alternative hypothesis. As the number of alternative hypotheses increases, however, *MDB*'s become smaller, whereas *MIB*'s become larger.

The theory presented so far is for a single round of data snooping. In practice, however, the data snooping is applied iteratively in the process of estimation, identification, and adaptation. First, the least-squares residual vector is estimated and Baarda's w-test statistics are computed by (4). Then, the detector given by (7) is applied to identify the most likely outlier. The identified outlier is then excluded from the dataset and the least-squares estimation adjustment is restarted without the rejected observation. Then, Baarda's w-test (4) as well as the detector (7) are again computed. Obviously, if redundancy permits, this procedure is repeated until no more (possible) outliers can be identified. This procedure is called iterative data snooping procedure - *IDS* (Teunissen; 2006).

In the case of IDS, a reliability measure cannot be easily computed for quality control purposes. Consequently, *MIB* is valid only for the case where data snooping is run once, and they cannot be used as a diagnostic tool for IDS. Because an analytical formula is not easy to compute, a MMC should be run to obtain the *MIB* for IDS. The MMC allows insights into these cases where analytical solutions are extremely complex to fully understand, are doubted for one reason or another, or are not available (Rofatto et al.; 2018b).

Recent studies by Rofatto et al. (2017) showed how to extract the probability levels associated with Baarda's *IDS* procedure by MMC. Furthermore, they introduced two new classes of wrong decisions for *IDS*, which they called over-identification. One is the probability of *IDS* flagging simultaneously the outlier and good observations. Second is the probability of *IDS* flagging only the good observations as outliers (more than one) while the outlier remains in the dataset. Obviously, these two new false decisions could occur during the iterative process of estimation, identification, and exclusion, as is the case of *IDS*.

3 MIB based on Monte Carlo Method

The probability levels associated with *IDS* are not easy to study using analytical models owing to the paucity or lack of practically computable solutions (closed form or numerical). Therefore, identifying an outlier is still a bottleneck in geodesy. On the other hand, a MMC method can almost always be run to generate system histories that yield useful statistical information on system operation and performance measures as pointed out by Altiok and Melamed (2007).

A geodetic network are typically composed by distances and angles measurements. Generally, the random errors of good measurements are normally distributed with expectation zero. In order to have normal random errors, uniformly distributed random number sequences (produced by the Mersenne Twister algorithm, for example) are transformed into a normal distribution by using the Box–Muller transformation (Box and Muller; 1958). Box–Muller has been used in geodesy for MMC (Lehmann; 2012).

A procedure based on the MMC is applied to compute the probability levels of *IDS* as follows (summarised as a flowchart in Fig. 2).

In the first step, the design matrix $\boldsymbol{A} \in \mathbb{R}^{n \times u}$ and the co-variance matrix of the measurements $\boldsymbol{\Sigma_{yy}} \in \mathbb{R}^{n \times n}$ are entered; then, the significance level α and the magnitude intervals of simulated outliers are defined.

The magnitude intervals of outliers are based on a standard deviation of measurements (e.g. $|3\sigma \text{ to } 9\sigma|$, where σ is the standard deviation of measurement. The random error vectors are artificially generated based on a multivariate normal distribution, because the assumed stochastic model for random errors is based on a matrix co-variance of the measurements. In this work, we use the Mersenne Twister algorithm to generate a sequence of PRNG and Box–Muller to transform it into a normal distribution. On the other hand, the magnitude of the outlier (one outlier at a time, $r = 1$) is selected based on magnitude intervals of the outliers for each Monte Carlo experiment. We use the continuous 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. Thus, the total error ϵ is a combination of the random errors and its corresponding outlier, which is given as as follows:

$$\epsilon = \boldsymbol{e} + \boldsymbol{c}_i \boldsymbol{\nabla}_i \tag{17}$$

Where: $\boldsymbol{e} \in \mathbb{R}^n$ is the PRNG from normal distribution, i.e. $\boldsymbol{e} \sim \mathcal{N}(0, \boldsymbol{\Sigma_{yy}})$, $\boldsymbol{c}_i$ consists exclusively of elements with values of 0 and 1, where 1 means that an outlier of magnitude $\boldsymbol{\nabla}_i$ affects an *i-th* measurement, and 0 otherwise.

After the total error has been generated, the least-squares residuals vector $\hat{\boldsymbol{e}}_0$ is computed using equation (18):

$$\hat{\boldsymbol{e}}_0 = \boldsymbol{R}\epsilon, \; with\; \boldsymbol{R} = \boldsymbol{I} - \boldsymbol{A}(\boldsymbol{A}^\top \boldsymbol{W}\boldsymbol{A})^{-1}\boldsymbol{A}^\top \boldsymbol{W} \tag{18}$$

In the equation (18), we have $\boldsymbol{R} \in \mathbb{R}^{n \times n}$ as the redundancy matrix, $\boldsymbol{W} = {\sigma_0}^2 \boldsymbol{\Sigma}_{yy}^{-1} \in \mathbb{R}^{n \times n}$ the weight matrix of the measurements, where ${\sigma_0}^2$ is the variance scalar factor, and $\boldsymbol{I} \in \mathbb{R}^{n \times n}$ the identity matrix (Koch; 1999).

For *IDS*, the hypothesis of (2) for one outlier is assumed and the corresponding test statistic is computed according to (4). Then, the maximum test statistic value is computed according to (7). After identifying the observation suspected as the most likely outlier, it is typically excluded from the model, and least-squares estimation and data snooping are applied iteratively until there are no further outliers identified in the dataset. The procedure should be performed for m experiments of random error vectors with each experiment contaminated by an outlier.

If m is the total number of MMC experiments, we count the number of times that the outlier is correctly identified (denoted as n_{CI}), i.e. $\max |{w_i}^{\nu}| > \sqrt{k}$ for $\nu = \{1, \ldots, m\}$. Then, the probability of correct identification (P_{CI}) can be approximated as follows (Rofatto et al.; 2018b):

$$P_{CI} \approx \frac{n_{CI}}{m} \tag{19}$$

The error probabilities are also approximated as follow:

$$P_{MD} \approx \frac{n_{MD}}{m} \tag{20}$$

$$P_{WE} \approx \frac{n_{WE}}{m} \tag{21}$$

$$P_{over+} \approx \frac{n_{over+}}{m} \tag{22}$$

$$P_{over-} \approx \frac{n_{over-}}{m} \tag{23}$$

Where:

- n_{MD} is the number of experiments in which the *IDS* does not detect the outlier;
- P_{MD} represents the type II error, also referred to as missed detection probability;
- n_{WE} is the number of experiments in which the *IDS* procedure flags a non-outlying observation while the '*true*' outlier remains in the dataset;
- P_{WE} represents the type III error, also referred to as wrong exclusion probability;
- n_{over+} is the number of experiments where the *IDS* identifies correctly the outlying observation and others;
- P_{over+} corresponds to the probability of *over+*;
- n_{over-} represents the number of experiments where the *IDS* identifies more than one non-outlying observation, whereas the '*true outlier*' remains in the dataset;
- P_{over-} corresponds to the probability of *over−* class;

In practice, as the magnitudes of outliers are unknown, one can define the probability of the correct identification in order to find the *MIB* for a given application. In the next section, the procedure based on MMC for the computation of *MIB* is applied in a geodetic network. The relationship between detection by *MDB* and identification by *MIB* is also studied.

4 An example of the Monte Carlo Method applied to the reliability analysis of geodetic networks

As an example, the procedure based on MMC experiments for the computation of probability levels of *IDS* is applied to the simulated closed-levelling network given by Rofatto et al. (2018b), with one control (fixed) point (A) and three points

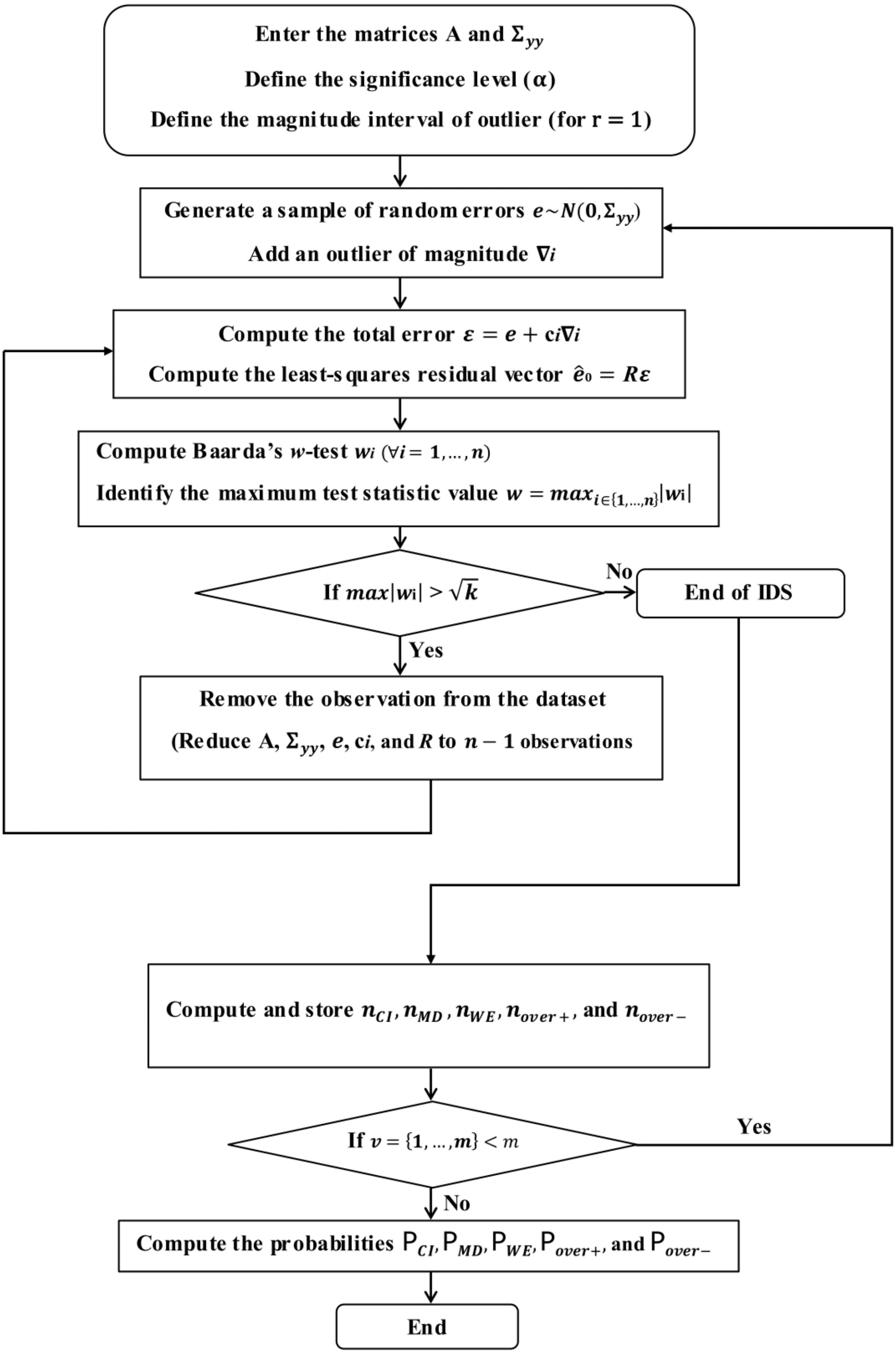

Figure 2: Flowchart of the procedure based on MMC for computation of the probability levels of IDS for each measurement (Rofatto et al.; 2018b).

Table 1: *MDB* and *MIB* for each significance level $\alpha_{(\%)}$ and for a power of $\gamma = 0.8_{(80.0\%)}$

Measurement		$\alpha_{0.1\%}$	$\alpha_{1\%}$	$\alpha_{5\%}$	$\alpha_{10\%}$
1	MDB	5.3σ	4.4σ	3.6σ	3.2σ
	MIB	5.5σ	4.8σ	4.7σ	6.5σ
2	MDB	6.6σ	5.4σ	4.4σ	4.0σ
	MIB	6.8σ	6.0σ	5.8σ	$> 9\sigma$
3	MDB	6.6σ	5.4σ	4.4σ	4.0σ
	MIB	6.8σ	6.0σ	5.8σ	6.5σ
4	MDB	5.3σ	4.4σ	3.6σ	3.2σ
	MIB	5.5σ	4.8σ	4.7σ	$> 9.0\sigma$
5	MDB	6.6σ	5.4σ	4.4σ	4.0σ
	MIB	6.8σ	6.0σ	5.8σ	7.0σ
6	MDB	5.3σ	4.4σ	3.6σ	3.2σ
	MIB	5.5σ	4.8σ	4.7σ	6.0σ

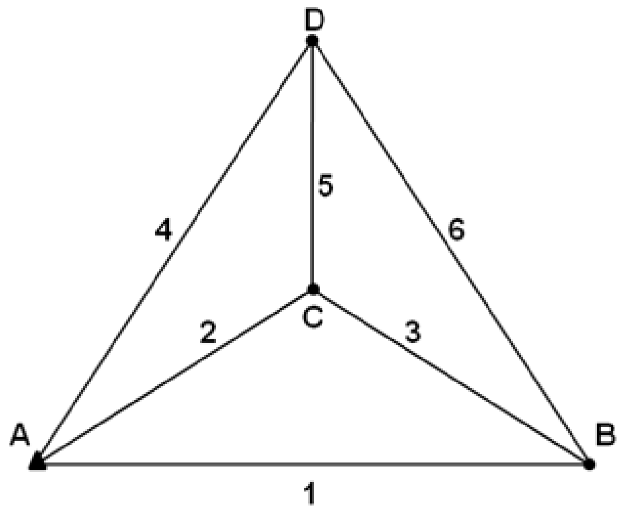

Figure 3: Simulated geodetic levelling network

with unknown heights (B, C, and D), totalling four minimally constrained points (Fig. 3). The simulated geodetic network has a minimal number of redundant measurements that lead the identification of a single outlier.

It is important to mention that geodetic network presents a minimum configuration to identify at least one single 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 number of possible identifiable outliers should be equal to the minimal number of redundant measurements across each and every point, minus one.

There are $n = 6$ measurements, $u = 3$ unknowns, and $n - u = 3$ redundant measurements in this network. Therefore, the geodetic network would be able to identify one outlier. The measurements 1, 2, 3, 4, 5, and 6 are assumed normally distributed, uncorrelated, and with nominal precision (a prior standard deviation σ) of $\pm 8mm$, $\pm 5.6mm$, $\pm 5.6mm$, $\pm 8mm$, $\pm 5.6mm$, and $\pm 8mm$, respectively. The magnitude interval of outlier is from the minimum 3σ to maximum 9σ, with an interval rate of 0.1σ. Here, positive and negative outliers are considered for each measurement. Four values were considered for the significance level: $\alpha = 0.001(0.1\%)$, $\alpha = 0.01(1\%)$, $\alpha = 0.05(5\%)$ and $\alpha = 0.1(10\%)$. We ran 10,000 MMC experiments for each measure and for each outlier magnitude interval, totalling 12,960,000 numerical experiments.

Figure 4) shows the power of the test, type II and III errors of *IDS*, and (Figure 5) the over-identification probabilities, for the case where there is a single outlier contaminating the measurements. In general, the larger magnitude of the outlier, the higher the success rate (i.e. power of the test). It can be noted that the type III error is the smallest for $\alpha = 0.001$ and largest for the type II error. Furthermore, it is rare for an outlier of small magnitude, say 3σ to 4σ, to be identified on that network.

In general, for the simulated network, the smaller α, the larger is the β. On the other hand, the smaller α, the smaller type III error (κ). For two classes of over-identification probabilities, in general, the influence of committing the over-identification+ and over-identification− is directly related to probability level α: the larger α, the larger the over-identifications case. Note that for $\alpha = 0.001$, the over-identification cases are practically absent.

Besides that, the *MDB* were computed for each measurement and for the four significance level described above. The relationship between *MDB* and *MIB* is showed in the Tab. 1. The higher the level of significance α, the higher is the probability of detecting it, i. e. the smaller the *MDB*. This relationship, however, does not work for *MIB*. The MIB is slightly larger than the MDB for that geodetic network, except for the significance level of 10%, for which the *MIB* is approximately two times larger than the *MDB*. Therefore, due to the low redundancy of measurements in the network, it is not recommended to use a significance level of 10% for outlier identification proposals.

This example shows how to compute for the IDS case based on the MMC. Obviously, should be computed for a given probability of correct identification (γ) and significance level (α).

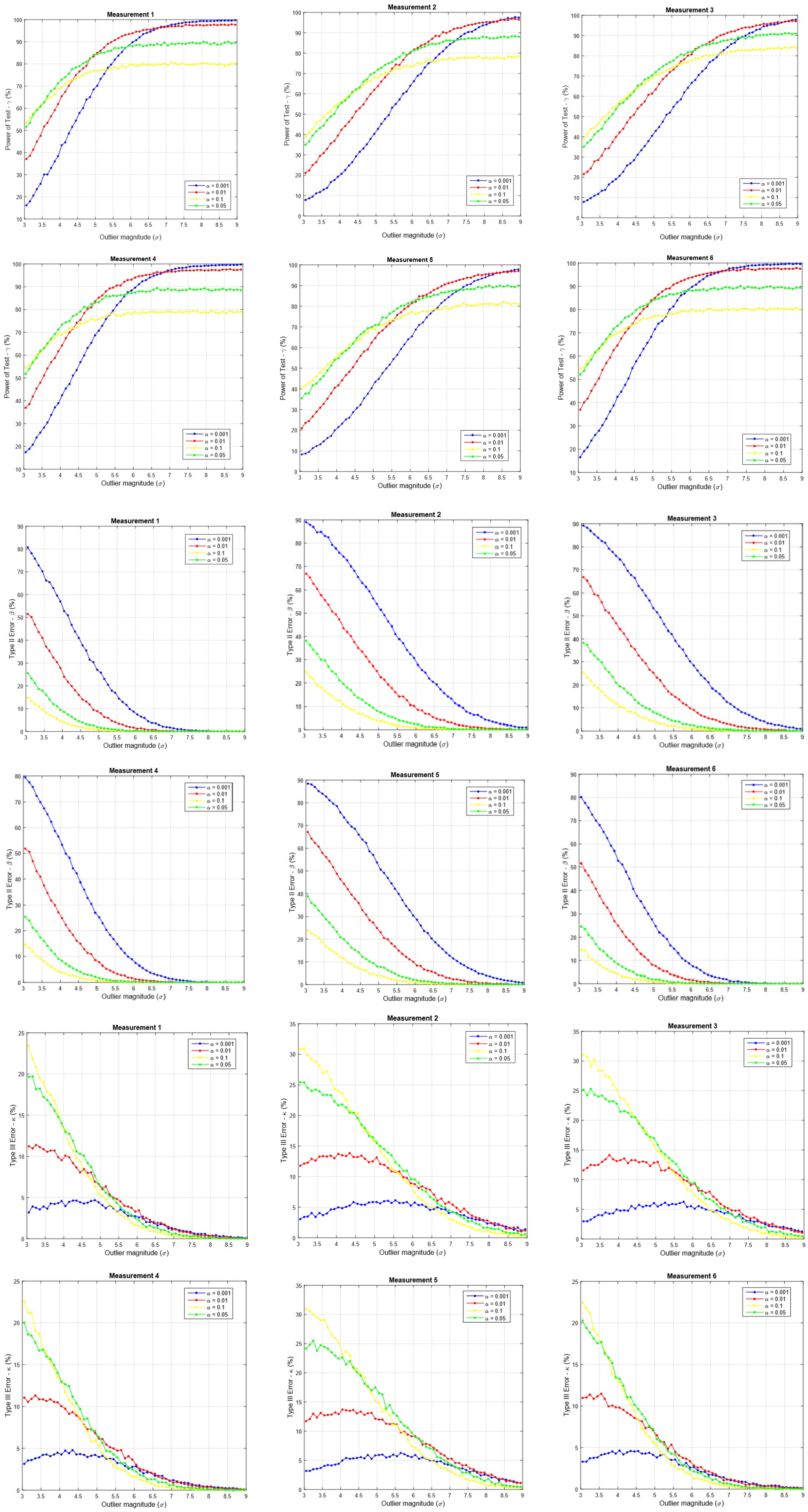

Figure 4: Power of the test, type II and type III error for each significance level α.

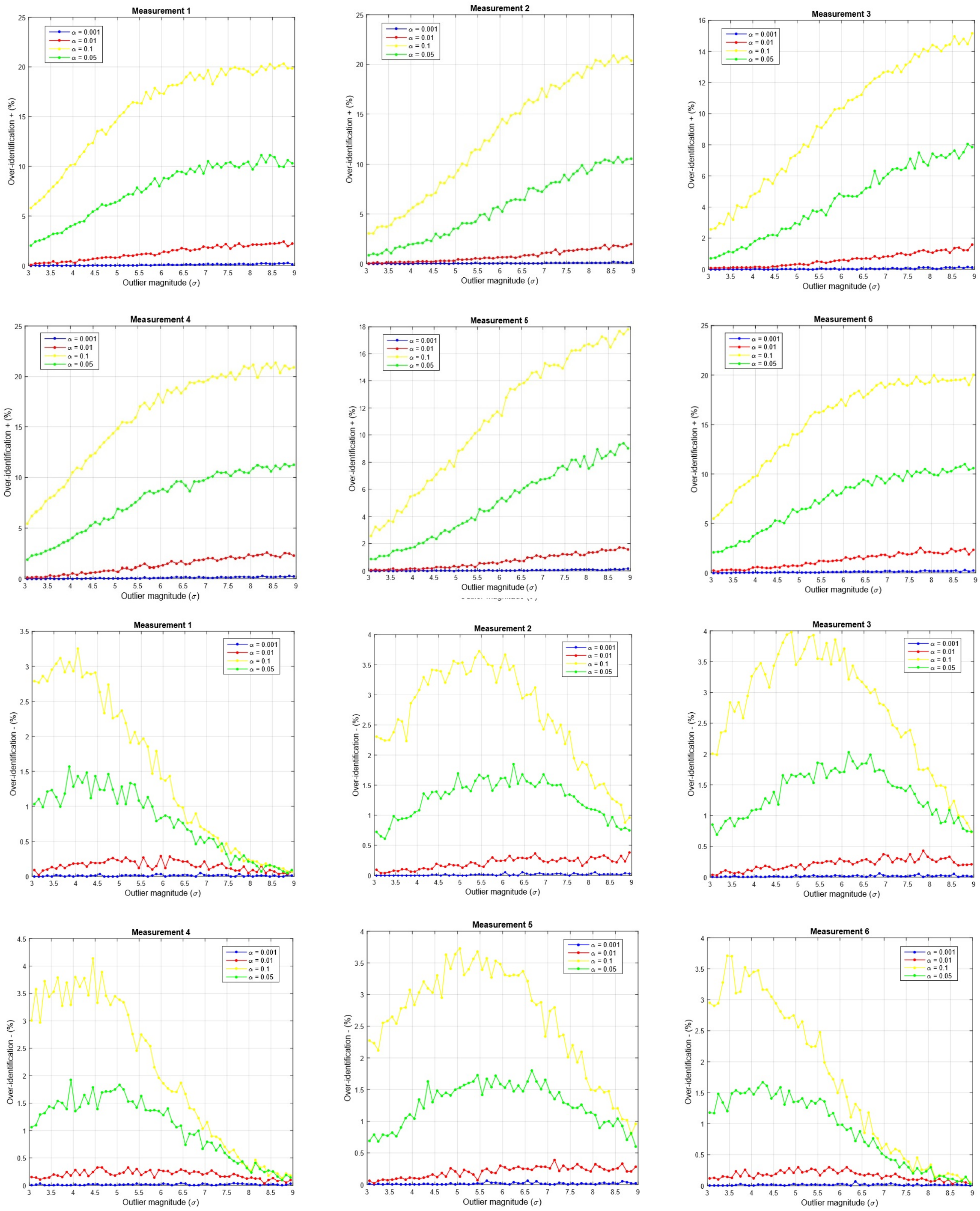

Figure 5: Over-identification probabilities for each significance level α

5 Final Remarks

In this study, we highlighted that Monte Carlo method (MMC) is a primary tool for deriving solutions to complex problems. We used the Monte Carlo method as a key tool for studying the IDS procedure. We emphasized that, the method discards the use of real measurements. Actually, it is assumed that the random errors of the good measurements are normally distributed, and therefore can be artificially generated by means of a PRNG. Thus, in fact, the only needs are the geometrical network configuration (given by design matrix); the uncertainty of the observations (which can be given by nominal standard deviation of the equipment); and the magnitude intervals of the outliers.

We also highlighted that in contrast to the well-defined theories of reliability, the IDS procedure is a heuristic method, and therefore, there is no theoretical reliability measure for it. Hence, an analytical model with tractable solution is unknown, and therefore, one needs to resort to MMC. Based on the work by Rofatto et al. (2018b), we showed how to find the probability levels associated with IDS and how to obtain its for each observation by means of the MMC for a given correct identification probability and significance level.

Acknowledgements

The authors would like to acknowledge the support from FAPEMIG (Fundação de Amparo à Pesquisa do Estado de Minas Gerais, research project 2018/7285). The authors also like to extend our gratitude to the anonymous referee for some valuable comments on a previous version of this text.

References

Altiok, T. and Melamed, B. (2007). *Simulation Modeling and Analysis with Arena*, 1 edn, Academic Press.

Arnold, S. (1981). *The theory of linear models and multivariate analysis*, 1 edn, Wiley.

Aydin, C. (2012). Power of global test in deformation analysis, *Journal of Surveying Engineering* **138**(2): 51–56. https://doi.org/10.1061/(ASCE)SU.1943-5428.0000064.

Baarda, W. (1968). A testing procedure for use in geodetic networks., *Publ. on geodesy, New Series* **2**(5).

Box, G. E. P. and Muller, M. E. (1958). A note on the generation of random normal deviates, *The Annals of Mathematical Statistics* **29**(2): 610–611. https://doi:10.1214/aoms/1177706645.

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*, Vol. Series A, Deutsche Geodätische Kommision, Munich, Germany, pp. 79–103.

Gamerman, D. and Lopes, H. F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, 2 edn, Chapman and Hall/CRC.

Hawkins, D. M. (1980). *Identification of Outliers*, 1 edn, Springer Netherlands. https://doi.org/10.1007/978-94-015-3994-4.

Hekimoglu, S. and Koch, K. R. (1999). How can reliability of the robust methods be measured?, *in* M. O. Altan and L. Gründig (eds), *Third Turkish-German joint geodetic days*, Vol. 1, Istanbul Technical University, Istanbul, Turkey, pp. 179–196.

Imparato, D., Teunissen, P. J. D. and Tiberius, C. C. J. M. (2018). Minimal detectable and identifiable biases for quality control, *Survey Review* . https://doi.org/10.1080/00396265.2018.1437947.

Kargoll, B. (2007). *On the theory and application of model misspecification tests in geodesy*, Doctoral thesis, University of Bonn, Landwirtschaftliche Fakultät, German, Bonn.

Klein, I., Matsuoka, M. T., Guzatto, M. P. and Nievinski, F. G. (2017). An approach to identify multiple outliers based on sequential likelihood ratio tests, *Survey Review* **49**(357): 449–457. https://doi.org/10.1080/00396265.2016.1212970.

Koch, K. R. (1999). *Parameter estimation and hypothesis testing in linear models.*, 2 edn, Springer.

Koch, K. R. (2015). Minimal detectable outliers as measures of reliability, *Journal of Geodesy* **89**(5): 483–490. https://doi.org/10.1007/s00190-015-0793-5.

Lehmann, R. (2012). Improved critical values for extreme normalized and studentized residuals in gauss-markov models, *Journal of Geodesy* **86**(12): 1137–1146. https://doi.org/10.1007/s00190-012-0569-0.

Lehmann, R. (2013). On the formulation of the alternative hypothesis for geodetic outlier detection, *Journal of Geodesy* **87**(4): 373–386.

Lehmann, R. and Scheffler, T. (2011). Monte carlo based data snooping with application to a geodetic network, *Journal of Applied Geodesy* **5**(3-4): 123–134. https://doi.org/10.1515/JAG.2011.014.

Matsumoto, M. and Nishimura, T. (1998). Mersenne twister: A 623-dimensionally equidistributed uniform pseudo-random number generator, *ACM Transactions on Modeling and Computer Simulation* **8**(1): 3–30. https://dl.acm.org/citation.cfm?id=272991.

Rofatto, V. F., Matsuoka, M. T. and Klein, I. (2017). An attempt to analyse baarda's iterative data snooping procedure based on monte carlo simulation, *South African Journal of Geomatics* **6**(6): 416–435. http://dx.doi.org/10.4314/sajg.v6i3.11.

Rofatto, V. F., Matsuoka, M. T. and Klein, I. (2018a). Design of geodetic networks based on outlier identification criteria: An example applied to the leveling network, *Bulletin of Geodetic Sciences* **24**(2): 152–170. http://dx.doi.org/10.1590/s1982-21702018000200011.

Rofatto, V. F., Matsuoka, M. T., Klein, I., Veronez, M., Bonimani, M. L. and Lehmann, R. (2018b). A half-century of baarda's concept of reliability: A review, new perspectives, and applications, *Survey Review* . https://doi.org/10.1080/00396265.2018.1548118.

Stigler, S. M. (2002). *Statistics on the Table: The History of Statistical Concepts and Methods*, 1 edn, Harvard University Press.

Teunissen, P. J. G. (1998). Minimal detectable biases of gps data, *Journal of Geodesy* **72**(4): 236–244. https://doi.org/10.1007/s001900050163.

Teunissen, P. J. G. (2006). *Testing Theory: an introduction*, 2 edn, Delft University Press.

Teunissen, P. J. G. (2018). Distributional theory for the dia method, *Journal of Geodesy* **91**(1): 59–80. https://doi.org/10.1007/s00190-017-1045-7.

Xu, P. (2005). Sign-constrained robust least squares, subjective breakdown point and the effect of weights of observations on robustness, *Journal of Geodesy* **79**(1): 146–159.

Yang, L., Wang, J., Knight, N. and Shen, Y. (2013). Outlier separability analysis with a multiple alternative hypotheses test, *Journal of Geodesy* **87**(6): 591–604. https://doi.org/10.1007/s00190-013-0629-0.