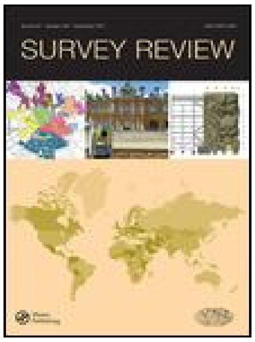



Survey Review

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

# On evaluation of different methods for quality control of correlated observations

I. Klein, M. T. Matsuoka, M. P. Guzatto, S. F. de Souza & M. R. Veronez

**To cite this article:** I. Klein, M. T. Matsuoka, M. P. Guzatto, S. F. de Souza & M. R. Veronez (2015) On evaluation of different methods for quality control of correlated observations, Survey Review, 47:340, 28-35, DOI: 10.1179/1752270614Y.0000000089

**To link to this article:** https://doi.org/10.1179/1752270614Y.0000000089



Published online: 08 Jul 2014.

# On evaluation of different methods for quality control of correlated observations

**I. Klein[1], M. T. Matsuoka[1], M. P. Guzatto[1], S. F. de Souza[1] and M. R. Veronez[2]**

This paper evaluates, compares, and discusses different methods for quality control in geodetic data analysis in the general scenario of correlated observations and multiple outliers. The investigated methods are the data snooping procedure, the statistical tests for multiple outliers, the recently proposed quasi-accurate detection of outliers method for correlated observations, the Danish method for correlated observations, the robust estimator for correlated observations based on bifactor equivalent weights, and the robust estimator for correlated observations based on a local sensitivity downweighting strategy. To evaluate these methods, outliers between 3$\sigma$ and 9$\sigma$ magnitude (positives and/or negatives) are randomly generated and added to some observations ($\sigma$ being the respective standard deviation of the observation) in two different global navigation satellite system (GNSS) networks that contain correlated observations. For each network, 15 000 scenarios are performed, 5000 with one outlier, 5000 with two outliers, and 5000 with three outliers, using Monte–Carlo simulations. The investigated methods have advantages and limitations, and the discussions and conclusions about the experiments are accurately presented.

**Keywords:** Quality control, Multiple outliers, Correlated observations, GNSS networks

## Introduction

Quality control in geodetic measurements has been widely investigated since Baarda's pioneering work (Baarda, 1968). Two of the most used strategies consist of statistical testing procedures (see, e.g. Baarda, 1968; Pope, 1976; Kok, 1984; Wang and Chen, 1994; Schaffrin, 1997; Knight *et al.*, 2010; Yang *et al.*, 2013) and robust estimation methods (see, e.g. Huber, 1964; Rousseeuw and Leroy, 1987; Yang, 1994; Yang *et al.*, 2002; Guo *et al.*, 2010).

However, these strategies have some limitations. For example, most of the statistical testing procedures that are used are based on the assumption of a single outlier (Knight *et al.*, 2010), and most of the available robust estimation methods are only applicable for uncorrelated observations (Guo *et al.*, 2010). In practice, the observations in the data set can be correlated and contain more than one outlier, as in the adjustment of global navigation satellite system (GNSS) networks. Thus, other approaches have been proposed, such as the quasi-accurate detection of outliers method (QUAD) for correlated observations (Guo *et al.*, 2007), statistical tests for multiple outliers (see, e.g. Knight *et al.*, 2010; Baselga, 2011), and robust estimators for correlated observations (see, e.g. Wieser and Brunner, 2002; Yang *et al.*, 2002; Guo *et al.*, 2010).

These methods are relatively new and require further investigation, as noted by Yang *et al.* (2002) and Knight *et al.* (2010). In this context, this paper evaluates, compares, and discusses different methods for quality control in the (general) scenario of correlated observations and multiple outliers. The analysed methods in this study are the data snooping procedure (Baarda, 1968; Teunissen, 2006), the Danish method scheme for correlated observations (Wieser and Brunner, 2002), the robust estimation scheme for correlated observations using Huber's $\psi$ function (Yang *et al.*, 2002), the QUAD method for correlated observations (Guo *et al.*, 2007), the robust estimation scheme for correlated observations proposed in Guo *et al.* (2010), and the statistical tests for multiple outliers based on the likelihood ratio (Teunissen, 2006; Knight *et al.*, 2010).

To evaluate these methods, outliers between 3$\sigma$ and 9$\sigma$ magnitude (positives and/or negatives) are randomly generated and added to some observations ($\sigma$ being the respective standard deviation of the observation) in two different GNSS networks that contain correlated observations. For each network, 15 000 scenarios are performed, 5000 with one outlier, 5000 with two outliers, and 5000 with three outliers, using Monte-Carlo simulations. The investigated methods have advantages and limitations, and the discussions and conclusions about the experiments are accurately presented.

Quality control should include reliability analysis (such as minimal detectable bias – MDB, and reliability

[1]Laboratory of Geodetic Researches (LAGEO), Department of Geodesy, Federal University of Rio Grande do Sul (UFRGS), Porto Alegre City, 91501-970, Post Office box 15001, Rio Grande do Sul State, Brazil
[2]Remote Sensing and Digital Cartography Laboratory (LASERCA), Universidade do Vale do Rio dos Sinos (UNISINOS), 93022-000, São Leopoldo City, Rio Grande do Sul State, Brazil

*Correspondence author, email ivandroklein@gmail.com


Received 22 October 2012; accepted 9 January 2014

DOI 10.1179/1752270614Y.0000000089

measures), but the robust estimation methods do not show the clear connections with quality measures such as MDBs. Thus, the reliability analysis is outside the scope of this paper (for more details, see, e.g. Baarda, 1968; Wang and Chen, 1994; Knight *et al.*, 2010; Yang *et al.*, 2013).

## Theoretical overview

This section presents a brief theoretical background on the investigated methods in this work. For more details, see the cited references.

### Data snooping procedure

Data snooping is the best established method for identifying outliers in geodetic data analysis (Lehmann, 2012). Initially proposed by Baarda, data snooping assumes that only one outlier is present in the data set. Therefore, in practice, the adjustment and test procedures are continued until there is no outlier (Berber and Hekimoglu, 2003). For correlated data, to each observation (i), the test statistic is given as follows (Teunissen, 2006)

$$T_{\mathrm{i}} = \frac{v^{\mathrm{T}} w c_{\mathrm{i}} c_{\mathrm{i}}^{\mathrm{T}} W \mathbf{v}}{c_{\mathrm{i}}^{\mathrm{T}} W \Sigma_{\mathrm{v}} W \mathbf{c}_{\mathrm{i}}} \tag{1}$$

in which $\mathbf{v}$ is the vector of residuals (computed by a least squares solution), $\mathbf{W}$ is the weight matrix of the observations, $\boldsymbol{\Sigma}_{\mathbf{v}}$ is the covariance matrix of the residuals, and $\mathbf{c}_{\mathrm{i}}$ is the canonical unit vector with 1 as its $i$th entry. The $i$th observation is flagged as an outlier if $T_{\mathrm{i}} > K_{\alpha}$, in which $K_{\alpha}$ is the critical value of a $\chi^2$ distribution, with 1 degree of freedom and a significance test level of $\alpha$. Because all observations are considered suspicious outliers, it was decided that the observation in which the test statistic $T_{\mathrm{i}}$ value is the largest is the outlier, in the case that two or more observations exceed the critical value $K_{\alpha}$ (Berber and Hekimoglu, 2003). For more details of this testing procedure, see, e.g. Baarda, 1968; Teunissen, 2006.

### Danish method for correlated observations

The Danish method is a robust estimator for uncorrelated observations, initially proposed by Krarup *et al.* (1980). The method works by performing a least squares adjustment using the a priori weight matrix (**W**). The process is repeated iteratively, altering the weight matrix until convergence is achieved (Knight and Wang, 2009).

For correlated observations, the Danish method can be performed in the reweighted least squares for correlated observations (RLSCO) scheme. The elements of the equivalent weight matrix ($\mathbf{W}_{\mathbf{k}}$) for the k-iteration of the RLSCO scheme are determined as follows (Wieser and Brunner, 2002)

$$\begin{cases} \bar{q}_{\mathrm{i,j(k)}} = \begin{cases} q_{\mathrm{i,j}} e^{\left(\frac{|\tau_{\mathrm{i(k)}}|}{c}\right)}, & i=j,\ \tau_{\mathrm{i\ (k)}} > c\,. \\ q_{\mathrm{i,j}}, & \text{otherwise} \end{cases} \\ W_{\mathrm{k}} = \bar{Q}_{\mathrm{k}}^{-1} \end{cases} \tag{2}$$

in which $\bar{q}_{\mathrm{i,j\ (k)}}$ is the (i,j) element of the cofactor matrix in the k-iteration ($\bar{\boldsymbol{Q}}_{\mathrm{k}}$), $q_{\mathrm{i,j}}$ is the (i,j) element of the a priori cofactor matrix (see, e.g. Teunissen, 2003), $\tau_{\mathrm{i(k)}}$ is the 'local sensitivity indicator' of the $i$th observation at the k-iteration ($\tau_{\mathrm{i}} = \sqrt{T_{\mathrm{i}}}$, see, e.g. Guo *et al.*, 2010), $\sigma_0^2$ is the variance factor, $\mathbf{W}_{\mathbf{k}}$ is the equivalent weight matrix at the k-iteration, and $c$ is a threshold value, usually taken as 3. Here, only the absolute value of $\tau_{\mathrm{i}}$ is used, but no statistical test is applied.

The iterations start with the a priori weight matrix **W** (conventional least squares estimation), and in each iteration the quantity $\tau_{\mathrm{i}}$ is computed for all observations using the residuals of the current iteration. The value of $\tau_{\mathrm{i}}$ is then compared with a threshold value $c=3$, and the equivalent weight of the corresponding observation is modified according to equation (2), if $\tau_{\mathrm{i}} > c$ (Wieser and Brunner, 2002). To enable convergence of the algorithm and backward testing, the covariance matrix of the residuals is not updated to compute $\tau_{\mathrm{i}}$ during the iterations, and the iterations of the RLSCO are terminated when the equivalent weights have stopped significantly changing. For more details of the proposed scheme, see Wieser and Brunner (2002).

### Robust estimator for correlated observations based on bifactor equivalent weights

The robust estimator for correlated observations that was proposed by Yang *et al.* (2002) is based on a bifactor reduction model of the elements of the weight matrix (**W**). In this bifactor reduction model, the equivalent weight matrix of the observations ($\mathbf{W}_{\mathbf{k}}$) can be determined as follows

$$W_{\mathrm{k}} = \begin{bmatrix} (\gamma_{11})^{1/2} & 0 & \dots & 0 \\ 0 & (\gamma_{22})^{1/2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & (\gamma_{\mathrm{nn}})^{1/2} \end{bmatrix} W \begin{bmatrix} (\gamma_{11})^{1/2} & 0 & \dots & 0 \\ 0 & (\gamma_{22})^{1/2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & (\gamma_{\mathrm{nn}})^{1/2} \end{bmatrix} \tag{3}$$

in which $\gamma_{\mathrm{ii}}$ are the reduction factors of the weight elements. Therefore, a reduction function similar to Huber's $\psi$ function can be chosen as (Yang *et al.*, 2002)

$$\gamma_{\mathrm{ii}} = \begin{cases} 1, & |\bar{v}_{\mathrm{i}}| \le c \\ \dfrac{c}{|\bar{v}_{\mathrm{i}}|}, & |\bar{v}_{\mathrm{i}}| > c \end{cases}. \tag{4}$$

in which $\bar{v}_{\mathrm{j}}$ is the stantardised residual of the $i$th observation (see, e.g. Baarda, 1968), and $c$ is a constant that is chosen to be 1.0–1.5. The equivalent weight matrix ($\mathbf{W}_{\mathbf{k}}$) that is determined by this bifactor reduction model is symmetric and keeps the original correlation coefficients of the observations unchanged. For more details of the method, see Yang *et al.* (2002).

### QUAD method for correlated observations

The QUAD method, initially proposed by Ou (1999) for uncorrelated observations, is based on the functional relationship existing between the observations and the (unknown) true errors.

The principle of the method consists in selecting $r$ appropriate observations (with $r > u$, in which $u$ is the number of unknown parameters), called 'quasi-accurate observations' (QAOs), and estimating the real errors of these observations, under the restriction that the norm of these (estimated) errors is minimum (Ou, 1999). Therefore, the key to the method is how to correctly select the QAOs.

For correlated observations, the selection of QAOs can be formulated using the 'local sensitivity indicators' of the observations. Therefore, the $i$th observation will be selected as quasi-accurate if the following condition is fulfilled (Guo *et al.*, 2007)

$$\tau_i < k\left(\sigma^2\right)^{1/2} \tag{5}$$

in which $\tau_i$ is the local sensitivity indicator of the $i$th observation ($\tau_i = \sqrt{T_i}$; see equation (1) for $T_i$), $\sigma^2$ is the factor of variance of the observations, and $k$ is a constant that serves as a threshold value and usually can be taken from the interval [1.5, 2.5]. Only the absolute value of $\tau_i$ is used, but no statistical test is applied.

Because the variance factor is unknown in many practical applications, the parameter $\sqrt{\sigma^2}$ may be replaced with its robust least median of squares estimator (Guo *et al.*, 2007)

$$\tilde{\sigma} = 1{\cdot}4826\left(\underset{i}{\text{median}}\ \tau_i^2\right)^{1/2} \tag{6}$$

Therefore, after selecting QAOs, to minimise their (estimated) real errors, a vector of unknown parameters can be estimated as

$$\mathbf{x}_r = (A_r^T W_r \mathbf{A}_r)^{-1} A_r^T W_r \mathbf{l}_r \tag{7}$$

in which $\mathbf{x_r}$ is the vector of the estimated parameters, $\mathbf{A_r}$ is a partitioned sub-matrix of the design matrix $\mathbf{A}$ corresponding to the QAOs, $\mathbf{W_r}$ is the inverse of the partitioned sub-matrix of the cofactor matrix corresponding to the QAOs, and $\mathbf{l_r}$ is the sub-vector of the full vector of observations ($\mathbf{l}$), also corresponding to the selected QAOs. With this estimated vector of parameters (considering only the QAOs), the residuals of all the observations can be determined as

$$v = A\, x_r - \mathbf{l} \tag{8}$$

If the QAOs are selected reasonably, the estimates of the associated residuals with the (potential) outliers would be far away from the bulk of the others. This interesting phenomenon is metaphorically called 'hive-off' (Guo *et al.*, 2007). Based on the 'hive-off' phenomenon, the $q$ outlying observations can be identified. Thus, the gross errors contaminating these $q$ observations can be estimated by the least squares principle, and then the vector of unknown parameters ($\mathbf{x}$) can be estimated, minimising the influence of the outlying observations and resulting in (Teunissen, 2006)

$$\begin{cases} \nabla = (C_q^T \mathbf{R} \mathbf{C}_q)^{-1} C_q^T \mathbf{R} l \\ x = (A^T W A)^{-1} A^T W (l - \mathbf{C}_q \nabla)' \end{cases} \tag{9}$$

in which $\nabla$ is the vector of (estimated) $q$ outliers, $\mathbf{C}_q$ is a matrix containing $q$ canonical unit vectors $\mathbf{c}_i$ (each one referred to an outlying observation), and $\mathbf{R}$ is the so-called 'redundancy matrix' (see, e.g. Guo *et al.*, 2007; Schaffrin, 1997). For more details on the QUAD method, see Ou (1999), and Guo *et al.* (2007).

## Robust estimator for correlated observations based on local sensitivity downweighting strategy

Another approach to robust estimation with correlated observations was recently proposed in Guo *et al.* (2010). The downweighting strategy is performed in a similar way to equation (3), but the reduction factors that are used are based on the test statistics $\tau_i = \sqrt{T_i}$ (see equation (1) for $T_i$), also called local sensitivity indicators. Thus, the reduction factors of the weight elements can be determined as Guo *et al.* (2010)

$$\gamma_{ii} = \begin{cases} 1, & |\tau_i| \le z_{\alpha/2} \\ \dfrac{z_{\alpha/2}}{|\tau_i|}, & |\tau_i| > z_{\alpha/2} \end{cases}. \tag{10}$$

in which $z_{\alpha/2}$ is the upper $\alpha$-percentage point of the standard normal distribution ($\alpha$ being the significance test level), and for routine scientific data, can be taken as a constant from the interval [1.0, 2.0]. This downweighting strategy is computationally inexpressive because the necessary information can be obtained without extra calculations (Guo *et al.*, 2010).

## Generalised test for multiple (simultaneous) outliers

Last, the generalised test for multiple outliers, based on the likelihood ratio (being the most uniformly powerful test), has the following test statistic (Teunissen, 2006)

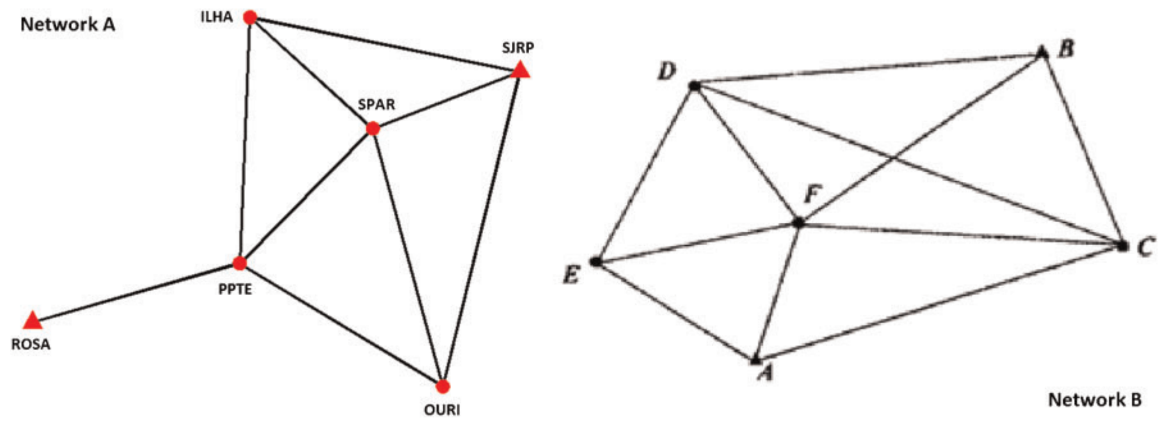
$$T_q = \frac{v^T W \mathbf{C}_q C_q^T W v}{C_q^T W \Sigma_v W \mathbf{C}_q} \tag{11}$$

in which the number of outliers considered, $q$, must satisfy the inequality $1 \le q \le f$, in which $f$ represents the degrees of freedom of the adjustment. The $q$ observations are flagged as outliers (simultaneously) if $T_q > K_\alpha$, in which $K_\alpha$ is the critical value of the $\chi^2$ distribution, with $q$ degrees of freedom and significance test level $\alpha$. Because there are $C_q^{\,n}$ combinations of the $\mathbf{C_q}$ matrix that can be formed for $q$ outliers, there are also $C_q^{\,n} T_q$ test statistics (Knight *et al.*, 2010). It should be noted that the test statistic $T_i$ in equation (1) is a particular case of this generalised test statistic, when $q=1$.

Unfortunately, in practice, the true number of outliers is unknown, and all that can be obtained is an estimate of the reasonable maximum number of outliers to be encountered (Knight *et al.*, 2010). Similarly to the QUAD method, the gross errors contaminating the q outlying identified observations can be estimated by the least squares principle, and then the vector of unknown parameters can be estimated, minimising the influence of the outlying observations (see equation (9)). For more details on the test for multiple outliers, see Teunissen (2006) and Knight *et al.* (2010).

# Experiments and results

In this study, two GNSS baseline networks are analysed (network A and network B). In network A, the stations 'ROSA' and 'SJRP' are considered control (fixed) stations, and stations 'PPTE', 'SPAR', 'OURI', and 'ILHA' are considered points of unknown three-dimensional (3D) position. In network B, the stations 'A' and 'B' are considered control (fixed) stations, and stations 'C', 'D', 'E', and 'F' are considered points of unknown 3D position (see Fig. 1). For each network, there are 12




1 GNSS networks

measured baseline vectors (with 3 baselines 'repeated' in network A and 1 baseline 'repeated' in network B). Thus, there are 36–12=24 redundant observations for each one.

For network A, the processed GNSS data contain 6 h independent sessions, and all the stations belong to the Brazilian Network for Continuous Monitoring of GNSS systems, with the data available (free of charge) at the following electronic page: http://www.ibge.gov.br/home/geociencias/geodesia/rbmc/rbmc.shtm?c=7.

The 12 independent considered baselines are PPTE–SPAR, PPTE–ILHA, PPTE–OURI, SJRP–SPAR, SJRP–ILHA, SJRP–OURI, SPAR–ILHA, SPAR–OURI, ROSA–PPTE, and the 'repeated occupations' ILHA–SPAR, OURI–SPAR, and SPAR–SJRP. The correlation coefficients between components of the same baseline lie in the magnitude interval of [0.2237, 0.7962]. For more details on network A, such as the precision and length of the baselines, see Klein *et al.* (2012).

For network B, the data were obtained from Ghilani and Wolf (2006), and the 12 independent considered baselines are AC, AE, BC, BD, DC, DE, FA, FC, FE, FD, FB, FB, and the 'repeated occupation' BF. The correlation coefficients between components of the same baseline lie in the magnitude interval of [0.0085, 0.0117]. For more details on the network, see Ghilani and Wolf, 2006.

To compare the methods of quality control, outliers between $3\sigma$ and $9\sigma$ magnitude (positives and/or negatives) are randomly generated and added to some observations ($\sigma$ being the respective standard deviation of the observation; see Ghilani and Wolf, 2006; Klein *et al.*, 2012). For each network, 15 000 scenarios are performed, 5000 with one outlier, 5000 with two outliers, and 5000 with three outliers, using Monte–Carlo simulations. It should be mentioned that, before the insertion of outliers, all the methods were applied in both networks, and no outliers were found.

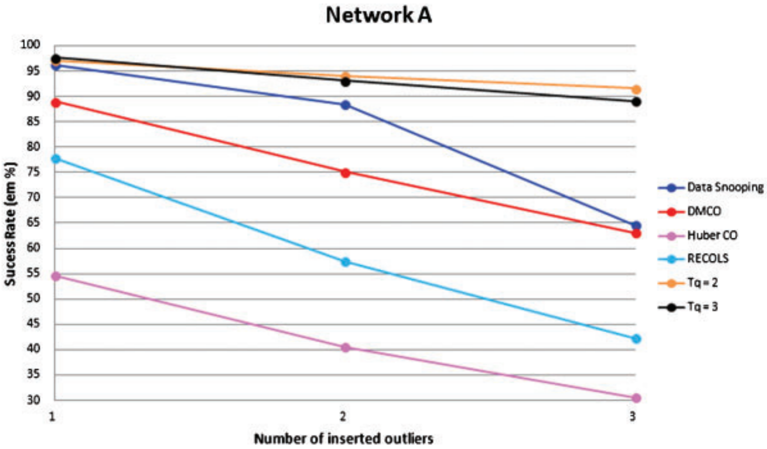
Concerning the data snooping procedure (DS), the significance level is taken as $\alpha=0.001$ (the typically adopted value, see, e.g. Baarda, 1968), so the critical value of the $\chi^2$ distribution for the test statistic $T_i$ (see equation (1)) is $K_\alpha=10.83$.

For the Danish method for correlated observations (DMCO), according to Wieser and Brunner (2002), the constant $c$ in equation (2) is taken as $c=3$, the factor of variance is considered known and taken as $\sigma_0^2=1$, and the convergence criterion is when the difference between the estimated parameters at two consecutive iteration steps is less than 1 mm. For the robust estimation for correlated observations based on bifactor equivalent weights using Huber's $\psi$ function (Huber CO), the same convergence criterion of the Danish method scheme for correlated observations is considered, and the constant $c$ in equation (4) is taken as $c=1.5$. For the robust estimation for correlated observations with the sensitivity based downweighting strategy given by equation (10) (RECOLS), the critical value is taken as $z_{\alpha/2}=1.5$, and the same convergence criterion of the other robust estimators in this study is considered [see Guo *et al.* (2010) and Yang *et al.* (2002)].

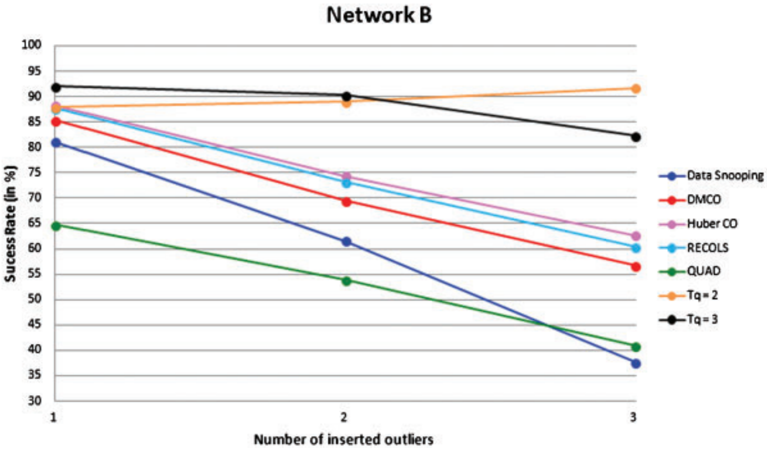
For the QUAD method, the constant $k$ in equation (5) is taken as $k=1.5$, and the parameter $\sqrt{\sigma^2}$ is replaced with its robust least median of squares estimator, given by equation (6). When rank deficiency occurs, the constant $k$ is systematically increased by 0.25 until the rank deficiency disappears. Because the QUAD method requires visual analysis of the hive-off phenomenon to identify the outlying observations, empirical data were used to produce the threshold that observations exceeding nine times the median of the residuals would be considered contaminated by outliers in the performed simulations.

Finally, to test multiple outliers, it is assumed that the reasonable maximum number of outliers is $q=3$ (less than 10% of the total of $n=36$ observations). Thus, the tests are performed for $q=2$ and 3 outliers (disregarding the combinations of suspicious observations that result in rank deficiency). The critical values for $q=2$ and 3 outliers are given by $K_\alpha=11.83$ and 12.45 respectively, and they were obtained considering the same test power (fixed in $\gamma=0.8$) of the data snooping procedure (for more details, see, e.g. Baarda, 1968; Teunissen, 2006; Knight *et al.*, 2010).

The DS and QUAD methods are considered successful when only the contaminated observation(s) is (are) correctly identified and excluded. For the DMCO, Huber CO and RECOLS, as by Knight and Wang (2009), the method is considered successful when, at the end of the iterations, only the contaminated observation(s) shows residual(s) higher than three standard deviation(s) of the observation(s). The tests for multiple outliers are considered 'successful' when the contaminated observations show the higher test statistic value, and it exceeds the critical value. When the number of outliers is smaller than the number considered in the test




**2 Number of inserted outliers versus success rate of methods for Network A**




**3 Number of inserted outliers versus success rate of methods for Network B**

(e.g. 2 outliers in the test for $q=3$ outliers), the test is considered successful if the higher test statistic corresponds to the contaminated observations (correctly) plus any other(s), and when the number of outliers is higher than the number considered in the test (e.g. 3 outliers in the test for $q=2$ outliers), the test is considered successful if the higher test statistic corresponds to some contaminated observations (correctly).

Table 1 (see also Fig. 2) shows the successful rate (number of outliers correctly identified/number of experiments) of the methods for network A, and Table 2 shows the same for network B (see also Fig. 3). Figure 4 shows the total success rate of the methods for both networks (for all 15 000 simulations with one, two and three inserted outliers).

## Discussion

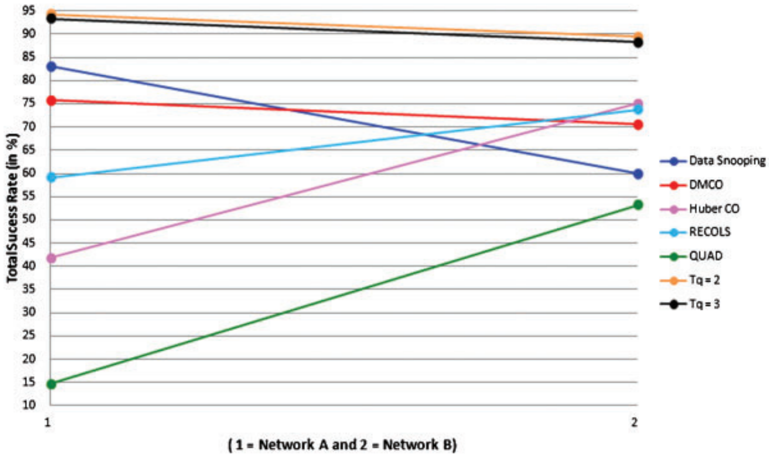
It can be concluded from an analysis of Tables 1 and 2 (and Figs. 2 and 3), that, in general, the higher the number of outliers, the lower the efficiency of the methods for both networks. The statistical tests for $T_{q=2}$ and $T_{q=3}$ outliers have higher success rates for both networks (between 82 and 98% for all tested cases),

**Table 1 Successful rate (in %) of DS, DMCO, Huber CO, RECOLS, QUAD, $T_{q=2}$ and $T_{q=3}$ (for network A)**

| Scenario | DS | DMCO | Huber CO | RECOLS | QUAD | $T_{q=2}$ | $T_{q=3}$ |
|---|---|---|---|---|---|---|---|
| 1 outlier | 96.24 | 88.94 | 54.68 | 77.80 | 29.06 | 97.14 | 97.64 |
| 2 outliers | 88.45 | 75.06 | 40.56 | 57.38 | 10.56 | 94.12 | 93.18 |
| 3 outliers | 64.58 | 63.14 | 30.62 | 42.34 | 4.50 | 91.54 | 89.14 |
| Total | 83.09 | 75.71 | 41.95 | 59.17 | 14.71 | 94.27 | 93.32 |

**Table 2 Successful rate (in %) of DS, DMCO, Huber CO, RECOLS, QUAD, $T_{q=2}$ and $T_{q=3}$ (for network B)**

| Scenario | DS | DMCO | Huber CO | RECOLS | QUAD | $T_{q=2}$ | $T_{q=3}$ |
|---|---|---|---|---|---|---|---|
| 1 outlier | 81.10 | 85.34 | 88.20 | 87.72 | 64.76 | 87.94 | 92.02 |
| 2 outliers | 61.50 | 69.58 | 74.34 | 73.14 | 53.98 | 88.94 | 90.30 |
| 3 outliers | 37.58 | 56.66 | 62.70 | 60.52 | 40.98 | 91.74 | 82.36 |
| Total | 60.06 | 70.53 | 75.08 | 73.79 | 53.24 | 89.54 | 88.23 |




**4 Total success rate of methods for both networks**

whereas the QUAD method has the lowest success rate for both networks and the largest variation (between 65 and 4% for the cases tested). However, the tests for multiple outliers only indicate the observations that are most likely to be contaminated, considering the number of fixed outliers (which in practice is unknown). Nevertheless, for the cases in which two outliers are inserted in Network A and B, the success rate of the test for $T_{q=2}$ outliers is 94.12 and 88.94 respectively, whereas for the cases in which three outliers are inserted in Network A and B, the success rate of the test for $T_{q=3}$ outliers is 89.14 and 82.36 respectively.

The main limitation of the QUAD method is the dependence of visual analysis for the hive-off phenomenon. As described previously, the hive-off phenomenon is substituted for a constant criterion, which limits the performance of the method in the simulations. For every network and number of inserted outliers, a suitable constant criterion must be empirically selected for better performance. Thus, the QUAD method is more difficult to analyse in several simulations than the other tested methods. Additionally, the QUAD method can locate multiple outliers correctly, but because it depends on the hive-off phenomenon, it cannot be efficient for outliers of small magnitude, as noted by Klein (2012). However, the QUAD method has a (relatively) low computational cost and does not have one of the main limitations of the tests for multiple outliers, which is the correct definition of the number of considered outliers ($q$). Thus, the QUAD method can be used to locate outliers of high magnitude simultaneously, before the rigorous testing procedures to locate remaining outliers in the data set.

Concerning robust estimation methods, the DMCO shows the best results for both networks, and the performance is between 5 and 10% better for Network A (in which the correlations of the observations are higher) than for Network B, whereas the methods Huber CO and RECOLS show better performance in network B (in which the correlations of the observations are smaller). The Huber CO and RECOLS methods perform similarly for network B, whereas RECOLS performs better than Huber CO for network A (see Fig. 4). The Huber CO scheme uses the standardised residuals of the observations, and the RECOLS scheme uses the local sensitivity indicators of the observations, so that it becomes equal to the standardised residuals in the case of uncorrelated observations. Thus, both strategies show similar results for network B because the low correlation coefficients between the components of the same baselines, whereas RECOLS outperforms Huber CO for network A because of the higher correlation coefficients between the components of the same baselines.

It should also be noted that Huber CO uses the standardised residuals of the observations, whereas DMCO and RECOLS use the test statistics $\tau_i$, which provides the most uniform powerful tests. Most likely, DMCO outperforms RECOLS because of the downweighting strategy: the threshold value for the $\tau_i$ test statistics is given by $c=3.0$, instead of $z_{\alpha/2}=1.5$ (consequently, the probability of downweighting good observations is lower). Moreover, the downweighting strategy is an exponential function (see equation (2)), instead of the proportional factor $z_{\alpha/2}/\tau_i$ of the latter (in which the weights of the observations are further reduced in the former). However, the DMCO does not keep the original correlation coefficients of the observations unchanged, unlike Huber CO and RECOLS. Most likely, the null or small covariances of the observations of the analysed GNSS networks (especially for the case of network B) do not affect the performance of the DMCO. More studies are required to address these topics.

The data snooping (DS) procedure shows a success rate better than 60% in all simulated cases, except for the case of three outliers in network B (in which the success rate is 37.58%). The tests for $T_{q=2}$ and $T_{q=3}$ outliers show better performance than DS for all simulated cases, but the DS does not have the limitation that the number of considered outliers must be fixed, being applied iteratively. Nevertheless, data snooping and the tests for multiple outliers can be susceptible to some pitfalls, such as the masking and/or swamping phenomenon (see, e.g. Gui *et al.*, 2010; Klein, 2012). As an example, considering network B, if a systematic error of $+20$ cm is added to the $\Delta Z$ components of the baselines FA, FC, FE, FD, and FB (excluding the baseline BF from the model) for the data snooping procedure, the maximum test statistic results in $T_i=3.87$ for the observation $\Delta X$ AE. Thus, no observation is flagged as an outlier at the significance level of $\alpha=0.001$. However, for $T_{q=5}$, the maximum test statistic results in $T_{q=5}=88.21$ for this quintet of observations ($\Delta Z$ FA, FC, FE, FD, FB). Most likely, the test for $T_{q=5}$ outliers correctly identified all the contaminated observations because network B contains more $\Delta Z$ components relative to the unknown points C, D, and E, whereas the data snooping procedure considers only one suspicious observation at a time, i.e. the systematic (common) error in multiple observations cannot be detected by data snooping.

Another important point is that the data snooping procedure, DMCO, QUAD, RECOLS, and the tests for multiple outliers make use of the test statistic $T_q$, or similar, the test statistic $\tau_i = \sqrt{T_i}$ in the case of a single considered outlier, and Huber CO uses the standardised residuals, which become equal to the test statistic $\tau_i$ in the case of uncorrelated observations. Thus, the constants that are used as threshold values can be related to the significance levels of test statistics. For example, consider the data snooping procedure using the test statistic $\tau_i$ (that in the null hypothesis follows the normal distribution; see, e.g. Teunissen, 2006), if the significance level of the test is taken by $\alpha=0.01$ (instead of $\alpha=0.001$), the critical value would be given by 2.33. In some cases for network A (in which DS originally fails), with this new significance level, all the outlying observations would be correctly identified and removed from the model. Thus, using constants for threshold values lying in the interval [1.5, 3.0] for the test statistic $\tau_i$, the success rate of the data snooping procedure can be increased. The choice of the significance level of the data snooping procedure is intrinsically related to the success of the test (see, e.g. Lehmann, 2012).

Concerning the tests for multiple outliers and the data snooping procedure, DS has the disadvantage of the iteration adjustment procedure each time an observation is flagged as an outlier, and the tests for multiple outliers have the disadvantage of the (possible) rank deficiency for some combinations of suspicious observations, in addition to the problem of how to reasonably determine the number of outliers present in the observations. As an example, for network A, consider the case in which three outliers of $3\sigma$, $6\sigma$, and $9\sigma$ magnitude are added to the $\Delta Z$ components of the baselines SJRP–OURI, PPTE–OURI and ILHA–SPAR respectively. If the alternative hypothesis of the test for $q=2$ outliers (the pair of observations $\Delta Z$ ILHA–SPAR and $\Delta Z$ PPTE–OURI had outliers) becomes the null hypothesis, and the adjustment is performed by estimating these outliers using equation (9), and the alternative hypothesis of the test for $q=3$ outliers (the trio of observations $\Delta Z$ ILHA–SPAR, $\Delta Z$ PPTE–OURI and $\Delta Z$ SJRP–OURI had outliers) becomes another alternative hypothesis, i.e., another parameter is added to the model (the gross error in the observation $\Delta Z$ SJRP – OURI) using the test statistic for this new case (in which $q=1$) results in $T_i=10.33$ and is identical to the data snooping procedure, the observation $\Delta Z$ SJRP–OURI is not flagged as an outlier (at the significance level of $\alpha=0.001$). Thus, the test for $q=3$ outliers does not produce a significantly different result for the test for $q=2$ outliers, and there is no testing procedure that is 100% efficient to locate multiple outliers using a least squares adjustment, as also noted by Baselga (2011). In general, the similar performances of the tests for $q=2$ and $q=3$ outliers in all simulated cases also appointed it (see, for e.g. Fig. 4).

Finally, analysing the last row of Tables 1 and 2 and Fig. 4, it can be concluded that the tests for multiple outliers and the DMCO show success rates that are less sensitive to the different values of the network correlation coefficients than the other methods.

## Conclusions

Many methods of quality control for geodetic data adjustment have been developed and investigated since the pioneering work by Baarda (1968). However, these methods still require more investigation, especially in the case of multiple outliers and/or correlated observations. Thus, the goal of this paper was the evaluation, comparison, and discussion of recent methods for quality control in the general case of correlated observations and multiple outliers. The methods performed and analysed were the data snooping procedure (Baarda, 1968; Teunissen, 2006), the Danish method scheme for correlated observations proposed by Wieser and Brunner (2002), the robust estimator for correlated observations based on bifactor equivalent weights (Yang *et al.*, 2002), the QUAD method for correlated observations (Guo *et al.*, 2007), the robust estimator for correlated observations based on local sensitivity downweighting strategy (Guo *et al.*, 2010), and the statistical tests for multiple outliers based on the likelihood ratio (Teunissen, 2006; Knight *et al.*, 2010).

To compare the methods of quality control, outliers between $3\sigma$ and $9\sigma$ magnitude (positives and/or negatives) are randomly generated and added to some observations ($\sigma$ being the respective standard deviation of the observation; see Ghilani and Wolf, 2006; Klein *et al.*, 2012). For each network, 15 000 scenarios were performed, 5000 with one outlier, 5000 with two outliers, and 5000 with three outliers, using Monte–Carlo simulations.

Concerning the robust analysed estimators, the scheme proposed by Wieser and Brunner (2002) (DMCO in this paper) provides the best results for both networks, unlike the scheme proposed by Guo *et al.* (2010) (RECOLS in this paper), which also uses the local sensitivity indicators of the observations. The best performance of the former is most likely because of the computational aspects (with a larger threshold value, which reduces the probability of downweighting 'good' observations, and with the downweighting strategy as an exponential function instead of the proportional factor of the latter, which further reduced the weights of the observations). The Huber CO and RECOLS methods show similar results for network B because of the low correlation coefficients between the components of the same baselines, whereas the RECOLS outperforms the Huber CO for network A because of the higher correlation coefficients between the components of the same baselines.

The QUAD method for correlated observations can locate multiple outliers correctly, but because it depends on the hive-off phenomenon, it cannot be efficient for outliers of small magnitude. Moreover, after QAOs, and depending on the choice of the constant $k$, rank deficiency may occur in the system. However, the QUAD method has a (relatively) low computational cost and does not require consideration of the correct definition of the number of outliers ($q$), one of the main limitations of the tests for multiple outliers. Thus, the QUAD method can be performed to locate outliers of high magnitude simultaneously, before the rigorous testing procedures, to locate remaining outliers in the data set. In this study, because of the many simulations, the hive-off phenomenon is substituted for a constant criterion, which limited the performance of the QUAD method.

The test for multiple outliers provides the best results for both networks. Moreover, it can locate systematic (common) error in multiple observations that cannot be identified by the data snooping procedure. However, its main limitation is how to determine a reasonable number of outliers to be considered. Additionally, the (relatively) high computational cost, the possibility of rank deficiency (depending on the suspicious observations considered), and some pitfalls, such as the swamping and masking phenomenon (see, for e.g. Gui *et al.*, 2010), support the assertion noted by Baselga (2011) that there is no testing procedure that is 100% efficient in locating multiple outliers in the framework of a conventional least squares adjustment.

Finally, the tests for multiple outliers and the DMCO show success rates that are less sensitive to the different values of the correlation coefficients of the networks than the other methods.

Future studies should consider the effect of covariance on the robust estimators (e.g. the influence of keeping the original correlation coefficients unchanged or not in the downweighting strategy), other approaches to select the quasi-accurate observations with the QUAD method (e.g. to avoid the rank deficiency or the incorrect selection of contaminated observations), the selection of the significance level of the data snooping procedure (this has a large influence on the outlier identification), and how to determine a reasonable number of outliers to be considered for the tests of multiple outliers.

## Acknowledgements

The authors thank CAPES and FAPERGS for financial support provided to the first author (PhD scholarship), CNPq for financial support provided to the second author (proc. n.307472/2009-4 and proc. n.477914/2012-8), and IBGE for the GNSS data.

## References

Baarda, W., 1968. A testing procedure for use in geodetic networks. Netherlands Geodetic Commission, Publications on Geodesy, New Series, 2, pp. 1–60.

Baselga, S., 2011. Non existence of rigorous tests for multiple outlier detection in least squares adjustment. *Journal of Surveying Engineering*, 137(3), pp. 109–12.

Berber, M. and Hekimoglu, S., 2003. What is the reliability of conventional outlier detection and robust estimation in trilateration networks?. *Survey Review*, 37(290), pp. 308–18.

Ghilani, C. D. and Wolf, P. R., 2006. *Adjustment computations: spatial data analysis*. 4th ed. New York: John Wiley & Sons.

Gui, Q., Li, X., Gong, Y., Li, B. and Li, G., 2011. A Bayesian unmasking method for locating multiple gross errors based on posterior probabilities of classification variables. *Journal of Geodesy*, 85(4), pp. 191–203.

Gui, Q., Li, G. and Ou, J., 2005. Robust-biased estimation based on quasi-accurate detection. *Journal of Surveying Engineering*, 131(3), pp. 67–72.

Guo, J. F., Ou, J.K., and Wang, H.T., 2007. Quasi-accurate detection of outliers for correlated observations. *Journal of Surveying Engineering*, 133(3), pp. 129–33.

Guo, J., Ou, J. and Wang, H., 2010. Robust estimation for correlated observations: two local sensitivity-based downweighting strategies. *Journal of Geodesy*, 84(4), pp. 243–50.

Huber, P. J., 1964. Robust estimation of a location parameter. *Annals of Mathematical Statistics*, 35, pp. 73–101.

Klein, I., Matsuoka, M.T., de Souza, S.F. and Collischonn, C., 2012. Design of geodetic networks reliable against multiple outliers (in Portuguese). *Boletim de Ciências Geodésicas*, 18(3), pp. 480–507.

Klein, I., 2012. *Quality control on the adjustment of geodetic observations (in Portuguese)*. Dissertation, Universidade Federal do Rio Grande do Sul, Brazil.

Knight, N. L. and Wang, J., 2009. A comparison of outlier detection procedures and robust estimation methods in GPS positioning. *Journal of Naviagtion*, 62, pp. 699–709.

Knight, N. L., Wang, J., and Rizos, C., 2010. Generalised measures of reliability for multiple outliers. *Journal of Geodesy*, 84(10), pp. 625–35.

Kok, J. J., 1984. On data snooping and multiple outlier testing. NOAA Technical Report NOS NGS. 30, U. S. Department of Commerce, MD, USA.

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

Lehmann, H., 2012. Improved critical values for extreme normalized and studentized residuals in Gauss–Markov models. *Journal of Geodesy* (online first), doi:10.1007/s0019001205690

Ou, J. K., 1999. A new method of identifying and locating gross errors: Quasi-accurate detection. *Chinese Science Bulletin*, 44(23), pp. 2200–2204.

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.

Schaffrin, B., 1997. Reliability measures for correlated observations. *Journal of Surveying Engineering*, 123(3), pp. 126–37.

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.

Wang, J. and Chen, Y. Q., 1994. On the localizability of blunders in correlated coordinates of junction points in densification networks, *Australian Journal of Geodesy, Photogrammetry and Surveying*, 60, pp. 109–19.

Wang, J. and Chen, Y. Q., 1994. On the reliability measures of observations. *Acta Geodaetica et Cartographica Sinica* (English Edition), *Journal of Chinese Society of Geodesy, Photogrammetry and Cartography*, pp. 42–51.

Wieser, A. and Brunner, F. K., 2002. Short static GPS sessions: robust estimation results. *GPS Solutions*, 5(3), pp. 70–79.

Yang, Y., 1994. Robust estimation for dependent observations. *Manuscripta Geodaetica*, 19, pp. 10–17.

Yang, Y., Song, L. and Xu, T., 2002. Robust estimator for correlated observations based on bifactor equivalent weights. *Journal of Geodesy*, 76(6–7), pp. 353–358.

Yang, L., Wang, J., Knight, N.L. and Shen, Y., 2013. Outlier separability analysis with a multiple alternative hypotheses test. *Journal of Geodesy*, 87(6), pp. 591–604.