Towards Digital Earth
— Proceedings of the International Symposium on Digital Earth
Science Press ,1999

A Multiple Outlier Detection Algorithm for Instantaneous Ambiguity Resolution For Carrier Phase-Based GNSS Positioning

Liwen Dai Shaowei Han Chris Rizos

School of Geomatic Engineering
The University of New South Wales
Sydney NSW 2052, Australia
Tel: 61-2-93854189, Fax 61-2-93137493
Emial: liwen@unsw.edu.au

ABSTRACT In GPS/GLONASS pseudo-range and carrier phase observations, the main errors are caused by cycle slips, multipath effects, residual atmospheric biases, orbital errors, inter-channel biases and random noise. Because of these errors instantaneous ambiguity resolution with state-of-the-art GPS techniques becomes quite difficult, and even fails if observations are seriously contaminated, which is unfortunately not an unlikely event in kinematic applications. On the other hand, outlier detection becomes complicated, even impossible, because the double-differenced observations are not independent. In this paper an investigation of the stochastic model for GPS and GLONASS carrier phase and pseudo-range measurements is reported on, and a multiple outlier detection algorithm based on correlation analysis theory for the correlated double-differenced observations is proposed. Any observation errors, including outliers, affects the residuals through the reliability matrix in the least-squares adjustment. The correlation coefficient between column vector of the reliability matrix and residual vector is considered critical information, which can reflect the relationship between the true error of the correlated observations and the residuals. It replaces the standardised residual for the detection and analysis of the outliers in the correlated double-differenced observations. Tests have been carried out using dual-frequency GPS/GLONASS and GPS-only receivers, and single-frequency GPS/GLONASS receivers, over distances ranging up to 10 kilometres. The results indicate that after applying the proposed algorithm the success rates of ambiguity resolution using single epoch data can be significantly improved up to 98.3%, and the kinematic positioning accuracy is very high in the case of a large number of redundant satellites

KEY WORDS GPS/GLONASS, Outlier, Multiple Outlier Detection, Correlation Analysis, Instantaneous Ambiguity Resolution, Reliability Matrix, and Real-time Stochastic Model

.

1. Introduction

The centimetre accuracy of GPS kinematic positioning can be theoretically achieved in real-time due to the millimetre resolution of the carrier phase observable with the unknown ambiguities recovered. In practice, however, reliable and correct ambiguity resolution depends on observations to a large number of satellites which constrains its applications, making it difficult to address positioning applications in areas where the number of visible satellites is limited. The most obvious way to increase the number of satellites is to combine the observations from the GPS and GLONASS systems. On the other hand, significant errors and biases may exist in GPS/GLONASS pseudo-range and carrier phase observations, which are caused by cycle slips, multipath effects, residual atmospheric biases, orbital errors, inter-channel biases and random noise. Because of these errors and biases instantaneous ambiguity resolution with state-of-the-art GPS techniques becomes quite difficult, and even fails if observations are seriously contaminated, which is unfortunately not an unlikely event in kinematic applications. For ambiguity resolution and positioning purposes, the minimum number of satellites is 5 for GPS or GLONASS, 6 for combined GPS and GLONASS with the separation formulation, respectively. If more satellites are tracked, some of these observations can be removed if they are suspected to have been contaminated by the outliers or biases. There are three cases involved in outlier or bias detection in GPS/GLONASS data processing. The first case is that if the ambiguity–floated solution can not pass the χ^2-distribution statistic test, one or more outliers probably exist in the pseudo-range observations. The second case is that if ambiguity resolution fails, the carrier phase observations maybe also contain one or more outliers or biases. The last case is that if the ambiguity resolution still fails even though the two cases mentioned above are considered, the positioning solutions can still be provided using the previous epoch's fixed ambiguities if there are enough satellites without cycle slips. The cycle slips can be considered significant outliers. In order to detect and eliminate the effect of the outliers on the estimated results, many methods have been proposed in the literature. In general, all these methods can be divided into two categories.

One is that the gross errors are considered to have the same variance but different expectation as compared with normal observations. The other is that the gross errors are considered to have the same expectation but different variance. In the first mentioned case, outliers are included in the functional model. Based on some statistics and criteria with a level of significance, the outliers can be located and removed by comparing the estimated value of statistic with the critical value (Baarda , 1968; Förstner, 1983; Kubik, 1982; Li, 1988). The method based on the first case was firstly proposed by Baarda (1968), and has been applied in many data processing examples. The problem for this case is that the multiple outliers can not be located directly, especially for the highly correlated observations. In the second case, outliers are included in the stochastic model. This means that the gross observations should have the bigger variance (or smaller weight). After the appropriate weight or variance matrix has been assigned to the corresponding gross observations, the influence of the outliers on the parameter estimation can be eliminated or mitigated with a few iterations (Zhou at el., 1997; Shi, 1998). Unfortunately the reasonable variance or weight matrix for the observations containing the outliers, especially for the highly correlated observations, is not easy to determine.

Detection and estimation of multiple outliers in combined GPS and GLONASS kinematic positioning with only a single epoch of data raises a number of challenges. Firstly, due to the double difference operator, the observations become highly (mathematically) correlated. For single epoch data processing, the spatial correlation still exists. Hence, outlier detection becomes complicated, even impossible, although the temporal correlation is absent. Secondly, the instantaneous existence of a few outliers may be possible, but it is quite difficult to locate them because of mutual effects of the several outliers and influences of the satellite geometry. Thirdly, the outliers to detect here are probably relatively small in magnitude, typically at the level of a few centimetres for carrier phase observations and a few metres for pseudo-range observations. Fourthly, the computation time should be as short as possible because the algorithms are designed for real-time applications and the large number of observations that have to be processed, especially for the dual-frequency GPS and GLONASS data.

In this paper an investigation of the stochastic model for GPS and GLONASS carrier phase and pseudo-range measurements is reported on, and a multiple outlier detection algorithm based on correlation analysis theory for correlated double-differenced observations is proposed. Any observation errors, including outliers, affects the residuals through the reliability matrix in the least-squares adjustment. If the outliers exist, there are strong correlation between the residuals and the column vector of the reliability matrix related to the outlier. The correlation coefficient between column vector of the reliability matrix and residual vector is considered critical information, which reflects the relationship between the true errors of the correlated observations and the residuals. It replaces the standardized residual for the detection and analysis of the outliers in the correlated double-differenced observations. Tests have been carried out in order to demonstrate the efficiency and performance of the proposed test procedures through the case study examples of the dual-frequency GPS/GLONASS and GPS-only receivers, and single-frequency GPS/GLONASS receivers, over distances ranging up to 10 kilometres.

2. Mathematical Model and Ambiguity Resolution

High quality results, estimated from the application of least-squares techniques, require the use of an optimal functional, and associated stochastic model. However, the stochastic model is dependent on the choice of the functional model. Hence, for a different choice of functional model, a different stochastic model may be needed.

- ***Functional Model***

Due to the different frequencies for the different GLONASS satellites, the difference of two receiver clock parameters is not zero, as in the commonly used double-differenced procedure for GPS. To overcome this problem, single-differenced pseudo-range observations have to be introduced (Pratt et al., 1998; Leick, 1998; Zhodzishsky, 1998; Kozlov et al., 1997; Wang, 1998, Han at el., 1999). The double-differenced GPS pseudo-range observables and the single-differenced GLONASS pseudo-range observables have been identified as an optimal functional model in combined GPS and GLONASS positioning (Rapoport, 1997; Wang, 1998; Han at el., 1999). The linearisation of the double-differenced (or single-differenced) pseudo-range and carrier phase observations can be represented by the following system of equations:

$$V = B_C \hat{X}_C + B_N \hat{X}_N - L \tag{1}$$

where X_C is the $t \times 1$ real-valued parameter vector which includes coordinate and clock bias parameters (and any other real-valued parameters which need to be estimated); X_N is the $m \times 1$ integer-valued parameter vector; L is the $n \times 1$ difference vector formed by subtracting computed values from the carrier phase and pseudo-range observations. V is the residual vector. The proposed functional model improves performance because the ambiguity resolution process is insensitive to the residual clock biases and the inter-channel biases, hence reliable and correct ambiguities can be recovered.

- ***Stochastic Model***

GPS and GLONASS observations are affected by several kinds of errors and biases. When forming the double-differences, the main biases are caused by multipath effects, residual atmospheric errors, orbital errors, and inter-channel biases. Due to insufficient knowledge concerning these physical phenomena, the above biases cannot be rigorously accounted for through functional modelling. The stochastic model has to therefore account for both the observation noise and the unmodelled residual biases. The well-known elevation dependent stochastic model is often used, which may be represented as an exponential function or an inverse of the sine of the satellite elevation angle (El-Rabbany, 1994; Jin, 1995). However, constant coefficients can only reflect error characteristics of the GPS receiver rather than the unmodelled residual biases, which most probably are related to the observing environment. Based on the fact that residual series of least-squares estimation contains sufficient information concerning the observation noise and biases, a more rigorous stochastic model is derived. The general least-squares linearised observation equation and the criteria can be modelled as:

$$V_i = B_i X_i - L_i \tag{2}$$

$$V_i^T D_i^{-1} V_i = \text{Minimum} \tag{3}$$

where V_i is the $n \times 1$ vector of and residuals at epoch i; L_i is the $n \times 1$ difference vector formed by subtracting computed values from the carrier phase and pseudo-range observations. B_i is the design matrix relating to the measurements L_i; X_i is the estimated unknown parameters; and D_i is the variance-covariance matrix of the measurements.

Based on the minimum quadratic form of the residuals, the least-squares estimated parameter $\hat{X}_i$ can be derived:

$$\hat{X}_i = (B_i{}^T D_i^{-1} B_i)^{-1} B_i D_i^{-1} L_i \tag{4}$$

Substituting equation (4) into equation (2), the estimated residuals can be written as:

$$V_i = (B_i (B_i{}^T D_i^{-1} B_i)^{-1} B_i D_i^{-1} - E) L_i \tag{5}$$

The variance-covariance matrix can then be represented as:

$$D_i = Q_{V_i} + B_i (B_i{}^T D_i^{-1} B_i)^{-1} B_i{}^T \tag{6}$$

where Q_{V_i} is the variance-covariance matrix of the residuals. Due to the similarity of the observation environments, the residuals of the observations show a high degree of temporal and spatial correlation, at least in the short term. In other words, the residual series can be considered as a wide-sense stationary process for short periods (a few minutes). The actual variance-covariance matrix of the residuals can then be estimated from the previous residual series, whose ambiguity sets have already been fixed to the correct values, using the following equation:

$$Q_{V_i} = \frac{1}{N} \sum_{k=1}^{N} V_{i-k} V_{i-k}^T \tag{7}$$

where N is the width of the moving window, which is empirically chosen as 10 in the following experiment.

In equation (6), the variance-covariance matrix of the measurements cannot be estimated directly. An iterative procedure becomes necessary. The initial (or default) variance-covariance matrix is determined by using the empirical model. Based on the previous measurement residuals, the variance-covariance matrix of the measurements can be rigorously estimated in real-time from equations (6, 7). Normally iterating twice is enough. The default stochastic model should be used at the beginning of the data processing, or for a new satellite, or after a long data gap.

The stochastic model in this paper not only reflects the stochastic characteristics of the observation noise, but also the residual biases due to multipath, the atmospheric delays, the inter-channel biases and the orbital error remaining after double-differencing both the carrier phase and pseudo-range observations. With the help of the estimated variance-covariance matrix, the reliability of ambiguity resolution and the accuracy of the real-time kinematic positioning results can be significantly improved.

- ***Ambiguity Resolution, Validation and Fault Detection***

Equation (1), combining carrier phase and pseudo range observations, can be used to estimate the real-valued parameter vector, including user receiver coordinates, ambiguities and clock bias, and their variance-covariance matrix. The associated stochastic model is derived from the residual series over the previous epoch series according to equations (6, 7). In the case of the ambiguity-floated solution (considering X_N as a real-valued parameter vector), estimates $\hat{X}_C$ and $\hat{X}_N$ are obtained using the standard least-squares procedure with aposteriori variance vector $\hat{\sigma}_0^2 = \frac{V^T PV_{Float}}{n-t-m}$,

where $V^T PV_{Float}$ is the quadratic form of the residuals. Reliable results at this step are dependent on the appropriateness of the stochastic model of the observations with respect to the functional model. The following rejection regions should be employed in order to check the fidelity of the stochastic and functional models:

$$V^T PV_{Float} \geq \sigma_0^2 \cdot \xi_{\chi^2_{n-t-m};1-\alpha/2} \tag{9a}$$

$$V^T PV_{Float} \leq \sigma_0^2 \cdot \xi_{\chi^2_{n-t-m};\alpha/2} \tag{9b}$$

where $\xi_{\chi^2_{n-t-m};\alpha/2}$ and $\xi_{\chi^2_{n-t-m};1-\alpha/2}$ are the lower and upper boundaries of the $1-\alpha$ confidence interval for the χ^2 -distribution statistic with n-t-m degrees of freedom. This test is used to test the pseudo-range observation quality because Ω_{Float} is only dependent on pseudo-range observations (Han & Rizos, 1997). If the Ω_{Float} is rejected by equation (9a), the outlier detection procedures should be applied because the outliers may exist in the pseudo-range observations, which are caused by multipath or system biases, or the stochastic model does not reflect the actual accuracy of the observation. If the Ω_{Float} is rejected by equation (9b), a check should be made to determine whether there are enough redundant observations, or the stochastic model does not reflect the actual accuracy of the observation.

The LAMBDA procedure is then implemented to search the integer ambiguity set (Teunissen, 1994; Han & Rizos, 1995). The validation criteria test suggested by Han (1997), and the ratio test, are implemented. If both tests are passed, the ambiguity resolution is assumed to be correct. The quadratic form of the residuals $\Omega_{Fix,k}$ corresponding to the ambiguity-fixed solution should be compatible with σ_0^2, represented by the condition:

$$\sigma_0^2 \cdot \xi_{\chi^2_{n-t};\alpha/2} \leq \Omega_{Fix,k} \leq \sigma_0^2 \cdot \xi_{\chi^2_{n-t};1-\alpha/2} \tag{10}$$

where $\xi_{\chi^2_{n-t};\alpha/2}$ and $\xi_{\chi^2_{n-t};1-\alpha/2}$ are the lower and upper boundary of the $1-\alpha$ confidence interval for the χ^2 -distribution statistic with n-t degrees of freedom. If $\Omega_{Fix,k}$ is rejected, the corresponding integer vector will be rejected. In this case, using the outlier detection procedures, a satellite with the outlier should be removed so as to attempt to fix the ambiguities again, or the previous fixed ambiguities (without cycle slips) should be introduced in order to generate reliable positioning results.

In order to ensure that the ambiguity resolution is correct, global information should be used for dual-frequency data (Han, 1997). As is well known, the Total Electron Content (TEC) of the path through the ionosphere has very strong correlation in space and time. The TEC value for the neighbouring epoch should therefore be very close and this information will be considered as the basis for a global test. The difference between the double-differenced ionospheric delay on L1 and L2 carrier phase observations is defined as Δ_{ion}. If the integer ambiguities are resolved correctly, the Δ_{ion} sequence should change smoothly. Otherwise, a jump will occur due to wrong ambiguity resolution. The criterion $\delta\Delta_{ion} < 5.0\text{cm}$ is used for fault detection in this paper. If the $\delta\Delta_{ion} > 5.0\text{cm}$, the ambiguity sets are considered to have been fix to the wrong values. Hence, the corresponding ambiguities should be rejected.

3. Outlier Detection Algorithm Using Correlation Analysis

Based on the above analysis, adaptation procedures should be applied if the equations (9a, 9b), or the fault detection criteria for dual-frequency observation data are accepted, or the equation (10) is rejected. For ambiguity resolution and positioning purposes the minimum number of satellites is 5 for GPS or GLONASS, and 6 for combined GPS and GLONASS with the separation formulation. If more satellites are tracked, some of these observations, which are suspected as being contaminated by outliers or biases, can be removed so that: (1) the instantaneous ambiguity resolution can be obtained with maximum success rate, and

(2) the position solutions can still be output using the fixed ambiguities from the previous epoch if ambiguity resolution fails and there are enough satellites without cycle slips. The objective of a multiple outlier detection algorithm is: (1) to judge whether any outlier is present; (2) to determine which observations should be identified correctly as containing outliers; and (3) to make an acceptable decision about outliers and to estimate the corresponding effect on the final solution if outliers can not be uniquely identified. All the above three problems are important for a successful multiple outlier detection algorithm. However, the emphasis of this paper will be on the former two issues. Before a discussion on these, the classical outlier detection algorithms will be reviewed.

3.1. Existing Outlier Detection Methods

Baarda's data snooping theory assumes that only one outlier is present in the observations (Baarda, 1968). Applying a series of one-dimensional tests, that is, testing consecutively all residuals, is the standard data snooping strategy. Baarda's test belongs to the group of un-studentized tests that assume that the apriori variance of unit weight is known. The test statistic is written as:

$$n_{\mathrm{i}} = \frac{\mathrm{V_i}}{\delta_0 \mathrm{Q}_{\mathrm{vv_{ii}}}} \in \mathrm{N}(0,1) \tag{11}$$

The critical value can be determined from the normal distribution with a significance level of α. If α is assigned to 5%, the value is 1.96. It can be seen that the critical value for this test is independent of the degrees of freedom. Hence Pope (1976) introduced the τ test, which belongs to the group of studentized tests. It makes use of the aposteriori variance of unit weight as estimated from the observations (Leick, 1995). The test statistic is:

$$\tau_{\mathrm{i}} = \frac{\mathrm{V_i}}{\hat{\delta}_0 \mathrm{Q}_{\mathrm{vv_{ii}}}} \in \tau(\mathrm{n-r}) \tag{12}$$

The symbol $\tau(\mathrm{n-r})$ denotes the τ distribution with n-r degrees of freedom. It is related to the Student's t distribution by:

$$\tau(\mathrm{n-r}) = \frac{\sqrt{\mathrm{n-r}}\ \mathrm{t_{n-r-1}}}{\sqrt{n-r-1+t_{n-r-1}^2}} \tag{13}$$

Both the data snooping and τ test procedures are based on statistical theory. For correlated observations with small redundancy numbers, outliers can not be detected correctly and located exactly.

3.2.Relationship Between True Errors and Residuals

The residuals can be considered as indicators of the errors and are therefore useful information to reflect outliers. In order to locate the outliers, the relationship between the true errors and residuals should be firstly analysed. The true error ε can be defined as:

$$\varepsilon = \tilde{L} - L \tag{14}$$

where $\tilde{L}$ and L are vectors of the true values and the observations, respectively. Neglecting the subscript, the relationship between the true errors and the residuals can be derived from equations (5, 6) as follows:

$$-V = R\ \varepsilon \tag{15}$$

$$R = Q_{vv}\ D^{-1} \tag{16}$$

where R is called the reliability matrix. In general, R is an asymmetric matrix. Nevertheless, it is symmetric when the observations are independent and have the same variance. The elements in the R matrix are usually denoted by r_{ij} .

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \ldots & r_{2n} \\ \vdots & \vdots & \ldots & \vdots \\ r_{n1} & r_{n2} & \ldots & r_{nn} \end{bmatrix} \tag{17}$$

The R is only related to the satellite geometry and the variance-covariance matrix of the observations. However, there is no relationship with the observations and the corresponding errors. As it is well known that R is an idempotent matrix, hence, its rank is equal to its trace, that is:

$$Rank(R) = tr(R) = \sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{r_{ii}} = \mathrm{r < n} \tag{18}$$

Equation (18) indicates that the sum of the main diagonal elements in the R matrix is equal to the number of the redundancy. It should be emphasised that because the R matrix is not a non-negative definite matrix for highly correlated observations. The r_{ii} (I=1,2, …, n) is also called the observation redundancy component, which may take a negative value for some special conditions.

Due to the rank deficiency of the R matrix, the true errors can not be computed directly by inverting R using equation (15). If the R_i and ε are defined as:

$$R_i = [\, r_{1i} \quad r_{2i} \quad \cdots \quad r_{ni} \,]^T \quad (i = 1,2,\cdots,n) \qquad (19)$$

$$\varepsilon = [\, \varepsilon_1 \quad \varepsilon_2 \quad \cdots \quad \varepsilon_i \quad \cdots \quad \varepsilon_n \,]^T \qquad (20)$$

Hence, equation (15) can be written as:

$$-V = R_1\varepsilon_1 + R_2\varepsilon_2 + \cdots + R_n\varepsilon_n \qquad (21)$$

Any observation error ε_i affects the residuals through the column vector R_i of the reliability matrix in the least-squares adjustment. It is easily seen that the R_i vector controls the influence of the true errors on the residuals. The more the component of the R_i , the more the residuals. On the other hand, the more the true errors ε_i , the more the residuals. Any observation error ε_i probably contributes to all the residuals. The biggest influence of one error on the residuals is likely to occur to other residuals.

3.3.Outlier Detection Based on Correlation Analysis

Suppose that the outliers exist in the former k observations. Equation (21) can be simplified because the normal errors can be neglected:

$$-V \approx R_1\varepsilon_1 + R_2\varepsilon_2 + \cdots + R_k\varepsilon_k \qquad (22)$$

Equation (22) shows that the residual quantities come mainly from the outliers contributions. It also shows that there is significant mathematical correlation between the vector of the residuals and the column vectors of the reliability sub-matrix relating to the outliers.

The correlation coefficient is considered critical information, which reflects the relationship between the true outliers of the correlated observations and the residuals. Figure 1 shows the degree of correlation between the column vector R_7 of the reliability matrix and the residuals with an artificial outlier, in the case of combined GPS and GLONASS positioning. Multiple outlier detection procedures for large-scale network adjustment with high observation redundancy, based on correlation analysis, was firstly proposed by Shi (1998). In GNSS kinematic positioning, however, it will involve variable satellite geometry and lower observation redundancy.

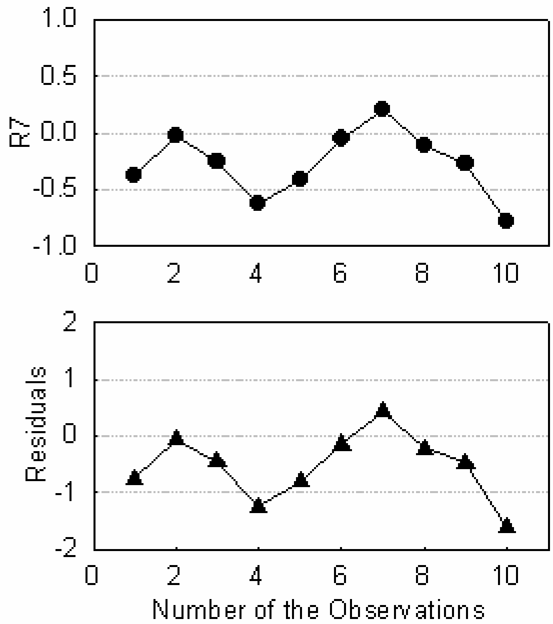

Figure 1. Degree of correlation between the column vector of the reliability matrix and the residuals.

In order to describe mathematically the quantity of this degree of correlation between them, two constant variables a and b are introduced so that the difference d between R_j vector and V vector can be minimised:

$$d = Min(\frac{1}{n}\sum_{i=1}^{n}(v_i - a - br_{ij})^2) \qquad (23)$$

where v_i (i=1,…,n) are the components of the V vector. The partial derivatives of equation (23) are:

$$\frac{\partial d}{\partial a} = -\frac{2}{n}\sum_{i=1}^{n}(v_i - a - br_{ij}) = 0 \qquad (24)$$

$$\frac{\partial d}{\partial b} = -\frac{2}{n}\sum_{i=1}^{n}(v_i - a - br_{ij})r_{ij} = 0 \qquad (25)$$

According to equations (24, 25), the parameters a and b can be obtained:

$$b = \frac{\sum_{i=1}^{n}(r_{ij} - \bar{r}_j)(v_i - \bar{v})}{\sum_{i=1}^{n}(r_{ij} - \bar{r}_j)^2} \qquad (26)$$

$$a = \bar{v} - b\,\bar{r}_j \qquad (27)$$

where $\bar{v}$ and $\bar{r}_j$ are the average values of the V and R_j vectors, respectively. Substituting equations (26, 27) into equation (23), the difference d can be written as:

$$d = \frac{1}{n}\sum_{i=1}^{n}(v_i - \bar{v})^2(1-\rho_{R_j,V}^2) \tag{28}$$

$$\rho_{R_j,V} = \frac{\sum_{i=1}^{n}(r_{ij}-\bar{r}_j)(v_i-\bar{v})}{\sum_{i=1}^{n}(r_{ij}-\bar{r}_j)^2\sum_{i=1}^{n}(v_i-\bar{v})^2} \tag{29}$$

In equation (29), the greater the correlation coefficient $\rho_{R_j,V}$, the less the difference d. The correlation coefficient can range in value from -1 to +1. If the absolute correlation coefficient value is close to 1, there is a strong correlation relationship between the R_j and V vectors. This also means that the contribution to the residuals come mainly from the observation relating to the R_j vector. If an outlier occurs, the observation corresponding to R_j has the largest possibility. If multiple outliers occur, they make a combined contribution to the residuals. There are also high correlation coefficients between the column vectors of the reliability matrix related to the outliers and the residuals. Based on the above analysis, outliers can be located through analysing $\rho_{R_j,V}$.

According to the correlation analysis theory, the following statistic for the correlation coefficient test can be obtained:

$$\rho_{R_j,V} = \frac{t_c}{\sqrt{t_c^2+n-2}} \tag{30}$$

where t_c is the critical value associated with the t-distribution with a given significant level and n-2 degrees of freedom. If the correlation coefficient is great than the given critical value, it is significant. It indicates the biggest possibility that the observation(s) relating to the significant correlation coefficient(s) should be suspected as having been contaminated by outliers. An observation with the largest correlation coefficient should be removed at each iteration. Hence, multiple outliers can be located through an iterative procedure. After the outlers have been located, the outliers should be put back into the observations to test them one by one, so that any mis-flagged outliers can be restored.

3.4. Summary of Outlier Detection Procedures

In this paper a multiple outlier detection procedure for GNSS instantaneous ambiguity resolution and positioning can be summarised as follows:

(1) After computing the ambiguity-float solution using the estimated stochastic model from the residuals, test whether the $V^T PV_{Float}$ is less than a detection threshold defined by apriori σ_0^2 with degree of freedom (n-t-m) and significant level 1-α. If the test is accepted, go to next step, otherwise go to (3). It should be emphasized that if there are less than the necessary number of satellites when some satellites are removed, go to step (4).

(2) Attempt ambiguity resolution, and apply validation and fault detection procedures. If one ambiguity set can pass all the tests, output the ambiguity-fixed solution, keep the corresponding residuals and go to step (1) to process the next epoch of data.

(3) After computing the correlation coefficients between the residuals and the column vectors of the reliability matrix, test whether the maximum correlation coefficient is bigger than a detection threshold with degrees of freedom (n-2) and significant level 1-α. If true, delete the satellite relating to the maximum correlation coefficient and go to step (1). Otherwise, go to step (4).

(4) After introducing the previous fixed ambiguity set and restoring all the deleted satellites, test according to equation (10) using only carrier phase observations. If the test passes, the positioning solution is output. Otherwise, an iterative process, with the satellite relating to the maximum correlation coefficient removed, would be repeated until the test passes, or less than the necessary double-differenced carrier phase observables can be formed. If less than the necessary number of double-differenced carrier phase observables can be formed, the process will be repeated again after some eliminated satellites are restored and the reference satellite is deleted. If all the adaptive procedures still fail, the adaptive procedure is considered to have finally failed. Following this, go to step (1) to process the next epoch of data.

It should be emphasised that if the previous ambiguities, fixed to the wrong values, are introduced in step (4), the wrong ambiguities can easily pass the test according to equation (10). It will lead to seriously wrong positioning solutions. Hence it should be ensured that the introduced ambiguity sets are the correct ones. In this paper, if the ratio value for the ambiguity validation is greater than 3, the ambiguities can only be introduced in the next epoch. If the outliers, such as cycle slips, exist in the observations, it is almost impossible to pass the test in equation (10). On the other hand, in order to obtain a precise positioning solution, the satellite geometry should satisfy at least the given conditions (PDOP<5, in this paper).

4. Experiments

The proposed procedures, including the multiple outlier detection algorithm based on correlation theory and the real-time stochastic modelling, were tested during two sets of experiments using dual-frequency GPS/GLONASS and GPS-only receivers, and single-frequency GPS/GLONASS receivers.

Table 1. Details of the test data sets.

Name	Length (m)	GPS/GLN satellites	Total Epochs	Receivers	Survey Date
A1	12	8-5/7-3	14362	GG24	12.5.99
A2	2873	9-5/7-4	4012	GG24	11.5.99
A3	4053	9-6/5-3	6868	GG24	10.5.99
A4	10494	8-6/2-3	120	Z18	12.1.99

The first is a set of static experiments using data f rom the dual-frequency GPS/GLONASS Ashtech Z18 receivers (with 30 second sampling rate) and single-frequency GPS/GLONASS Ashtech GG24 receivers (with 1 second sampling rate). The reference GG24 receiver was set up on the Mather Pillar, on the roof of the Geography and Surveying building, at The University of New South Wales. The rover GG24 receiver was set up at different sites, which included the same roof nearby to the reference receiver, at Coogee Beach, at Maroubra Beach, and at the La Perouse Beach. The baseline name, baseline length, number of satellites, observation span (total number of epochs) are given in Table 1. The positioning results can be easily checked from the repeatability of the baseline vectors for the different sessions. In the case of all the data sets the cut-off elevation angle was set to 15 degrees during the processing.

All static baselines were processed in only single epoch mode. Data processing has been carried out for all the static data using two different strategies. One strategy only applies the empirical stochastic model dependent on the satellite elevation angle, and the other one applies the multiple outlier detection procedure and the real-time stochastic model estimated using the residual series from previous epochs as proposed in this paper. The results are listed in columns 3-5 in Table 2 and Table 3. The third column is the number (and percentage) of epochs for which ambiguity resolution is successful on an epoch-by-epoch basis. The fourth column is the number of epochs (and percentage) which do not pass the validation criteria test. The fifth column is the number of epochs (and percentage) which pass the validation criteria tests, but for which the result is incorrect. In Table 3, there are three sub-columns in column 3, where the first sub-column stands for the total success rate. Case 1 means the ambiguities can be successfully resolved directly; and case 2 means that the ambiguities can be resolved successfully or whose correct positioning results can be output after the outlier detection algorithm is applied. It can be seen that using the elevation-dependent empirical stochastic model the success rates for ambiguity resolution range from 81.4% to 100.0%. It should be emphasised that because of the redundant observations and precise pseudo-range data, the ambiguities for the dual-frequency GPS/GLONASS data from the Z18 receivers can be fixed easily to the correct ones in the single epoch mode even though the baseline length is over 10km. It also shows that quite a large percentage of the epochs (0.2%) at baseline A1 give the wrong ambiguity resolution results. After applying the multiple outlier detection procedure and the real-time stochastic model estimated using the residuals from the previous epochs (in this case 10 epochs), the success rates of ambiguity resolution range from 99.3% to 100.0%. No wrong ambiguity resolution results or incorrect positioning results, estimated using fixed ambiguities that are introduced from the previous epoch, are accepted. It is also noted that the multiple outlier detection procedure is responsible for quite a large percentage of epochs (7.6% for A1, 2.2% for A2, 5.1% for A3) whose ambiguities are recovered correctly or whose positioning results are output correctly. The results indicate that single-epoch ambiguity resolution can achieve up to 99.3% success rate with redundant GPS and GLONASS satellite observations. The conclusion that can be drawn is that the multiple outlier detection algorithm based on correlation analysis theory and the estimated stochastic model from the residuals are, in theory, rigorous and, in practice, very powerful.

In order to demonstrate the power of the multiple detection procedure based on correlation analysis theory, the 9th epoch in the A1 data containing 5 GPS satellites and 7 GLONASS satellites is analysed. In this epoch, the ambiguities can not be fixed correctly. Hence, the previous epoch's fixed ambiguities are introduced. 5 artificial outliers are added to the GPS 3 (0.5 cycle), 19 (1 cycle) and 27 (1 cycle) satellites and the GLONASS 38 (1 cycle) and 48 (.2 cycle) satellites. Table 4 lists the relevant numerical results of the multiple outlier detection step by step. For each iteration, three sub-rows are included. The first sub-row is the residual value (in cycles); the second sub-row is the standardised residual; the third sub-row is the correlation coefficient between the residual vector and the column vector of the reliability matrix. The 12th, 13th and 14th columns give the quadratic form of the residuals and the corresponding upper boundary value of the $\rho_{R_j,V}$ and the χ^2-distribution statistic, respectively. From Table 4 it is clearly seen that in the first iteration the maximum residual and standardised residual exists for satellites 42 and 39. However, they have no any artificial outlier actually. It should also be noted that the maximum standardised residual is less than 2. The outliers can not be located through the data snooping and τ test procedures. Fortunately, the outliers can be located through the correlation coefficient test with the application of the iterative process.

Table 2. Single-epoch solution using the elevation-dependent empirical stochastic model.

Name	Total Epochs	Correct (%)	Reject (%)	Wrong (%)
A1	14362	11689(81.4%)	2639(18.4%)	33(0.2%)
A2	4012	3335(83.1%)	677(16.9%)	0(0.0%)
A3	6868	5723(83.3%)	1145(16.7%)	0(0.0%)
A4	120	120(100.0%)	0(0.0%)	0(0.0%)

At each iteration the satellite containing the biggest outliers will be removed until the χ^2-distribution test can pass or the maximum $\rho_{R_j,V}$ is less than the critical value with the significance level (1-α) and the degrees of freedom (n-2). At last, the identified outliers values are estimated and given in the last rows. The conclusion can be clearly drawn that the outlier detection procedure based on the correlation analysis theory can indeed locate multiple outliers. It is especially powerful when only one outlier occurs. The proposed algorithm for multiple outlier detection can work well not only for independent observations, but also for highly correlated observations.

Table 3. Single-epoch solution using the multiple outlier detection algorithm and the real-time stochastic model derived using residuals from previous epochs, as proposed in this paper.

Name	Total Epochs	Correct (%)			Reject (%)	Wrong (%)
		Total	Case 1	Case 2		
A1	14362	14362 (100.0%)	13273 (92.4%)	1089 (7.6%)	0 (0.0%)	0 (0.0%)
A2	4012	4012 (100.0%)	3922 (97.8%)	90 (2.2%)	0 (0.0%)	0 (0.0%)
A3	6868	6821 (99.3%)	6474 (94.3%)	347 (5.1%)	47 (0.7%)	0 (0.0%)
A4	120	120 (100.0%)	120 (100.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)

The second is a kinematic experiment carried out on 29 April 1999 using two GG24 GPS/GLONASS single-frequency receivers and three dual-frequency Leica SR399 GPS receivers. One GG24 receiver and one Leica SR399 were set up at the reference site. The other GG24 receiver and the two Leica GPS receivers were mounted on a car. The trajectory of the rover receivers is shown in Figure 2. The reason for using three rover receivers is as a mutual check on whether the derived positioning results are correct. The experiment started on the roadside of the M4 motorway, Sydney, which is nearby to the reference site. After the first 40 minutes in static mode, the car moved along the motorway and the Great Western Highway, finishing the experiment in static mode again for 15 minutes. This concluded a single loop. A total of two loops were completed with 1Hz data rate. The number of observed satellites is plotted in Figure 3.

Table 4. The test results of multiple outlier detection algorithm with 5 artificial outliers introduced.

Iterations		SV 3 (-0.5)	SV 19 (1.0)	SV 27 (1.0)	SV 31	SV 39	SV 40	SV 38 (1.0)	SV 42	SV 48 (0.2)	SV 52	V^TPV	$\rho_{R,V}$ (1-0.05)	Upper Boundary $\chi^2(1-0.05)$
1	V	0.63	-0.64	0.13	0.30	0.56	0.76	-0.16	**0.83**	-0.22	0.40	1386.5	**0.549**	12.59
	R	0.84	1.67	0.18	0.35	**1.78**	0.94	0.92	0.89	0.24	0.19			
	ρ	0.36	**-0.61**	-0.22	0.07	0.21	0.32	-0.31	0.40	-0.37	0.12			FAIL
2	V	**0.85**		-0.36	0.55	0.44	0.40	-0.17	0.44	-0.04	0.67	652.09	**0.582**	11.07
	R	1.54		0.66	0.87	**1.95**	0.68	1.34	0.65	0.07	0.44			
	ρ	0.48		-0.43	0.13	0.28	0.34	**-0.59**	0.35	-0.46	0.31			FAIL
3	V	0.46		**-0.57**	-0.08	0.05	0.37		0.36	-0.32	-0.12	203.72	**0.621**	9.49
	R	1.46		**1.77**	0.25	0.69	1.05		0.88	0.80	0.14			
	ρ	0.66		**-0.84**	-0.02	0.19	0.41		0.42	-0.35	-0.15			FAIL
4	V	**0.37**			-0.13	-0.01	-0.03		-0.07	**-0.37**	0.01	66.14	**0.669**	7.82
	R	**1.85**			0.59	0.32	0.17		0.34	1.46	0.01			
	ρ	**0.87**			-0.28	0.05	0.03		-0.09	-0.76	0.08			FAIL
5	V				0.11	-0.03	0.02		-0.02	**-0.19**	0.13	11.45	**0.729**	5.99
	R				1.31	1.67	0.18		0.22	**1.69**	0.48			
	ρ				0.33	-0.79	0.33		0.08	**-0.96**	0.54			FAIL
6	V				-0.01	0.00	0.02		-0.02		0.08	0.46	0.805	3.84
	R				1.12	0.57	1.03		0.63		1.16			
	ρ				-0.93	-0.90	0.63		-0.12		0.94			**PASS**
Estimated Outliers		**-0.48**	**1.00**	**1.11**				**0.99**		**0.29**				

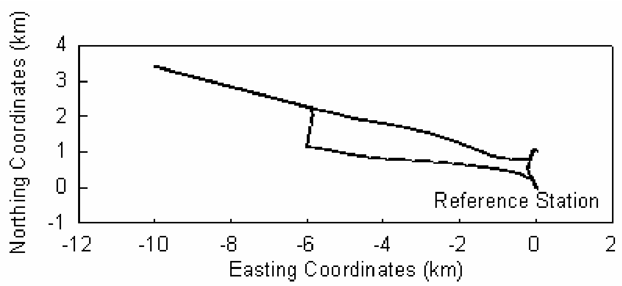

Figure 2. Trajectory of the rover receivers relative to the reference site.

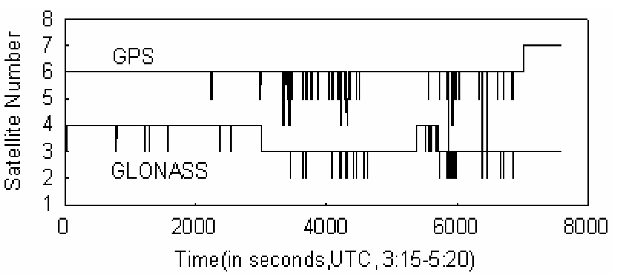

Figure 3. Number of observed satellites.

In this kinematic experiment the processed results are shown in Table 5. The constant distance (about 60cm) between the two Leica receivers and one GG24 receiver are used to check whether or not the kinematic positioning results are correct. If distance differences between the Leica rover receivers and the GG24 rover receiver exceed some specified tolerance value (10cm in this experiment), the ambiguities are considered to have been fixed to the wrong values. From Table 5 it can be seen that using the elevation-dependent empirical stochastic model the success rate for ambiguity resolution is only 62.4% for the GG24 receivers, and 87.7% and 87.8% for the Leica receivers. The percentage of rejected epochs is 33%, 12.3% and 12.2%, respectively. It also shows that quite a large percentage of epochs (4.6%) in B3 from the GG24 receiver give the wrong ambiguity resolution results. Table 6 lists the processing results after applying the multiple outlier detection procedure and the real-time stochastic model. The success rates of single-epoch ambiguity resolution, or correct positioning results, can be significantly improved to 98.2%, 99.3% and 99.9%, respectively. No wrong ambiguity resolution results or incorrect positioning results, estimated by using fixed ambiguities that are introduced from the previous epoch, are accepted. It is also seen that the multiple outlier detection procedure contributes a quite large percentage of epochs (4.4% for B1, 5.2% for B2, 19.7% for B3) whose ambiguities are fixed correctly or whose positioning results are output correctly. The results indicate that single-epoch ambiguity resolution or correct positioning results can be achieved for up to a 99.3% success rate. The results also indicate that the proposed multiple detection procedure has the ability of detecting the outliers even in correlated observations.

Table 5. GG24 kinematic positioning results using a single epoch of data.

Baseline Name	Total Epochs	Correct (%)	Reject (%)	Wrong (%)
B1 Leica 1	5767	5762 (87.8%)	705 (12.2%)	0 (0.0%)
B2 Leica 2	5720	5019 (87.7%)	701 (12.3%)	0 (0.0%)
B3 GG24	7572	4721 (62.4%)	2501 (33.0%)	350 (4.6%)

Table 6. Single-epoch solution using the multiple outlier detection strategy and the real-time stochastic model derived using residuals from previous epochs.

Name	Total Epochs	Correct (%)			Reject (%)	Wrong (%)
		Total	Case 1	Case 2		
B1	5767	5760 (99.9%)	5510 (95.5%)	259 (4.4%)	7 (0.1%)	0 (0.0%)
B2	5720	5682 (99.3%)	5383 (94.1%)	299 (5.2%)	38 (0.7%)	0 (0.0%)
B3	7572	7436 (98.2%)	5945 (78.5%)	1491 (19.7%)	136 (1.8%)	0 (0.0%)

5. Concluding Remarks

A multiple outlier detection procedure based on the correlation analysis theory for integrated GPS/GLONASS positioning (or GPS alone) has been introduced in this paper. A real-time stochastic model estimated using residuals from the previous epochs has also been applied. The multiple outlier detection procedure can locate rapidly and correctly the outliers or biases even in the case of highly correlated observations. It is especially powerful for only one outlier with a small value and for small degrees of freedom. It can improve significantly the ambiguity resolution success rate and the number of valid kinematic positioning solution epochs. The real-time stochastic model estimated from the residuals can significantly improve the ambiguity resolution success rate, and the accuracy of the final solutions.

The results indicate that using the proposed outlier detection procedure and the real-time stochastic model results in almost a 100% success rate for single-epoch solutions, based on four static dual-frequency and single-frequency GPS/GLONASS experiments. The single-epoch solution for kinematic positioning using dual-frequency GPS-only receivers and single-frequency GPS/GLONASS receivers could achieve up to a 98.3% success rate over 10km baseline lengths.

This algorithm has been designed for real-time applications. Although the data has been post-processed, all computations were carried out in a simulated real-time processing mode.

Acknoledgements

The authors would like to thank Matt Higgins for providing the dual-frequency Z18 data and Mr. Cheng Horngyue for collecting the GG24 data. Discussions with Mr. Cheng Horngyue and Mrs. He Ying have been very helpful.

References

Baarda, W. (1968), A testing procedure for use in geodetic networks. Netherlands Geodetic Commission, *Publications on Geodesy*, New Series, Vol.2, No.5.

El-Rabbany, A.E-S. (1994), The effect of physical correlations on the ambiguity resolution and accuracy estimation in GPS differential positioning. *Ph.D. Dissertation*, Dept. of Geodesy & Geomatics Eng., Tech. Rept. No.170, University of New Brunswick, Canada, 161pp.

Förstner, W. (1983), Reliability and discernability of extended Gauss-Markov models. *DGK, Reighe Helft Nr.98*.

Han, S., L. Dai and C. Rizos (1999), A new data processing strategy for combined GPS/GLONASS carrier phase-based positioning. Proc. *12th Int. Tech. Meeting of the Satellite Division of the U.S. Inst. of Navigation GPS ION'99*, Nashville, Tennessee, 14-17 Sept.

Han, S. (1997), Quality control issues relating to ambiguity resolution for real-time GPS kinematic positioning. *Journal of Geodesy*, 71(6):351-361.

Han, S. & C. Rizos (1995), A new method for constructing multi-satellite ambiguity combinations for improved ambiguity resolution. Proc. *8th Int. Tech. Meeting of the Satellite Division of the U.S. Inst. of Navigation GPS ION'95*, Palm Springs, California, 12-15 Sept., 1145-1153.

Jin, X.X. (1995), The change of GPS code accuracy with satellite elevation. Proc. *4th Int. Conf. on Differential Satellite Navigation Systems*, Bergen, Norway, 24-28 April.

Kubik K., (1982), An error theory for the Danish method. Symposium of Comm. III of ISP, Helsinki.

Kozlov, D. (1997), Instant RTK cm with low cost GPS+GLONASS^{TM} C/A receivers. Proc. *10th Int. Tech. Meeting of the Satellite Division of the U.S. Inst. of Navigation GPS ION'97*, Kansas City, Missouri, 16-19 Sept., 1559-1569.

Leick, A. (1995), *GPS Satellite Surveying*, John Wiley &

Sons, Inc., pp560.

Leick, A. (1998), GLONASS satellite surveying, *Journal of Surv. Eng.*, Vol. 121, 91-99.

Li, D. (1988), Error processing and reliability theory. *Press for Surveying and Mapping* (Chinese)

Pope, A.J. (1976), The statistics of residuals and the detection of outliers. NOAA Technical Report NOS 65 NGS 1, National Technical Information Center, NOAA, Silver Spring, MD.

Pratt, M., B. Burke & P. Misra (1998), Single-epoch integer ambiguity resolution with GPS-GLONASS L1-L2 data. Proc. *11th Int. Tech. Meeting of the Satellite Division of the U.S. Inst. of Navigation GPS ION'98*, Nashville, Tennessee, 15-18 Sept., 389-398.

Rapoport L. (1997) General purpose kinematic/static GPS/GLONASS postprocessing engine. Proc. *10th Int. Tech. Meeting of the Satellite Division of the U.S. Inst. of Navigation GPS ION'97*, Kansas City, Missouri, 16-19 Sept., 1757-1772.

Shi, C. (1998), Large scale GPS network adjustment & analysis theory and its application. *Ph.D. Dissertation*, The School of Geoscience and Engineering Surveying, Wuhan Technical University of Surveying and Mapping, pp.123 (Chinese).

Teunissen, P.J.G. (1994), A new method for fast carrier phase ambiguity estimation. Proc. *IEEE Position Location & Navigation Symposium PLANS94*, Las Vegas, Nevada, 11-15 April, 562-573.

Wang, J., (1998), Mathematical models for combined GPS and GLONASS positioning. Proc. *11th Int. Tech. Meeting of the Satellite Division of the U.S. Inst. of Navigation GPS ION'98*, Nashville, Tennessee, 15-18 Sept, 899-907.

Zhodzishsky, M., M. Vorobiev, A. Khvalkov & J. Ashjaee (1998), Real-time kinematic (RTK) processing for dual-frequency GPS/GLONASS. Proc. *11th Int. Tech. Meeting of the Satellite Division of the U.S. Inst. of Navigation GPS ION'98*, Nashville, Tennessee, 15-18 Sept., 1325-1331.

Zhou, J., Y. Huang, Y. Yang and J. Ou (1997), *Robust Least Square Estimation*. Press. for Huazhong University of Science and Techlonogy (Chinese).