

Local Statistical Testing in Quality Control of GNSS Observations

Ahmed El-Mowafy

Department of Spatial Sciences, Curtin University, Australia

Abstract

The paper discusses GNSS statistical-testing methods for detection and identifications of observation outliers at the estimation “current” epoch, defined here as local testing. First, detection methods are discussed including testing the probability distribution (likelihood) of the residuals, testing using dynamic control limits of the range, the mean and standard deviation of the residuals. A method is proposed examining the difference between the mean and the median of the residuals. To identify the satellites with faulty measurements, several methods were investigated, including checking the likelihood of the residuals at the present epoch, and in a time series. A test is presented utilizing the control limits of the residuals' moving range for each satellite. Testing of the proposed methods was carried out using only GPS phase measurements in the kinematic mode. Results show that the proposed methods are efficient for detection and identification of large errors/outliers. However, the performance degrades with error values less than 5 cycles and when using small significance levels.

Keywords: Quality Control, Statistical Testing, Positioning, GNSS.

1. Introduction

Quality control (QC) of a system generally refers to checking that the system can deliver outputs that meet application requirements. This comprises many aspects such as QC of observations, QC of system mathematical model, system functionality, reliability (integrity), capabilities, accuracy, precision, robustness, response time, etc. In practice, positioning by GNSS is based on computing the user location from satellite-to-user range observations, which are biased and contain random errors. Thus, quality control of the observations should be performed whenever a new set of observations are used.

Quality control of observations in positioning by GNSS is needed to attain optimality of the solution, where errors such as outliers or blunders, e.g. receiver clock

jumps or ionospheric scintillation, which are not specified in the functional and stochastic models, must be correctly detected and removed. Detection and identification methods can also be formulated for testing other error forms, such as slips or errors in the dynamic state modelling. However, this study only addresses detection and identification of outliers. This process can be performed based on statistical hypothesis testing (Kim and Langley, 2001). One should note that validation through hypothesis testing relies on redundancy and the use of good known or well-approximated stochastic models. The presence of redundant observations allows us to assess whether the model and observed data are statistically consistent. In principle, statistical observation testing allows us to decide on the basis of the collected observations and used model, whether or not an anomaly is present (Teunissen and Kleusberg, 1998). However, due to the finiteness of the available samples in GNSS positioning, no definite statistical decision can be made in some cases (Kuang, 1996).

Statistical observation testing involves three tasks. Firstly, detection of the presence of outliers. Secondly, identification of individual observations that carry these outliers. Finally, once outliers are captured; one should either exclude the observations associated with them or modify modelling of the system parameters to adopt these errors.

In this paper, statistical testing is discussed for the case of testing measurements at the current epoch. This testing procedure is defined here as *local* testing. Parameters from measurements from previous epochs may be used according to the type of test under consideration, but under the assumption that these measurements were previously tested in the same manner and are error free. One should not confuse this procedure with “*global*” testing presented in the literature (see for instance, Teunissen and Salzmann, 1989, Teunissen 1990, Teunissen and Kleusberg, 1998). The former case tries to detect outliers that may take place in the current epoch, while in the latter case of “*global*” testing, it is assumed that outliers take place or

started in an earlier epoch and testing is performed at a later epoch (the present epoch).

Local testing can be applied in Kalman filtering or in epoch-by-epoch or sequential least squares estimation. In this study, several tests are presented. The set of the presented tests include traditional tests used in geodesy, in addition to quality control tests that are used in the industry, formulated for testing GNSS observations. In addition, newly developed simple tests are presented.

2. Local Detection Testing

Local testing of observations for the presence of outliers can be performed by examining the statistical behaviour of the observational residuals. These are defined as the discrepancy between measured values of the observables and their “modelled” values computed from the estimated unknowns. The observational residuals indicate the extent to which the measurements agree with the model. Measurement residuals can be statistically tested epoch-by-epoch. In traditional statistical testing, a test statistic is formed and the null-hypothesis H_0 is proposed to form a reference level, which is tested against an alternative-hypothesis H_a . The null-hypothesis H_0 indicates a fault-free situation, i.e. no

outliers exist in the measurements or no missmodelling, and the distributional assumptions meet the reality. The alternative-hypothesis H_a indicates the presence of outliers, and can be used to describe the type of misspecifications in the models (Baarda, 1968; Teunissen and Kleusberg, 1998; Leick, 2004, Hwang and Brown, 2008). Assuming the use of a correct measurement model, a 'detection' test can be applied to check for the presence of outliers in the whole data set at the estimation epoch. Detection testing requires the presence of at least one degree of freedom. Several detection tests are presented in the following subsections. A traditional detection test is given first followed by the contribution of this paper in presenting additional statistical tests and analysing their performance.

2.1 Likelihood testing

Using the Gauss-Markov observation model, the measurement y_i can be modelled as a function (h) of the unknowns (\mathbf{x}). The measurement residuals can be computed from the estimated unknowns as:

$$r_i = y_i - h(\mathbf{x}) \quad (1)$$

A local 'detection' test statistic can be formulated including all measurements at the epoch of estimation as follows (Teunissen *et al.*, 2008):

$$T = \mathbf{r}^T \mathbf{C}_y^{-1} \mathbf{r} \quad (2)$$

Where \mathbf{C}_y is the covariance matrix of the observations, and \mathbf{r} is the vector of residuals.

When no outliers are present in the measurements, the residuals can be assumed normally distributed with a zero mean, which underline the null hypothesis (H_0). In this case, the weighted sum of their squares should have a central Chi-square distribution assuming a known standard deviation. In the case of experiencing outliers, the residuals grow dramatically and the sum of their squares does not follow a central Chi-square distribution. This can be formulated as the alternative hypothesis H_a . Thus, the outliers detection test can then be given as (Vaniček and Krakwski, 1986):

$$H_0: T \leq \chi_{1-\alpha,df}^2 \quad (3)$$

$$H_a: T > \chi_{1-\alpha,df}^2 \quad (4)$$

Where $\chi_{1-\alpha,df}^2$ is the Chi-square value for $1-\alpha$, where α is the significance level, i.e. probability of Type I error, representing false alarm and df degrees of freedom, which equals $(n-u)$, where n is the number of observations and u denotes the number of unknowns.

In addition to the likelihood test, simple detection tests can be formed. In the following tests, it is assumed that the residuals of phase observations at one epoch are of the same type, and under H_0 , have an expected value (equals zero) and follow a normal distribution. Thus, uni-variant statistical testing methods can be applied. Clearly, this is also applicable when considering the residuals of phase observations for the same satellite between epochs.

2.2 Testing the mean of residuals between the current and previous epochs

Assume that at the computation epoch we have n_2 phase residuals with a mean value (\bar{r}_2) and a standard deviation (S_2), and for the previous epoch, the phase residuals of the same set of satellites has a mean value (\bar{r}_1) and standard deviation (S_2). We wish to test the assumption that the residuals unknown population means μ_1 and μ_2 do not exceed a threshold value Δ_0 . In the absence of outliers, this threshold can be assumed based on the expected value of the combined random and quasi-random errors. However, at subsequent epochs, the differences due to ionospheric and tropospheric errors, and multipath, should be minimal, leaving only errors due to noise. Thus, Δ_0 can be selected as a few millimetres. In one approach, we can consider that the two samples come from similar populations of equal variance, i.e. $\sigma_1^2 = \sigma_2^2$, with a pooled estimator of σ^2 , denoted by S_p^2 computed as (Montgomery, 2005):

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \quad (5)$$

The test statistic can then be formulated as:

$$t_0 = |\bar{r}_1 - \bar{r}_2| / -\Delta_0 \quad (6)$$

and the detection test criterion will be:

for H_0 : $|\mu_1 - \mu_2| = |\Delta_0|$,

$$-S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} t_{\alpha/2, n_1+n_2-2} < t_0 \leq S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} t_{\alpha/2, n_1+n_2-2} \quad (7)$$

and for H_a : $|\mu_1 - \mu_2| > |\Delta_0|$,

$$t_0 > S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} t_{\alpha/2, n_1+n_2-2} \quad (8)$$

where $t_{\alpha/2, n_1+n_2-2}$ is the t -distribution value with $(n_1 + n_2 - 2)$ degrees of freedom. The alternative hypothesis, if proven, will indicate the presence of large errors or outliers in the data. Since this test is applied sequentially in time, moving from one error-free epoch to a next epoch will indicate that the suspected data are from the latter. However, it should be noted that the presence of outliers in the tested data set would amplify the value of S_p , which may result in Type II error i.e. accepting H_0 when it is false. To rectify this shortcoming, a population standard deviation (σ) can be used, which can be selected from experience or by studying a large number of data sets and based on the expected level of noise, quasi-random errors, and dynamics. The test criterion in this case can be formulated as:

$$H_0: |\bar{r}_1 - \bar{r}_2| = \Delta_0, \\ -\sigma Z_{\alpha/2, n_1+n_2-2} < t_0 \leq \sigma Z_{\alpha/2, n_1+n_2-2} \quad (9)$$

$$H_a: |\bar{r}_1 - \bar{r}_2| > \Delta_0, \\ t_0 > \sigma Z_{\alpha/2, n_1+n_2-2} \quad (10)$$

where $Z_{\alpha/2, n-1}$ denotes the upper $\alpha/2$ percentile point of the standardized normal distribution (Z) with (n_1+n_2-2) degrees of freedom.

2.3 Testing Using Control Limits

For data collected for a time window of m epochs, where the last epoch is the current one, \bar{r} and R can be computed, where \bar{r} is the average value of the mean of the residuals (\bar{r}) and R is the range of residuals, which is computed at epoch k from:

$$R_k = (r_{max} - r_{min})_k \quad (11)$$

where $(r_{max})_k$ and $(r_{min})_k$ are the maximum and minimum residuals at epoch k , and:

$$\bar{\bar{r}} = \frac{\sum_{k=1}^m n_k \bar{r}_k}{\sum_{k=1}^m n_k} \quad (12)$$

Assuming that the residuals are uncorrelated and stationary (i.e. vary around a fixed mean, e.g. zero, in a stable or predictable manner and its stochastic properties does not significantly change over time), observations may contain outliers if:

$$R_k < D_3 \bar{R} \quad \text{or} \quad R_k > D_4 \bar{R} \quad (13)$$

and/or if

$$\bar{r}_k < \bar{\bar{r}} - A_2 \bar{R} \quad \text{or} \quad \bar{r}_k > \bar{\bar{r}} + A_2 \bar{R} \quad (14)$$

where A_2 , D_3 , and D_4 are control constants that vary with the size of the sample n . These control constants are tabulated in quality control references, e.g. Montgomery, 2005. \bar{R} denotes the mean value of the ranges within the selected time window. Here $(D_3 \bar{R})$ and $(\bar{\bar{r}} - A_2 \bar{R})$ are defined as the lower control limits (LCL), $(D_4 \bar{R})$ and $(\bar{\bar{r}} + A_2 \bar{R})$ are defined as the upper control limits (UCL) for testing R_k and \bar{r}_k , respectively.

The \bar{r} test monitors between-sample variability, and the R test measures within-sample variability. The \bar{r} and \bar{R} are computed within a pre-set window, e.g. 5-10 epochs. For post-mission applications, the data within the specified window is examined and the test is performed rolling the selected window between epochs such that the current epoch's data are included as the last data set. If an out-of-control value for \bar{r} or R is detected, the presence of large errors/outliers in the data of this epoch is suspected.

The control limits defined in Eq. 13 and 14 assume that the standard deviation for different samples is approximately the same. However, if this assumption is not proven, use of standardized R and \bar{r} can handle this situation. For instance, for m number of samples, the standardized R_k^s and \bar{r}_k^s values at epoch k will read:

$$R_k^s = \frac{R_k}{\bar{R}_{1 \rightarrow m}} \quad (15)$$

$$\bar{r}_k^s = \frac{\bar{r}_k - \bar{r}_{nom(1 \rightarrow m)}}{\bar{R}_{1 \rightarrow m}} \quad (16)$$

Where $\bar{R}_{1 \rightarrow m}$ and $\bar{r}_{nom(1 \rightarrow m)}$ are the average range and nominal value for \bar{r} ; respectively, for the samples under consideration. The observations may contain outliers if:

$$R_k^s < D_3 \quad \text{or} \quad R_k^s > D_4 \quad (17)$$

and for the \bar{r} test, the outliers are expected if:

$$\bar{r}_k^s < -A_2 \quad \text{or} \quad \bar{r}_k^s > +A_2 \quad (18)$$

When using the \bar{r} test, usually large shifts can be detected with small sample sizes, e.g. 4-5 observations per epoch, whereas small shifts need large sample sizes to be detected. Therefore, by using only GPS, where the number of observed satellites is usually less than 10, we expect the test to be effective only for detection of large outliers. When measurements from other GNSS are used in addition to GPS, e.g. GLONASS, Galileo and QZSS, the efficiency of this test for detection of small biases will significantly improve. On the other hand, the efficiency of the range method (\bar{R} test) drops dramatically with the increase of the sample size. Therefore, for samples that have more than 10 measurements, the R test can be replaced by testing the sample standard deviation (S). In this test, a set of data can be considered affected by large errors/outliers if for epoch k :

$$S_k < B_3 \bar{S} \quad \text{or} \quad S_k > B_4 \bar{S} \quad (19)$$

and/or if

$$\bar{r}_k < \bar{r} - A_3 \bar{S} \quad \text{or} \quad \bar{r}_k > \bar{r} + A_3 \bar{S} \quad (20)$$

where for measurements from m epochs, with variable sample size (n_k), and k varies from 1 to m , the tested standard deviation would be:

$$\bar{S} = \left(\frac{\sum_{k=1}^m (n_k - 1) S_k^2}{\sum_{k=1}^m n_k - m} \right)^{\frac{1}{2}} \quad (21)$$

where B_3 , B_4 and A_3 are the tabulated control factors that vary with the size of the sample (Montgomery, 2005). These factors can be determined as follows: consider \bar{S}/c_4 is an unbiased estimator of σ for the residuals population, where:

$$c_4 \approx \frac{4(n-1)}{4n-3} \quad (22)$$

then:

$$B_3 = 1 - \frac{3}{c_4} \sqrt{1-c_4^2} \quad (23)$$

$$B_4 = 1 + \frac{3}{c_4} \sqrt{1-c_4^2} \quad (24)$$

$$A_3 = \frac{3}{c_4 \sqrt{n}} \quad (25)$$

2.4 Testing the difference between the mean and the median

For outlier detection in positioning using GNSS, the number of observations is usually limited at a single epoch. If no outliers are present, the mean value should be very close to the median. In case of the presence of an outlier in one satellite observation (or more up to $n-u$), the mean of the absolute values of residuals will be biased, with a value at the decimetre to metre level. On the other hand, the median of the absolute values of the residuals of phase data should be small and unaffected by the presence of these outliers. Therefore, a comparison between the mean and the median of the absolute values of the residuals can give a quick indication about the existence of outliers. The test static, at each epoch, can be formulated with a null hypothesis H_0 stating: $\bar{r}_{abs} \approx MD_r$, where \bar{r}_{abs} denotes the mean of the absolute values of the residuals, and MD_r is their median. The test reads:

$$H_0: \quad \left| \bar{r}_{abs} - MD_r \right| \leq Z_{\alpha/2, n-1} \frac{\sigma_{dr}}{\sqrt{n}} \quad (26)$$

$$H_a: \quad \left| \bar{r}_{abs} - MD_r \right| > Z_{\alpha/2, n-1} \frac{\sigma_{dr}}{\sqrt{n}} \quad (27)$$

The null hypothesis, with outlier-free observations, assumes the difference ($\bar{r}_{abs} - MD_r$) to have a noise-like behaviour with normal distribution and zero mean. $Z_{\alpha/2, n-1}$ denotes the upper $\alpha/2$ percentage point of the standardized normal distribution with $(n-1)$ degrees of freedom and σ_{dr} is the standard deviation of the population of the difference ($\bar{r}_{abs} - MD_r$) that can be selected based on studying several data sets of a large number of outlier-free observations. Alternatively, $Z_{\alpha/2, n-1}$ and σ_{dr} in Eq. 26 and 27 can be replaced with the $t_{\alpha/2, n-1}$ value and the standard deviation of the above difference (s_{dr}); respectively, where s_{dr} is computed from the collected observations. If the difference between the mean and the median does not pass the test, this implies the presence of large errors. The test works efficiently

when the errors in phase measurements are larger than two cycles. When it is only one cycle, the test might not be conclusive and further testing will be needed.

3. Identification Testing

If the detection test fails, an identification test with more specific alternative hypotheses needs to be performed for failure identification of individual observations. This requires the presence of at least two degrees of freedom.

3.1 Testing individual observations

We start here by briefly describing the method given by Teunissen *et al.*, 2008. With a known covariance matrix of the observations (\mathbf{C}_y), the null hypothesis based on Gauss-Markov model and assuming residuals of normal distribution and zero expectation, can be formulated as:

$$H_0: E(\mathbf{y}) = \mathbf{h}(\mathbf{x}) \quad (28)$$

If outliers, represented as the vector Δ , are present, the alternative hypothesis will read:

$$H_a: E(\mathbf{y}) = \mathbf{h}(\mathbf{x}) + \mathbf{W}_y \Delta \quad (29)$$

where \mathbf{W}_y is a matrix with columns comprising the canonical unit vector aligning errors with their observations. For instance, if an outliers δ_i in observation i exist, where $i=1$ to n , then:

$$w_{yi} = [0, 0, 1_i, 0, \dots, 0]^T \quad (30)$$

If no more outliers are described, the remaining elements of the \mathbf{W}_y matrix will be populated with zeros. The test statistic can be formed as:

$$T_k = \mathbf{r}^T \mathbf{C}_y^{-1} \mathbf{W}_y (\mathbf{W}_y^T \mathbf{C}_y^{-1} \mathbf{C}_r \mathbf{C}_y^{-1} \mathbf{W}_y)^{-1} \mathbf{W}_y^T \mathbf{C}_y^{-1} \mathbf{r} \quad (31)$$

The null hypothesis describes a central Chi-square distribution, whereas, an alternative hypothesis would describe a non-central Chi-square distribution, with an offset λ that can be determined from the value of the error Δ . The test is formed as:

$$H_0: T_{df} \sim \chi^2(df, 0), \quad H_a: T_{df} \sim \chi^2(df, \lambda) \quad (32)$$

Thus, reject H_0 if $T_{df} > \chi^2_{\alpha}$ where $(1 \leq df \leq n-u)$. When $df=1$, the *data snooping* procedure (Baarda, 1968) can be applied, where the observations are checked one-by-one, and the matrix \mathbf{W}_y becomes a vector \mathbf{w}_y . The test statistic then reads (Teunissen *et al.*, 2008):

$$T_k = (\mathbf{w}_y^T \mathbf{C}_y^{-1} \mathbf{C}_r \mathbf{C}_y^{-1} \mathbf{w}_y)^{-1} \mathbf{w}_y^T \mathbf{C}_y^{-1} \mathbf{r} \quad (33)$$

The above test can be further simplified by formulating the test statistic using the standardized residual (r_i/σ_i), assuming that the unbiased standardized residuals are normally distributed. The test can be formed as (Kuusniemi, 2005):

$$H_0: r_i/\sigma_i \leq Z_{1-\frac{\alpha}{2}} \quad (34)$$

$$H_a: r_i/\sigma_i > Z_{1-\frac{\alpha}{2}} \quad (35)$$

where σ_i denotes the standard deviation of residuals, which can be estimated from the covariance matrix \mathbf{C}_r , and $Z_{1-\frac{\alpha}{2}}$ is the standard normal distribution with $N(0,1)$.

3.2 Testing of individual satellite residuals between epochs using MR

This proposed simple test is performed by computing the moving range of residuals MR of phase observations for each satellite individually. The MR of residuals between the epochs k (the present epoch) and $k-1$ for the satellite under consideration can be computed as:

$$MR_i = |r_k - r_{k-1}| \quad (36)$$

The mean value of MR within a pre-set time window, e.g. 5-10 epochs, defined as \overline{MR} , is used in forming the test statistic such that:

$$H_0: 0 < MR_k < D_4 \overline{MR} \quad (37)$$

$$H_a: MR_k > D_4 \overline{MR} \quad (38)$$

Where D_4 is a control factor that varies with the size of the sample and it is tabulated in quality control references, e.g. Montgomery, 2005. Since two residuals are considered when computing MR , the number of samples reduces to 2 and D_4 is 3.267. If MR_k does not pass the test and the alternative hypothesis is accepted, then the satellite residual at the present epoch (k) is suspected of having an outlier.

3.3 Testing the time series of individual residuals

In this test, the phase residual for each satellite can be screened over a number of epochs. Since one observable is considered in this case, the χ^2 distribution converts to the Gaussian distribution. The standard deviation of the tested sample (s) is used and the test becomes equivalent to the Student's t_α test. Thus, for a residual r_i of satellite observation i computed at the current epoch, the failure identification test statistic can read:

$$H_0: \frac{r_i - \bar{r}}{s} \leq t_{\alpha,df} \quad (39)$$

$$H_a: \frac{r_i - \bar{r}}{s} > t_{\alpha,df} \quad (40)$$

Where \bar{r} is the residual mean value computed over the selected window, and $t_{\alpha,df}$ is computed based on df degrees of freedom extracted from the length of the data used and equals $m-1$.

Before excluding a specific observation, one has to consider the correlation between a possible blunder that appears in the residual of one observation with residuals from other observations as the adjustment may smooth out multiple gross errors across an entire data set. This may lead to the possibility of smoothing out a specific large error throughout its neighbouring observations. Therefore, re-iterating the adjustment process by taking out or re-inputting one or more of those observations at a time may be necessary in order to locate the right observations containing gross errors (Kuang, 1996, Kuusniemi, 2005). In one approach, the satellites can be ranked in a descending order according to the value of their test statistic. The LS or filtering is iterated in a recursive manner using a subset of measurements excluding the measurement corresponding to the residual with the biggest standardized value (r_i/σ_i) that does not pass the above tests, and detection and identification testing is repeated. If the tests still do not pass, re-selection is carried out, moving the subset along the list. The excluded observation has to be reused to account for correlation between residuals.

4. Testing and Results

To evaluate the performance of the presented statistical observation testing methods, a test was performed on the 19 November 2008, in the Edinburgh Oval Park at Curtin University of Technology, Perth, Australia. Only carrier phase observations were considered. Their ambiguities were first estimated from previous epochs and considered fixed during testing. The test was carried out for 2290 continuous epochs with one-second intervals in a walking mode. A Sokkia GSR2700ISX dual-frequency receiver was used for data collection and data processing was performed using Kalman filtering. The number of observed satellites ranged between 9 and 11. During one part of the test period, the receiver went under trees or close to a tree canopy, which resulted in experiencing some cycle slips and interruption of data.

To establish a reference for assessment of detection and identification capabilities, the data were first cleaned from possible outliers. First, cycle slips were detected and fixed using the triple difference method presented by Kim and Langley, 2002. The data were next screened using commercial software and were proven free of outliers. Next, some artificial large errors, resembling

outliers, were implanted into these data at specific epochs to test if the presented methods would be able to detect them. The implanted errors ranged from 1 to 50 cycles of phase data of L1 frequency. Out of the total 2290 epochs, 222 epochs in four discrete periods were uploaded with outliers. To test the ability of detection and identification of outliers in multiple observations at the same epoch, outliers were inserted in the phase observations of more than one satellite at a time, where the number of 'infected' satellites ranged from one to five. Their distribution along the time series of the data is illustrated in Fig. 1.

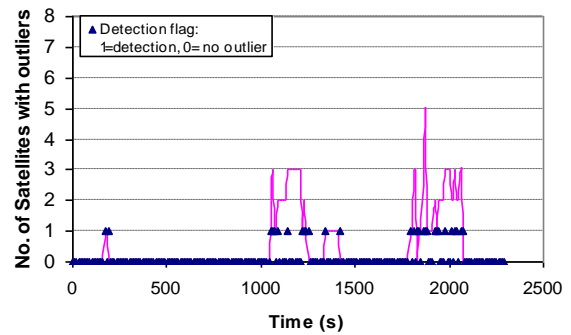


Figure 1: Number of satellites with faulty measurements

4.1 Results of detection testing

Table 1 summarizes the results of independently running different detection tests for the data at hand in the case of implementing large errors of 25 and 50 cycles. The Chi-square test (Eq. 3 & 4) proved to be efficient as the observations with the implanted large errors were detected at all epochs, as seen in the second column in the table. Similarly, the method of testing the difference between the mean of the residuals between two consecutive epochs (Eq. 8), which is given in the third column in Table 1, also proved to perform well. In this method, a population standard deviation σ of 2 cm was used, which was selected based on analysing the outlier-free residuals. Fig. 1 also shows an error indexing system that was used during testing. An index of 1 was given if a correct detection of faulty data is made when they first spotted, whereas an index of 0 was used to indicate if the data were clear of outliers, or if there is no change in the error-performance of the satellites detected with outliers. As can be seen from Fig. 1, all data with the implanted large errors were correctly flagged.

The tests applying the control limits that dynamically change according to the data at hand, also performed well, as shown in the fourth column of Table 1. This include checking whether the range of residuals R and the mean values of the residuals \bar{r}_k lie within the dynamic control limits set by the mean value of the ranges \bar{R} (Eq. 13 and 14). For these tests, a window of five epochs was selected for computation of the mean

values. One should note that these tests require cleaning the data from outliers, once detected, before examining data for the following epochs. Fig. 2 and 3 show the performance of these tests in detecting the 'faulty' measurements at the start of the four periods that contain them. As can be seen, the R and \bar{r}_k values exceeded the control limit at the beginning of each of these periods. For better illustration, and as an example, the test values for the first period that has outliers are shown enlarged next to the control chart plot in Fig. 3. The results of the test checking the sample standard deviation S against the control limit set by the mean value \bar{S} (Eq. 19 and 20) was shown in the fifth column in Table 1. The test was also successful in detecting the data sets containing the implanted outliers. Fig. 4 illustrates the control chart for this test, which depicts the time sequence of the control limits and the tested standard deviations. The figure shows that at the beginning of each period containing outliers, S was greater than the upper control limit and detection of implanted errors was thus possible.

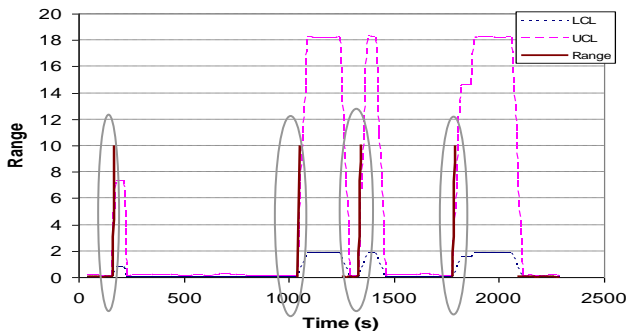


Figure 2: Detection using control limits of the range

When testing the difference between the mean and the median of the absolute values of the residuals (Eq. 26 and 27), the standard deviation of the residuals' population σ was taking 2 cm, which was used in setting the test threshold. Results demonstrated that the test was successful in detecting all epochs that contained data of errors larger than two cycles.

To illustrate the performance with smaller errors, e.g. errors that range between 1 cycle and 5 cycles (0.2m to 1m). Results of the likelihood test (Eq. 3 and 4) are given as an example. For this test, the covariance matrix of the observations plays an important role that may change the test output. A zenith direction variance of $(0.005\text{m})^2$ and uncorrelated observations with variances that are satellite elevation angle dependent were used.

Table 2 shows the percentage of successful detection using the likelihood test for different error levels and with three significance levels: 0.05, 0.01 and 0.001. The latter is of particular interest for safety-of-life applications, such as airborne navigation (see El-

Mowafy, 2008). The test was successful for detection of all errors that have a magnitude more than 4 cycles, even with the critical significance level of 0.001. However, the power of detection decreases with the decrease of the significance level for small errors that range from one to three cycles. It is worth mentioning that the same conclusion holds when rerunning the detection test after using both phase and code data. Computation of the minimal detectable bias, as described by Teunissen and Kleusberg (1998) can help in identifying the threshold for error detection.

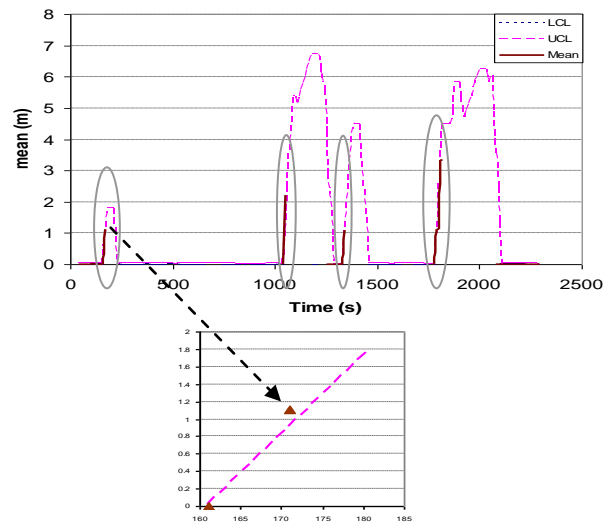


Figure 3: Detection using control limits of the mean

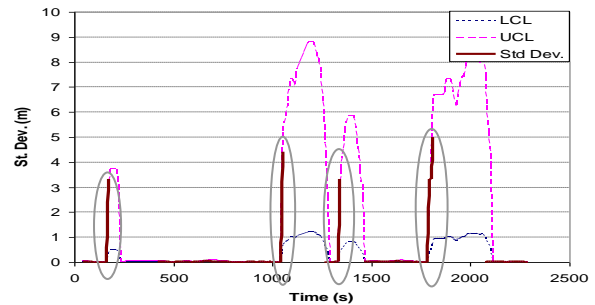


Figure 4: Detection using control limits of the std. dev. (S)

Table 2: Performance of detection testing of phase observations at various levels (success percentage%)

α	1 m	0.8 m	0.6 m	0.4 m	0.2 m
0.05	100	100	100	91.07	41.10
0.01	100	100	96.43	83.93	21.43
0.001	100	100	94.64	67.86	8.93

Table 1: Detection methods for all satellite residuals in one epoch (large cycles)

Method	$\mathbf{r}^T \mathbf{C}_y^{-1} \mathbf{r} > \chi_{1-\alpha,df}^2$			$ \bar{r}_1 - \bar{r}_2 - \Delta_0 > \sigma Z_{\alpha/2, n_1+n_2-2}$			<ul style="list-style-type: none"> • $R_k < D_3 \bar{R}$ or $R_k > D_4 \bar{R}$ • $\bar{r}_k < \bar{r} - A_2 \bar{R}$ or $\bar{r}_k > \bar{r} + A_2 \bar{R}$ 			<ul style="list-style-type: none"> • $S_k < B_3 \bar{S}$ or $S_k > B_4 \bar{S}$ 		
	A*	B*	C*	A	B	C	A	B	C	A	B	C
% Ratio	100	0	0	100	0	0	100	0	0	100	0	0

- * A: % Detection of faulty satellites
- B: % Miss detection of faulty satellites
- C: % False alarm

4.2 Results of identification testing

When a set of data containing large errors are detected at a certain epoch, identification testing is performed to identify the specific satellites with bad measurements. The test employing standardized residuals as the test statistic (Eq. 34 and 35) was first performed. For the test data at hand, the method performed only well for large blunders, e.g. 25 cycles, with a detection rate of almost 100%. For errors that were 25 cycles or less, there were several cases of miss-detection, but there were no observed cases of false alarm. The overall performance of this method was 90.5% success rate as shown in the second column in Table 3.

Results of the moving range test MR (Eq. 37 and 38) showed a success in detecting 97.3% of all implanted large errors and identifying their specific satellite observations. However, there were also several false alarms raised, i.e. indicating the existence of an outlier while there were none. For the test at hand, where measurements of 9 satellites were collected for a period of 2290 epochs with 222 cases of implanted large errors, 216 cases were detected, 6 were missed, and there were 27 cases of false alarm. The results are given in the third column in Table 4. Examples of the results of this test are illustrated in Fig. 5 and 6, which show the control limits and the MR values during testing of satellite number 2 and satellite number 27. The former was outlier free since MR values were always less than the dynamically changing UCL values, whereas the latter satellite had several outliers that were successfully detected as the MR values exceeded the UCL.

Table 3: Results of Identification tests for all satellites

Method	$r_i/\sigma_i \leq Z_{1-\frac{\alpha}{2}}$			$MR_i > D_4 \bar{MR}$			$\frac{r_i - \bar{r}}{s} \leq t_\alpha$		
	A*	B*	C*	A*	B*	C*	A*	B*	C*
% Ratio	90.5	9.4	0	97.3	2.7	0.7	99.0	1.0	0
No. of cases	201	21	0	216	6	27	220	2	0

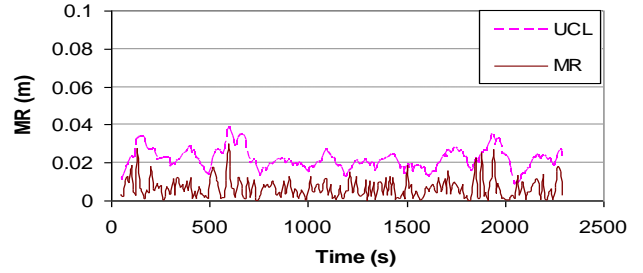


Figure 5: MR test for a satellite free from blunders/outliers

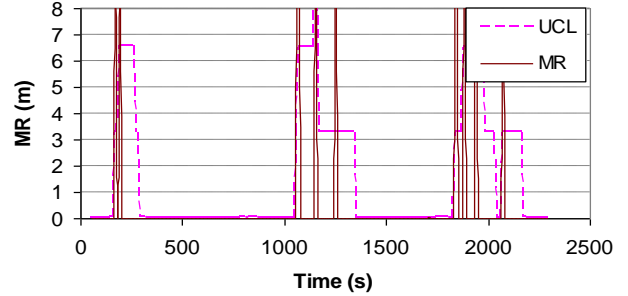


Figure 6: MR test for a satellite with multiple blunders

When examining the time sequence of individual standardized residuals (Eq. 39 and 40), a standard deviation of 2 cm was used with a mean value of the residuals assumed 0. The length of data included in computation of the test statistic was selected as 10 epochs. Results of the test at hand, checking each satellite observation individually in its time series, show that all observations containing large errors were identified except the data of two epochs, with no false alarms, which means a success rate of 99% as shown in the last columns in Table 3.

Results of checking individual phase observations using the data snooping test statistic given in Eq. 33 and the test in Eq. 32 are shown in Table 4. The table shows the success of identification of outliers as a percentage of all observations containing outliers. Errors from 1 cycle up to 50 cycle (0.2m to 10m) were tested with three significance levels, which are 0.05, 0.01 and 0.001. In each test case, the same error level was implanted to all selected epochs. The elevation-angle dependent model was used to create the observation covariance matrix.

Fig. 7 and 8 show the results for $\alpha=0.05$ and $\alpha=0.001$, respectively. In all tests, there were no cases of false alarm. As can be seen from Table 4 and the figures, the success of identification for the data at hand is high with large errors, and it decreases with smaller errors. This also has a direct relation with the chosen significance level as it sets the test threshold. For instance, with $\alpha = 0.05$, which corresponds to a confidence interval of 95%, the success rate was 100% when the blunders were 4m (approximately 21 cycles) or higher. When $\alpha = 0.01$, the 100% success rate was reached when the error was 6 m (32 cycle) or higher. For $\alpha = 0.001$, which is needed for critical applications, the 100% was only reached with 10 m errors. Table 4 also shows that with $\alpha = 0.05$ and 0.01, the test was unable to detect errors of less than 5 cycles and 15 cycles, respectively. With $\alpha = 0.001$, errors below 27 cycles were never detected.

As can be seen from Eq. 33, the test is also dependent on the model of the observation covariance matrix used. To investigate this, the test was repeated using an equal-weight observation covariance matrix and results from the two covariance models were compared. Results of using equal-weight covariance matrix are given in Table 5. When comparing the identification performance listed in Table 4 with that of Table 5, some differences can be seen. Unexpectedly, results of identification testing when using an equal-weight covariance model outperforms the elevation-angle dependent model for the test data at hand. This can be partially explained by noting that large errors in our test were not necessarily implanted at satellites of low elevation angles rather randomly distributed to mimic a case of unusual or untraditional fault occurrence. Nevertheless, the conclusion is that the performance of identification of individual faulty observations can significantly change according to the observation covariance model used.

Table 4: Identification testing with elevation-angle dependent variances (Success %)

α	10 m	9 m	8 m	7 m	6 m	5 m	4 m	3 m	2 m	1 m	0.8 m	0.6 m	0.4 m	0.2 m
0.05	100	100	100	100	100	100	100	98.1	20.4	0	0	0	0	0
0.01	100	100	100	100	100	88.9	22.3	0	0	0	0	0	0	0
0.001	100	98.1	88.9	27.8	20.4	0	0	0	0	0	0	0	0	0

Table 5: Identification testing with equal weight variances (Success %)

α	10 m	9 m	8 m	7 m	6 m	5 m	4 m	3 m	2 m	1 m	0.8 m	0.6 m	0.4 m	0.2 m
0.05	100	100	100	100	100	100	100	100	20.4	0	0	0	0	0
0.01	100	100	100	100	100	94.4	47.2	4.6	0	0	0	0	0	0
0.001	100	100	94.4	73.1	23.1	4.6	0	0	0	0	0	0	0	0

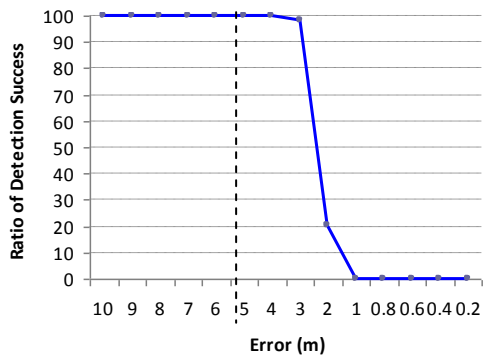


Figure 7: Success of detection ($\alpha=0.05$)

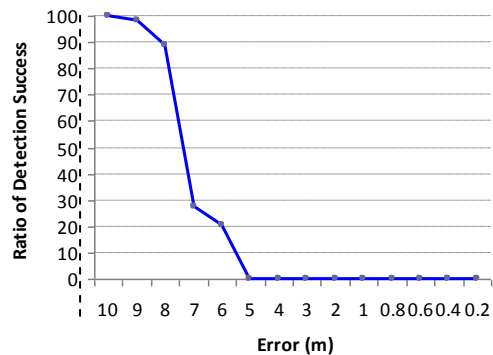


Figure 8: Success of detection ($\alpha=0.001$)

5. Conclusions

The paper presents a wide range of local statistical tests that can be applied for detection and identification of outliers in GNSS measurements. Results showed that all detection methods that screen all satellite measurements performed well for large errors. This includes likelihood testing, testing the difference between the mean of the residuals for subsequent epochs, testing the difference between the mean and the median, and testing using control limits on the range, the mean and the standard deviation of the residuals.

For the identification testing, presented methods include testing the likelihood of the standardized residuals, checking individual observations in a data snooping approach, or using the control limits on the moving range of residuals of each satellite, and checking the time series of individual residuals. For these tests, the identification ratio was very high with large errors, but the success rate drops rapidly when the blunders get smaller. In the Moving Range test, there were cases of false alarm. The performance is also dependent on the selected significance level, as the success of detection is decreased with the decrease in the significant level. Finally, it was shown that the identification performance can significantly change with the observation-covariance matrix model used, and thus accurate models should always be used.

References

- Baarda W (1968), *A Testing Procedure for Use in Geodetic Networks*, Netherlands Geodetic Commission, Publication on Geodesy, 2(5), Delft, Netherlands.
- El-Mowafy, A. (2008), *Improving the Performance of RTK-GPS Reference Networks for Positioning in the Airborne Mode*, *Navigation*, 55(3), 215-224.
- Hwang, P. and R.G. Brown (2008), *From RAIM to NIORAIM*, *Inside GNSS*, 3 (4), 24-33.
- Hofmann-Wellenhof B, Lichtenegger H, Collins J (2001), *GPS theory and practice*, Springer Verlag, Wien, New York, 5th revised edition, 208-212.
- Kim D, Langley RB (2001), *Quality Control Techniques and Issues in GPS Applications: Stochastic Modelling and Reliability Test*, Proceedings of the 2001 International Symposium on GPS/GNSS (the 8th GNSS Workshop), Jeju Island, Korea, 7-9 November 2001.
- Kim D, and Langley RB (2002), *Instantaneous Real-Time Cycle-Slip Correction for Quality Control of GPS Carrier-phase Measurements*, *Navigation*, 49(4), 205-222.
- Kuang, S. (1996), *Geodetic Network Analysis and Optimal Design: Concepts and Applications*, Ann Arbor Press Inc., Chelsea, Michigan, 1996.
- Kuusniemi H (2005), *User-Level Reliability and Quality Monitoring in Satellite-Based Personal Navigation*, Ph.D. Dissertation, Institute of Digital and Computer Systems, Tampere University of Technology, Finland, 71-105.
- Leick A (2004), *GPS Satellite Surveying*, John Wiley and Sons, Inc., Hoboken, New Jersey, 3rd edition.
- Montgomery D C (2005), *Introduction to Statistical Quality Control*, 5th ed., John Wiley & Sons, NY.
- Teunissen P J G, and M.A. Salzmann (1989), *A Recursive Slippage Test for use in State-Space Filtering*, *Manuscripta Geodaetica*, 14: 383-390.
- Teunissen P J G (1990), *Quality Control in Integrated Navigation Systems*, *IEEE AES Magazine*, 5(7): 35-41.
- Teunissen P J G, Kleusberg A (1998), *Quality control and GPS - GPS for Geodesy*, 2nd Edition, Springer, NY, 271-318.
- Teunissen, PJG, DG Simons, and CCJM Tiberius (2008), *Probability and Observation Theory*, Lecture Notes AE2-E01, Faculty of Aerospace Engineering, Delft University of Technology.
- Vaniček, P. and E. Krakwski (1986), *Geodesy the Concepts*, 2nd edition, North Holland, NY, pp 1-669.

Biography

Dr. Ahmed El-Mowafy is a Senior Lecturer at Curtin University, Australia, where he is conducting research in the area of precise positioning using multi-CNSS, PPP, quality control and validation of measurements and system models.