Survey Review

ISSN: 0039-6265 (Print) 1752-2706 (Online) Journal homepage: http://www.tandfonline.com/loi/ysre20

An approach to identify multiple outliers based on sequential likelihood ratio tests

I. Klein, M. T. Matsuoka, M. P. Guzatto & F. G. Nievinski

To cite this article: I. Klein, M. T. Matsuoka, M. P. Guzatto & F. G. Nievinski (2017) An approach to identify multiple outliers based on sequential likelihood ratio tests, Survey Review, 49:357, 449-457, DOI: 10.1080/00396265.2016.1212970

To link to this article: http://dx.doi.org/10.1080/00396265.2016.1212970

Published online: 05 Aug 2016.

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
http://www.tandfonline.com/action/journalInformation?journalCode=ysre20

Download by: [Ivandro Klein] **Date:** 20 September 2017, At: 05:55

Check for updates

An approach to identify multiple outliers based on sequential likelihood ratio tests

I. Klein*[1], M. T. Matsuoka[2], M. P. Guzatto[1] and F. G. Nievinski[3]

One of the main challenges in the quality control of geodetic measurements is the reliable identification of multiple outliers. Within this context, the goal of this paper is to present a procedure designated here as Sequential Likelihood Ratio Tests for Multiple Outliers (SLRTMO). To verify its performance, a levelling network was simulated involving one, two and three (simultaneous) outliers. Also a GNSS network involving one and two (simultaneous) outliers was analysed. Results showed that SLRTMO is efficient for single and multiple outliers, simulated with magnitude greater than five standard deviations, *with a mean success rate of 79.6% for these cases*. Furthermore, the maximum number of outliers to be tested has to be defined according to the redundancy of the network so as to ensure the performance of SLRTMO.

Keywords: Multiple outliers, Sequential testing procedure, Maximum likelihood ratio, Geodetic networks

Downloaded by [Ivandro Klein] at 05:55 20 September 2017

Introduction

Quality control to identify gross errors (outliers) in geodetic measurements has been widely investigated since the pioneering work of Baarda (1968). In the context of least-squares estimation (LSE), statistical testing procedures for identification of outliers are based either on the assumption of a single outlier (see e.g. Baarda, 1968; Pope, 1976; Berber and Hekimoglu, 2003; Lehmann, 2012) or multiple (simultaneous observation) outliers (see e.g. Ding and Coleman, 1996; Gui *et al.*, 2010; Knight *et al.*, 2010; Baselga, 2011a; Lehmann and Lösler, 2016). Statistical testing procedures for a single outlier, such as data snooping and τ test, are already well developed in the literature thus outside the scope of this paper. For the same reason, so are robust estimation methods (Huber, 1964; Krarup *et al.*, 1980; Rousseeuw and Leroy, 1987; Xu 2005; Guo *et al.*, 2010). Statistical testing procedures for multiple outliers in geodetic measurements include: the detection via redundancy contributions of observations (Ding and Coleman, 1996); maximum likelihood ratio tests (Teunissen, 2006; Knight *et al.*, 2010); the Bayesian unmasking method based on posterior probabilities of classification variables (Gui *et al.*, 2010); and the exhaustive search procedure based on multiple least-squares adjustment process (Baselga, 2011a).

Some concerns about the statistical testing procedures for multiple outliers are, on the one hand, that different combinations of outliers lead to the same least-squares observation residuals (Baselga, 2011b); and, on the other hand, the question of how to determine the number of outliers existing in the observations (Knight *et al.*, 2010). This paper focusses on the latter issue, developing a procedure to determine the number of outlying observations and identifying their locations, designated as 'Sequential Likelihood Ratio Tests for Multiple Outliers' (SLRTMO).

Although the statistical testing procedures can also be applied for the identification of systematic errors (see e.g. Koch, 1999; Teunissen, 2006), systematic errors are outside the scope of this paper. Here, it is assumed that outliers are observations contaminated by gross errors (blunders), following the statement of Lehmann (2013) that in Geodesy, 'outliers are most often caused by gross errors and gross errors most often cause outliers'. It is also important to mention that there is no method of quality control completely efficient in the framework of the LSE (Baselga, 2011b; Klein *et al.*, 2015). Furthermore, the statistical tests for multiple outliers may require a large computational effort; for example, if $n = 300$ and $q_{\max} = 3$ (n is the number of observations; $q_{\max}$ is the maximum number of simultaneous outliers), there are $\binom{300}{1} + \binom{300}{2} + \binom{300}{3} = 4,500,250$ possible q-combinations of outlying observations (i.e. alternative hypotheses). For a more complete review about these questions, see Teunissen (2006), Knight *et al.* (2010), Baselga (2011a, 2011b).

During the peer-review of the present article, Lehmann and Lösler (2016) was published making several contributions about the multiple outlier problem. Initially, it was cast as a model selection problem. Then an information-theory approach based on the Aitken information criterion was considered. An alternative approach based on sorted p-values (the probability that the test statistic is at least as extreme as observed) was also considered, which seems not unlike Holm's procedure for controlling the family-wise error rate (Holm, 1979). Both approaches were compared to the common consecutive data snooping.

[1]Department of Civil Construction, Federal Institute of Santa Catarina (IFSC), Florianopolis, SC 88020-300, Brazil
[2]Institute of Geography, Federal University of Uberlandia (UFU), Monte Carmelo, MG 38500-000, Brazil
[3]Department of Geodesy, Federal University of Rio Grande do Sul (UFRGS), Porto Alegre, RS 91501-970, Brazil

*Corresponding author, email: ivandroklein@gmail.com

© 2016 Survey Review Ltd
Received 27 January 2016; accepted 9 July 2016
DOI 10.1080/00396265.2016.1212970

Downloaded by [Ivandro Klein] at 05:55 20 September 2017

Results were different in each case but a best approach was not elected, a conclusion which would require extensive Monte Carlo simulations under controlled conditions.

The main differences between the SLRTMO and the tests presented in Teunissen (2006) and Knight *et al.* (2010) is that the null hypothesis is sequentially updated. In fact, in every step of SLRTMO just one additional outlying observation is tested in the alternative hypothesis, as will be seen in sections 'Two-step incremental likelihood ratio tests for multiple outliers' and 'Multi-step sequential likelihood ratio tests for multiple outliers'. Alternatively, Lehmann and Lösler (2016) use the testing procedure of Knight *et al.* (2010), making use of the *p*-value concept to decide between the null and multiple alternative hypotheses. The rest of the paper is organised as follows. Section 'Theoretical overview' brings a theoretical background about LSE and statistical tests for multiple outliers. Sections 'Two-step incremental likelihood ratio tests for multiple outliers' and 'Multi-step sequential likelihood ratio tests for multiple outliers' present the method proposed (SLRTMO) for determining the number and the identification of the outlying observations. Section 'Experiments and results' contains the experiments setup and results, which are then discussed in the next section. Finally, last section offers conclusions and recommendations for future studies.

Theoretical overview

This section briefly reviews the theoretical background about LSE and statistical tests based on the likelihood ratio.

Least-squares estimation

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

$$\mathbf{v} = \mathbf{A}\mathbf{x} - \mathbf{y} \tag{1}$$

in which $\mathbf{v}$ is the vector of n observation residuals, $\mathbf{A}$ is the design or Jacobian matrix, $\mathbf{x}$ is the vector of u unknown parameters and $\mathbf{y}$ is the vector of n observations (assuming $n > u$). The LSE for the unknown parameters ($\hat{\mathbf{x}}$) in these cases is given by (Koch, 1999)

$$\hat{\mathbf{x}} = (\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{y} \tag{2}$$

in which $\mathbf{W}$ is the weight matrix of the observations, taken as $\mathbf{W} = \boldsymbol{\sigma}_0^2\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}$, where $\boldsymbol{\sigma}_0^2$ is the variance factor and $\boldsymbol{\Sigma}_{\mathbf{y}}$ is the covariance matrix of the observations (see Koch, 1999; Lehmann, 2012); if $\boldsymbol{\Sigma}_{\mathbf{y}}$ is diagonal, one speaks of weighted LSE (WLSE); if it is full, generalised LSE (GLSE).

If there are only random errors in the observations, the LSE is the best linear unbiased estimator (BLUE) for the unknown parameters; if the observational errors follow the multivariate normal distribution with mean $\boldsymbol{\mu} = [\mathbf{0}]$ and covariance matrix $\boldsymbol{\Sigma}_{\mathbf{y}}$, the LSE coincides with the maximum likelihood estimator (Teunissen, 2003). However, if there are systematic and/or gross errors (blunders) in the observations, LSE is no longer optimal, because they are not robust estimators, that is, insensitive to outliers (Huber, 1964; Rousseeuw and Leroy, 1987). One way to restore optimality is applying statistical testing procedures for multiple outliers, as described in the next subsection.

Statistical testing procedures for multiple outliers based on maximum likelihood ratio

Assuming normally distributed observation errors, the generalised statistical test for multiple outliers is formulated by the following hypotheses (Teunissen, 2006)

$$\begin{aligned} &H_0\colon\ \mathrm{E}\{\mathbf{y}\} = \mathbf{A}\mathbf{x} \quad \text{vs.} \\ &H_{\mathrm{A}}\colon\ \mathrm{E}\{\mathbf{y}\} = \mathbf{A}\mathbf{x} + \mathbf{C}_{q,k}\nabla_{\boldsymbol{q},\boldsymbol{k}};\quad \nabla_{\boldsymbol{q},\boldsymbol{k}} \neq 0, \end{aligned} \tag{3}$$

Respectively, H_0 is the null hypothesis (namely, absence of outliers in the observations) and H_{A} is an alternative hypothesis (presence of q outlying observations in $\mathbf{y}$ at certain known locations). The quantity $\mathbf{C}_{q,k}$ is the outlier model matrix (with dimension $n \times q$), defining the k-th combination of q outlying observations considered in $\mathbf{y}$; $\nabla_{\boldsymbol{q},\boldsymbol{k}}$ is the corresponding vector of q outliers. For example, if $n = 9$ and $q = 2$, then a possible outlier model is $\mathbf{C}_{q,k} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}^{\mathrm{T}}$ and $\nabla_{\boldsymbol{q},\boldsymbol{k}} = \begin{bmatrix} \nabla_{(4)} \\ \nabla_{(7)} \end{bmatrix}$ (one outlier in each observation y_4 and y_7). For more details about outlier models, see e.g. Knight *et al.* (2010) and Lehmann and Lösler (2016).

After postulating a value for q (more about that below), there are $c = \binom{n}{q}$ alternative hypotheses (H_A), one for each possible outlier model given by the matrix $\mathbf{C}_{q,k}$, whose test statistic $T_{q,k}$ is (Teunissen, 2006):

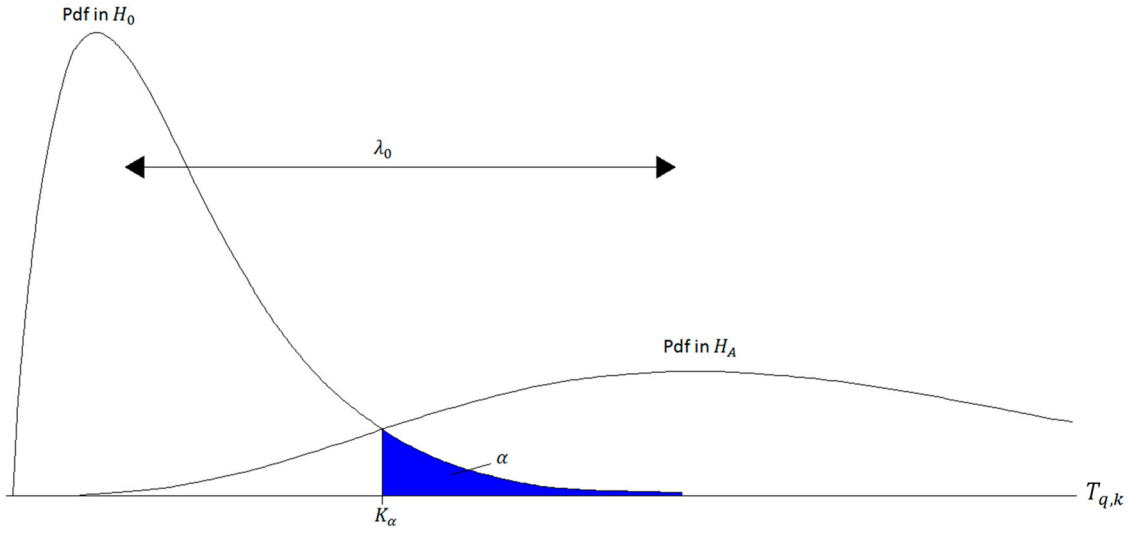
$$T_{q,k} = \hat{\mathbf{v}}^{\mathrm{T}}\mathbf{W}\mathbf{C}_{q,k}(\mathbf{C}_{q,k}^{\mathrm{T}}\mathbf{W}\boldsymbol{\Sigma}_{\hat{\mathbf{v}}}\mathbf{W}\mathbf{C}_{q,k})^{-1}\mathbf{C}_{q,k}^{\mathrm{T}}\mathbf{W}\hat{\mathbf{v}} \tag{4}$$

where $\hat{\mathbf{v}}$ and $\boldsymbol{\Sigma}_{\hat{\mathbf{v}}}$ are the estimated residuals vector and *a posteriori* covariance matrix of the estimated residuals computed via LSE, see Koch (1999) and Teunissen (2003); the *a priori* observation covariance matrix $\boldsymbol{\Sigma}_{\mathbf{y}}$ involved in the weight matrix $\mathbf{W}$ remains the same in all cases.

Under the null hypothesis, the test statistic $T_{q,k}$ follows the central chi-squared distribution with q degrees of freedom; under the alternative hypothesis, the test statistic $T_{q,k}$ follows the non-central chi-squared distribution with q degrees of freedom and non-centrality parameter λ_0 (see Fig. 1). Therefore, one particular k-th combination of q tested observations is considered flagged as suspected outliers if $T_{q,k} > \chi^2_{(q,\alpha)}$, where $\chi^2_{(q,\alpha)}$ is the chi-squared critical value at a given significance level α. For more details about these relationships and related probability levels, see Koch (1999) and Teunissen (2006). Additionally, if more than one combination satisfies the inequality above, the one with largest test statistic is picked. In other words, among all possible combinations, the one combination of observations considered as outlying observations is that whose test statistic $T_{q,k}$ satisfies the following two conditions (Knight *et al.*, 2010)

$$\begin{cases} T_{q,k} > \chi^2_{(q,\alpha)} \\ T_{q,k} \text{ is maximum for all } \mathbf{C}_{q,k} \text{ matrices} \end{cases} \tag{5}$$

The difficulty with the statistical testing procedures based on maximum likelihood ratio is they identify the q most probable outlying observations by considering q fixed, i.e. assuming the number of outliers to be known. Therefore, these procedures do not estimate the number of outliers present in the observations (Knight *et al.*, 2010). For example, all that can be obtained are the



1 Pdf (probability density function) of $T_{q,k}$ in H_0 and in H_A (adapted from Knight *et al.* 2010)

Downloaded by [Ivandro Klein] at 05:55 20 September 2017

maximum $T_{q,k}$ statistics for pre-determined q (e.g. $q = 2$, $q = 4$, $q = 4$ and so on), but the true value of q itself is not estimated. In view of these issues, the next section presents an approach to determine the number of outlying observations and identify their locations.

It is important to mention that there are other statistical testing procedures to identify multiple outliers in the observations, e.g. those presented in Guo *et al.* (2010) and Baselga (2011a). Also, statistical tests for a single outlier such as data snooping can also be iteratively applied in the context of multiple outliers (see Baarda, 1968; Berber and Hekimoglu, 2003; Klein *et al.*, 2015). The statistical testing procedures described here lead to the uniformly most powerful tests, being the data snooping procedure a particular case when $q = 1$ and the candidate outlier matrices $\mathbf{C}_{q,k} = \mathbf{c}_i = \begin{bmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{bmatrix}^{\mathrm{T}}$ for $i = 1, 2, \ldots, n$, i.e. are the i-th column of the n-by-n identity matrix (see Baarda, 1968; Teunissen, 2006; Lehmann, 2012).

Two-step incremental likelihood ratio tests for multiple outliers

The approach proposed in this paper for determining the number of outlying observations and their locations is formulated based on a maximum likelihood ratio. Consider a null hypothesis H_0 for the parameters of the population probability distribution of a vector of observations $\mathbf{y}$. Consider further an alternative hypothesis H_A for these parameters, constructed in a way that H_0 is a subset of H_A. In this general case, the maximum likelihood ratio between H_0 and H_A is given by (Larson, 1974)

$$\lambda(\mathbf{y}) = \frac{\max\, p(\mathbf{y}|H_0)}{\max\, p(\mathbf{y}|H_A)} \tag{6}$$

in which $\max\, p(\mathbf{y}|H_0)$ is the maximum of the probability density function (pdf) of $\mathbf{y}$ under H_0 and $\max\, p(\mathbf{y}|H_A)$ is the maximum of the pdf of $\mathbf{y}$ under H_A. As the null hypothesis is defined so that its sample space is contained in the sample space of the alternative hypothesis, the ratio in equation (6) lies in the interval of $0 \le \lambda(\mathbf{y}) \le 1$ (Teunissen, 2006). Furthermore, as the ratio in equation (6) tends to have a value near 1 when the likelihood of H_0 is high, the test criterion for the maximum likelihood ratio is given by (Larson, 1974)

$$\text{Do not reject } H_0 \text{ if } \lambda(y) \ge c \tag{7}$$

in which $c > 0$ is the critical value for the test according to the significance level α stipulated (for more details, see Larson, 1974; Teunissen, 2006). It is also important to mention that the equation (4) in 'Theoretical overview' section can be derived from equations (6) and (7) in this section, as those statistical testing procedures are all based on maximum likelihood ratio, see Teunissen (2006).

In the present approach, the null hypothesis H_0 is defined as specifying $q_{(H0)}$ outliers in the vector of observations $\mathbf{y}$. In contrast, the alternative hypothesis H_A is defined specifying $q_{(H0)} + 1$ outliers in $\mathbf{y}$. For example, H_0 can define observations y_5 and y_9 as outliers, while H_A can define observations y_5, y_9 and y_{14} as outliers. Again, note that H_0 must be a subset of H_A. Therefore, in a general case, these hypotheses can be expressed as

$$\begin{aligned} H_0&: \mathrm{E}\{\mathbf{y}\} = \mathbf{A}\mathbf{x} + \mathbf{C}_{q,k_{(H0)}}\nabla_{q,k_{(H0)}} \quad \text{vs.} \\ H_A&: \mathrm{E}\{\mathbf{y}\} = \mathbf{A}\mathbf{x} + \mathbf{C}_{q_{(H0)}+1,k}\nabla_{q_{(H0)}+1,k}, \end{aligned} \tag{8}$$

in which $\mathbf{C}_{q(H0),\mathrm{k}}$ is an outlier model matrix (with dimension $n \times q_{(H0)}$) an $\nabla_{q(H0),k}$ is the respective vector of $q_{(H0)}$ outliers; $\mathbf{C}_{q(H0)+1,\mathrm{k}}$ and $\nabla_{q,H0)+1,\mathrm{k}}$ are the corresponding terms for H_A.

Considering the maximum of the pdf of $\mathbf{y}$ under H_0 and H_A, the maximum likelihood ratio λ between the two hypotheses is expanded as (Teunissen, 2006):

$$\begin{aligned} &\frac{\max\, p(\mathbf{y}|\mathbf{x}, \nabla_{q_{(H_0),k}})}{\max\; p(\mathbf{y}|\mathbf{x}, \nabla_{q_{(H_0)}+1,k})} \\ &= \frac{(2\pi)^{\frac{-n}{2}}|\boldsymbol{\Sigma}_{\mathbf{y}}|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\,\mathbf{v}_{H_0}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\,\mathbf{v}_{H_0}\right\}}{(2\pi)^{\frac{-n}{2}}|\boldsymbol{\Sigma}_{\mathbf{y}}|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\,\mathbf{v}_{H_A}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\,\mathbf{v}_{H_A}\right\}} \\ &= \exp\left\{-\frac{1}{2}(\mathbf{v}_{H_0}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\,\mathbf{v}_{H_0} - \mathbf{v}_{H_A}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\,\mathbf{v}_{H_A})\right\} \end{aligned} \tag{9}$$

which $\mathbf{v}_{H0}$ is the vector of observation residuals under

H_0 and $\mathbf{v}_{HA}$ is the vector of residuals under H_A. It is assumed that observations are normally distributed with mean $E\{\mathbf{y}\} = \mathbf{Ax} + \mathbf{C}_{q(H0),k}\nabla_{q(H0),k}$ and $E\{\mathbf{y}\} = \mathbf{Ax} + \mathbf{C}_{q(H0)+1,k}\nabla_{q(H0)+1,k}$, respectively. The scalar $|\boldsymbol{\Sigma}_\mathbf{y}|$ is the determinant of $\boldsymbol{\Sigma}_\mathbf{y}$.

The ratio in equation (9) is smaller than a positive constant c if and only if the exponential argument in the right-hand side of equation (9) is greater than a positive constant K_α. Thus, the maximum likelihood ratio test of H_0 against H_A finally becomes (Teunissen, 2006):

$$\text{Reject } H_0 \text{ if: } \hat{\mathbf{v}}_{H_0}^T \boldsymbol{\Sigma}_\mathbf{y}^{-1} \hat{\mathbf{v}}_{H_0} - \hat{\mathbf{v}}_{H_A}^T \boldsymbol{\Sigma}_\mathbf{y}^{-1} \hat{\mathbf{v}}_{H_A} > K_\alpha \qquad (10)$$

in which K_α is the critical value for the test according to the significance level α; it must be determined according to the chi-squared distribution with one degree of freedom, because the quadratic variable in the left side of equation (10) follows the chi-squared distribution and the alternative hypothesis H_A contains one more outlier (model parameter) than the null hypothesis H_0.

Thus, considering that the LSE coincides with the maximum likelihood estimator when the vector of observations $\mathbf{y}$ is normally distributed and the weight matrix is defined as $\mathbf{W} = \boldsymbol{\Sigma}_\mathbf{y}^{-1}$, the observation residuals in equations (9) and (10) can be obtained augmenting the Jacobian matrix $\mathbf{A}$ with the outlier model matrix $\mathbf{C}_{q,k}$ (and, likewise, the parameters vector $\mathbf{x}$ with the outlier vector ∇), as

$$\begin{cases} \hat{\mathbf{v}}_{H_0} = \mathbf{y} - [\mathbf{A}\vdots\mathbf{C}_{q(H0)}]\begin{bmatrix}\hat{\mathbf{x}}_{H_0}\\ \hat{\nabla}_{H_0}\end{bmatrix}, \text{ with } \begin{bmatrix}\hat{\mathbf{x}}_{H_0}\\ \hat{\nabla}_{H_0}\end{bmatrix} \\ \qquad = [(\mathbf{A}\vdots\mathbf{C}_{q(H0),k})^T \mathbf{W}(\mathbf{A}\vdots\mathbf{C}_{q(H0),k})]^{-1}[\mathbf{A}\vdots\mathbf{C}_{q(H0),k}]^T\mathbf{Wy} \\ \hat{\mathbf{v}}_{H_A} = \mathbf{y} - [\mathbf{A}\vdots\mathbf{C}_{q(H0)+1,k}]\begin{bmatrix}\hat{\mathbf{x}}_{H_A}\\ \hat{\nabla}_{H_A}\end{bmatrix}, \text{ with } \begin{bmatrix}\hat{\mathbf{x}}_{H_A}\\ \hat{\nabla}_{H_A}\end{bmatrix} \\ \qquad = [(\mathbf{A}\vdots\mathbf{C}_{q(H0)+1,k})^T \mathbf{W}(\mathbf{A}\vdots\mathbf{C}_{q(H0)+1,k})]^{-1}[\mathbf{A}\vdots\mathbf{C}_{q(H0)+1,k}]^T\mathbf{Wy} \end{cases} \qquad (11)$$

Multi-step sequential likelihood ratio tests for multiple outliers

Based on the relationships above, a sequential statistical testing procedure to determine the number and identify the location of the outlying observations is established as follows (summarised as a flowchart in Fig. 2). In the first step, define the maximum number of outliers to be considered (q_{max} – see below for details) and the significance level to be adopted (α). Then, assume the hypothesis of equation (3) for $q = 1$ (one outlier), and compute the corresponding test statistics $T_{q=1,k}$ according to equation (4), identifying the maximum test statistic value (max $T_{q=1,k}$). If this value exceeds the critical value at the significance level adopted, i.e. if max $T_{q=1,k} > K_\alpha$, reject the null hypothesis – there are no outliers in the observations. Otherwise, the null hypothesis is not rejected and the testing ends with no outliers detected.

In case the null hypothesis was rejected, the sequential testing procedure resumes. It proceeds computing the test statistics for $q = 2$ (two outlying observations) according to equation (4), and identifying the maximum test statistic value (max $T_{q=2,k}$). If the observation with max $T_{q=1,k}$ is also contained in the case of max $T_{q=2,k}$, i.e. if the one observation flagged initially is among one of the pairs flagged subsequently – the hypotheses of equation (8) for $q_{(H0)} = 1$ and $q_{(H0+1)} = 2$ are defined accordingly. Otherwise, if H_0 is not a subset of H_A, the testing ends and only the observation with max $T_{q=1,k}$ is flagged as outlier. Thus, if H_0 is indeed a subset of H_A, observation residuals are computed for H_0 and H_A using equation (11), and the test criterion of the equation (10) is applied. If the null hypothesis is not rejected, the testing ends and only the one observation corresponding to max $T_{q=1,k}$ is flagged as outlier.

In case the null hypothesis is rejected again, a new step with $q = 2$ and $q + 1 = 3$ is started in the exactly same way as before. This sequential testing procedure is repeated until any of the following conditions is met: the current null hypothesis fails to be rejected; H_0 is no longer a subset of the current best alternative hypothesis; or the iteration reaches q_{max} (i.e. until the maximum number of outliers is fully inspected). In other words, the sequential testing procedure consists in determining whether the additional outlying observation, considered in every new alternative hypothesis, is statistically significant or not, in terms of its impact on the quadratic form of residuals. These approach leads to the number of outliers and the outlying observations applying a sequential testing procedure based on likelihood ratio, thus designated as SLRTMO.

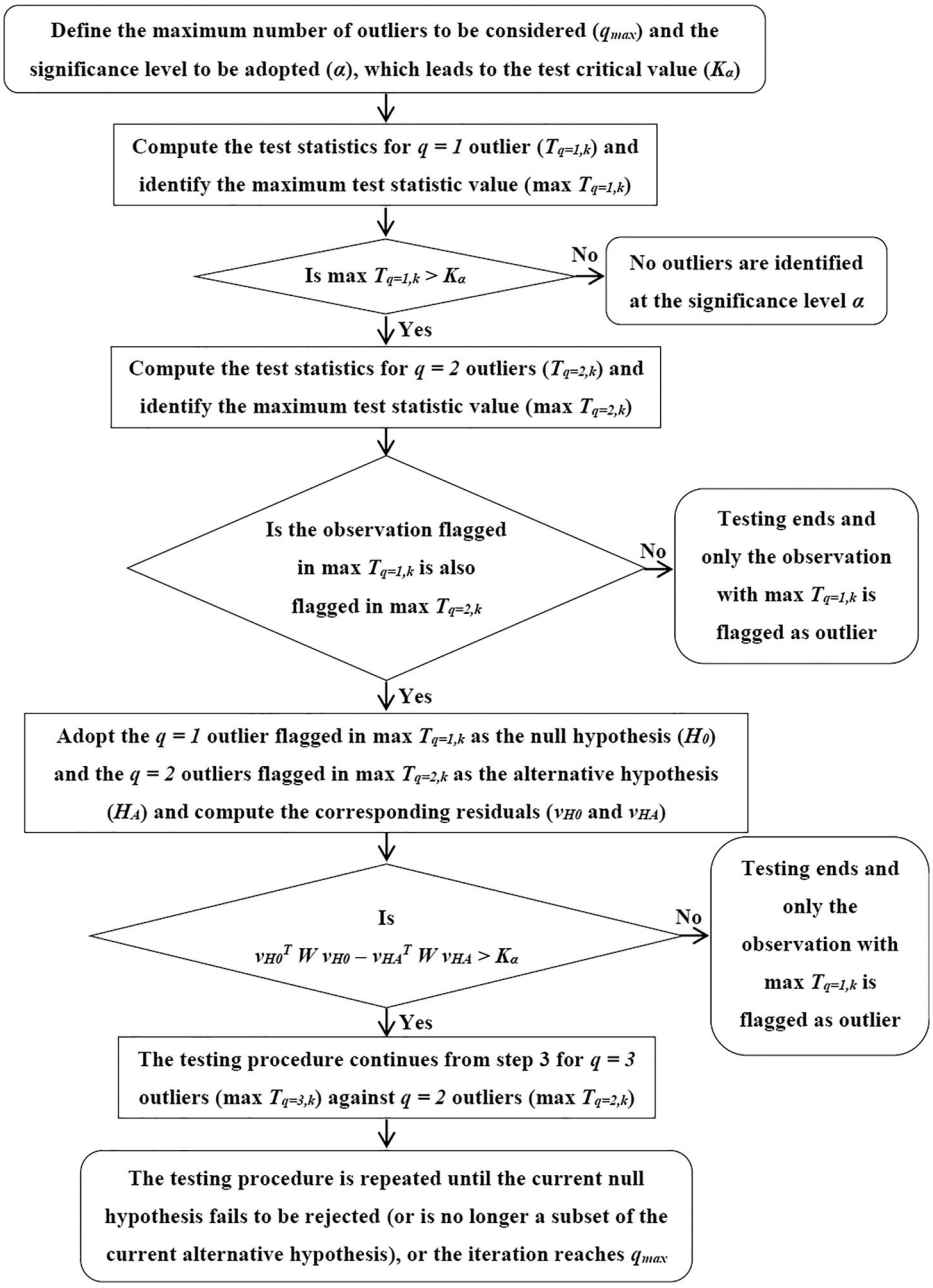
The main differences between SLRTMO and the tests presented in Teunissen (2006) and Knight *et al.* (2010) is that the null hypothesis is sequentially updated; furthermore, in every step just one additional outlying observation is in fact tested in the current alternative hypothesis. The reliability measures of SLRTMO could be obtained through the multiple outlier context presented in Knight *et al.* (2010).

Choosing the maximum number of outliers to be considered

The value of q_{max} needs to be chosen so as to avoid rank deficiency or tests statistics with the same computed value. It is proposed to define q_{max} equal to the minimal redundancy number (degrees of freedom) of the problem in a local scale, as defined below. Consider a point with unknown 2D position (see Fig. 3). Two observations such as horizontal distances (HD) and azimuths (Az) lead to a unique solution for the two plane coordinates (x,y). Three observations would lead to different solutions and allow the detection of an inconsistency between them. Four observations would lead to different solutions and the identification of one outlying observation, and so on. Thus, in a general case, the value for q_{max} is equal to the minimal number of redundant observations across each and every point, minus one. In the case that each unknown station is involved in at least five observation equations (thus, three redundant observations), then $q_{max} = 3 - 1 = 2$.

This strategy follows the statement of 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'. In more complex problems, the choice of q_{max} is more difficult and its strategy could not be applicable, but the value of q_{max} always must be smaller than the global degrees of freedom of the problem, that is: $q_{max} < n - u$. To verify the performance of the SLRTMO procedure, the next section presents experiments and results.

Downloaded by [Ivandro Klein] at 05:55 20 September 2017



2 Flowchart of the SLRTMO

Downloaded by [Ivandro Klein] at 05:55 20 September 2017

Experiments and results

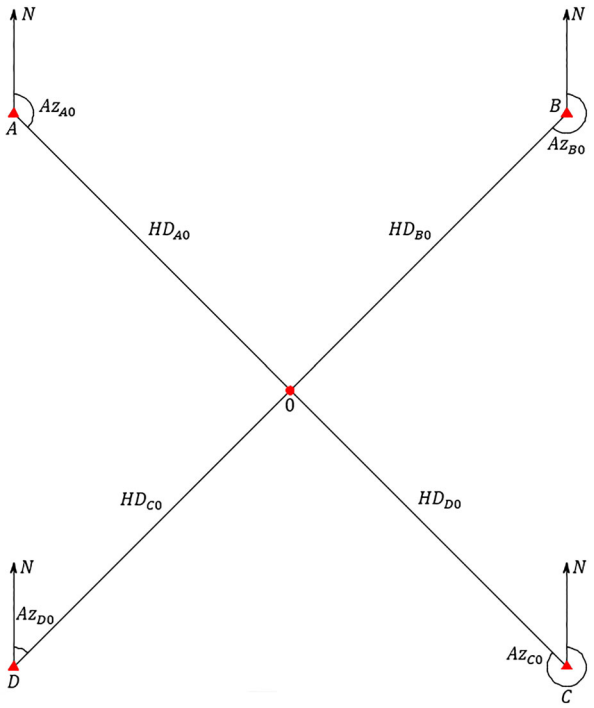
Levelling network

In this study, a simulated closed levelling network is analysed (network A), with one control (fixed) point, and 13 points with unknown heights, totalling fourteen minimally constrained points (see Fig. 4). For each pair of points, there are two or four height difference measurements. Thus, there are $n = 17 \times 2 + 2 \times 4 = 42$ observations, $u = 13$ unknowns, and $n - u = 42 - 13 = 29$ redundant observations in this simulation. All observations are assumed with a nominal precision (*a priori* standard deviation) of $\sigma = \pm 2$ mm and also uncorrelated. Each unknown point is involved in at least six different observations; thus, the local-scale redundancy equals five, and it is assumed that the maximum number of outliers to be tested is $q_{\max} = 5 - 1 = 4$.

To analyse the testing procedure proposed here (SLRTMO), 27 scenarios are ran for network A. Each scenario has a unique combination of number of outliers, significance level and magnitude of outliers. We ran 1000 experiments for each scenario, totalling 27,000 Monte Carlo simulations.

Random errors and outliers are synthetically generated and added to the observations. Random errors are generated using the normal distribution and outliers are generated using the uniform distribution. Positive and negative outliers are clipped between 3σ and 5σ, 5σ and 7σ, and 7σ and 9σ magnitude intervals in each experiment. Thus outliers are observations contaminated by gross errors.

The significance level is varied, taken as $\alpha = 0.001$ (0.1%), $\alpha = 0.005$ (0.5%) and $\alpha = 0.01$ (1%). The corresponding critical values of the χ^2 distribution are



3 Redundant observations to determine the unknown position (x,y) of the point 0

$K_{0.001} = 10.828$, $K_{0.005} = 7.879$ and $K_{0.01} = 6.635$. Finally, the number of outliers is varied: 9000 experiments with one outlier, 9000 with two outliers and 9000 with three outliers.

Table 1 shows the success rate (fraction of outliers correctly identified, averaged over all experiments of a given scenario) and the misidentification rates for one outlier. Tables 2 and 3 show the analogous rates for two and three outliers, respectively. The misidentifications rates are divided in three classes: fewer (fewer outlier(s) than simulated or none), more (more outliers than simulated) and wrong (correct number of outlier but wrong identification).

GNSS network

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

Random errors and outliers are synthetically generated and added to the observations. In total, 18,000 experiments are performed using Monte Carlo simulations. There are 1000 repetitions of each of 18 unique scenarious, obtained varying the significance level α, the magnitude interval of outliers, and the number of outliers: 9000 experiments with one outlier and 9000 with two outliers.

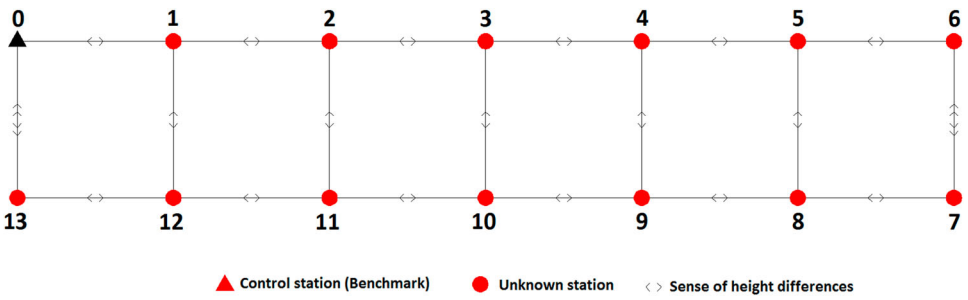
Each unknown station is involved in at least four baseline vectors; thus, the local-scale redundancy equals three. Therefore, the maximum number of outliers to be tested is $q_{\max} = 3 - 1 = 2$. However, the true number of outliers is also 2 (at most), so the maximum number of outliers to be tested is set to $q_{\max} = 3$, to avoid overoptimistic results.

Table 4 shows the success rate (fraction of outliers correctly identified, averaged over all experiments) and the misidentification rates for one outlier. Table 5 shows analogous rates for two outliers. The classes of misidentifications rates are the same as for network A.

Discussion

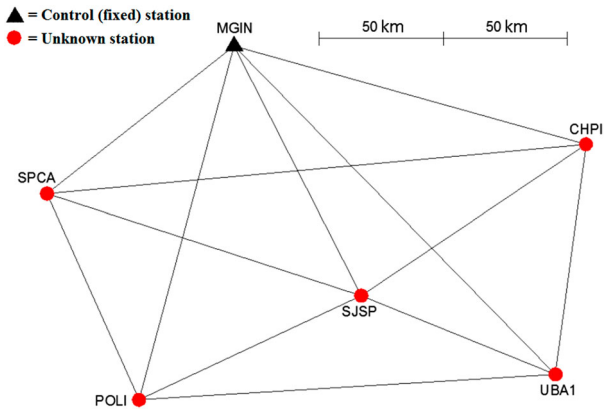
Tables 1–5 show that, in general, the greater the number of outliers, the lower is the efficiency of SLRTMO for both networks. Furthermore, in general, the greater the magnitude of outliers, the greater is the efficiency of SLRTMO, also for both networks. These results are not unexpected, as larger and less numerous outliers are generally easier to detect. However, results show no direct relationship between success rate and significance level (α) for both networks analysed. It appears that higher values for α are not recommended for outliers of greater magnitude and that lower values for α are not recommended for outliers of smaller magnitude, but results are not consistent for both networks. In practice, neither the number nor the magnitude of outliers are known. Therefore, these results show the importance of a correct choice of α, as pointed out by Lehmann (2012); it also highlights the challenges in controlling the error rate in multiple hypotheses tests.

Regarding the three classes of misidentification rates, in general, an increase in the magnitude interval of outliers, leads to an increase in the over-identification rate (more



4 Levelling network analysed (network A)

Downloaded by [Ivandro Klein] at 05:55 20 September 2017



5 GNSS network analysed (network B)

outliers being identified than simulated) and a decrease in the under-identification rate (fewer outliers being identified than simulated). This fact is due to the error propagation among all residuals. The rate of cases with correct number of outliers but with wrong identification also decreases when increasing the magnitude interval of outliers.

Considering all results, the mean success rate is 62.1% for network A and 77.2% for network B. In general, the SLRTMO is efficient for outliers greater than 5σ, with a mean success rate of 76.7% for network A and 82.5% for network B. For outliers smaller than 5σ, the mean success rate is 32.9% for network A and 66.4% for network B. Therefore, with an appropriate choice of q_{max}, results show that SLRTMO can locate multiple outliers greater than 5σ with high success rate. However, the choice of significance level α, that also affects the efficiency of SLRTMO, requires further investigation.

Conclusions

Many methods of quality control for geodetic data analysis have been developed and investigated since the pioneering work of Baarda (1968). However, these methods still deserve further investigation, especially in the case of multiple outliers. Thus, the goal of this paper was to propose a sequential testing procedure to locate multiple outliers based on maximum likelihood ratio, designated here as SLRTMO.

SLRTMO shows high success rates in the experiments of a simulated levelling network and a GNSS network for single and multiple outliers randomly generated between five and nine standard deviations. However, the efficiency of SLRTMO significantly decreases for outliers

Table 1 Success rate and misidentification rates for one outlier in network A

α	Magnitude interval of outliers											
	3σ–5σ				5σ–7σ				7σ–9σ			
	Success/ %	Fewer/ %	More/ %	Wrong/ %	Success/ %	Fewer/ %	More/ %	Wrong/ %	Success/ %	Fewer/ %	More/ %	Wrong/ %
0.001	44.2	49.5	2.6	3.7	89.3	5.6	4.1	1	96.2	0.1	3.7	0
0.005	53.0	25.6	12.9	8.5	79.5	2.2	17.3	1	82.9	0	17.1	0
0.01	50.8	15.2	23.5	10.5	66.1	0.7	32.4	0.8	66.5	0	33.5	0

Table 2 Success rate and misidentification rates (in %) for two outliers in network A

α	Magnitude interval of outliers											
	3σ–5σ				5σ–7σ				7σ–9σ			
	Success/ %	Fewer/ %	More/ %	Wrong/ %	Success/ %	Fewer/ %	More/ %	Wrong/ %	Success/ %	Fewer/ %	More/ %	Wrong/ %
0.001	26.1	69.2	0.9	3.8	83.1	11.7	2.9	2.3	93.5	0.6	4.5	1.4
0.005	32.7	6.3	50.7	10.3	76.4	4.6	16.7	2.3	83.4	0.4	15.4	0.8
0.01	35.9	30.6	20.2	13.3	66.7	3	27.4	2.9	68.1	0.6	30	1

Table 3 Success rate and misidentification rates for three outliers in network A

α	Magnitude interval of outliers											
	3σ–5σ				5σ–7σ				7σ–9σ			
	Success/ %	Fewer/ %	More/ %	Wrong/ %	Success/ %	Fewer/ %	More/ %	Wrong/ %	Success/ %	Fewer/ %	More/ %	Wrong/ %
0.001	11.2	85.6	0.3	2.9	68.5	24.9	2.5	4.1	86.2	5	6.2	2.6
0.005	18.6	4.3	65.8	11.3	70.7	10.5	13	5.8	78.5	4.2	14.5	2.7
0.01	23.5	48.5	12.3	15.7	59.9	7.4	28.1	4.6	65.5	3.4	28.4	2.7

Downloaded by [Ivandro Klein] at 05:55 20 September 2017

Table 4 Success rate and misidentification rates for one outlier in network B

α	Magnitude interval of outlier											
	3σ–5σ				5σ–7σ				7σ–9σ			
	Success/ %	Fewer/ %	More/ %	Wrong/ %	Success/ %	Fewer/ %	More/ %	Wrong/ %	Success/ %	Fewer/ %	More/ %	Wrong/ %
0.001	80.5	14.4	2.5	2.6	96.7	0.1	2.7	0.5	96	0	4	0
0.005	76.1	6.6	13.4	3.9	83.5	0	16.4	0.1	85.5	0	14.5	0
0.01	66.2	3.8	25.9	4.1	74	0	25.9	0.1	72.2	0	27.8	0

Table 5 Success rate and misidentification rates for two outliers in network B

α	Magnitude interval of outliers											
	3σ–5σ				5σ–7σ				7σ–9σ			
	Success/ %	Fewer/ %	More/ %	Wrong/ %	Success/ %	Fewer/ %	More/ %	Wrong/ %	Success/ %	Fewer/ %	More/ %	Wrong/ %
0.001	57.6	35.3	1.5	5.6	91.3	4.2	2.7	1.8	92.2	3.1	3.8	0.9
0.005	62	20.3	9.8	7.9	78.5	4.4	16	1.1	80.3	4.2	14.6	0.9
0.01	56.3	13.8	22.2	7.7	70.8	3.2	24.9	1.1	69.4	5	24.8	0.8

Downloaded by [Ivandro Klein] at 05:55 20 September 2017

smaller than five standard deviations. The optimal value for the significance level changes for each network.

The maximum number of outliers to be considered must be equal to the minimal number of redundant observations in each point across the network minus one so as to avoid rank deficiency or test statistics with the same computed value and to maintain good computational performance.

Future studies should consider various issues: the performance of the SLRTMO in cases of over-constrained systems and linearised (originally non-linear) models; the most adequate criteria for the selection of the significance level α, which greatly affects the efficiency of SLRTMO; the development of reliability measures; and the method performance in different networks with various geometry and varying redundancy.

It is also important to mention that the several discussions about multiple alternative hypotheses testing presented in Lehmann and Lösler (2016) are based on the assumption of a unique null hypothesis, defined as having no outliers in the observations. Future studies must consider the approach in which the null hypothesis is sequentially updated, as in SLTMO presented here.

Acknowledgements

The second author thanks CNPq for financial support provided (303306/2012-2, 477914/2012-8, 305599/2015-1).

ORCiD

F. G. Nievinski http://orcid.org/0000-0002-3325-1987

References

Baarda, W., 1968. *A testing procedure for use in geodetic networks (Netherlands geodetic commission, publications on geodesy, New series)* [online]. Delft, vol. 2, 1–60. Available from: http://www.ncgeo.nl/phocadownload/09Baarda.pdf

Baselga, S., 2011a. Exhaustive search procedure for multiple outlier detection. *Acta Geodaetica et Geophysica Hungarica*, 46 (4), 401–416.

Baselga, S., 2011b. Nonexistence of rigorous tests for multiple outlier detection in least-squares adjustment. *Journal of surveying engineering*, 137 (3), 109–112.

Berber, M. and Hekimoglu, S., 2003. What is the reliability of conventional outlier detection and robust estimation in trilateration networks? *Survey review*, 37 (290), 308–318.

Ding, X. and Coleman, R., 1996. Multiple outlier detection by evaluating redundancy contributions of observations. *Journal of geodesy*, 70 (8), 489–498.

Gui, Q., *et al*., 2010. A Bayesian unmasking method for locating multiple gross errors based on posterior probabilities of classification variables. *Journal of geodesy*, 85 (4), 651–659.

Guo, J., Ou, J., and Wang, H., 2010. Robust estimation for correlated observations: two local sensitivity-based downweighting strategies. *Journal of geodesy*, 84 (4), 243–250.

Holm, S., 1979. A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, 6 (2), 65–70.

Huber, P.J., 1964. Robust estimation of a location parameter. *The annals of mathematical statistics*, 35, 73–101.

Klein, I., 2014. *Proposal of a new method for geodetic networks design (in Portuguese)*. Doctoral thesis. Universidade Federal do Rio Grande do Sul, Brazil.

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

Knight, N.L., Wang, J., and Rizos, C., 2010. Generalised measures of reliability for multiple outliers. *Journal of geodesy*, 84 (10), 625–635.

Koch, K.R., 1999. *Parameter estimation and hypothesis testing in linear models*. 2nd ed. Berlin: Springer.

Krarup, T., Juhl, J., and Kubik, K., 1980. Götterdämmerung over least squares. *In*: *Proc. 14th Congress of the International Society of Photogrammetry Vol. B3*, Hamburg, Germany, 369–378.

Larson, H.J., 1974. *Introduction to probability theory and statistical inference*. 2nd ed. New York: John Wiley & Sons.

Lehmann, R., 2012. Improved critical values for extreme normalized and studentized residuals in Gauss–Markov models. *Journal of geodesy*, 86 (12), 1137–1146.

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

Lehmann, R. and Lösler, M., 2016. Multiple outlier detection: hypothesis tests versus model selection by information criteria. *Journal of surveying engineering* [online]. Available from: http://ascelibrary.org/doi/abs/10.1061/(ASCE)SU.1943-5428.0000189

Pope, A.J., 1976. *The statistics of residuals and the detection of outliers*. NOAA Technical Report NOS 65 NGS 1, U.S. National Geodetic Survey, MD, USA.

Rousseeuw, P.J. and Leroy, A.M., 1987. *Robust regression and outlier detection*. New York: John Wiley and Sons.
Teunissen, P.J.G., 2003. *Adjustment theory: an introduction*. Delft: Delft University Press.
Teunissen, P.J.G., 2006. *Testing theory: an introduction*. 2nd ed. Delft: VSSD.
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.

Downloaded by [Ivandro Klein] at 05:55 20 September 2017