Offset detection in GPS position time series using multivariate analysis

Amiri-Simkooei, A. R.; Hosseini-Asl, M.; Asgari, J.; Zangeneh-Nejad, F.

DOI
10.1007/s10291-018-0805-z

Publication date
2019

Document Version
Final published version

Published in
GPS Solutions

Citation (APA)
Amiri-Simkooei, A. R., Hosseini-Asl, M., Asgari, J., & Zangeneh-Nejad, F. (2019). Offset detection in GPS position time series using multivariate analysis. *GPS Solutions*, *23*(1), Article 13. https://doi.org/10.1007/s10291-018-0805-z

GPS Solutions (2019) 23:13
https://doi.org/10.1007/s10291-018-0805-z

ORIGINAL ARTICLE

Offset detection in GPS position time series using multivariate analysis

A. R. Amiri-Simkooei[1,2] · M. Hosseini-Asl[1] · J. Asgari[1] · F. Zangeneh-Nejad[1]

Received: 24 May 2018 / Accepted: 14 November 2018 / Published online: 21 November 2018

Abstract

Proper analysis and subsequent interpretation of GPS position time series is an important issue in many geodetic and geophysical applications. The GPS position time series can possibly be contaminated by some abrupt changes, called offsets, which can be well compensated for in the functional model. An appropriate offset detection method requires proper specification of both functional and stochastic models of the series. Ignoring colored noise will degrade the performance of the offset detection algorithm. We first introduce the univariate analysis to identify possible offsets in a single time series. To enhance the detection ability, we then introduce the multivariate analysis, which considers the three coordinate components, north, east and up, simultaneously. To test the performance of the proposed algorithm, we use synthetic daily time series of three coordinate components emulating real GPS time series. They consist of a linear trend, seasonal periodic signals, offsets and white plus colored noise. The average detection power on individual components, either north, east or up, are 32.3 and 47.2% for the cases of white noise only and white plus flicker noise, respectively. The detection power of the multivariate analysis increases to 70.8 and 87.1% for the above two cases. This indicates that ignoring flicker noise, existing in the structure of the time series, leads to lower offset detection performance. It also indicates that the multivariate analysis is more efficient than the univariate analysis for offset detection in the sense that the three coordinate component time series are simultaneously used in the offset detection procedure.

Keywords Time series analysis · Offset detection · Variance component estimation · Multivariate analysis

Introduction

In many geodetic and geophysical studies, GPS position time series are used in various applications such as plate tectonics, glacial isostatic rebound, crustal deformation and earthquake dynamics (Segall and Davis 1997). Proper analysis of the time series is thus a prerequisite for subsequent geodetic and geophysical interpretations. An appropriate functional model of GPS time series consists of a linear trend, possible periodic signals (mainly annual and semi-annual signals), and probabilistic offsets. Other unmodeled effects can best be described as a combination of white and colored noise.

✉ A. R. Amiri-Simkooei
amiri@eng.ui.ac.ir

1 Department of Geomatics Engineering, Faculty of Civil Engineering and Transportation, University of Isfahan, Hezar-Jarib Ave., Isfahan 81746-73441, Iran

2 Aircraft Noise and Climate Effects (ANCE), Faculty of Aerospace Engineering, Delft University of Technology, Delft, The Netherlands

Offset detection, also known as data segmentation or homogenization in literature, aims at detecting abrupt changes (offsets) in a signal. Such a problem has been widely investigated in many scientific areas such as in climate and meteorology to homogenize temperature and precipitation series (Beaulieu et al. 2008; Gazeaux et al. 2011), in biology for the detection of chromosomal aberrations (Olshen et al. 2004; Picard et al. 2005), in image processing (Pham et al. 2000) and in geodetic and geophysical applications.

In geophysical studies, this problem appears in particular in GPS permanent station coordinates, which are affected by offsets. Possible offsets are considered as one of the main sources of systematic errors introducing biases into GPS time series. They can have long-lasting effects on the estimation of site velocities. Williams (2003a) discusses the role of offsets on the site velocity estimation of GPS time series and investigates its bias induced by the position and magnitude of the offsets. The correct detection and adaptation of offsets is thus an important issue, which has been paid much attention by many researchers over the past decades for which we may refer to Williams et al. (2004), Perfetti (2006), Borghi et al. (2012), Vitti (2012), Gazeaux et al. (2013), and Montillet et al. (2015).

There are two main reasons that can cause offsets: (1) actual crustal movements, for example, due to earthquakes, and (2) artificial events. Artificial offsets can occur because of environmental, equipment malfunction and change, or human errors. Human-induced offsets tend to be related to changes in processing strategy, model changes, or input information such as elevation cutoff angle and reference frame. Equipment changes are due to changing the receiver, antenna, monument, or radome at a site (Johansson et al. 2002). Interaction of human and equipment changes can also introduce offsets into GPS time series. Such an interaction may include reporting an incorrect date or incorrect information in site logs or processing input files (Williams 2003a).

There is an ongoing research in the field of offset detection and estimation. Chen and Tiao (1990) proposed a random level shift (RLS) model to detect unknown offsets. This model can only detect unexpected offsets with large magnitude. It is thus not applicable to the offsets with small magnitude (Chen and Tiao 1990). Williams (2003a) proposed an offset detection algorithm using the change detection methods. This method can be employed to detect offsets with small magnitudes. The algorithm provides an estimate about the time instant of the offset along with its magnitude (Williams 2003a). This algorithm is, however, based on the assumption that there is only white noise in the error term of the series. This may not be a valid assumption for the case of GPS position time series for which colored noise also exists in the series (Zhang et al. 1997; Williams et al. 2004; Amiri-Simkooei et al. 2007; Khodabandeh et al. 2012). Kenyeres and Bruyninx (2004) identify and estimate offsets for position time series in the EUREF permanent network. Perfetti (2006) addresses the detection of the existing offsets in the Italian GPS Fiducial Network. This method succeeded to detect 70% of the known offsets in the coordinate components of the stations, but it failed to detect offsets of some stations due to the presence of colored noise of the series (Perfetti 2006).

To detect and estimate offsets, one may employ a univariate or multivariate time series analysis. If in a linear model, instead of one observation vector having a single design and covariance matrix, there exist several observation vectors with identical design matrices and identical covariance matrices, the model is referred to as a multivariate linear model (Koch 1999; Amiri-Simkooei 2007, 2009). A multivariate analysis might, for instance, be used to model the three coordinate components at a single station simultaneously. Signal detection and noise assessment of multivariate GPS time series have been the subject of research over the last decade (Amiri-Simkooei 2009, 2013; Amiri-Simkooei and Asgari 2012; Amiri-Simkooei et al. 2017a). In the present contribution, a similar analysis is performed to the problem of offset detection in GPS position time series.

We aim at detecting the possible offsets using the univariate and multivariate analysis. A mathematical foundation for the offset detection using the multivariate time series analysis is presented. The idea behind the proposed method, either in the univariate or in the multivariate analysis, originates from the works of Baarda (1968), Teunissen (2000), and Teunissen et al. (2005) of which some misspecifications such as blunders are detected using the statistical tests in the functional model. To detect possible offsets, two hypotheses testing on two functional models are put forward. In the null hypothesis, we assume that there is no offset, whereas in the alternative hypothesis there is at least one. The two functional models are compared with each other by means of the statistical test, which is referred to the generalized likelihood ratio (GLR) test. It allows one to decide between the original model under the null hypothesis and the extended model under the alternative hypothesis. The correct offset detection requires a proper estimation of noise components of the series. Towards this end, the least-squares variance component estimation (LS-VCE) can be used (Teunissen 1988; Teunissen and Amiri-Simkooei 2008; Amiri-Simkooei 2007, 2009).

The next two sections provide the mathematical foundation of the univariate and multivariate time series analyses. For both cases, the offset detection procedure is explained in details. A later section presents a few simulation case studies on GPS time series to investigate the efficacy of the proposed method. Finally, we make some conclusions in the last section.

Univariate GPS position time series

For the univariate time series analysis, there exists only one observation vector such as a daily GPS position time series of one component, either north, east or up. For our application, the univariate time series analysis consists of the following two steps: (1) functional and stochastic model identification, and (2) offset detection and validation.

Functional and stochastic model identification

For an appropriate GPS position time series analysis, the functional and stochastic models should be correctly specified. We briefly explain the identification of the functional and stochastic models for an individual coordinate component of GPS time series.

The functional model of an individual coordinate component, namely either of the north, east or up components, is of the form:

$$\mathrm{E}(y) = Ax, \tag{1}$$

where E is the expectation operator, y is the m-vector of time series observations, e.g., daily GPS position of one component, x is the n-vector of the unknown parameters and A is the $m \times n$ design matrix. Hereinafter, the observation vector is denoted by $y(t)$, where t refers to the time instant. If a linear trend plus q periodic sinusoidal signals properly describe the deterministic behavior of the series, the functional model is of the form:

$$\mathrm{E}(y(t)) = y_0 + vt + \sum_{k=1}^{q} \left(a_k \cos \omega_k t + b_k \sin \omega_k t\right), \tag{2}$$

where y_0 and v are the intercept and the slope (site velocity or rate) of the line fitted to the series, respectively. The two trigonometric terms cos and sin together represent a sinusoidal wave with, in general, a non-zero initial phase. Examples of such periodic patterns include annual and semi-annual signals, and the GPS draconitic periodic signal and its higher harmonics (Ray et al. 2008; Amiri-Simkooei et al. 2007). The unknown vector x consists of the intercept y_0, the slope v and the coefficients (amplitudes) and a_k and b_k. Proper identification of the periodic signals is the task of the spectral analysis methods for which we may refer to the least squares harmonic estimation (LS-HE), developed and applied to time series by Amiri-Simkooei (2007, 2013) and Amiri-Simkooei et al. (2007, 2017a). An algorithm for identification of other misspecifications—offset detection for instance—can similarly be developed by the LS-HE theory. This is the subject of discussion in the present contribution.

The stochastic model describes the statistical properties of observable vector y by means of a covariance matrix. An appropriate stochastic model leads to the best linear unbiased estimation (BLUE) of the unknown parameters. For many geodetic applications, however, the covariance matrix of observables is only partly known because it can be expressed as an unknown linear combination of a few known cofactor matrices,

$$\mathrm{D}(y) = Q_y = Q_0 + \sum_{k=1}^{p} \sigma_k Q_k, \tag{3}$$

where D is the dispersion operator, $\sigma_k,\ k = 1, \ldots, p$ are the unknown variance components, and $Q_k,\ k = 1, \ldots, p$ are some known $m \times m$ cofactor matrices; Q_0 is the known part of the stochastic model. The estimation of these unknown variances σ_k is referred to as variance component estimation for which we employ the least squares variance component estimation (LS-VCE) in the present contribution (Teunissen 1988).

In the case of GPS position time series, the covariance matrix Q_y is composed of white noise plus power-law colored noise—flicker noise for instance. It is then of the form:

$$Q_y = \sigma_{\mathrm{w}}^2 Q_{\mathrm{w}} + \sigma_{\mathrm{f}}^2 Q_{\mathrm{f}}, \tag{4}$$

where the white noise cofactor matrix $Q_1 = Q_{\mathrm{w}} = I_{\mathrm{m}}$ is an identity matrix of size m and $Q_2 = Q_f$ is the cofactor matrix of flicker noise based on the Hosking noise structure introduced and used by Williams (2003b), Langbein (2004), and Williams et al. (2004). The LS-VCE method can be used to assess the noise components of the GPS position time series in an iterative manner. LS-VCE has many attractive features for which we refer to Teunissen (1988), Teunissen and Amiri-Simkooei (2008), and Amiri-Simkooei (2007). The variance components are estimated as $\hat{\sigma} = N^{-1} l$, where N is a $p \times p$ matrix, l is a p-vector and $\hat{\sigma} = \left[\hat{\sigma}_1, \hat{\sigma}_2, \ldots, \hat{\sigma}_p\right]^{\mathrm{T}}$ is a p-vector of unknown variances to be estimated. The entries of N and l are:

$$n_{ij} = \frac{1}{2} \mathrm{tr}\left(Q_i Q_y^{-1} P_A^{\perp} Q_j Q_y^{-1} P_A^{\perp}\right) \tag{5}$$

and

$$l_i = \frac{1}{2} \hat{e}^{\mathrm{T}} Q_y^{-1} Q_i Q_y^{-1} \hat{e}, \tag{6}$$

where $P_A^{\perp} = I - A(A^{\mathrm{T}} Q_y^{-1} A)^{-1} A^{\mathrm{T}} Q_y^{-1}$ is an orthogonal projector (Teunissen 2000) and $\hat{e} = P_A^{\perp} y$ denotes the m-vector of the least squares residuals.

Offset detection and validation

GPS position time series can be perturbed by offsets occurring at time instants that are either known, e.g., due to the documented equipment changes, or unknown. The offset may degrade the accuracy and reliability of the parameters estimated for the series—the estimated velocities or rates for instance (Williams 2003a). To overcome such a problem, a proper offset detection algorithm needs to be developed.

After identifying the most recent functional model, explained in the previous section, we may now employ new statistical tests to detect possible offsets. Two hypotheses testing on the functional model are put forward. In the null hypothesis, it is assumed that there is no offset, whereas in the alternative hypothesis there is at least one. This idea originates from the works of Baarda (1968), Teunissen (2000), and Teunissen et al. (2005) in which some misspecifications in the functional model were detected using the statistical tests. Later, this idea was used in many geodetic applications (Perfetti 2006; Amiri-Simkooei et al. 2017b). It is noted that the classical method for identification of model misspecifications is formulated when the estimation and testing are treated individually. Teunissen (2018) proposed a new and more elegant detection, identification and adaptation (DIA) estimator that combines estimation with testing. The aim of such DIA estimator is to introduce a unifying framework that captures the combined estimation and testing schemes of the DIA method. The DIA method has a wide range of applications in geodetic community. The application in the present contribution on offset detection is considered to be an important issue, which will be the subject of research in the future.

In the classical representation of DIA, the two functional models under the two hypotheses are defined as

$$\text{Model 1 } \left(H_0\right) : \mathrm{E}(y) = Ax \tag{7}$$

versus

$$\text{Model 2 } \left(H_a\right) : \mathrm{E}(y) = [A \,\vdots\, a_j] \begin{bmatrix} x \\ x_j \end{bmatrix} = Ax + a_j x_j, \tag{8}$$

where j runs from 1 to m and $x = \left[y_0, v,\ a_1,\ b_1 \dots\ a_q,\ b_q\right]^{\mathrm{T}}$ is the unknown vector. In Model 2, the x_j is the offset magnitude corresponding to the augmenting m-vector a_j defined as

$$a_j = \begin{bmatrix} H_j(t_1) \\ \vdots \\ H_j(t_m) \end{bmatrix} \tag{9}$$

with the Heaviside step function $H_j(t_i)$ as

$$H_j(t_i) = \begin{cases} 0 & t_i < t_j \\ 1 & t_i \geqslant t_j,\ i = 1, \dots, m \end{cases} . \tag{10}$$

Model 1, given in (7), is considered as the basis or the nominal model. The basic model can also include some disturbances or anomalies, represented by the column vector a_j and its corresponding unknown parameter x_j in Model 2 (j ranges from 1 to m). They are caused by the possible offsets that invalidate the basis model. The epoch at which the null hypothesis powerfully tends to be rejected (namely t_{off}) indicates the presence of an offset in the time series. To this end, the following maximization problem is used:

$$t_{\mathrm{off}} = \arg\max_{t_j} P(t_j)\ , \tag{11}$$

where $P(t_j)$, called the offset power, is obtained for the epoch j, with $j = 1, \dots, m$, from the following equation (Amiri-Simkooei 2013):

$$P\left(t_j\right) = \hat{e}_0^{\mathrm{T}} Q_y^{-1} a_j \left(a_j^{\mathrm{T}} Q_y^{-1} P_A^{\perp} a_j\right)^{-1} a_j^{\mathrm{T}} Q_y^{-1} \hat{e}_0, \tag{12}$$

which can be developed and rewritten for the column vector a_j as

$$P\left(t_j\right) = \frac{\left(\hat{e}_0^{\mathrm{T}} Q_y^{-1} a_j\right)^2}{a_j^{\mathrm{T}} Q_y^{-1} P_A^{\perp} a_j},\ \ j = 1, \dots, m\ , \tag{13}$$

where $\hat{e}_0 = P_A^{\perp} y$ is the least-squares residuals under the null hypothesis. In (13), when the time series contains white noise plus flicker noise, the covariance matrix of time series observations reads $Q_y = \sigma_{\mathrm{w}}^2 I + \sigma_{\mathrm{f}}^2 Q_{\mathrm{f}}$, which becomes available after estimating the variance components σ_w^2 and σ_f^2 using the LS-VCE method.

Because an analytical evaluation of the above maximization problem is complicated, one has to be satisfied with numerical evaluation. That is, we need to compute the column vector a_j at each epoch (for $j = 1, \dots, m$) using (9) and (10). One can, therefore, obtain the power of the offset values for different alternative hypotheses (i.e., $j = 1, \dots, m$). The epoch at which $P\left(t_j\right)$ gets its maximum value, say epoch k, is recognized as a candidate at which possibly an offset has occurred ($t_{\mathrm{off}} = t_k$). The power at this epoch is:

$$\underline{T}_1 = P\left(t_k\right) = \frac{\left(\underline{\hat{e}}_0^{\mathrm{T}} Q_y^{-1} a_k\right)^2}{a_k^{\mathrm{T}} Q_y^{-1} P_A^{\perp} a_k}. \tag{14}$$

As a next stage, one has to validate the detected offset of the time series. In other words, one has to test whether or not the detected offset is significant. To test H_0 against H_a, given in (7) and (8), respectively, the test statistic in (14) can be used (Teunissen et al. 2005). Under the null hypothesis, the test statistics has a central chi-squared distribution with one degree of freedom, i.e., $\underline{T}_1 \sim \chi^2(1, 0)$. With the significance level α, the null hypothesis is accepted if $\underline{T}_1 < \chi_\alpha^2(1, 0)$. This indicates that the offset detected in the previous section is not significant. If the test statistic exceeds the critical value of the chi-squared distribution, the hypothesis will be rejected in the significance level α (i.e., $\underline{T}_1 > \chi_\alpha^2(1, 0)$). This indicates that there is a significant offset occurred at this epoch.

The above distribution assumption is based on the known covariance matrix Q_y. In the case that the covariance matrix is unknown, its variance components are to be estimated by LS-VCE. When dealing with only one variance component, the chi-squared distribution is to be replaced by a Fisher distribution. When there exist at least two variance components, the distribution of the above test statistic becomes complicated. Our observations show, however, that in case of GPS time series when the number of observations m is much larger than the number of unknowns n, the above distributional assumption is still valid to a good approximation.

The previous steps can then be repeated to find yet new offsets (if there is any). This is accomplished by adding a new column a_{off} to the matrix A. The old design matrix A should then be replaced with the new one as $A \leftarrow [A \,\vdots\, a_{\mathrm{off}}]$. The previous steps are repeated by employing the new design matrix. A new time instant of the offset t_{off} can then be detected and tested. The above steps are repeated until the null hypothesis is accepted.

Multivariate GPS position time series

The univariate analysis of the previous section employs one observation vector. To enhance the detection ability of the offsets, one may use several observation vectors having identical structure. This is the case, for example, when analyzing the daily GPS time series of the three coordinate components of north, east and up of a station, simultaneously. A multivariate linear model, also known as a repeated linear model, is in fact an extension of the univariate linear model. The multiple observation vectors have identical design and covariance matrices in the multivariate model. The multivariate analysis of the GPS time series is also divided into two steps: (1) functional and stochastic model identification, and (2) offset detection and validation.

Functional and stochastic model identification

For r time series, the multivariate functional model of the series is of the form:

$$\mathrm{E}(\mathrm{vec}(Y)) = (I_r \otimes A)\mathrm{vec}(X) \tag{15}$$

with the multivariate stochastic model characterized as (Amiri-Simkooei 2009)

$$\mathrm{D}(\mathrm{vec}(Y)) = \Sigma \otimes Q, \tag{16}$$

where Y and X are, respectively, the $m \times r$ and $n \times r$ matrices of time series observations and unknown parameters, vec is the vector operator, and $\otimes$ is the Kronecker product. The $m \times n$ matrix A is the design matrix of a single time series obtained from (1) and (2). This matrix is thus assumed to be identical for all series. The $r \times r$ matrix Σ expresses the correlation among the series, while the $m \times m$ matrix Q characterizes the temporal correlation of the series. The matrix Q can be expressed as $Q = s_\mathrm{w} Q_\mathrm{w} + s_\mathrm{f} Q_\mathrm{f}$, where the white noise cofactor matrix $Q_\mathrm{w} = I_m$ is an identity matrix and Q_f is the cofactor matrix of flicker noise. The contribution of the two noise components is determined through the variance factors s_w and s_f. The matrix Σ along with the factors s_w and s_f can simultaneously be estimated using the multivariate noise assessment techniques. In particular, use is made of the multivariate variant of LS-VCE presented by Amiri-Simkooei (2009).

Offset detection and validation

To implement the multivariate offset detection method, we may use the most recent functional model expressed in (15). The following hypotheses testing on two functional models are then put forward. In the null hypothesis, it is assumed that there is no offset, whereas in the alternative hypothesis there is at least one, common at the same time instant for multiple time series. The two functional models are then of the form:

$$\text{Model 1}: \mathrm{E}(\mathrm{vec}(Y)) = (I_r \otimes A)\mathrm{vec}(X) \tag{17}$$

versus

$$\text{Model 2}: \mathrm{E}(\mathrm{vec}(Y)) = \left(I_r \otimes A\right)\mathrm{vec}(X) + (I_r \otimes a_j)\mathrm{vec}(x_j). \tag{18}$$

The r-vector of offsets is $x_j = [x_j^1, \ldots, x_j^r]$ in which $x_j^i,\ i = 1, \ldots, r$ denotes the offset magnitude of the ith time series. The corresponding design matrix (column vector) a_j can be derived from (9) and (10). The structure introduced in the augmenting matrix $I_r \otimes a_j$ indicates that the offset time instant is the same for all series, but their magnitudes are different through elements of x_j. Here, again, we aim at identifying the time instant at which the offset power becomes maximum (t_off and its corresponding a_off). The same idea as in the univariate analysis is also employed here. Towards this end, the following maximization problem is used:

$$t_\mathrm{off} = \arg\max_{t_j} P(t_j)\ , \tag{19}$$

where $P(t_j)$, called the multivariate offset power, is obtained for each of the epochs (i.e., $j = 1, \ldots, m$) from the following equation (Amiri-Simkooei 2013):

$$P\left(t_j\right) = \mathrm{tr}\left(\hat{E}_0^T Q^{-1} a_j \left(a_j^T Q^{-1} P_A^\perp a_j\right)^{-1} a_j^T Q^{-1} \hat{E}_0 \Sigma^{-1}\right), \tag{20}$$

which can be developed and rewritten as

$$P\left(t_j\right) = \frac{a_j^\mathrm{T} Q^{-1} \hat{E}_0 \Sigma^{-1} \hat{E}_0^\mathrm{T} Q^{-1} a_j}{a_j^T Q^{-1} P_A^\perp a_j},\ \ j = 1, \ldots, m, \tag{21}$$

where $\hat{E}_0 = P_A^\perp Y$ is the $m \times r$ least-squares residual matrix under the null hypothesis. As already mentioned, matrices Σ and Q are to be estimated using the LS-VCE method. The preceding equation, with $\hat{\Sigma} = \hat{E}_0^\mathrm{T} Q^{-1} \hat{E}_0 / (m-n)$ (see Amiri-Simkooei 2009), can be further reformulated as

$$P\left(t_j\right) = b\,\frac{a_j^T Q^{-1} P_{\hat{E}_0} a_j}{a_j^T Q^{-1} P_A^\perp a_j} = b\,\frac{\left\|P_{\hat{E}_0} a_j\right\|_{Q^{-1}}^2}{\left\|P_A^\perp a_j\right\|_{Q^{-1}}^2}, \tag{22}$$

where j runs from 1 to m, $b = m - n$ is the redundancy (degrees of freedom) of the univariate functional model, $\|\cdot\|_{Q^{-1}}^2 = (\,.\,)^\mathrm{T} Q^{-1}(\,.\,)$ denotes the squared norm of a vector, and $P_{\hat{E}_0} = \hat{E}_0(\hat{E}_0^T \mathrm{Q}^{-1} \hat{E}_0)^{-1} \hat{E}_0^T Q^{-1}$ is an orthogonal projector.

Again, the numerical evaluation of the maximization problem in (19) is required. That is, we need to compute the column vector a_j at each epoch (for $j = 1, \dots, m$) using (9) and (10). One can, therefore, obtain the power of the offset values for different alternative hypotheses (i.e., $j = 1, \dots, m$) by employing (21) or (22). The epoch at which $P(t_j)$ gets its maximum value, say epoch k, is recognized as a candidate at which possibly an offset has occurred (t_{off}). The power at this epoch becomes:

$$\underset{-}{T}_r = P(t_k) = P_{\max} = b\,\frac{\left\| P_{\hat{E}_0} a_k \right\|^2_{Q^{-1}}}{\left\| P_A^{\perp} a_k \right\|^2_{Q^{-1}}}. \tag{23}$$

One then has to validate the detected offsets of the multivariate analysis. We need to test whether or not the detected offset is statistically significant. Under the null hypothesis, the test statistics has a central chi-squared distribution with r degrees of freedom, i.e., $\underline{T}_r \sim \chi^2(r, 0)$. With the significance level α, the null hypothesis is accepted if $\underline{T}_r < \chi^2_\alpha(r, 0)$. This indicates that the offset detected is not significant. If the test statistic exceeds the critical value of the chi-squared distribution, the hypothesis will be rejected in the significance level α (i.e., $\underline{T}_r > \chi^2_\alpha(r, 0)$). This indicates that there is a significant offset occurred at this epoch. The same procedure is repeated for identifying other possible offsets.

Numerical results and discussion

To investigate the performance of the proposed method, synthetic GPS time series are used in this section. Two issues are highlighted in this section. (1) We investigate the impact of an appropriate stochastic model of the series on the offset detection algorithm. (2) The superiority of the multivariate analysis over the univariate analysis is also highlighted in this section. The significance level of the hypothesis testing is considered to be $\alpha = 0.001$. This will then result in the critical values of the univariate and multivariate test $\chi^2_\alpha(1, 0) = 10.83$ and $\chi^2_\alpha(3, 0) = 16.27$, respectively. The significance level α is the probability of incorrect decision when indeed the null hypothesis is true. Therefore, the smaller the significance level is, the more likely the hypothesis testing will detect a small offset. This may, however, lead also to an increase in detecting incorrect offsets that was not simulated but were detected as an offset by the method.

We create synthetic time series of three coordinate components of a permanent GPS station. The data are time series spanning 10 years of daily coordinate positions. They consist of a linear trend, the annual and semi-annual signals, offsets, and white and flicker noise. The parameter settings characterizing the synthetic data sets are summarized in Table 1. We only consider the first and second harmonics of the annual signal. Higher harmonics such as tri-annual signals were also included in the series. However, there was no significant change in the final results. The multivariate analysis can in principle handle the statistical correlation among the three coordinate components. Such correlation is, however, absent in real GPS time series and hence it was ignored here (see Amiri-Simkooei et al. 2007). For each time series, the covariance matrix of the series is constructed based on the white and flicker noise amplitudes specified in this table. A random error vector of normal distribution is then simulated using the Cholesky decomposition of the covariance matrix. The simulated error, consisting of both white and colored noise, is then added up to the deterministic model explained above.

Table 1 Parameter settings of simulated data sets employed in functional and stochastic models

Component	North	East	Up
Parameter			
WN amplitude (mm)	1.5	1.5	3
FN amplitude (mm/year$^{1/4}$)	3	3	6
Annual amplitude (mm)	2	2	3
Semi-annual amplitude (mm)	1	1	2
v (mm/year)	5	5	1
y_0 (mm)	10	10	10
Offset magnitude (mm)	[1–3]	[1–3]	[2–6]

The next step is to introduce offsets. The magnitudes of offsets are also generated randomly, which range from 1 to 3 mm for the north and east components and from 2 to 6 mm for the up component (Table 1). Each offset is simulated using a uniform distribution in the above ranges. Each simulated time series has 11 offsets occurred at the fixed times instants of multiples of 300, i.e., at epochs 300, 600, …, 3300. Figure 1 illustrates one typical example of the simulated time series along with the position of the simulated offsets. We did other tests considering the efficiency of the solutions when the offsets are at the beginning or at the end of the series. The results did not change significantly.

The goal now is to identify the offsets included intentionally in the time series at the known epochs. The advantage of the simulated data compared with real data is that the magnitude and position of the offsets are perfectly known. This will make some statistical analysis about the detection ability of the proposed method using both univariate and multivariate analyses described in the previous section. The offsets are detected in four cases explained in

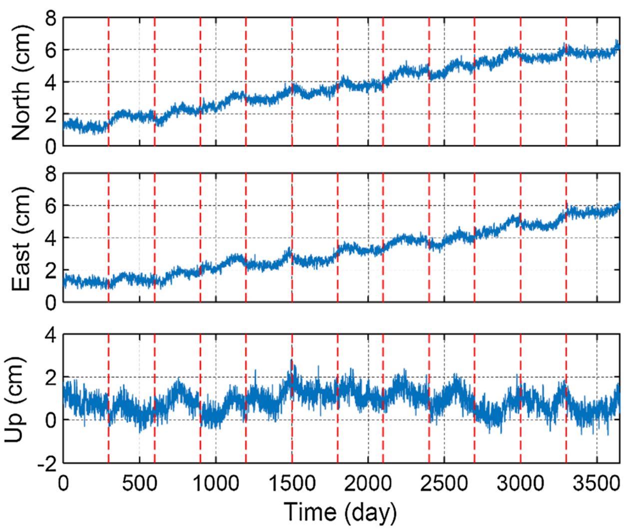

Fig. 1 Typical example of synthetic time series with settings described in Table 1; offsets are at epochs 300, 600, …, 3300 days (vertical dashed lines)

Table 2 Four cases of offset detection: univariate vs. multivariate analysis; simple stochastic model vs. realistic stochastic model

Case	Type of analysis		Stochastic model	
	Univariate	Multivariate	White	Flicker
I	✓	–	✓	–
II	✓	–	✓	✓
III	–	✓	✓	–
IV	–	✓	✓	✓

Table 2. In Case I, the coordinate components are treated individually using the univariate analysis. The covariance matrix of the series is composed of only white noise in this case. This is, however, not a realistic noise model because the simulated time series contain both white and flicker noise. Case II is similarly performed by the univariate analysis, but the covariance matrix is considered to be realistic as it considers both white and flicker noise. Cases III and IV are similarly specified for the multivariate analysis.

The data of the north, east and up components of a GPS station are simulated for 500 runs based on the parameter settings in Table 1. Under the above four cases, we implement the offset detection algorithm of the previous section. The results are presented for Cases I and II (Fig. 2) and Cases III and IV (Fig. 3) within a 3-day window for the univariate and multivariate analyses, respectively. In the univariate analysis, the three coordinate components are treated individually for the north, east and up component, whereas, in the multivariate analysis, the observation vector consists of three time series of the north, east and up components at the same time. A few observations are highlighted:

- The same significance level $\alpha = 0.001$ was used in the hypothesis testing of all cases. An algorithm is said to be reliable if it identifies the correct positions of all 11 offsets of the time series. False alarm (Type I error) and missed detection (Type II error) are two kinds of errors in the hypotheses testing. With the specified α, Cases II and IV, considering white plus flicker noise, identifies more or less 11 offsets, although some of them are incorrect causing Type I error. This, however, does not hold for Cases I and III, consisting of only white noise, because they fail to stop after identifying 11 offsets. Our observations indicate that these two cases identify more than 20 offsets per time series, most of which are incorrect. This indicates that an even smaller α is to be set for Cases I and III to detect on average 11 offsets. Therefore, to make a fair comparison among all cases, we just analyze the first 11 detected offsets of each case.
- The right frames of Figs. 2 and 3, when compared with the left frames, indicate that the realistic noise model can significantly improve the offset detection procedure. For the univariate analysis, Case II has higher detection ability compared with Case I (49 vs. 39%). A similar situation also holds for the multivariate analysis of Cases IV and III (87 vs. 71%). This indicates that ignoring the existing colored noise (e.g., flicker noise) of the time series leads to a sub-optimal offset detection method. Therefore, a realistic covariance matrix of time series has a direct impact on the results of the offset detection.
- Figure 3, when compared with Fig. 2, indicates that the multivariate analysis has higher detection ability than the univariate analysis. For the white noise model, the offset ability detection of Case III is much larger than that of Case I (71 vs. 39%). Similarly, Case IV has higher detection ability compared with Case II (87 vs. 48%). The multivariate analysis has thus much better performance than the univariate analysis. This is because in the multivariate analysis three coordinate time series contribute simultaneously in the offset detection procedure.

The above discussions indicate that a reliable offset detection method should take an appropriate noise model into consideration. For GPS position time series, the noise characteristics are described as a combination of white plus flicker noise (Zhang et al. 1997; Mao et al. 1999; Williams et al. 2004; Amiri-Simkooei et al. 2007). Therefore, a reliable automatic method, prior to offset detection, should estimate the amplitudes of different noise

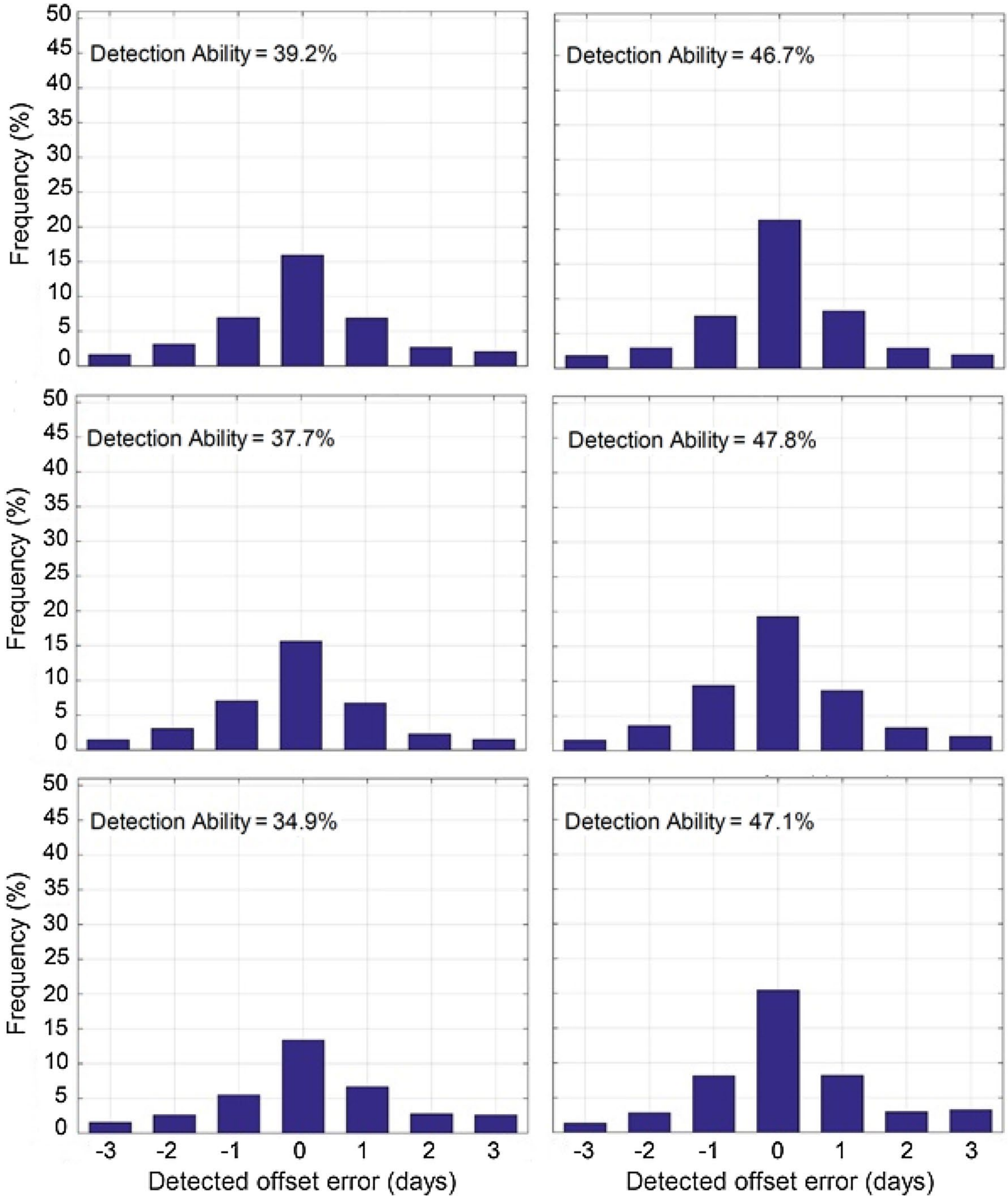

Fig. 2 Percentage histogram of correctly detected offsets within a 3-day window in univariate analysis of 500 independent runs for north (top row), east (middle row) and up (bottom row) components. Only white noise (left column), white noise plus flicker noise (right column)

components of the series. In addition, because an offset usually affects the three coordinate components of a station simultaneously, a properly performed offset detection algorithm should take this advantage into account. The multivariate analysis can thus provide higher detection ability than the univariate analysis because it considers the contribution of the simultaneous offsets of the three components. Figure 4 shows the scatter plot of the detected offsets on the 500 independent runs using the multivariate analysis vs. their true values at epochs 300, 600, …, 3300 days. The bottom frame is less scattered than the top frame, and hence providing better results. This indicates, among all possibilities, the multivariate model with a proper noise model consisting of both white

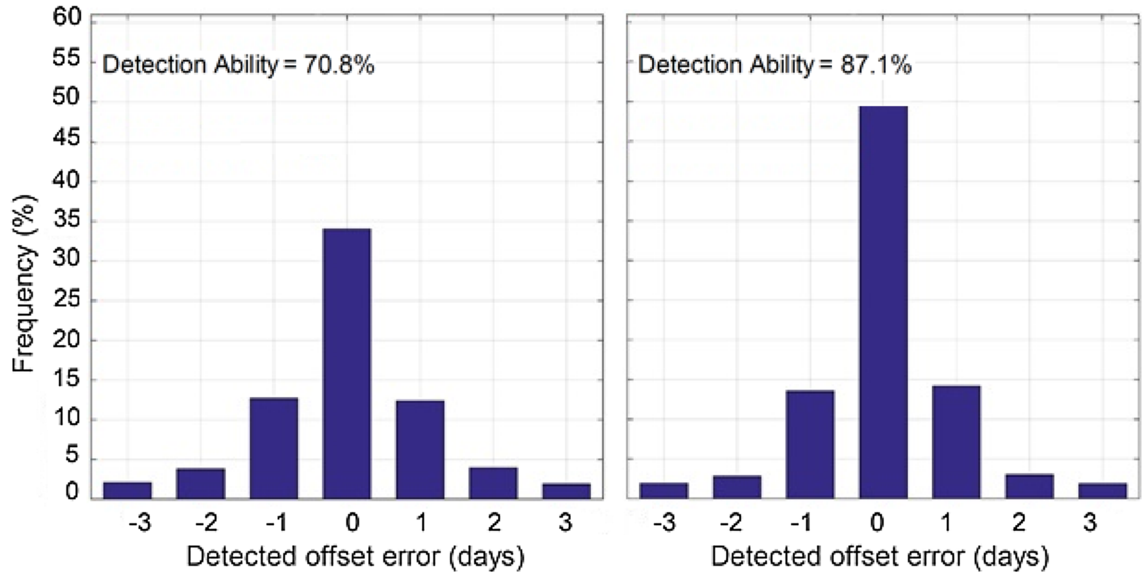

Fig. 3 Percentage histogram of correctly detected offsets within a 3-day window in multivariate analysis of 500 independent runs; Case III of only white noise (left), Case IV of white noise plus flicker noise (right)

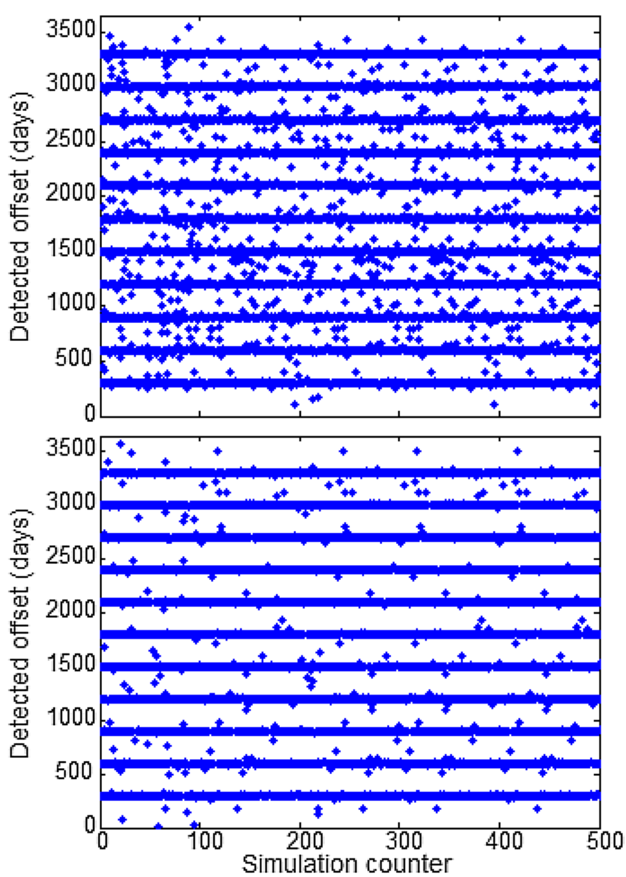

Fig. 4 Scatter plot of detected offsets on 500 independent runs using multivariate analysis vs. their true values at epochs 300, 600, …, 3300 days; White noise only (top), white noise plus flicker noise (bottom)

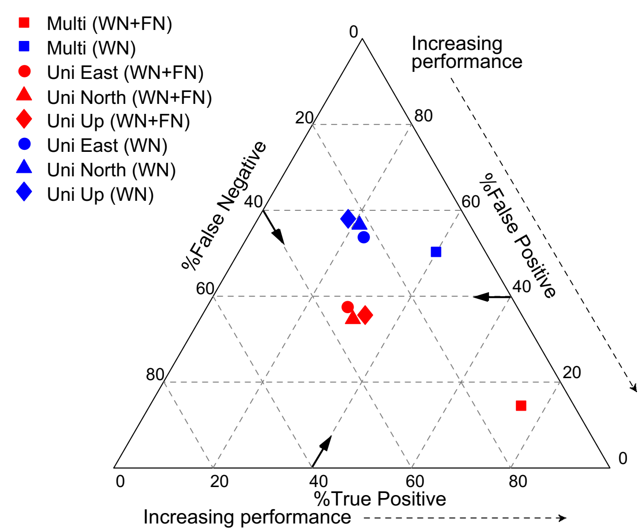

Fig. 5 Ratio between the three performance measures of the solutions (TP, FP and FN)

noise and flicker noise (Case IV) provides the best results. This is also what we can observe when comparing the different frames in Figs. 2 and 3.

Finally, to consider the overall performance of the solutions, the true positive (TP), false positive (FP), and false negative (FN) counts are also computed within a 3-day window. A "True Positive" defines an offset that was originally simulated and also detected by the solution. A "False Positive" (type I error) refers to an offset that was not simulated, but it was reported to be present. Finally, an offset that was simulated but could not be detected is

Table 3 Detection ability of the correct detection of offsets over all simulated offsets (in percent)

Case	Multivariate	Univariate (E)	Univariate (N)	Univariate (U)
WN + FN	31.6	15.9	13.5	12.8
WN	24.0	13.0	13.3	12.6

referred to as a "False Negative", or type II error (Gazeaux et al. 2013). Figure 5 illustrates the percentages of the three variables TP, FP and FN by their position in an equilateral triangle. The method has a perfect performance if it appears on the bottom right corner of the triangle. This indicates that the multivariate analysis with considering white and flicker noise has the best performance among other methods.

We now aim at investigating the efficacy of the solutions at small offsets. To this end, the offsets magnitudes were halved, i.e., 0.5–1.5 mm for horizontal components and 1–3 mm for up component. The other settings remain the same as the first experiment, provided in Table 1. Table 3 gives the power of the correct detection of offsets over all synthetic time series having small offsets. Here again, the multivariate analysis has the higher detection power than the univariate analysis. Again, the realistic noise model can improve the offset detection procedure.

Finally, to investigate the performance of the solutions in the presence of random walk noise, the synthetic GPS time series of the north, east and up components were simulated on 100 independent runs. Parameter settings of simulated data sets employed in the functional model are provided in Table 1. The covariance matrix of the series, composed of white, flicker and random walk noise, is constructed based on the white and flicker noise amplitudes specified in Table 1. The random walk noise amplitude is considered to be 0.25 mm/year$^{1/2}$ for the east and north components, and 0.75 mm/year$^{1/2}$ for the up component; the average values are reported by Amiri-Simkooei et al. (2017a). The offsets are then detected under four cases over three different stochastic models (WN, WN + FN and WN + FN + RWN). The results are provided in Table 4. The multivariate analysis has again higher detection power than the univariate analysis. Also, the realistic noise model can improve the offset detection procedure in both cases of the univariate and multivariate analyses.

Concluding remarks

It is well known that the GPS position time series can be disrupted by offsets. Subsequently, the accuracy of the estimated parameters such as site velocities is degraded. Proper analysis of time series in general and a reliable offset detection method in particular are thus essential issues to be considered. To this end, one requires realistic and proper functional and stochastic models of the series. A proper functional model for the GPS position time series analysis includes a linear trend, periodic signals, probabilistic offsets, and blunders. A realistic stochastic model of the GPS position time series should best model the noise components of the data such as white noise and flicker noise. For this purpose, one may use the least squares variance component estimation (LS-VCE) to estimate the noise components.

We proposed a mathematical foundation for offset detection in the GPS time series, which can be applied to the univariate or multivariate time series analyses. For this purpose, two hypotheses testing on two functional models were put forward. In the null hypothesis, we assumed that there is no offset, whereas in the alternative hypothesis there is at least one. The two functional models were compared by means of the statistical test, called the generalized likelihood ratio (GLR) test. Using the GLR test, one can decide between the original model under the null hypothesis (a model without offset) and the extended model under the alternative hypothesis.

The performance of the multivariate time series analysis was compared with the univariate analysis through simulated data sets. The results indicated that a proper selection of the noise components of the data has a significant impact on the correct detection of offsets. This indicates that ignoring the colored noise results in an inaccurate offset detection method. The results indicated that the multivariate analysis of the time series is more efficient than the univariate analysis for offset detection in the sense that the three coordinate components of north, east and up of a station simultaneously contributed in the offset detection procedure.

Table 4 Detection power (in percent) of simulated offsets over three different stochastic models

Case	Multivariate	Univariate (E)	Univariate (N)	Univariate (U)
WN + FN + RWN	47.1	23.5	26.7	18.3
WN + FN	27.6	18.5	20.6	12.0
WN	15.2	13.9	12.4	7.8

References

Amiri-Simkooei AR (2007) Least-squares variance component estimation: theory and GPS applications. PhD thesis, Mathematical Geodesy and Positioning. Faculty of Aerospace Engineering, Delft University of Technology, Delft, Netherlands

Amiri-Simkooei AR (2009) Noise in multivariate GPS position time series. J Geodesy 83(2):175–187

Amiri-Simkooei AR (2013) On the nature of GPS draconitic year periodic pattern in multivariate position time series. J Geophys Res Solid Earth 118(5):2500–2511

Amiri-Simkooei AR, Asgari J (2012) Harmonic analysis of total electron contents time series: methodology and results. GPS Solut 16(1):77–88

Amiri-Simkooei AR, Tiberius CCJM, Teunissen PJG (2007) Assessment of noise in GPS coordinate time series: methodology and results. J Geophys Res Solid Earth 112(B7):413

Amiri-Simkooei AR, Mohammadloo TH, Argus DF (2017a) Multivariate analysis of GPS position time series of JPL second reprocessing campaign. J Geodesy 91(6):685–704

Amiri-Simkooei AR, Alaei-Tabatabaei SM, Zangeneh-Nejad F, Voosoghi B (2017b) Stability analysis of deformation-monitoring network points using simultaneous observation adjustment of two epochs. J Surv Eng 143(1):04016020

Baarda W (1968) A testing procedure for use in geodetic networks. Technical report, Netherlands Geodetic Commission, publ. on Geodesy, New series, Vol. 2(5), Delft

Beaulieu C, Seidou O, Ouarda T, Zhang X, Boulet G, Yagouti A (2008) Intercomparison of homogenization techniques for precipitation data. Water Resour Res 44(2):W02425. https://doi.org/10.1029/2006WR005615

Borghi A, Cannizzaro L, Vitti A (2012) Advanced techniques for discontinuity detection in GNSS coordinate time-series. In: An Italian case study, geodesy for planet earth, International Association of Geodesy Symposia, 136, pp 627–634

Chen C, Tiao GC (1990) Random level-shift time series models, ARIMA approximations and level-shift detection. J Bu Econ Stat 8(1):83–97

Gazeaux J, Flaounas E, Naveau P, Hannart A (2011) Inferring change points and nonlinear trends in multivariate time series: application to West African monsoon onset timings estimation. J Geophys Res 116:D05, 101

Gazeaux J, Williams SDP, King MA, Bos M, Dach R, Deo M, Moore AW, Ostini L, Petrie E, Roggero M, Teferle FN, Olivares G, Webb FH (2013) Detecting offsets in GPS time series: first results from the detection of offsets in GPS experiment. J Geophys Res Solid Earth 118(5):2397–2407

Johansson JM et al (2002) Continuous GPS measurements of postglacial adjustment in Fennoscandia. 1. Geodetic results. J Geophys Res Solid Earth 107(B8):2157

Kenyeres A, Bruyninx C (2004) EPN coordinate time series monitoring for reference frame maintenance. GPS Solut 8(4):200–209

Khodabandeh A, Amiri-Simkooei AR, Sharifi MA (2012) GPS position time-series analysis based on asymptotic normality of M-estimation. J Geodesy 86(1):15–33

Koch K-R (1999) Parameter estimation and hypothesis testing in linear model, 2nd edn. Springer, Berlin, p 333

Langbein J (2004) Noise in two-color electronic distance meter measurements revisited. J Geophys Res Solid Earth 109:B4

Mao A, Christopher GA, Timothy H (1999) Noise in GPS coordinate time series. J Geophys Res 104(B2):2797–2816

Montillet JP, Williams SDP, Koulali A, McClusky SC (2015) Estimation of offsets in GPS time-series and application to the detection of earthquake deformation in the far-field. Geophys J Int 200(2):1207–1221

Olshen AB, Venkatraman ES, Lucito R, Wigler M (2004) Circular binary segmentation for the analysis of array based DNA copy number data. Biostatistics 5(4):557–572

Perfetti N (2006) Detection of station coordinates discontinuities within the Italian GPS Fiducial Network. J Geodesy 80(7):381–396

Pham D, Xu C, Prince J (2000) A survey of current methods in medical image segmentation. Annu Rev Biomed Eng 2:315

Picard F, Robin S, Lavielle M, Vaisse C, Daudin J (2005) A statistical approach for CGH microarray data analysis. BMC Bioinform 6:27

Ray J, Altamimi Z, Collilieux X, vanDam T (2008) Anomalous harmonics in the spectra of GPS position estimates. GPS Solut 12(1):55–64

Segall P, Davis JL (1997) GPS applications for geodynamics and earthquake studies. Annu Rev Earth Planet Sci 25(1):301–336

Teunissen PJG (1988) Towards a least-squares framework for adjusting and testing of both functional and stochastic models. In: Internal research memo, Geodetic Computing Centre, Delft. A reprint of original 1988 report is also available in 2004, Series on mathematical Geodesy and Positioning, No. 26

Teunissen PJG (2000) Testing theory an introduction. In: Series on Mathematical Geodesy, Delft University Press, VSSD Publisher, p 156

Teunissen PJG (2018) Distributional theory for the DIA method. J Geodesy 92(1):59–80

Teunissen PJG, Amiri-Simkooei AR (2008) Least-squares variance component estimation. J Geodesy 82(2):65–82

Teunissen PJG, Simons DG, Tiberius CCJM (2005) Probability and observation theory. Delft University of Technology, Delft

Vitti A (2012) SIGSEG: A tool for the detection of position and velocity discontinuities in geodetic time-series. GPS Solut 16(3):405–410

Williams SDP (2003a) Offsets in global positioning system time-series. J Geophys Res Solid Earth 108(B6):2310

Williams SDP (2003b) The effect of coloured noise on the uncertainties of rates estimated from geodetic time series. J Geodesy 76(9–10):483–494

Williams SDP, Bock Y, Fang P, Jamason P, Nikolaidis RM, Prawirodirdjo L, Miller M, Johnson DJ (2004) Error analysis of continuous GPS position time-series. J Geophys Res Solid Earth 109(B3):412

Zhang J, Bock Y, Johnson H, Fang P, Williams S, Genrich J, Wdowinski S, Behr J (1997) Southern California permanent GPS geodetic array: error analysis of daily position estimates and site velocities

A. R. Amiri-Simkooei graduated from the Mathematical Geodesy and Positioning Group at the Delft University of Technology, The Netherlands. His research area focused on least-squares variance component estimation (LS-VCE) with applications to GPS data. He is currently a professor at the Department of Geomatics Engineering, University of Isfahan, Iran.

M. Hosseini-Asl is a Ph.D. student at the University of Tehran, Tehran, Iran. She completed her M.Sc. degree in 2012 in Geodesy at University of Isfahan, Isfahan, Iran. Her interests are mainly focused on GPS time series analysis.

F. Zangeneh-Nejad completed her Ph.D. in 2018 in Geodesy at University of Tehran, Tehran, Iran. Her principal research interests are GNSS-based positioning and GNSS processing techniques. She has also worked extensively on the applications of the LS-VCE method in time series analysis.

J. Asgari graduated with a Ph.D. degree in Geodesy from ENSG, France. His research focused on GNSS positioning techniques with an emphasis on Precise Point Positioning (PPP). His research interests are satellite positioning, PPP, and ionospheric studies. He is currently an assistant professor at the Department of Geomatics Engineering, University of Isfahan, Iran.