Survey Review

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ysre20

An artificial neural network-based critical values for multiple hypothesis testing: data-snooping case

Vinicius Francisco Rofatto, Marcelo Tomio Matsuoka, Ivandro Klein, Maria Luísa Silva Bonimani, Bruno Póvoa Rodrigues, Caio Cesar de Campos, Mauricio Roberto Veronez & Luiz Gonzaga da Silveira Jr.

To cite this article: Vinicius Francisco Rofatto, Marcelo Tomio Matsuoka, Ivandro Klein, Maria Luísa Silva Bonimani, Bruno Póvoa Rodrigues, Caio Cesar de Campos, Mauricio Roberto Veronez & Luiz Gonzaga da Silveira Jr. (2022) An artificial neural network-based critical values for multiple hypothesis testing: data-snooping case, Survey Review, 54:386, 440-455, DOI: 10.1080/00396265.2021.1968176

To link to this article: https://doi.org/10.1080/00396265.2021.1968176

Published online: 08 Sep 2021.

Submit your article to this journal

Article views: 97

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=ysre20

Check for updates

An artificial neural network-based critical values for multiple hypothesis testing: data-snooping case

Vinicius Francisco Rofatto [1*], Marcelo Tomio Matsuoka [1,2,3], Ivandro Klein [4,5], Maria Luísa Silva Bonimani [2], Bruno Póvoa Rodrigues [2], Caio Cesar de Campos[1], Mauricio Roberto Veronez [3] and Luiz Gonzaga da Silveira Jr. [3]

Data Snooping is the most best-established method for identifying outliers in geodetic data analysis. It has been demonstrated in the literature that to effectively user-control the type I error rate, critical values must be computed numerically by means of Monte Carlo. Here, on the other hand, we provide a model based on an artificial neural network. The results prove that the proposed model can be used to compute the critical values and, therefore, it is no longer necessary to run the Monte Carlo-based critical value every time the quality control is performed by means of data snooping.

Keywords: Quality control, Reliability, Outlier detection, Monte Carlo, Artificial intelligence, Artificial neural network

1. Introduction

Outlier is an observation that has moved away from most likely value to the point of not belonging to the mathematical model (functional and stochastic) stipulated. Therefore, it is better to either not use it or not use it as observation (Lehmann 2013). Failure to identify an outlier can jeopardise the reliability level of a system. Because of its importance, outliers must be appropriately treated to ensure the quality of data analysis (Rofatto 2020b).

Data Snooping is one of the most best-established methods for processing observations contaminated by outliers. It has also become very popular and is routinely used in adjustment computations (Ghilani 2017). This testing procedure consists of screening each individual observation for the presence of an outlier. The test statistic employed in the data snooping is given by a normalised least-squares residual, and it is well-known as w-test (Baarda 1968). The w-test, which is based on a linear mean-shift model, can also be derived as a particular case of the generalised likelihood ratio test (Teunissen 2006).

In principle, w-test only makes a decision between the null, denoted by $\mathcal{H}_0$, and a single alternative hypothesis, say $\mathcal{H}_a$. The model under the null hypothesis is often formulated under the condition of absence of outliers, whereas the alternative model is proposed when there is an outlier in the dataset. In order to verify whether the alternative hypothesis is significant or not (i.e. whether we accept or reject the null hypothesis), the w-test statistic is compared with its critical values (i.e. the percentile of its probability distribution), which can be taken from well-known standard normal distribution. In that case, rejection of null hypothesis $\mathcal{H}_0$ automatically implies acceptance of the alternative hypothesis $\mathcal{H}_a$, and vice versa (Imparato *et al.* 2019).

In that case, the probability level associated with data snooping is restrict to type I error (rejection of a true null hypothesis) and type II error (acceptance of a false null hypothesis). The probability of committing the type I error is well-known significance level or type I error rate (denoted by α_0), which is controlled by the user. However, data snooping is by nature a procedure that involves multiple hypothesis testing (Teunissen *et al.* 2017; Zaminpardaz and Teunissen 2019). The multiple hypothesis testing problem occurs when a number of individual hypothesis tests are considered simultaneously. This means testing $\mathcal{H}_0$ against $\mathcal{H}_a^{(1)}, \mathcal{H}_a^{(2)}, \ldots, \mathcal{H}_a^{(n)}$ (Lehmann 2012).

To make it clearer, let's start by assuming that the individual tests are independent and the significance level for each test (α_i) be α_0; then the probability that one of the tests is rejecting the null hypothesis is given by Lehmann and Lösler (2016):

$$\alpha' = 1 - \prod_{i=1}^{n} (1 - \alpha_i) = 1 - (1 - \alpha_0)^n \qquad (1)$$

which α' is the probability of making one or more false

[1]Institute of Geography, Federal University of Uberlandia, Monte Carmelo, Brazil
[2]Graduate Program in Agriculture and Geospatial Information, Federal University of Uberlandia, Monte Carmelo, Brazil
[3]Advanced Visualization Laboratory (Vizlab), Graduate Program in Applied Computing, Unisinos University, São Leopoldo, Brazil
[4]Department of Civil Construction, Federal Institute of Santa Catarina, Florianopolis, Brazil
[5]Graduate Program in Geodetic Sciences, Federal University of Paraná, Curitiba, Brazil

*Email: Corresponding author, email vfrofatto@gmail.com

© 2021 Survey Review Ltd
Received 14 January 2021; accepted 29 July 2021
DOI 10.1080/00396265.2021.1968176

positives, or type I errors, when performing multiple hypotheses tests. It is known as family-wise error rate (FWER).

In this case, the significance level or the type I error rate of individual tests (α_0) no longer represents the error rate of the combined set of tests (denoted by α').

For example, let the significance level of individual tests be $\alpha_0 = 0.05$; if ten tests were run, the probability of rejecting at least one true null hypothesis would be 0.4, and if 50 tests were run for the same $\alpha_0 = 0.05$, then we would have $\alpha' = 0.92$. If multiple hypotheses are tested, the chance of observing a rare event increases, and therefore, the probability of incorrectly rejecting a null hypothesis (i.e. committing a type I error) increases (Mittelhammer *et al.* 2000). Figure 1 shows the behaviour of the FWER (α') in function of the number of the tests, and for $\alpha_0 = 0.01$, $\alpha_0 = 0.05$ and $\alpha_0 = 0.1$.

Therefore, we are interested in controlling α'. In this context, methods that deal with multiple alternative hypotheses are referred to as multiple comparison methods. Investigations on such methods are rare in geodetic applications, which, in a way, opens the way for research and applications. More details can be seen, for example, in Miller (1981), Simes (1986), Wright (1992), Sarkar and Chang (1997), Lehmann and Romano (2005) and Rom (2013).

One of the simplest methods employed to control the FWER is known as Šidák correction (Šidák 1967). The Šidák correction is derived by assuming that the individual tests are independent, as follows:

$$\alpha_0 = 1 - (1 - \alpha')^{\frac{1}{n}} \tag{2}$$

The goal of the Šidák correction in (2) is to adjust α_0 so that the significance level for the entire series of tests α' is warranted. Šidák correction produces a family-wise error rate of exactly α' when the tests are independent from each other and all null hypotheses are true. In that case, the critical values for w-test statistic can be computed as follows:

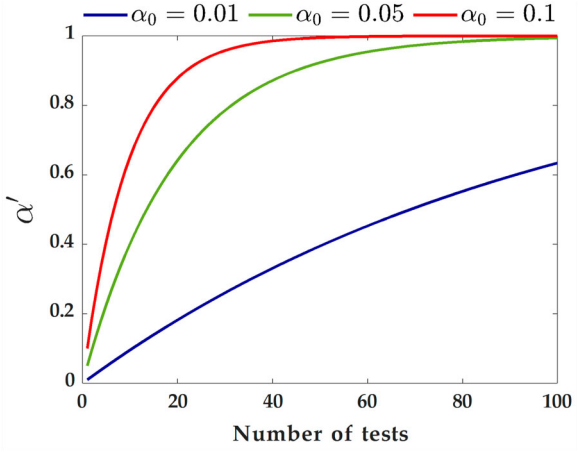
$$k_{sid} = \Phi_N^{-1}\left[\left(1 - \frac{\alpha'}{2}\right)^{\frac{1}{n}}\right] \tag{3}$$

where Φ_N^{-1} is the inverse of the standard normal cumulative distribution. The number two in the denominator is because the w-test is two-sided test. The critical values computed from Šidák correction for $\alpha' = 0.01$, $\alpha' = 0.05$ and $\alpha' = 0.1$ are displayed in Fig. 2. These critical values can be easily obtained in any software that has the implemented normal table in its routine. For example, the inverse of the standard normal cumulative distribution can be obtained directly by the 'norminv' function from Matlab.

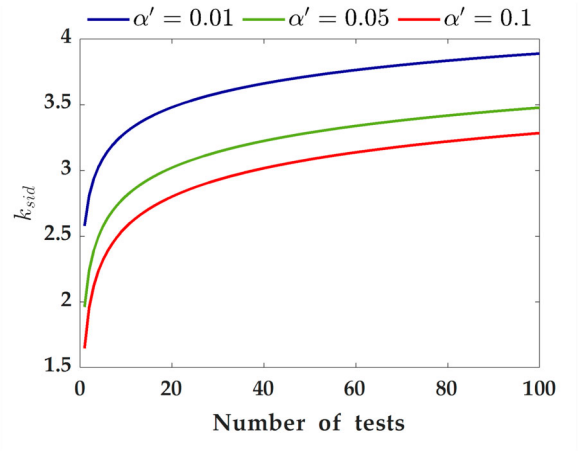
In practice, however, the mathematical model promotes correlation between w-test statistics. This means that we will always have some degree of correlation between the tests. If we neglect the correlation between the tests, we overestimate the critical values computed from Šidák correction in (3) (Lehmann 2013).

In this contribution, we apply a procedure based on Monte Carlo simulation in order to obtain the critical values that considers the correlation between the w-test statistics. The drawback is that every time data snooping is run, Monte Carlo method is used to compute the critical values (Rofatto 2020b). To overcome this issue, here, on the other hand, we first compute a series of critical values for a fixed number of observations with pre-fixed correlation between the w-test statistics by using Monte Carlo. Then, a Supervised Back Propagation Neural Network (SBPNN) architecture was trained and tested using such critical values databases. The purpose of the SBPNN model is to suppress the use of Monte Carlo method when the data snooping is in play. Furthermore, we provide a MATLAB's flexible network object type (called *SBPNN.mat*) which allows anyone to obtain the critical values quickly and easily, in addition to ensuring good control of the type I error rate when applying Data Snooping.

The paper is organised as follows. First, we provide a background of the Monte Carlo approach for user-controlling the FWER and we apply it to compute a series of critical values for some applicable FWER by considering a fixed number of observations with pre-fixed correlation between the outlier test statistics. Second, we present the methodology for the development of the SBPNN model for the computation of critical values. Next, the design of the experiments is presented and some criteria are given to compute the overall correlation between the outlier test statistics when a multivariate



1 Family-wise error rate (α') for $\alpha_0 = 0.01$, $\alpha_0 = 0.05$ and $\alpha_0 = 0.1$



2 Critical values computed from Šidák correction for $\alpha' = 0.01$, $\alpha' = 0.05$ and $\alpha' = 0.1$

problem is in play (here, specifically geodetic networks). Then, the SBPNN-based critical values are evaluated with respect to the Monte Carlo-based critical values in two different situations: (i) difference in terms of critical values and false positive rates; (ii) difference in terms of success rate of the data snooping for outlier identification and elimination. Finally, the concluding remarks are summarised at the end of this paper.

2. Monte Carlo approach for controlling the family-wise error rate

Let us start by assuming that random errors are normally distributed with expectation zero and the dispersion given by covariance matrix $\boldsymbol{Q}_e$, i.e.:

$$\mathcal{H}_0{:}\boldsymbol{e} \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{Q}_e) \tag{4}$$

This choice is further justified by both the central limit theorem and the maximum entropy principle. Some alternative observation error models can be found in Lehmann (2015) and Lichti *et al.* (2021).

However, the null hypothesis may not be fulfilled if the dataset are contaminated by outliers. Since it is not usually known which observations are outliers, a more appropriate alternative hypothesis would be: there is at least one outlier in the dataset (Lehmann 2012). Hence, the model in (4) becomes:

$$\mathcal{H}_a^{(i)}{:}\boldsymbol{e_i} \sim \boldsymbol{N}(\boldsymbol{c_i}\nabla_i, \boldsymbol{Q}_e), \quad \forall i = 1, \ldots, n \tag{5}$$

with $\boldsymbol{c_i} \in \mathbb{R}^{n\times 1}$ being a canonical unit vector, which consists exclusively of elements with values of 0 and 1, where 1 means that the ith outlier of magnitude ∇_i affects the observations, and 0 means otherwise.

These alternative hypotheses are formulated under the condition that the outlier acts as a systematic effect by shifting the random error distribution under $\mathcal{H}_0$ by its own value (Lehmann 2013). In other words, the presence of an outlier in a dataset can cause a shift of the expectation under $\mathcal{H}_0$ to a nonzero value. Therefore, hypothesis testing is often employed to check whether the possible shifting of the random error distribution under $\mathcal{H}_0$ by an anomaly is, in fact, an outlier or merely a random effect (Rofatto 2020b).

Now, we are interested in knowing which of the alternative hypotheses in (5) may lead to the rejection of the null hypothesis in (4) with a certain probability. This means testing $\mathcal{H}_0$ against $\mathcal{H}_a^{(1)}$, $\mathcal{H}_a^{(2)}$, $\mathcal{H}_a^{(3)}$, $\ldots$, $\mathcal{H}_a^{(n)}$. This is known as multiple hypothesis testing. In that case, the maximum test statistic reveals the most likely alternative hypothesis:

$$w = \max_{i\in\{1,\ldots,n\}} |\boldsymbol{w_i}| \tag{6}$$

where $\boldsymbol{w_i}$ is known as Baarda's w-test (Baarda 1968), and it is given as follows:

$$\boldsymbol{w_i} = \frac{\boldsymbol{c_i}^T\boldsymbol{Q}_e^{-1}\hat{\boldsymbol{e}}}{\sqrt{\boldsymbol{c_i}^T\boldsymbol{Q}_e^{-1}\boldsymbol{Q}_{\hat{e}}\boldsymbol{Q}_e^{-1}\boldsymbol{c_i}}}, \forall i = 1, \ldots, n \tag{7}$$

being $\hat{\boldsymbol{e}} \in \mathbb{R}^{n\times 1}$ the least squares solution for the estimated observation errors (residuals)

$$\hat{\boldsymbol{e}} = \boldsymbol{y} - \boldsymbol{A}\hat{\boldsymbol{x}} \tag{8}$$

where $\boldsymbol{y} \in \mathbb{R}^{n\times 1}$ is the vector of observations; $\boldsymbol{A} \in \mathbb{R}^{n\times u}$ is design matrix relating n observations and u parameters; and $\hat{\boldsymbol{x}} \in \mathbb{R}^{u\times 1}$ is the least squares solution for the estimated parameters vector

$$\hat{\boldsymbol{x}} = (\boldsymbol{A}^T\boldsymbol{W}\boldsymbol{A})^{-1}(\boldsymbol{A}^T\boldsymbol{W}\boldsymbol{y}) \tag{9}$$

with $\boldsymbol{W} \in \mathbb{R}^{n\times n}$ is the known matrix of weights, taken as $\boldsymbol{W} = \sigma_0{}^2\boldsymbol{Q}_e^{-1}$, where σ_0^2 is the variance factor; and $\boldsymbol{Q}_{\hat{e}} \in \mathbb{R}^{n\times n}$ is the covariance matrix of the estimated observation errors

$$\boldsymbol{Q}_{\hat{e}} = \sigma_0^2\boldsymbol{R}\boldsymbol{W}^{-1} \tag{10}$$

with $\boldsymbol{R} \in \mathbb{R}^{n\times n}$ being the well-known redundancy matrix

$$\boldsymbol{R} = \boldsymbol{I} - \boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{W}\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{W} \tag{11}$$

where $\boldsymbol{I} \in \mathbb{R}^{n\times n}$ is the identity matrix. The elements of the main diagonal of the redundancy matrix represent the redundancy of each observation, known as local redundancy (denoted by r_i).

The decision rule for this case is given by

$$\text{Otherwise,} \quad \begin{array}{ll} \text{Accept} & \mathcal{H}_0 \text{ if } \quad w \le k \\ \text{Accept} & \mathcal{H}_a^{(i)} \text{ if } \quad w > k \end{array} \tag{12}$$

where k represents the critical value.

Under the reality that the null hypothesis $\mathcal{H}_0$ in (4) is true, there is the probability of rejecting it in at least one test, which is known as family-wise error rate (FWER), as seen in previous section. The FWER is defined as:

$$FWER = \alpha' = \mathcal{P}\Big(\max|\boldsymbol{w_i}| > k \mid \mathcal{H}_0^{(i)}{:}true\Big), \quad \forall i = 1, \ldots, n \tag{13}$$

One way to control the FWER (also known as type I decision error rate) is through Šidák correction in (2). In that case, the critical value k can be computed from (3). As seen, however, this method neglects the inevitable dependencies between the test statistics w_i in (7) of each observation. This problem has already been addressed by Lehmann (2012). In order to guarantee the user-defined type I decision error rate (α') for data snooping, the critical value must be computed by Monte Carlo simulation, which is given step-by-step below (Rofatto 2020b):

(1) Given the both functional and stochastic models of the problem (i.e.the matrices $\boldsymbol{A}$ and $\boldsymbol{Q}_e$, respectively), compute the correlation matrix for the w-test statistics as

$$\mathcal{R}_w = \begin{pmatrix} \rho_{w_1,w_1} & \rho_{w_1,w_2} & \cdots & \rho_{w_1,w_n} \\ \rho_{w_2,w_1} & \rho_{w_2,w_2} & \cdots & \rho_{w_2,w_n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{w_n,w_1} & \rho_{w_n,w_2} & \cdots & \rho_{w_n,w_n} \end{pmatrix} \tag{14}$$

with $\mathcal{R}_w \in \mathbb{R}^{n\times n}$ and ρ_{w_i,w_j} being the correlation coefficient between the pairs of w-test statistics:

$$\rho_{w_i,w_j} = \frac{\boldsymbol{c_i}^T\boldsymbol{Q}_e^{-1}\boldsymbol{Q}_{\hat{e}}\boldsymbol{Q}_e^{-1}\boldsymbol{c_j}}{\sqrt{\boldsymbol{c_i}^T\boldsymbol{Q}_e^{-1}\boldsymbol{Q}_{\hat{e}}\boldsymbol{Q}_e^{-1}\boldsymbol{c_i}}\sqrt{\boldsymbol{c_j}^T\boldsymbol{Q}_e^{-1}\boldsymbol{Q}_{\hat{e}}\boldsymbol{Q}_e^{-1}\boldsymbol{c_j}}}, \quad \forall i = 1, \ldots, n \ \forall j = 1, \ldots, n \tag{15}$$

as presented by Förstner (1981). The main diagonal

elements of $\mathcal{R}_w$ will always be equal to 1, and the off-diagonal may assume values within the range $[-1, 1]$ (Knight *et al.* 2010; Yang 2013).
(2) The probability density function (pdf) assigned to the w-test statistics under $\mathcal{H}_0$-distribution is

$$(w_1, w_2, w_3, \ldots, w_n)^T \sim N(\mathbf{0}, \mathcal{R}_w) \tag{16}$$

as can be e.g. seen in Förstner (1981), Yang (2017) and Imparato *et al.* (2019).
(3) In order to have w-test statistics under $\mathcal{H}_0$, uniformly distributed random number sequences are produced by the Mersenne Twister algorithm, and then they are transformed into a normal distribution by using the Box–Muller transformation (Box and Muller 1958). Box–Muller has already been used in geodesy for Monte Carlo experiments (Lemeshko and Lemeshko 2005; Lehmann and Scheffler 2011; Lehmann 2012). Therefore, a sequence of m random vectors from the pdf assigned to the w-test statistics is generated according to (16). (Note: For the multivariate normal distribution we can use directly Matlab's pseudo-random number generator 'mvnrnd'.) In that case, we have a sequence of m vectors of the w-test statistics as follows:

$$\left[(w_1, w_2, w_3, \ldots, w_n)^{T^{(1)}}, \ldots, (w_1, w_2, w_3, \ldots, w_n)^{T^{(m)}}\right] \tag{17}$$

(4) Compute the test statistic by (6) for each sequence of w-test statistics. Thus, we have

$$\left(\max_{i\in\{1,\ldots,n\}} |\boldsymbol{w}_i|^{(1)}, \max_{i\in\{1,\ldots,n\}} |\boldsymbol{w}_i|^{(2)}, \ldots, \max_{i\in\{1,\ldots,n\}} |\boldsymbol{w}_i|^{(m)}\right) \tag{18}$$

(5) Sort in ascending order the maximum test statistic in Equation (18), getting a sorted vector $\tilde{w}$, such that

$$\tilde{w}^{(1)} < \tilde{w}^{(2)}, \tilde{w}^{(3)}, \ldots, < \tilde{w}^{(m)} \tag{19}$$

The sorted values $\tilde{w}$ in (19) provide a discrete representation of the cumulative density function (cdf) of the maximum test statistic max-w.
(6) Determine the critical value $\hat{k}$ as follows:

$$\hat{k} = \tilde{w}_{[(1-\alpha')\times m]} \tag{20}$$

where [.] denotes rounding down to the next integer that indicates the position of the selected elements in the ascending order of $\tilde{w}$. This position corresponds to a critical value for a stipulated overall false positive probability α'. This can be done for a sequence of values α' in parallel, as can be seen in Lehmann (2012).

3. A predictive model of critical values based on the neural networks concept

The use of neural network has been increasingly used in geodesy, e.g. in problems of coordinate transformation (Ziggah 2016); in the implementation of ionospheric forecasting model (Srivani *et al.* 2019); in the tropospheric delay prediction (Selbesoglu 2020). However, here it is the first time that it has been tested within the context of the critical values for the quality control.

Here, we develop a critical value predictive model based on the concept of neural networks. The proposed model was created by considering the known inputs as being overall significance level α', number of observations n and correlation between the tests $|\rho_{w_i,w_j}|$; and outputs are the critical values $\hat{k}$ computed according to the Monte Carlo, as described below.

3.1. Generation of a set of critical values from Monte Carlo for neural network

Here, critical values were computed according to the procedure described in Section (2) and stored for $\alpha' = 0.001$, $\alpha' = 0.01$, $\alpha' = 0.05$, $\alpha' = 0.1$ and $\alpha' = 0.5$. The number of the observations (n) were fixed for each α' with $n = 5$ to $n = 100$ by increment of 5. The correlation values were taken as absolute values (denoted by $|\rho_{w_i,w_j}|$). It is important to mention that the absolute value of the correlation coefficients between each pair of the w-test statistics is equivalent to the multiple correlation coefficient given by Anderson (1984). Thus, for each n the same correlation between all the pairs of w-test statistics were fixed from $|\rho_{w_i,w_j}| = 0.0$ to $|\rho_{w_i,w_j}| = 1.0$, by increment of 0.1, considering also taking into account the correlation $|\rho_{w_i,w_j}| = 0.999$.

For each combination of α', n and $|\rho_{w_i,w_j}|$, $m =$ 5,000,000 Monte Carlo experiments were run. This number of Monte Carlo experiments provides a low relative error ($< 0.1\%$) (Rofatto 2020a). In total, therefore, 1200 critical values for Data-Snooping test were computed and stored. For instance, Fig. 3 shows an example of the behaviour of the critical values for the extreme cases of having $\alpha' = 0.001$ and $\alpha' = 0.5$ in function of the number of observations n and $|\rho_{w_i,w_j}|$. These generated data are available at https://data.mendeley.com/datasets/77sfpx9b74/3.

In general, we can observe the following relationship ($\uparrow$ means increase and $\downarrow$ decrease):

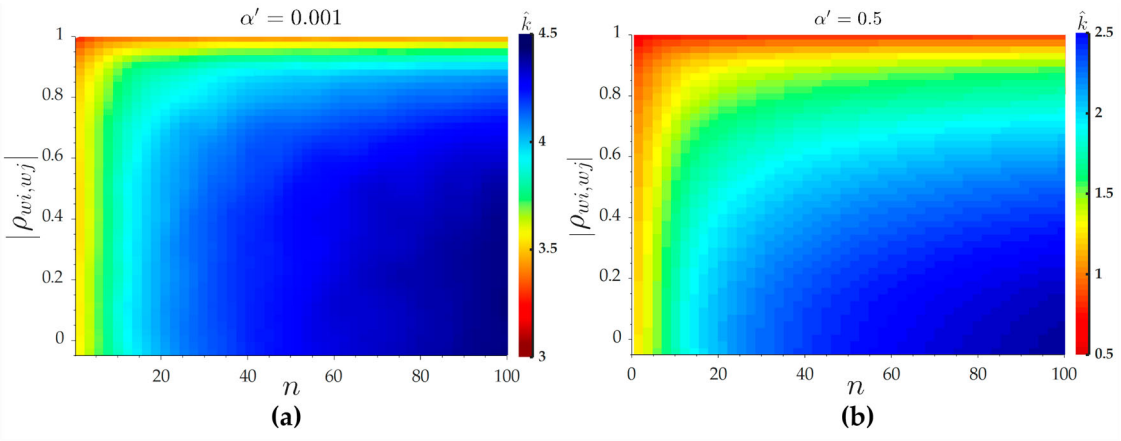
$$\uparrow |\rho_{w_i,w_j}| \Rightarrow \downarrow \hat{k} \text{ and } \uparrow n \Rightarrow \uparrow \hat{k} \text{ or } \downarrow |\rho_{w_i,w_j}| \Rightarrow \uparrow \hat{k} \text{ and } \downarrow n \Rightarrow \downarrow \hat{k} \tag{21}$$

$$\downarrow |\rho_{w_i,w_j}| \text{ and } \uparrow n \Rightarrow \uparrow \hat{k} \text{ or } \uparrow |\rho_{w_i,w_j}| \text{ and } \downarrow n \Rightarrow \downarrow \hat{k} \tag{22}$$

From Fig. 3, we also observe that the critical values do not depend on of the size of samples n when $|\rho_{w_i,w_j}| > 0.99$. This fact was also presented by Yang (2013). We also note that for $n > 30$ and for a given $|\rho_{w_i,w_j}|$, the critical values are virtually constant. These dataset are used to build a neural network for computation the critical values for any correlation matrix $\mathcal{R}_w$, as described below.

3.2. Development of a neural network model to predict critical values

From the critical values generated in the previous section, we build a neural network. The input variables consist of the overall significance level ($\alpha' = 0.001$, $\alpha' = 0.01$, $\alpha' = 0.05$, $\alpha' = 0.1$ and $\alpha' = 0.5$), number of observations ($n = 5$ to $n = 100$ by increment of 5) and correlation between the tests ($|\rho_{w_i,w_j}| = 0.0$ to $|\rho_{w_i,w_j}| = 1.0$, by increment of 0.1, by taking also into account the correlation $|\rho_{w_i,w_j}| = 0.999$); and the outputs (targets) are the corresponding critical values computed from the Monte Carlo ($\hat{k}$).



3 Critical values for Data-Snooping computed by Monte Carlo method for $\alpha' = 0.001$ (a) and $\alpha' = 0.5$ (b)

We randomly subdivided the dataset into 80% for network training and 20% for validation purposes. The inputs/targets values were normalised according to the following expression:

$$y_{norm} = 2 \times \left(\frac{y - y_{\min}}{y_{\max} - y_{\min}} \right) - 1 \tag{23}$$

being y_{norm} the inputs/targets mapped to interval $[-1,1]$, y the original inputs/targets data, $y_{\min}$ and $y_{\max}$ the minimum and maximum values computed for the original inputs/targets, respectively. In this case, the normalised function gives input/targets values between $[-1,1]$.

The log-sigmoid transfer function (activation function) was applied in the intermediary layers (hidden layers) in order to compute a layer's output from its net input. The logistic function adopts values between 0 and 1, as follows:

$$f(v_i) = \frac{1}{1 + e^{-v_i}} \tag{24}$$

where v is a linear combination of the inputs, weights and bias for a given neuron 'i', i.e.:

$$v_i = p_{1i} \times I_1 + p_{2i} \times I_2 + \ldots + p_{ni} \times I_n + \theta_i \tag{25}$$

with p being the weights parameters of the network, I the input data and θ the bias parameter for 1, …, n input data.

The training set was used to adjust the weights p and biases θ of the neural network to produce the desired outcome, i.e. the critical values ($\hat{k}$). During the learning stage, it is customary and convenient to minimise the mean squared error (MSE) between the targets ($t^{(i)}$) and estimated output ($O^{(i)}$) of an ANN, as follows:

$$\chi^2 = \sum_{i=1}^{m} (t^{(i)} - O^{(i)})^2 \rightarrow \min \tag{26}$$

The parameters to be estimated are the weights (p) and biases (θ). Note that $O^{(i)} = f(\boldsymbol{y}; \boldsymbol{x})$, where $\boldsymbol{y}$ is the input data and $\boldsymbol{x}$ are the unknown parameters (i.e. p and θ).

Generally, the transfer functions involved in an ANN are non-linear, as can be seen in the (24). Consequently, a nonlinear least squares method which iteratively minimises the sum of the squares of the residuals in (26) is required. Here, we use the Levenberg–Marquardt algorithm (LM), also known as the damped least-squares (DLS) method. The LM combines two minimisation methods: the gradient descent method and the Gauss-Newton method (Levenberg 1944). In the gradient descent method, the sum of the squared residuals (χ^2) is reduced by updating the parameters in the steepest-descent direction. In the Gauss-Newton method, the χ^2 is reduced by assuming the least squares function is locally quadratic, and finding the minimum of the quadratic form. The LM method acts more like a gradient-descent method when the parameters are far from their optimal value, and acts more like the Gauss-Newton method when the parameters are close to their optimal value (Gavin 2020).

Levenberg's contribution provided a 'damped version' of the minimisation problem for (27), as follows:

$$\left[\boldsymbol{J}^T\boldsymbol{J} + \lambda \boldsymbol{I}\right]\boldsymbol{h}_{lm} = \boldsymbol{J}^T\left[\boldsymbol{y} - f(\hat{\boldsymbol{x}})\right] \tag{27}$$

where $\boldsymbol{J}$ being the Jacobian matrix which contains first derivatives of the network residuals with respect to the weights and biases; $\boldsymbol{I}$ is the identity matrix, giving as the increment $\boldsymbol{h}_{lm}$ to the estimated parameter vector $\hat{\boldsymbol{x}}$.

The non-negative damping factor (λ) in (27) is adjusted at each iteration. If reduction of χ^2 is rapid, a smaller value can be used, bringing the algorithm closer to the Gauss–Newton algorithm, whereas if an iteration gives insufficient reduction in the residual, λ can be increased, giving a step closer to the gradient-descent direction. Note that the gradient of χ^2 with respect to $\hat{\boldsymbol{x}}$ is $\boldsymbol{g} = -2[\boldsymbol{J}^T(\boldsymbol{y} - f(\hat{\boldsymbol{x}})]^T$. This means that for large values of damping factor λ, the step will be taken approximately in the direction opposite to the gradient. In other words, if any iteration occurs to result in a worse approximation, i.e. $\chi^2(\boldsymbol{x} + \boldsymbol{h}_{lm}) > \chi^2(\boldsymbol{x})$, then λ is increased. On the other hand, as the solution improves, λ is decreased and the solution tends to go to the local minimum and, therefore, LM approaches to the Gauss-Newton method (Madsen *et al.* 2004; Marquardt 1963; Gavin 2020).

The performance of the model was evaluated on the validation set at the end of each iteration (epoch). The aim of the validation set was to improve the generalisation of the LM. When the network starts to over-fit the data, the performance on the validation set typically begins to decrease. When the validation error increases for a specified number of epochs the training is stopped, and the weights and biases at the minimum of the validation error are returned. This method for avoiding over-fitting

is called early stopping. The performance function is given by the mean squared error (mse), as follows:

$$mse = \sum_{i=1}^{m} \frac{(t^{(i)} - O^{(i)})^2}{m} \tag{28}$$

Training was configured to stop when any of these conditions were met:

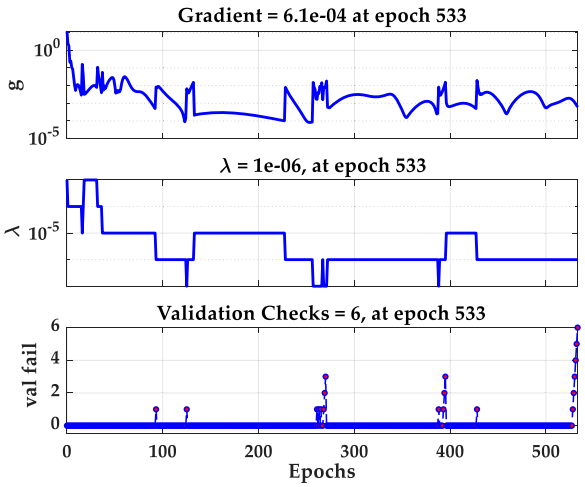
- The maximum number of epochs (repetitions/iteration) of 1,000 is reached.
- The performance gradient falls below 10^{-7}.
- $\lambda > \lambda_{\max} = 10^{10}$.
- The network performance on the validation samples fails to improve or remains the same for 6 epochs.

We used the command-line functionality of the Matlab's Neural Network toolbox (R2019b) for training and validation the ANN. Although we have configured a maximum of 1,000 epochs in the training stage, 527 epochs were sufficient in the validation stage, with $mse = 3.3 \times 10^{-4}$ (Figs. 4 and 5). The constructed ANN required two hidden layers, with each layer composed by four neurons (Fig. 6).

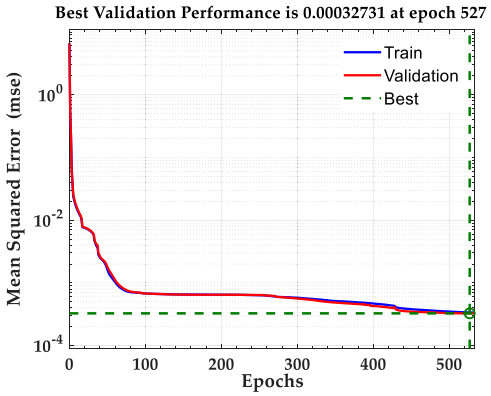
4. Experiments and results

The experiments were performed by considering two scenarios: first, we evaluate the critical values based on the Supervised Back-Propagation Neural Network model (Section 3); next, we apply the method from Rofatto (2020b) in order to analyse the success rate of outlier identification based on the critical values computed from the that proposed method. In latter case, therefore, we assume that the alternative hypotheses given in (5) is true.

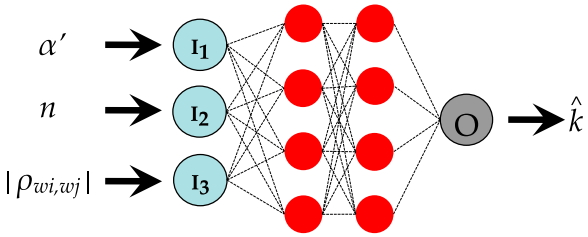
We considered ten geodetic networks for the experiments Fig. 7), namely: five levelling networks (A, B, C, D and E), two horizontal networks (F and G) and three GNSS (Global Navigation Satellite System) baseline networks by static relative positioning method (H, I and J). Table 1 summarises the characterisation of each geodetic network, listing the mean, maximum (max), minimum (min), and standard-deviation (*sd*) of the absolute correlations between all pairs of *w*-test statistics, largest eigenvalue of the correlation matrix in (14), number of observations (n), number of unknown parameters (u), and the mean redundancy ($\hat{r}$) in this order, respectively. The computation of the mean, max, min and *sd* only considered the absolute values of the off-diagonal elements related to $\mathcal{R}_w$ in (14), i.e. $|\rho_{w_i,w_j}|, \forall (i \neq j)$.



4 Performance of gradient (g) and damping factor (λ) as well as the number of validation fails



5 Best Validation Performance



6 Topology of the feed-forward neural network with Levenberg-Marquardt back-propagation algorithm for prediction of critical values

One of the SBPNN input data is the correlation coefficient between two *w*-test statistics, as can be seen in Fig. 6. However, the geodetic networks can be interpreted as a multivariate adjustment problem. Hence, we have not only one single correlation between two *w*-test statistics, but multiple correlations, being represented by the correlation matrix $\mathcal{R}_w$ in (14). In that case, we need to define a criterion to define which would be the most representative correlation of *w*-test statistics of the geodetic network for entry into SBPNN. The SBPNN-based critical values were obtained according to the following expressions:

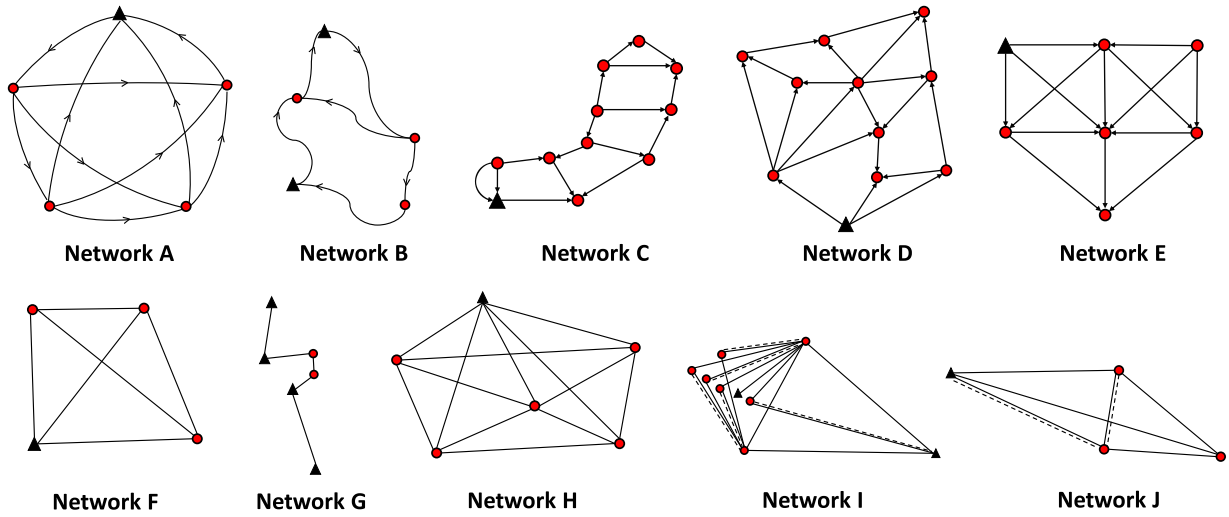
$$k_{SBPNN(mean)} = k_{SBPNN(mean|\rho_{w_i,w_j}|)} \tag{29}$$

$$k_{SBPNN(\max/mean)} = \frac{k_{SBPNN(\max|\rho_{w_i,w_j}|)} + k_{SBPNN(mean|\rho_{w_i,w_j}|)}}{2} \tag{30}$$

$$k_{SBPNN(\max/\min)} = \frac{k_{SBPNN(\max|\rho_{w_i,w_j}|)} + k_{SBPNN(\min|\rho_{w_i,w_j}|)}}{2} \tag{31}$$

$$k_{SBPNN(max_{eig})} = k_{SBPNN(pc)} \tag{32}$$

$$pc = \sqrt{\frac{\max_{eig}(\mathcal{R}_w)}{n}} \tag{33}$$



7 Geodetic Networks: observations are displayed in solid black line; repeated observations in dash line; the black triangles represent the control points, whereas the circles filled in red colour are points of unknown coordinates

Table 1 Geodetic networks characterisation

*Network	$mean\|\rho_{\boldsymbol{w}_i},\boldsymbol{w}_j\|$	$\max\|\rho_{\boldsymbol{w}_i},\boldsymbol{w}_j\|$	$\min\|\rho_{\boldsymbol{w}_i},\boldsymbol{w}_j\|$	[11] $sd\|\rho_{\boldsymbol{w}_i},\boldsymbol{w}_j\|$	[12] $\max_{eig}(\mathcal{R}_{\boldsymbol{w}})$	[13] n	[14] u	[15] $\hat{r}$
[1]A	0.2364	0.4146	0.0223	0.1415	1.9268	10	4	0.6000
[2]B	0.7053	1.0000	0.3643	0.2637	4.5937	6	3	0.5000
[3]C	0.2064	1.0000	0.0003	0.2763	3.4570	16	10	0.3750
[4]D	0.1682	0.7346	0.0006	0.1664	2.5886	20	10	0.5000
[5]E	0.1907	0.5460	0.0039	0.1489	1.9670	14	6	0.5714
[6]F	0.1150	0.3732	0.0000	0.0867	1.6034	19	6	0.6842
[7]G	0.3560	1.0000	0.0000	0.4491	3.9998	7	4	0.4286
[8]H	0.1271	0.6394	0.0001	0.1271	3.1219	39	15	0.6154
[9]I	0.0735	0.8818	0.0000	0.1428	4.2940	57	21	0.6316
[10]J	0.1612	0.7257	0.0008	0.1576	3.1590	24	9	0.6250

[1]Rofatto *et al.* (2018),[2]Knight *et al.* (2010),[3]Gemael (1994),[4]Suraci and de Oliveira (2010)[5,6]Ghilani (2017),[7,9]unpublished data,[8]Rofatto *et al.* (2017),[10]Bonimani (2021).[11] *sd*: standard-deviation.[12] $\max_{eig}(\mathcal{R}_{\boldsymbol{w}})$: maximum eigenvalue of the correlation matrix $\mathcal{R}_{\boldsymbol{w}}$ in (14).[13] n: number of observations.[14] u: number of unknown parameters.[15] $\hat{r}$: average redundancy = $\frac{n-u}{u}$.*The both matrices $\boldsymbol{A}$ and $\boldsymbol{Q_e}$ of each network are available at https://data.mendeley.com/datasets/77sfpx9b74/3.

where $k_{SBPNN(.)}$ is the overall neural-network-based critical value. The subscript within the parenthesis corresponds to each criterion, as follows:

- ($\max/mean$) corresponds to the mean of the critical values computed by the SBPNN-based critical value for the maximum ($\max|\rho_{\boldsymbol{w}_i},\boldsymbol{w}_j|$) and mean ($mean|\rho_{\boldsymbol{w}_i},\boldsymbol{w}_j|$) absolute values of the correlation between the w-test statistics;
- ($\max/\min$) the mean from the maximum ($\max|\rho_{\boldsymbol{w}_i},\boldsymbol{w}_j|$) and minimum ($\min|\rho_{\boldsymbol{w}_i},\boldsymbol{w}_j|$);
- ($mean$) corresponds to the the SBPNN-based critical value for the mean ($mean|\rho_{\boldsymbol{w}_i},\boldsymbol{w}_j|$); and
- ($\max_{eig}$) the SBPNN-based critical value obtained from the square root of the ration between the largest eigenvalue of the correlation matrix $\mathcal{R}_w$ (signs of correlations are considered) in (14) and number of observations n. The largest eigenvalue represents the maximum amount of information of the $\mathcal{R}_w$ and it can be easily obtained using Matlab's 'eigs' command. In the context of principal component analysis (PCA), the maximum eigenvalue represents the largest possible variance.

Obviously, the expressions above are computed for a given significance level (α') and number of observations (n), as can be seen in Fig. 6. In the next section, we evaluate each of these criteria. The goal is to investigate the extent to which the critical values computed by the Supervised Back-Propagation Neural Network (k_{SBPNN}) deviate from the critical values computed by Monte Carlo method.

4.1. Evaluation of critical values based on the supervised back-propagation neural network

The evaluation consisted of comparing the SBPNN-based critical values obtained for each of the criteria ($mean$), ($\max/mean$), ($\max/\min$) and ($\max_{eig}$) with respect to the Monte Carlo method. Monte Carlo-based critical values were computed from $m = 5{,}000{,}000$ sequence of random vectors, as described in Section 2. This number of simulations provides a numerical precision better than 0.1%, as demonstrated by Rofatto (2020a). Since Šidák approximation does not take into account the correlation of the w-test statistics, it can be used as a bound of the error of the critical values. Thus, the critical values computed from the Šidák approximation (k_{sid}) in (3) were also considered here.

At first, the type I error rates considered ranged from $\alpha' = 0.001$ to $\alpha' = 0.5$, with increment of 0.001, which

totalled 500 critical values for each method and for each geodetic network. Then, the relative errors (denoted by ε) were computed with respect to the Monte Carlo-based critical values. The mean of the relative error for each geodetic network (shown as blue bars) with the maximum and minimum values (shown as black lines) are displayed in Fig. 8.

In general, it is noted that the criterion based on the maximum eigenvalue ($\max_{eig}$) given by expression (33) is the one that best closest to the Monte Carlo. This means that the criterion ($\max_{eig}$) is best suited to describe the overall correlation of geodetic networks, and therefore serves as an input criterion for correlation between test statistics in the neural network. This criterion has a strong relationship with the associations given in (21) and (22). By extracting the maximum eigenvalue from the correlation matrix $\mathcal{R}_{w}$ in (14), we are capturing the most important and representative information of it. However, the critical value also depends on the number of observations. As previously shown, the critical value increases as the number of observations increases, but it decreases as the correlation increases. Therefore, the maximum eigenvalue is penalised by dividing the number of observations. Eigenvalues are given in quadratic values. For this reason, the criterion in (33) is computed by the square root of the ratio between the maximum eigenvalue and number of observations.

The results displayed in Fig. 8 are based on some critical values computed from type I error rates used in SBPNN training. To ensure a completely unbiased analysis, we also highlight the relative error of critical values from the type I error rate that were not part of the SBPNN development. In that case, we have considered the type I decision error rates of $\alpha' = 0.005$, $\alpha' = 0.025$, $\alpha' = 0.075$, $\alpha' = 0.15$ and $\alpha' = 0.35$ (Fig. 9).

Figure 9 shows that the higher the significance level (α'), the larger the error when applying the Šidák correction k_{sid}. This is due to the fact that the correlation between the w-test statistics is neglected by the Šidák correction and, therefore, it serves as an upper bound for analysing the error of the SBPNN method.

In the case of the Network (A) and for significance levels lower than 10% ($\alpha' < 0.1$), k_{SBPNN} for all criteria and Šidák correction (k_{sid}) provided critical values practically equal to those obtained by the Monte Carlo method, with error less than 3% (Fig. 9). However, the increase in the significance level (α') has enlarged the error in the case of Šidák correction, while the error for all criteria in the SPBNN method remained less than 3%. Similar results can be found for networks (E) and (F).

For the case of Network (B), the errors related to the SBPNN-based critical values for all criteria were less than Šidák correction (k_{sid}). In that case, SBPNN for $k_{SBPNN(\max/mean)}$, $k_{SBPNN(\max/\min)}$ and $k_{SBPNN(\max_{eig})}$ had errors less than 6.5%, while k_{sid} reached 39% ($\alpha' = 0.35$). Among the criteria considered, the one based on the mean value $k_{SBPNN(mean)}$ presented the highest error (22% for $\alpha' = 0.35$). Similar results can be found for network (G).

For the case of Network (C), SBPNN for $k_{SBPNN(\max/mean)}$ and $k_{SBPNN(\max/\min)}$ and for $\alpha' < 0.15$ had larger errors than Šidák correction (k_{sid}), with maximum error of the order of 16% ($\alpha' = 0.35$) and 14% ($\alpha' = 0.35$), respectively. On the other hand, $k_{SBPNN(max_{eig})}$ had error less than 7.5%. A similar relationship can be found for network (I), but both $k_{SBPNN(\max/mean)}$ and $k_{SBPNN(\max/\min)}$ with errors larger than the k_{sid} for all significance levels α'. In latter case, $k_{SBPNN(max_{eig})}$ had error less than 1.5%. Actually, there is the same behaviour for networks (D), (H) and (J), but the magnitude of the errors for both $k_{SBPNN(\max/mean)}$ and $k_{SBPNN(\max/\min)}$ are much smaller, with the maximum of the order of 5%.

In general, therefore, the criterion '$\max_{eig}$' based on the expression in (33) provides a better balance between the correlation and the number of observations for the computation of the critical values.

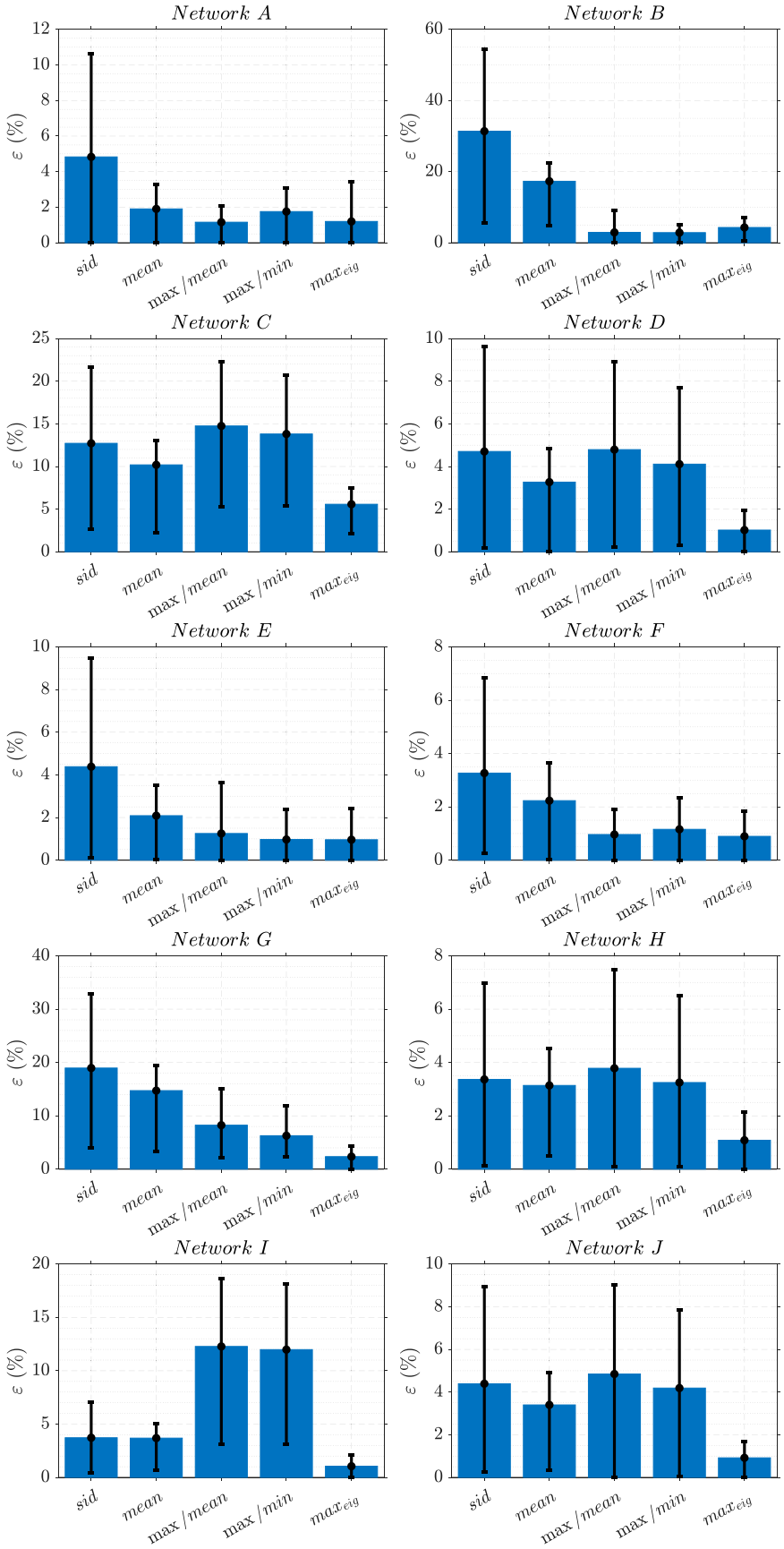
In order to have the true false positive rates from the SBPNN-based critical values for $\max_{eig}$, we also run a 200,000 Monte Carlo simulation in the absence of outlier and computed the number of times that an observation was identified as an outlier, when in fact there is none. This procedure has been applied within the context of reliability analysis (Klein 2021; Suraci 2021).

This same procedure was also applied to the critical value coming from Monte Carlo (k_{MC}) and Šidák correction (k_{sid}). In that case, we restrict ourselves to Network (A) and (B) without loss of generality. The Network (A) characterised by presenting low correlation between$w$-test statistics and good redundancy, whereas the Network (B) has low redundancy and high correlation (Fig. 10).

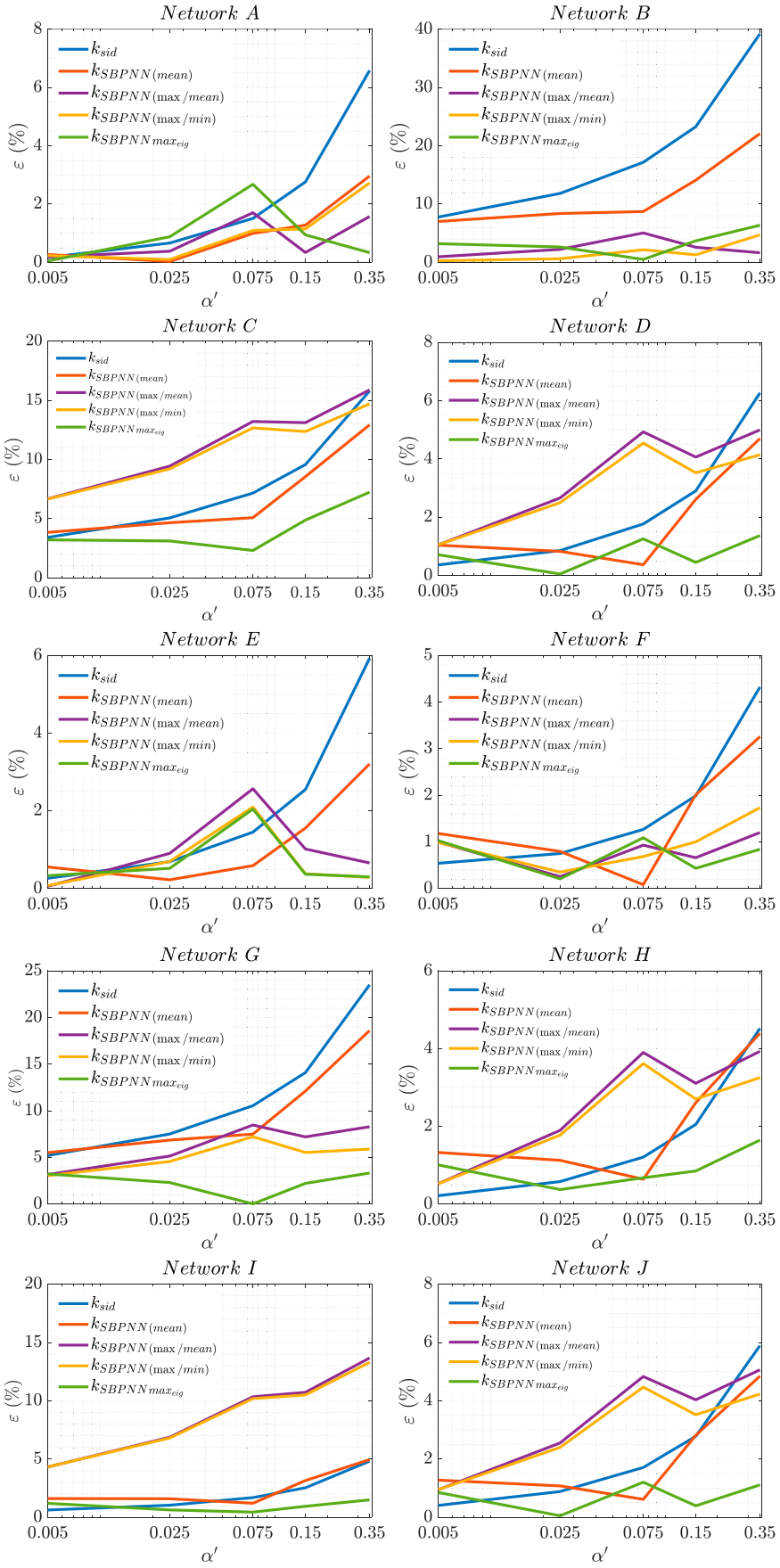
The critical values found are displayed in Table 2, with the true false positive rates in parentheses. It is important to highlight that 200,000 Monte Carlo simulation provide numerical accuracy better than 0.1% Rofatto (2020a). From Table 2, we observe that the critical values based on SBPNN by the $\max_{eig}$ criterion are always smaller than the k_{sid}.

The true false positive rates should be those provided by the user, i.e. in this case, $\alpha' = 0.005(0.5\%)$, $\alpha' = 0.025(2.5\%)$, $\alpha' = 0.075(7.5\%)$, $\alpha' = 0.15(15\%)$ and $\alpha' = 0.35(35\%)$. However, we note that the higher the FWER (α'), the larger the deviation of the true false positive when Šidák correction (k_{sid}) is applied (Table 2). Since Šidák correction does not take into account the dependence of the w-test statistics, the results suggest that it should only be used for systems with good local redundancy (on average, $r_i > 0.5$), low correlation between w-test statistics (on average, $|\rho_{\boldsymbol{w}_i,\boldsymbol{w}_j}| \leq 0.5$) and for low rates of individual false positives $\alpha_0 < 0.01(1\%)$, as can be also seen in Lehmann (2012), Imparato *et al.* (2019) and Rofatto (2020c). On the other hand, we observe that SBPNN is able to capture the correlation of the w-test statistics, and therefore it can be considered as a good approximation for the control of the false positive rates (α'), with the advantage of being computationally faster than the Monte Carlo method.

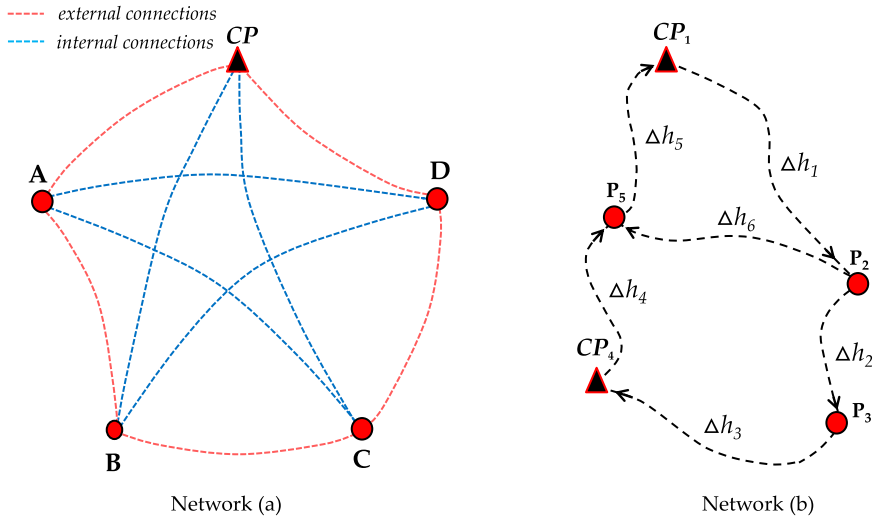
The computational cost related to the Monte Carlo method is mainly due to the generation of the sequence of the random vectors in (17) and the sorting of the values in (19). Note that $\mathcal{R}_{w}$ must be computed only once and also remember that the Monte Carlo-based critical value in (20) can be evaluated for a sequence of values α' in parallel. In our analyses, Intel Core i5-10300H @2.5GHz and Matlab (R2019b) were used for the experiments. The average time elapsed of the Monte Carlo method was in the order of 6 seconds for the largest



8 Relative error (ε) of the critical values (mean shown as blue bars and the both maximum and minimum values shown as black lines) for Šidák correction (*sid*) and SBPNN model based on each criterion of the correlation of *w*-test statistics (*mean*, max /*mean*, max /min and $\max_{eig}$) with respect to the Monte Carlo method



9 Relative error (ε) of the critical values for Šidák correction (*sid*) and SBPNN model based on each criterion of the correlation of *w*-test statistics (*mean*, max /*mean*, max /min and max$_{eig}$) with respect to the Monte Carlo method for $\alpha' = 0.005$, $\alpha' = 0.025$, $\alpha' = 0.075$, $\alpha' = 0.15$ and $\alpha' = 0.35$



10 Levelling geodetic networks: (a) Levelling network from Rofatto *et al.* (2018) with a low correlation between w-test statistics (Network A); (b) Levelling network from Knight *et al.* (2010) with a high correlation between w-test statistics (Network B)

Table 2 Critical values for levelling Network A and B

Network A				Network B			
α'	k_{sid}	k_{MC}	$k_{SBPNN(\max_{eig})}$	α'	k_{sid}	k_{MC}	$k_{SBPNN(\max_{eig})}$
0.005	3.48(0.5%)	3.47(0.5%)	3.48(0.5%)	0.005	3.34(0.2%)	3.09(0.5%)	3.2(0.3%)
0.025	3.02(2.3%)	3.00(2.4%)	2.98(2.7%)	0.025	2.87(1.1%)	2.56(2.5%)	2.63(2.1%)
0.075	2.67(6.6%)	2.63(7.4%)	2.56(9.1%)	0.075	2.50(2.9%)	2.13(7.4%)	2.14(7.2%)
0.15	2.43(12.5%)	2.36(15%)	2.33(15.8%)	0.15	2.24(5.7%)	1.81(15%)	1.87(13.1%)
0.35	2.11(25.8%)	1.95(35%)	1.94(35.4%)	0.35	1.89(12.5%)	1.34(35.1%)	1.42(30.4%)

network ($n = 57$ observations) and $m = 5,000,000$ Monte Carlo experiments. For this case, but for $m =$ 200,000, this computation takes less than a second. However, if we desired to effectively control the type I error of the Data Snooping procedure, the critical value should be updated every time an observation had been identified and removed from the dataset. Consequently, the correlation matrix $\mathcal{R}_w$ would be constantly updated, which would increase the computational cost of the Monte Carlo method. On the other hand, the neural network model provided here works as a closed-form expression and therefore their response is practically instantaneous, just as in the case of Šidák correction.

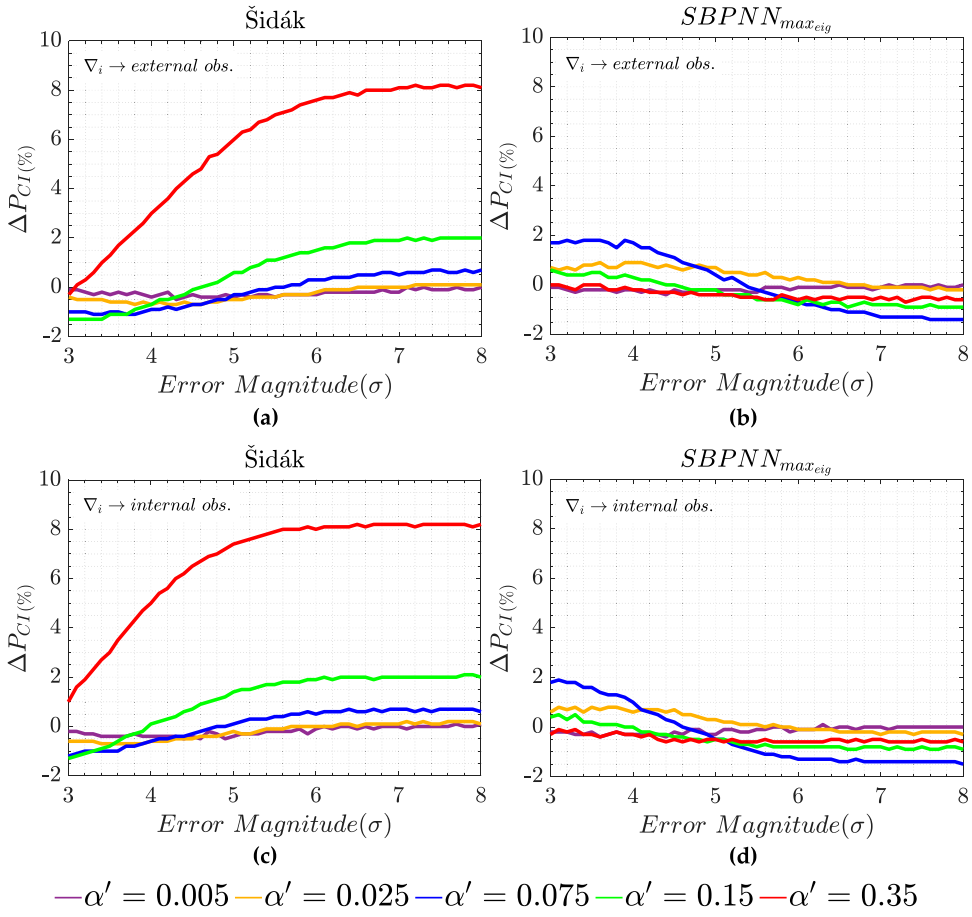
4.2. Evaluation of the success rate of data snooping based on the supervised back-propagation neural network-based critical values

Here, we also evaluate the success rate of the outlier identification based on the critical values computed from $k_{SBPNN(\max_{eig})}$. We used the procedure provided by Rofatto (2020b) to compute the probability of correct identification (denoted by $\mathcal{P}_{CI}$) when there is an error in the dataset. The probabilities of correct identification from the critical values of Šidák correction and SBPNN method were compared to those probabilities ($\mathcal{P}_{CI}$) based on the reference critical values from Monte Carlo for each error magnitude and false positive rates α' (Table 2). We arbitrarily defined the error magnitude from $|3\sigma|$ to $|8\sigma|$ for Network A and $|1\sigma|$ to $|12\sigma|$ for Network B, being σ the standard-deviation of each observation. The differences with respect to Monte Carlo ($\Delta\mathcal{P}_{CI}$) are displayed in Figs. 11 and 12 for Network A and Network B, respectively.

From Figs. 11 and 12, we observe that there are no significant differences between SBPNN and Monte Carlo methods. On the other hand, the higher the significance level, the larger the difference between Šidák correction and Monte Carlo. This difference also increases with the size of the error magnitude and it is also more expressive in the case of the Network B.

For systems with a high correlation between w-test statistics, as in the case of the Network B, the critical values based on SBPNN and Monte Carlo are smaller than Šidák correction (Table 2). Consequently, the type III error rate and overidentifications for Šidák correction are smaller than both Monte Carlo and SBPNN [details about the type III error rate and overidentification classes can be seen in Rofatto (2020b)], and therefore Šidák correction provides slightly smaller MIBs (Minimal Identifiable Bias) than Monte Carlo and SBPNN (Tables 3 and 4), although, as already discussed, it does not effectively control the rate of individual type I decision error.

It is also noted that all MIBs for $\alpha' = 0.35$ were lower than 80% for the case of Network A, as can be seen the success rate highlighted in parentheses in Table 3.



11 Differences of the probability of correct outlier identification and elimination ($\Delta\mathcal{P}_{CI}$) of the data snooping based on the critical values from Šidák correction (k_{sid}) and SBPNN method ($k_{SBPNN\max_{eig}}$) with respect do Monte Carlo-based critical values (k_{MC}) and for Network A

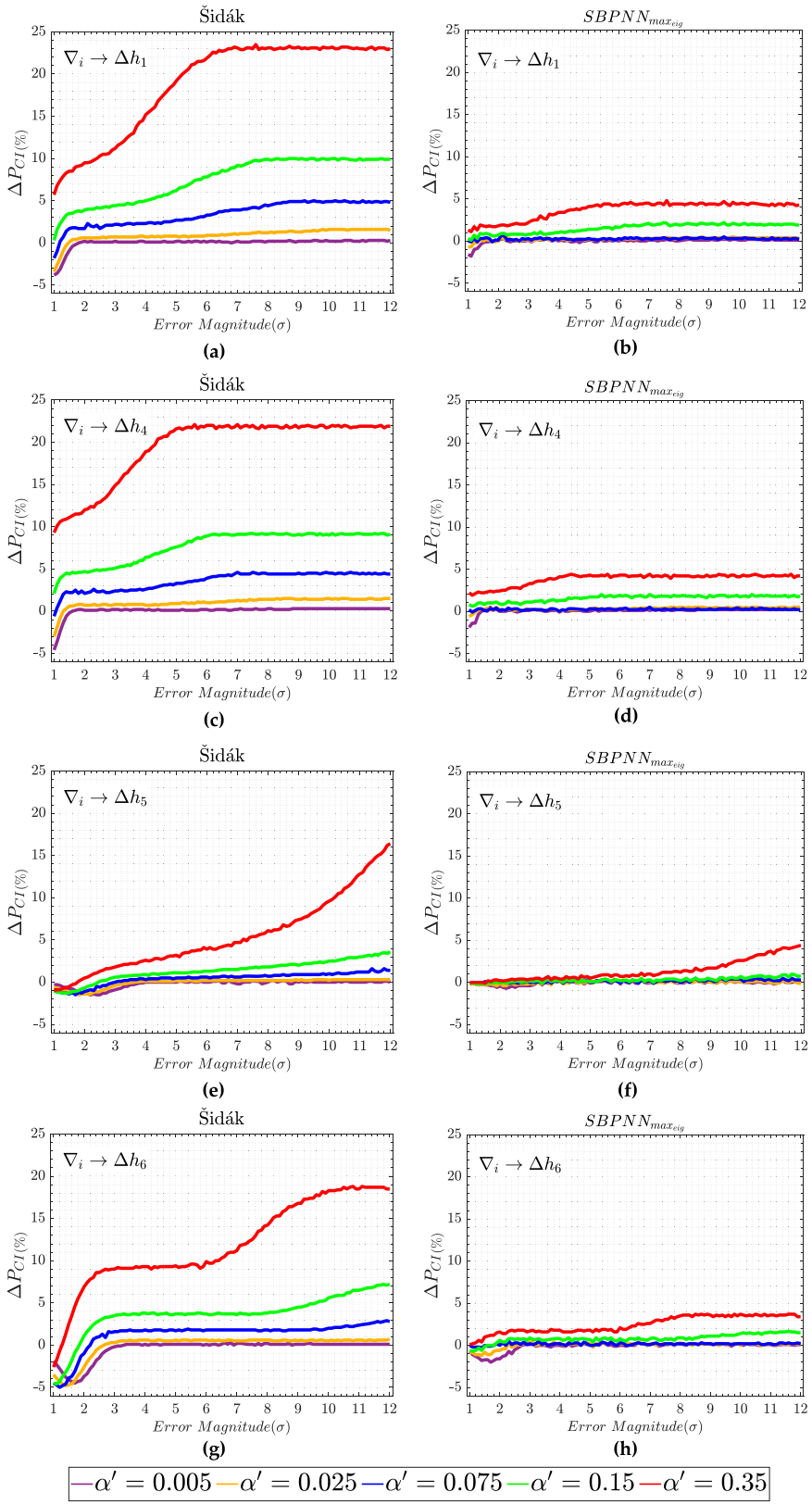
However, in the case of Network B, for that same $\alpha' = 0.35$, only MIBs based on the Šidák correction reached a success rate of 80% (success rate also highlighted in parentheses in Table 4).

In general, the critical values based on the SBPNN method were those that came closest to the Monte Carlo method in terms of outlier identification by the data snooping procedure (Figs. 11 and 12), with a maximum difference of 1.9% for the Network A and 4.8% for the Network B. By considering the sensitive indicator for outlier identification – MIB (Minimal Identifiable Bias) - for a success rate larger than 80%, SBPNN also provides a good approximation to the Monte Carlo (Tables 3 and 4).

5. Final remarks

Here, we have introduced a novel developed method for user-controlling the significance level (also called family-wise error rate or/ type I decision error rate) of the data snooping procedure through the Supervised Back Propagation Neural Network model (SBPNN). A Monte Carlo method was used to compute critical values for a fixed number of observations with prefixed correlation between the test statistics. Then, those critical values were used to train the SBPNN model. Following this, SPBNN-based critical values were tested into a geodetic networks adjustment problem. Geodetic network adjustment is a type of multivariate problem, in which there are multiple correlations between the outlier test statistics. On the other hand, SBPNN model is based only one single correlation between one pair of the outlier test statistics. To overcome that issue, we propose and test some criteria to extract only one single correlation. We find that criterion based on the ratio between the maximum eigenvalue of the correlation matrix and the number of observations is the most suitable to describe the overall correlation of geodetic networks, and therefore serves as the input for correlation between test statistics in the SPBNN model. To find the other approach in some practical sense, more experiences must be gained, including recent artificial intelligence methods.

In order to guarantee an unbiased evaluation, false positive rates other than the SPBNN training sample were also selected. Considering the critical values obtained by Monte Carlo as references, SBPNN method presented a mean relative error of 2% ($\pm$2%) and a maximum of 7%, whereas the Šidák correction about 9% ($\pm$9%) and maximum of 54%. Therefore, we observe that SBPNN is able to capture the dependence of the test statistics, and therefore it can be considered as good approximations for the control of the false positive rates (α'). Since Šidák correction does not take into account the correlation between the test statistics, we reinforce that its use for controlling the type I error rate



12 Differences of the probability of correct outlier identification and elimination ($\Delta\mathcal{P}_{CI}$) of the data snooping based on the critical values from Šidák correction (k_{sid}) and SBPNN method ($k_{SBPNN\,max_{eig}}$) with respect do Monte Carlo-based critical values (k_{MC}) and for Network B

should only be used for systems with high redundancy ($r_i > 0.5$), low correlation between w-test statistics ($|\rho_{w_i,w_j}| \leq 0.5$) and for low rates of individual false positives $\alpha_0 < 0.01(1\%)$.

We also evaluate the success rate of the data snooping in terms of outlier identification based on critical values computed from SBPNN. In that case, the maximum difference was 4.8% with respect to Monte Carlo-based

Table 3 MIB for Network A by considering a success rate $\geq$ 80%

	External Obs.				Internal Obs.		
α'	k_{sid}	k_{MC}	$k_{SBPNN(max_{eig})}$	α'	k_{sid}	k_{MC}	$k_{SBPNN(\max_{eig})}$
0.005	6.1	6.1	6.1	0.005	5.3	5.3	5.3
0.025	5.5	5.5	5.5	0.025	4.8	4.8	4.8
0.075	5.3	5.3	5.3	0.075	4.6	4.6	4.6
0.15	5.3	5.4	5.4	0.15	4.6	4.7	4.7
0.35	6.9 (77%)	6.3 (69%)	7.1 (69%)	0.35	5.5 (77%)	5.3 (69%)	6.3 (69%)

Table 4 MIB for Network B by considering a success rate $\geq$ 80%

	Δh_1			Δh_4			Δh_5			Δh_6		
α'	k_{sid}	k_{MC}	$k_{SBPNN(max_{eig})}$	k_{sid}	k_{MC}	$k_{SBPNN(\max_{eig})}$	k_{sid}	k_{MC}	$k_{SBPNN(\max_{eig})}$	k_{sid}	k_{MC}	$k_{SBPNN(\max_{eig})}$
0.005	3.7	3.7	3.7	2.5	2.6	2.6	11.2	11.3	11.2	5.7	5.7	5.6
0.025	3.7	3.8	3.8	2.6	2.7	2.6	11.3	11.4	11.3	5.7	5.8	5.7
0.075	3.8	4.1	4.1	2.7	2.9	2.9	11.4	11.7	11.5	5.8	6.2	6.1
0.15	4.0	4.8	4.5	2.8	3.4	3.2	11.5	12.6	12.2	6.1	6.8	6.6
0.35	4.5	4.7 (63%)	5.1 (68%)	3.2	3.8 (66%)	3.8 (70%)	12.4	11.3(63%)	11.3(67%)	6.6	7.4 (71%)	7.8(75%)

critical value. The results prove that SBPNN model can be used to compute the critical values and, therefore, it is no longer necessary to run the Monte Carlo-based critical value every time the quality control is performed through data snooping.

Here, random errors were treated as truly normal. This work may be extended in the future to the case where they are Laplacians, Cauchy, etc.

Finally, we provide a MATLAB's flexible network object type (called *SBPNN.mat*) that allows anyone to obtain the desired critical value with good control of type I error for the case where the random errors follow the normal distribution. We also provide a MATLAB's function (called 'kNN.m') which uses *SBPNN.mat* and the criterion $\max_{eig}$ to compute a desired critical value for the case where Gauss-Markov model is in play. Those interested in dataset of this work, *SBPNN.mat* and 'kNN.m' please send their request to the authors by e-mail or access https://data.mendeley.com/datasets/77sfpx9b74/3.

Funding

This research received financial support from the CNPq—Conselho Nacional de Desenvolvimento Científico e Tecnológico—Brasil (proc. nº 103587/2019-5).

Notes on contributors

Vinicius Francisco Rofatto received his PhD degree in Remote Sensing from the Federal University of Rio Grande do Sul, Brazil, in 2020. Currently, he is held lecturer and researcher positions at Federal University of Uberlandia, Monte Carmelo city, Minas Gerais state, Brazil. In 2018, he was appointed the Grand Master of Cartographers Order by the Brazilian Cartography Society. In 2021, he received the Honorable Mention of the CAPES Thesis 2021 Award in the field of Geosciences. His research interests include Estimation Theory, Quality Control, Computational Methods for solving linear and nonlinear geodetic problems, Monte Carlo Methods, Global Navigation Satellite Systems and Deep Learning in Computer Vision for geospatial applications.

Marcelo Tomio Matsuoka received his MSc and PhD degrees in Cartographic Sciences from the Paulista State University, Brazil, in 2003 and 2007, respectively. He has held lecturer and researcher positions (Undergraduate Surveying Engineering and Graduate Program in Agriculture and Geospatial Information) at Federal University of Uberlandia, Monte Carmelo city, Minas Gerais state, Brazil. In 2021, he received the Honorable Mention of the CAPES Thesis 2021 Award in the field of Geosciences. His research interests include Global Navigation Satellite Systems, Quality Control of Geodetic Networks and Monte Carlo and Intelligent Computational Methods Applications in Geodesy.

Ivandro Klein received his MSc and PhD degrees in Remote Sensing from the Federal University of Rio Grande do Sul state, Brazil, in 2012 and 2014, respectively. He has held lecturer and researcher positions at Federal Institute of Santa Catarina state, Florianópolis, and at the Graduate Program in Geodetic Sciences of Federal University of Paraná state, Curitiba, Brazil. In 2021, he received the Honorable Mention of the CAPES Thesis 2021 Award in the field of Geosciences. His research interests include design, adjustment and quality control of geodetic networks.

Maria Luísa Silva Bonimani received her Bachelor degree in Surveying and Cartographic Engineering from the Federal University of Uberlandia, Monte Carmelo City, Minas Gerais state, Brazil, in 2019. Currently, attending Master's Degree from the Graduate Program

in Agriculture and Geospatial Information at Federal University of Uberlandia, Monte Carmelo city, Minas Gerais state, Brazil, with emphasis on the Reliability Theory.

Bruno Póvoa Rodrigues received his MSc degree in Agriculture and Geospatial Information from the Federal University of Uberlandia, Monte Carmelo city, Minas Gerais state, Brazil, in 2021. He has developed his research in the field of artificial neural networks applied in agriculture.

Caio Cesar de Campos is a Surveying and Cartographic Engineering student at Federal University of Uberlandia, Monte Carmelo city, Minas Gerais state, Brazil. He is part of the scientific initiation program where he develops research in the area of artificial neural networks with applications in geodesy.

Mauricio Roberto Veronez received his PhD degree in Transport Engineering from the University of São Paulo, in 2004. Post-Doctorate in Geomatics from the University of São Paulo (EESC/USP, 2016). He is currently an academic member of commission 2 of FIG - International Federation of Surveyors. He has been a Research Productivity 2 Scholarship (SA: Architecture, Demography, Geography, Tourism and Urban and Regional Planning) since March 2014. Adjunct Professor II since 2005 working in the Graduate Programs in Applied Computing and Biology. Guides research in Geoinformatics, Digital Image Processing and Immersive 3D Visualization. He is currently part of the research team of the international project entitled Circum Arctic Geology for Everyone coordinated by UNIS (The University Center in Svalbard). He is the leader of the CNPq Research Group entitled: Digital Outcrop Modeling. He has experience in coordinating research, development and innovation projects in the area of Urban and Regional Planning.

Luiz Gonzaga da Silveira Jr received his master's and doctorate degrees in Electrical Engineering, with an emphasis on Computer Engineering, from UNICAMP, in 1996 and 2005, respectively. He is currently a member of ACM, ACM/SIGGRAPH, IEEE, IEEE Computer, SBC, adjunct professor of the Graduate Program in Applied Computing (PPGCA) at the University of Vale do Rio dos Sinos (UNISINOS). He is interested in the areas of visualization, computer graphics and computer vision, working mainly on the following topics: computer visualization and visual analysis for large volumes of data and information, virtual and augmented reality, graphic architectures for high-performance computing and visualisation applications in geosciences and transport.

ORCID

Vinicius Francisco Rofatto http://orcid.org/0000-0003-1453-7530
Marcelo Tomio Matsuoka http://orcid.org/0000-0002-2630-522X
Ivandro Klein http://orcid.org/0000-0003-4296-592X
Maria Luisa Silva Bonimani http://orcid.org/0000-0002-3314-3230
Bruno Póvoa Rodrigues http://orcid.org/0000-0002-5275-9049
Mauricio Roberto Veronez http://orcid.org/0000-0002-5914-3546
Luiz Gonzaga da Silveira Jr http://orcid.org/0000-0002-7661-2447

References

Anderson, T., 1984. *An introduction to multivariate statistical analysis.* 2nd ed. New York: Wiley.

Baarda, W., 1968. A testing procedure for use in geodetic networks. *Publication on geodesy, new series*, 2 (5), 1.

Bonimani, M.L.S., *et al.*, 2021. The effect of covariance between baseline components on the reliability of gnss networks: results for a highly redundancy network. *Revista brasileira de cartografia*, 73 (2), 666–684.

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. Available from: https://doi:10.1214/aoms/1177706645

Förstner, W., 1981. Reliability and discernability of extended gauss-markov models. *In*: *Proceedings of the International Symposium on Geodetic Networks and Computations*, no. 258. Deutsche Geod4tische Kommission, Reihe B. Available from: http://www.ipb.uni-bonn.de/pdfs/Forstner1981Reliability.pdf.

Gavin, H.P., 2020. *The levenberg-marquardt algorithm for nonlinear least squares curve-fitting problems.* Available from: http://people.duke.edu/ hpgavin/ce281/lm.pdf.

Gemael, C., 1994. *Introdução ao ajustamento de observações: aplicações geodésicas.* 1st ed. Curitiba, PR: Editora da UFPR.

Ghilani, C.D., 2017. *Adjustment computations: spatial data analysis.* 6th ed. Hoboken, NJ: John Wiley & Sons, Ltd.

Imparato, D., Teunissen, P., and Tiberius, C., 2019. Minimal detectable and identifiable biases for quality control. *Survey review*, 51 (367), 289–299. Available from: https://doi.org/10.1080/00396265.2018.1437947

Klein, I., *et al.*, 2021. An attempt to analyse iterative data snooping and l1-norm based on monte carlo simulation in the context of leveling networks. *Survey review*, 0 (0), 1–9.

Knight, N.L., Wang, J., and Rizos, C., 2010. Generalised measures of reliability for multiple outliers. *Journal of geodesy*, 84 (10), 625–635. Available from: https://doi.org/10.1007/s00190-010-0392-4

Lehmann, R., 2012. Improved critical values for extreme normalized and studentized residuals in Gauss–Markov models. *Journal of geodesy*, 86 (12), 1137–1146. Available from: 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., 2015. Observation error model selection by information criteria vs. normality testing. *Studia geophysica et geodaetica*, 59 (4), 489–504. Available from: https://doi.org/10.1007/s11200-015-0725-0

Lehmann, R., and Lösler, M., 2016. Multiple outlier detection: hypothesis tests versus model selection by information criteria. *Journal of surveying engineering*, 142 (4), 04016017:1–04016017:11.

Lehmann, E.L., and Romano, J.P., 2005. *Testing statistical hypotheses.* 3rd ed. New York: Springer-Verlag.

Lehmann, R., and Scheffler, T., 2011. Monte Carlo based data snooping with application to a geodetic network. *J. appl. geod.*, 5 (3-4), 123–134. Available from: https://doi.org/10.1515/JAG.2011.014

Lemeshko, B.Y., and Lemeshko, S.B., 2005. Extending the application of grubbs-type tests in rejecting anomalous measurements. *Measurement techniques*, 48 (6), 536–547. Available from: https://doi.org/10.1007/s11018-005-0179-9

Levenberg, K., 1944. A method for the solution of certain non-linear problems in least squares. *Quarterly of applied mathematics*, 2 (2), 164–168. Available from: http://www.jstor.org/stable/43633451

Lichti, D.D., Chan, T.O., and Belton, D., 2021. Linear regression with an observation distribution model. *Journal of geodesy*, 95 (2), 23.

Madsen, K., Nielsen, H.B., and Tinglef, O., 2004. *Methods for nonlinear least squares problems.* Available from: http://www2.imm.dtu.dk/pubdb/edoc/imm3215.pdf.

Marquardt, D.W., 1963. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the society for industrial and applied mathematics*, 11 (2), 431–441. Available from: https://doi.org/10.1137/0111030

Miller, R.G.J., 1981. *Simultaneous statistical inference.* 2nd ed. New York: Springer-Verlag.

Mittelhammer, R.C., Judge, G.G., and Miller, D.J., 2000. *Econometric foundations.* 1st ed. New York, NY: Cambridge University Press.

Rofatto, V.F., *et al.*, 2020a. A half-century of Baarda's concept of reliability: a review, new perspectives, and applications. *Survey review*, 52 (372), 261–277. Available from: https://doi.org/10.1080/00396265.2018.1548118

Rofatto, V.F., *et al.*, 2020b. A monte carlo-based outlier diagnosis method for sensitivity analysis. *Remote sensing*, 12 (5), 1. Available from: https://www.mdpi.com/2072-4292/12/5/860

Rofatto, V.F., *et al.*, 2020c. On the effects of hard and soft equality constraints in the iterative outlier elimination procedure. *PloS one*, 15

(8), 1–29. Available from: https://doi.org/10.1371/journal.pone.0238145

Rofatto, V., Matsuoka, M., 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 (3), 416–435. Available from: http://dx.doi.org/10.4314/sajg.v6i3.11

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

Rom, D.M., 2013. An improved hochberg procedure for multiple tests of significance. *The british journal of mathematical and statistical psychology*, 66 (1), 189–196. Available from: https://doi.org/10.1111/j.2044-8317.2012.02042.x

Sarkar, S.K., and Chang, C.K., 1997. The simes method for multiple hypothesis testing with positively dependent test statistics. *Journal of the american statistical association*, 92 (440), 1601–1608.

Selbesoglu, M.O., 2020. Prediction of tropospheric wet delay by an artificial neural network model based on meteorological and gnss data. *Engineering science and technology, an international journal*, 23 (5), 967–972. Available from: http://www.sciencedirect.com/science/article/pii/S2215098619316003

Šidák, Z., 1967. Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the american statistical association*, 62 (318), 626–633.

Simes, R.J., 1986. An improved bonferroni procedure for multiple tests of significance. *Biometrika*, 73 (3), 751–754. Available from: https://doi.org/10.1093/biomet/73.3.751

Srivani, I., Siva Vara Prasad, G., and Venkata Ratnam, D., 2019. A deep learning-based approach to forecast ionospheric delays for gps signals. *Ieee geoscience and remote sensing letters*, 16 (8), 1180–1184.

Suraci, S.S., and de Oliveira, L.C., 2010. Chebyshev norm adjustment: what to expect? case study in a leveling network. *Revista brasileira de geomática*, 7 (4), 172–185.

Suraci, S.S., *et al.*, 2021. Monte carlo-based covariance matrix of residuals and critical values in minimum l1-norm. *Mathematical problems in engineering*, 2021, 1.

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

Teunissen, P.J.G., Imparato, D., and Tiberius, C.C.J.M., 2017. Does RAIM with correct exclusion produce unbiased positions? *Sensors*, 17 (7), 1. Available from: https://www.mdpi.com/1424-8220/17/7/1508

Wright, S.P., 1992. Adjusted p-values for simultaneous inference. *Biometrics*, 48 (4), 1005–1013. Available from: http://www.jstor.org/stable/2532694

Yang, L., *et al.*, 2013. Outlier separability analysis with a multiple alternative hypotheses test. *Journal of geodesy*, 87 (6), 591–604. Available from: https://doi.org/10.1007/s00190-013-0629-0

Yang, L., *et al.*, 2017. Extension of internal reliability analysis regarding separability analysis. *Journal of surveying engineering*, 143 (3), 04017002.

Zaminpardaz, S., and Teunissen, P., 2019. DIA-datasnooping and identifiability. *Journal of geodesy*, 93 (1), 85–101. Available from: https://doi.org/10.1007/s00190-018-1141-3

Ziggah, Y.Y., *et al.*, 2016. Capability of artificial neural network for forward conversion of geodetic coordinates (ϕ, λ, h) to cartesian coordinates (x, y, z). *Mathematical geosciences*, 48 (6), 687–721. Available from: https://doi.org/10.1007/s11004-016-9638-x