

Weighted spatiotemporal filtering using principal component analysis for analyzing regional GNSS position time series

Weiwei Li · Yunzhong Shen · Bofeng Li

Received: 18 September 2014 / Accepted: 21 January 2015 / Published online: 6 February 2015
© Akadémiai Kiadó 2015

Abstract Principal component analysis (PCA) is a powerful tool for extracting common mode errors from the position time series of a regional station network determined by global navigation satellite system (GNSS). It is implicitly based on the assumption that a time series dataset contains temporally uniform white noise. Since the position time series of a regional station network are not uniform and could have data gaps, this paper develops a PCA-based weighted spatiotemporal filtering (WSF) approach by taking into account the positioning formal error of daily solution and the data gaps in time series. The position time series of 27 GNSS stations of the Crust Motion Observation Network of China are analyzed to demonstrate the performance of WSF approach, and also compared with the modified PCA technique in Shen et al. (J Geod 88:1–12, 2014). It shows that the WSF approach outperforms the modified PCA at 21, 19 and 17 out of the total 27 stations for north, east and up components, respectively. The average formal standard deviation of unit weight derived from WSF and modified PCA are 2.12, 2.42, 5.88 and 2.21, 2.52, 6.05 for north, east and up components, respectively; the relative improvements are 4.1, 4.0 and 2.8 %. Moreover, two simulations of a network with 4 stations are processed to show the performance of WSF. The results show that WSF provides better results for all coordinate components of all stations when the local effects are small or negligible. For cases when the local effects becoming larger, the WSF performs better than the modified PCA from the statistical point of view. From the real and synthetic time series analysis results, it is reasonable to conclude that the positioning formal error of daily solution should be considered in spatiotemporal filtering.

Keywords GNSS position time series · Principal component analysis · Common mode errors · Weighted spatiotemporal filtering

W. Li · Y. Shen (✉) · B. Li
College of Surveying and Geo-informatics, Tongji University, Shanghai,
People's Republic of China
e-mail: yzshen@tongji.edu.cn

W. Li
Center for Spatial Information Science and Sustainable Development,
Shanghai, People's Republic of China

1 Introduction

Many regional station networks have been established for monitoring the crust deformation and water vapor contents using global navigation satellite system (GNSS) (Wdowinski et al. 1997; Yuan et al. 2008; Nilsson and Elgered 2008; Lee et al. 2013; Ji et al. 2014). The dominant noises in GNSS position time series are white, flicker and random walk noises (Zhang et al. 1997; Langbein and Johnson 1997; Mao et al. 1999; Williams et al. 2004; Langbein 2004). Since the position time series of a regional GNSS network contain significant common mode errors (CMEs), several approaches, such as ‘stacking’ (Wdowinski et al. 1997), weighted ‘stacking’ (Nikolaidis 2002; Tian and Shen 2011) and principal component analysis (PCA) (Dong et al. 2006), have been developed for extracting the CMEs. The effects of un-modeled signals in the extracted CMEs have been discussed by Shen and Li (2014). The weighted ‘stacking’ approach takes into account of the positioning formal error of time series. Since its spatial response is uniform, it is usually used for small regional GNSS station networks. Recently the weighted ‘stacking’ approach has been improved by considering the correlation coefficients between the computed station and other stations (Tian and Shen 2011). The multivariate analysis approach of Amiri-Simkooei (2009) simultaneously computes the covariance of different stations for both white and colored noises in a unified model, this approach is theoretically rigorous, but computationally inefficient.

The PCA approach of Dong et al. (2006) reveals the different spatial responses for each station by its own data; thereby it can be used for processing relatively larger GNSS networks. However, it is implicitly assumed that the position time series contains only temporally uniform white noises in developing their PCA approach (Dong et al. 2006). When there are data gaps in the time series, the iterative computations are required for extracting CMEs. Ji and Herring (2013) introduced a Kalman filtering approach of filling data gap and the uncertainties in PCA analysis. Based on the principle that a time series can be reproduced with its principal components, Shen et al. (2014) recently proposed a modified PCA for extracting CMEs from the GNSS time series with missing data and contaminated by uniform white noises. However, the accuracies of daily solutions in a regional GNSS network are temporally and spatially varying. For a global GNSS network, the noises are also not uniform, and the solutions are usually more accurate in winter (Tian et al. 2013).

This paper develops a weighted spatiotemporal filtering (WSF) approach for extracting CMEs from the GNSS time series of a regional station network, which is contaminated by inhomogeneous noises and is with missing data. Compared to the modified PCA, the WSF approach considers the positioning formal error and the effects of missing data in constructing covariance matrix. Besides, the CMEs are calculated by multiplying principal components, eigenvectors of the covariance and the formal errors of the time series. The rest of paper is organized as follows. The Sect. 2 presents a brief review on the existing regional spatiotemporal filtering approaches. The following Sect. 3 presents the PCA-based WSF approach. Real and synthetic time series are processed and analyzed in Sect. 4, and conclusions are summarized in last Sect. 5.

2 Regional spatiotemporal filtering approaches

The position time series of each GNSS station is expressed as

$$y(t_i) = y_0 + bt_i + \sum_{f=1}^2 (c_f \sin(2\pi f t_i) + d_f \cos(2\pi f t_i)) + \sum_{j=1}^{n_g} g_j H(t_i - T_{gj}) + s(t_i) + \varepsilon(t_i), \quad (1)$$

where, $y(t_i)$ represents the position at epoch t_i (unit of year). y_0 , b the position and velocity parameters, c_f and d_f the periodic motion parameters ($f = 1, 2$ represent the annual and semi-annual terms, respectively). The offsets term $\sum_{j=1}^{n_g} g_j H(t_i - T_{gj})$ is caused by earthquakes, environmental, equipment changes or human intervention, in which g_j is the magnitude changed at epochs T_{gj} , n_g is the total number of offsets, and H is Heaviside step function. $s(t_i)$ and $\varepsilon(t_i)$ are the CMEs and other observation errors, respectively. The residual time series for station j can be expressed as

$$x(j, t_i) = s(j, t_i) + \varepsilon(j, t_i). \quad (2)$$

The weighted stacking approach proposed by Nikolaidis (2002) computes the regional filtering component with

$$s(t_i) = \frac{\sum_{j=1}^n (x(j, t_i) / \sigma_{(j,i)}^2)}{\sum_{j=1}^n (1 / \sigma_{(j,i)}^2)}, \quad (3)$$

where, $s(t_i)$ is the filtering component of epoch t_i , $\sigma_{(j,i)}^2$ is the variance of the observation at this epoch for station j . n is the number of sites for deriving CME in the regional network. This weighted stacking approach is further modified by considering the correlation between stations (Tian and Shen 2011), with the expression as

$$s(j, t_i) = \frac{\sum_{k=1}^n \frac{x(k, t_i)}{\sigma_{(k,i)}^2} \frac{r_{(j,k)}}{\max |r_{(j,k)}|}}{\sum_{k=1}^n \frac{1}{\sigma_{(k,i)}^2} \frac{r_{(j,k)}}{\max |r_{(j,k)}|}}, \quad (4)$$

where, $r_{(j,k)}$ is the correlation coefficient of station j and k , which is derived from all epochs of the time series in Tian and Shen (2011). The quantity $\max |r_{(j,k)}|$ does not affect the result in our opinion. Obviously, the CMEs derived from (4) are different for different stations. The time series after filtering are

$$\tilde{x}(j, t_i) = x(j, t_i) - s(j, t_i), \quad (i = 1, 2, \dots, m; j = 1, 2, \dots, n), \quad (5)$$

where $\tilde{x}(j, t_i)$ denotes the filtered series.

According to Dong et al. (2006), the CMEs derived by PCA can be summarized as follows. First, we form the covariance matrix \mathbf{B} by computing its i th row and j th column element $b(i, j)$ as

$$b(i, j) = \frac{1}{m-1} \sum_{k=1}^m x(i, t_k) x(j, t_k). \quad (6)$$

Since \mathbf{B} is a positive definite symmetric matrix, it can be decomposed with a diagonal matrix and an orthogonal matrix, the elements of diagonal matrix are called eigenvalues and the

column vectors of orthogonal matrix are called eigenvectors. Then computing the principal components with

$$a_k(t_i) = \sum_{j=1}^n x(j, t_i) v_k(j), \quad (k = 1, 2, \dots, n), \quad (7)$$

where, a_k is called the k th principal component (PC), $v_k(j)$ is the j th element of k th eigenvector. The original time series can be recovered from the PCs by

$$x(j, t_i) = \sum_{k=1}^n a_k(t_i) v_k(j), \quad (j = 1, 2, \dots, n). \quad (8)$$

If the eigenvalues are sorted in descending order, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$, and the spectral energy is dominated by the first p PCs, the original time series $x(j, t_i)$ can be mainly reconstructed by

$$x(j, t_i) = \sum_{k=1}^p a_k(t_i) v_k(j) + e(j, t_i), \quad (j = 1, 2, \dots, n), \quad (9)$$

where, $e(j, t_i)$ denotes the rest terms. If only the first PC satisfies the criterion of CME defined in Dong et al. (2006), the CME of the j th station is computed by

$$s(j, t_i) = a_1(t_i) v_1(j), \quad (i = 1, 2, \dots, m; j = 1, 2, \dots, n), \quad (10)$$

s is not only temporally correlated, but also different for stations. This filtering method can be used for processing large regional GNSS station network due to its non-uniform spatial response characteristics.

3 Weighted spatiotemporal filtering for regional GNSS time series

With n stations over m epochs ($m \geq n$), we can construct an $n \times m$ matrix \mathbf{x} for north, east and up coordinate components, respectively. The column of \mathbf{x} represents the coordinate components at a given epoch for all stations, whereas the row is the coordinate components of one station for all epochs. The j th row and k th column element $x(j, t_k)$ denotes the quantity of station j at epoch t_k . Equations from (6) to (10) are valid when the time series are complete and uniform, i.e. no data gaps and with the same formal error for each station and epoch.

Assumed $\sigma_{(j,k)}$ is the formal error for the time series $x(j, t_k)$ of station j at epoch t_k . In real observations, this formal error is station-and epoch-dependent. It is necessary to convert the time series to that with uniform formal error before using traditional PCA. Since $x(j, t_k)$ is demeaned and detrended, the converted time series $x'(j, t_k) = x(j, t_k)/\sigma_{(j,k)}$ must have the uniform formal error. Therefore, we can compute the ij th element $b'(i, j)$ of the covariance matrix \mathbf{B}' with

$$b'(i, j) = \frac{1}{m-1} \sum_{k=1}^m x'(i, t_k) x'(j, t_k) = \frac{1}{m-1} \sum_{k=1}^m \frac{x(i, t_k)}{\sigma_{(i,k)}} \frac{x(j, t_k)}{\sigma_{(j,k)}}. \quad (11)$$

Then we can compute the eigenvector v'_k of \mathbf{B}' , and the PCs with

$$a'_k(t_i) = \sum_{j=1}^n x'(j, t_i) v'_k(j), \quad (k = 1, 2, \dots, n). \quad (12)$$

If only the first PC is the common mode, then extracting the CME element (s'_w) of the converted time series with

$$s'_w(j, t_i) = a'_1(t_i)v'_1(j). \quad (13)$$

Then recovering the CME element $s_w(j, t_i)$ for the original time series by

$$s_w(j, t_i) = \sigma_{(j,i)}s'_w(j, t_i). \quad (14)$$

When data gaps exist, the covariance can only be determined by the valid observations. Let M_i and M_j denote the data sets of stations i and j , m_i and m_j denote the number of epochs. The elements of covariance matrix \mathbf{B}' can be expressed with

$$b'(i, i) = \frac{1}{m_i - 1} \sum_{k \in M_i} \frac{x(i, t_k)}{\sigma_{(i,k)}} \frac{x(i, t_k)}{\sigma_{(i,k)}}, \quad b'(i, j) = \frac{1}{m_{ij} - 1} \sum_{k \in M_i \cap M_j} \frac{x(i, t_k)}{\sigma_{(i,k)}} \frac{x(j, t_k)}{\sigma_{(j,k)}}, \quad (15)$$

where m_{ij} denotes the number of epochs of the intersect set $M_i \cap M_j$. Also PCs can be expressed with the converted series and eigenvector of \mathbf{B}' ,

$$a'_k(t_i) = \sum_{j \in S_i} x'(j, t_i)v'_k(j) + \sum_{j \notin S_i} x'(j, t_i)v'_k(j), \quad (k = 1, 2, \dots, n), \quad (16)$$

where S_i denotes the set of stations which have observations at epoch t_i . The second term in (16) can be recovered with the PCs, that is,

$$x'(j, t_i) = \sum_{k=1}^n a'_k(t_i)v'_k(j). \quad (17)$$

By substituting (17) into (16), we get the n equations of PCs as

$$\left. \begin{aligned} a'_1(t_i) &= \sum_{j \in S_i} x'(j, t_i)v'_1(j) + \sum_{j \notin S_i} [a'_1(t_i)v'_1(j) + a'_2(t_i)v'_2(j) + \dots + a'_n(t_i)v'_n(j)]v'_1(j) \\ a'_2(t_i) &= \sum_{j \in S_i} x'(j, t_i)v'_2(j) + \sum_{j \notin S_i} [a'_1(t_i)v'_1(j) + a'_2(t_i)v'_2(j) + \dots + a'_n(t_i)v'_n(j)]v'_2(j) \\ &\vdots \\ a'_n(t_i) &= \sum_{j \in S_i} x'(j, t_i)v'_n(j) + \sum_{j \notin S_i} [a'_1(t_i)v'_1(j) + a'_2(t_i)v'_2(j) + \dots + a'_n(t_i)v'_n(j)]v'_n(j) \end{aligned} \right\}.$$

In order to solve the PCs, we rewrite the above expression with the symbols of vector and matrix as follows

$$\mathbf{y}(t_i) = \mathbf{A}(t_i) \boldsymbol{\xi}(t_i), \quad (18)$$

$$\text{Where } \mathbf{y}(t_i) = \left(\sum_{j \in S_i} x'(j, t_i)v'_1(j) \quad \sum_{j \in S_i} x'(j, t_i)v'_2(j) \quad \sum_{j \in S_i} x'(j, t_i)v'_3(j) \quad \dots \quad \sum_{j \in S_i} x'(j, t_i)v'_n(j) \right)^T,$$

$$\mathbf{A}(t_i) = \begin{pmatrix} 1 - \sum_{j \notin S_i} v_1^2(j) & - \sum_{j \notin S_i} v'_1(j)v'_2(j) & - \sum_{j \notin S_i} v'_1(j)v'_3(j) & \dots & - \sum_{j \notin S_i} v'_1(j)v'_n(j) \\ - \sum_{j \notin S_i} v'_1(j)v'_2(j) & 1 - \sum_{j \notin S_i} v_2^2(j) & - \sum_{j \notin S_i} v'_2(j)v'_3(j) & \dots & - \sum_{j \notin S_i} v'_2(j)v'_n(j) \\ - \sum_{j \notin S_i} v'_1(j)v'_3(j) & - \sum_{j \notin S_i} v'_2(j)v'_3(j) & 1 - \sum_{j \notin S_i} v_3^2(j) & \dots & - \sum_{j \notin S_i} v'_3(j)v'_n(j) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ - \sum_{j \notin S_i} v'_1(j)v'_n(j) & - \sum_{j \notin S_i} v'_2(j)v'_n(j) & - \sum_{j \notin S_i} v'_3(j)v'_n(j) & \dots & 1 - \sum_{j \notin S_i} v_n^2(j) \end{pmatrix},$$

$$\boldsymbol{\xi}(t_i) = (a'_1(t_i) \ a'_2(t_i) \ a'_3(t_i) \ \dots \ a'_n(t_i))^T.$$

Since (18) is rank deficient, it is solved by introducing the following criteria

$$\min : \hat{\xi}^T(t_i) \mathbf{C}^{-1} \hat{\xi}(t_i),$$

and its solution is expressed as

$$\hat{\xi}(t_i) = \mathbf{C} \mathbf{A}^T(t_i) \left[\mathbf{A}^T(t_i) \mathbf{C} \mathbf{A}(t_i) \right]^{-1} \mathbf{y}(t_i), \quad (19)$$

where \mathbf{C} is the covariance matrix of $\xi(t_i)$, whose diagonal elements are the eigenvalues of the matrix \mathbf{B}' . For the details of deriving Eq. (19), one can refer to Shen et al. (2014). After the n PCs solved, the CMEs can be calculated using (13) and (14).

4 Real and synthetic time series analysis

The platform of GNSS data products and services created by China Earthquake Administration provides over 10 years position time series of 27 GNSS permanent stations of Crust Motion Observation Network of China (CMONOC) (Fig. 1) in China. All the position time series are computed with the GAMIT/GLOBK10.4 (see processing details in ftp://ftp.cgps.ac.cn/doc/processing_manual.pdf and download data in <http://www.cgps.ac.cn/>). The time series between 1999 and 2013 are used in this paper. The formal errors of the series in different stations and different epochs are not uniform, as shown in Fig. 2 for BJFS, DLHA, XNIN

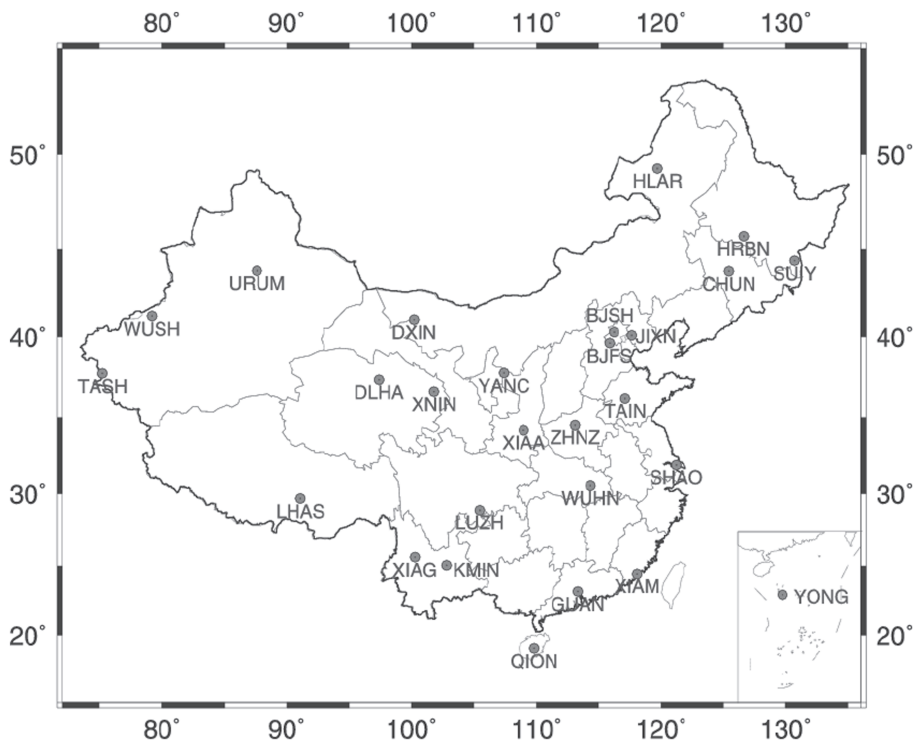


Fig. 1 Geographic locations of the 27 Stations in CMONOC

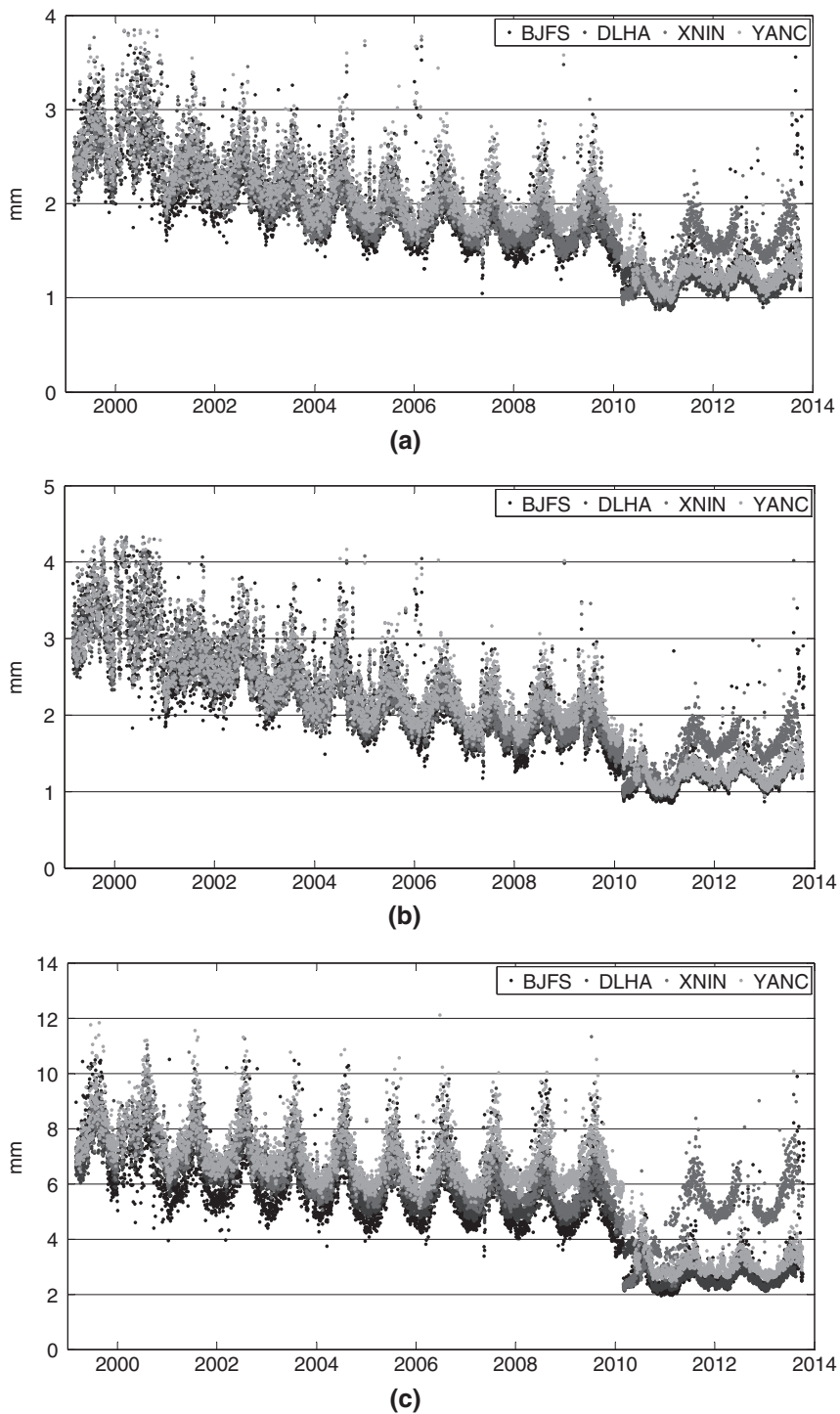


Fig. 2 Formal errors of north component (a) east component (b) up component (c)

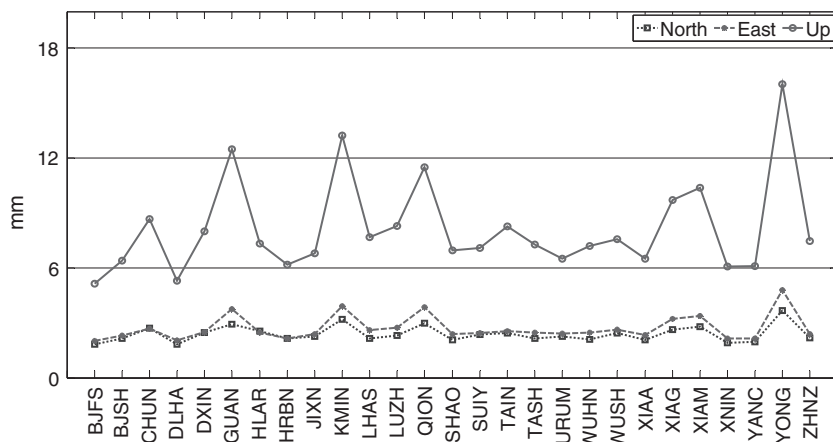


Fig. 3 Mean formal errors of 27 stations

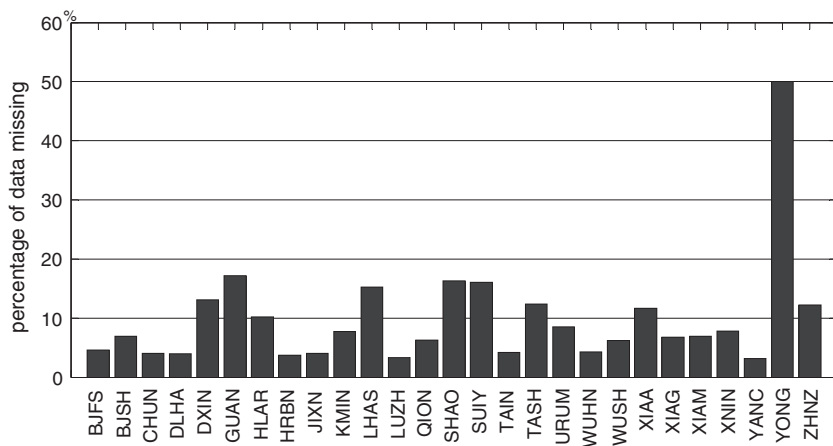


Fig. 4 Percentage of data missing for 27 stations

and YANC stations. The mean formal errors for all stations are presented in Fig. 3, which is clearly station-dependent. We can also see from Fig. 3 that the station YONG has the largest formal errors for all the three coordinate components. The percentage of data missing for each station is presented in Fig. 4, which shows that the missing data occurs at all stations, especially at the station YONG with about 50 % data missing.

We use our WSF approach proposed in Section 3.2 to process the time series of CMONOC. The eigenvalues of the covariance matrix formed with (15) are presented in Fig. 5 as a function of PC order, in which the first PC represents 39.2, 38.7, and 35.8 % of the total variance for north, east and up components, much larger than other PCs. According to the criterion of defining common mode proposed in Dong et al. (2006), we take the first PC as the common mode in this network. The spatial responses of these 27 stations are shown in bottom of Fig. 6, where the stations with larger response locate at the centre of CMONOC, while the responses for the stations in the west part are quite small. The first PCs solved with (19) are presented on the top of Fig. 6. We can see from Fig. 6 that the first PC series changes significantly.

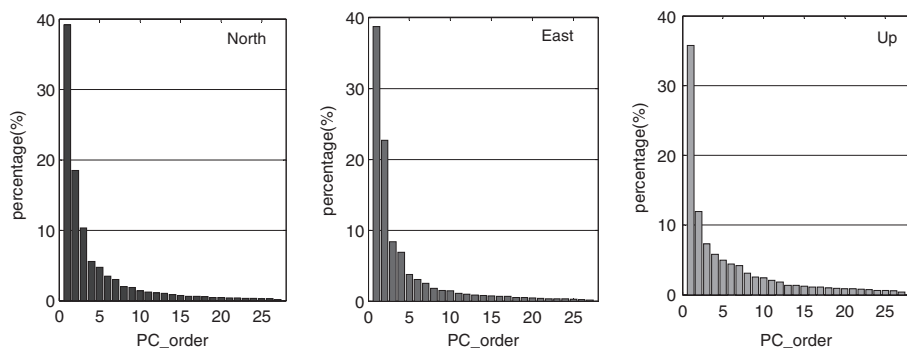


Fig. 5 Histogram of the eigenvalues as a function of PC orders in *three* components

Moreover, the first PC series of up components in Fig. 6c show significant annual patterns, which are probably the common annual signals, such as ionospheric delay.

In WSF, the weight of the GNSS time series for j th station at i th epoch is defined as

$$P(j, i) = \frac{\sigma_0^2}{\sigma_{(j,i)}^2}, \quad (i = 1, 2, \dots, m_j), \quad (20)$$

where, $\sigma_{(j,i)}^2$ is the variance of the GNSS time series for j th station at i th epoch, which is determined by GNSS daily solution; σ_0^2 is the variance of unit weight, and m_j is the number of epochs of j th station. However, in modified PCA the difference of station- and epoch-dependent variance $\sigma_{(j,i)}^2$ is ignored and unique weight is adopted for all epochs, thereby the sum of the weight for j th station is equal to m_j . With the condition that the sum of weight of WSF is equal to number of epoch m_j , one can derive σ_0^2 as

$$\sigma_0^2 = \frac{m_j}{\sum_{i=1}^{m_j} \frac{1}{\sigma_{(j,i)}^2}}. \quad (21)$$

By substituting (21) into (20) yields the weight $P(j, i)$ as

$$P(j, i) = \frac{m_j}{\sigma_{(j,i)}^2 \sum_{k=1}^{m_j} \frac{1}{\sigma_{(j,k)}^2}}, \quad (j = 1, 2, \dots, n; i = 1, 2, \dots, m_j). \quad (22)$$

If the CME computed by (14) is expressed with $\hat{s}_w(j, i)$ where the subscript w denotes WSF, we can estimate the formal standard deviation of unit weight $\hat{\sigma}_w$ for the j th station is as follows

$$\hat{\sigma}_w(j) = \sqrt{\frac{1}{m_j} \sum_{i=1}^{m_j} [(x(j, i) - \hat{s}_w(j, i))^2 P(j, i)]}, \quad (j = 1, 2, \dots, n). \quad (23)$$

Also, the formal standard deviation of unit weight $\hat{\sigma}_p$ for j th station in modified PCA is estimated with

$$\hat{\sigma}_p(j) = \sqrt{\frac{1}{m_j} \sum_{i=1}^{m_j} (x(j, i) - \hat{s}_p(j, i))^2}, \quad (j = 1, 2, \dots, n), \quad (24)$$

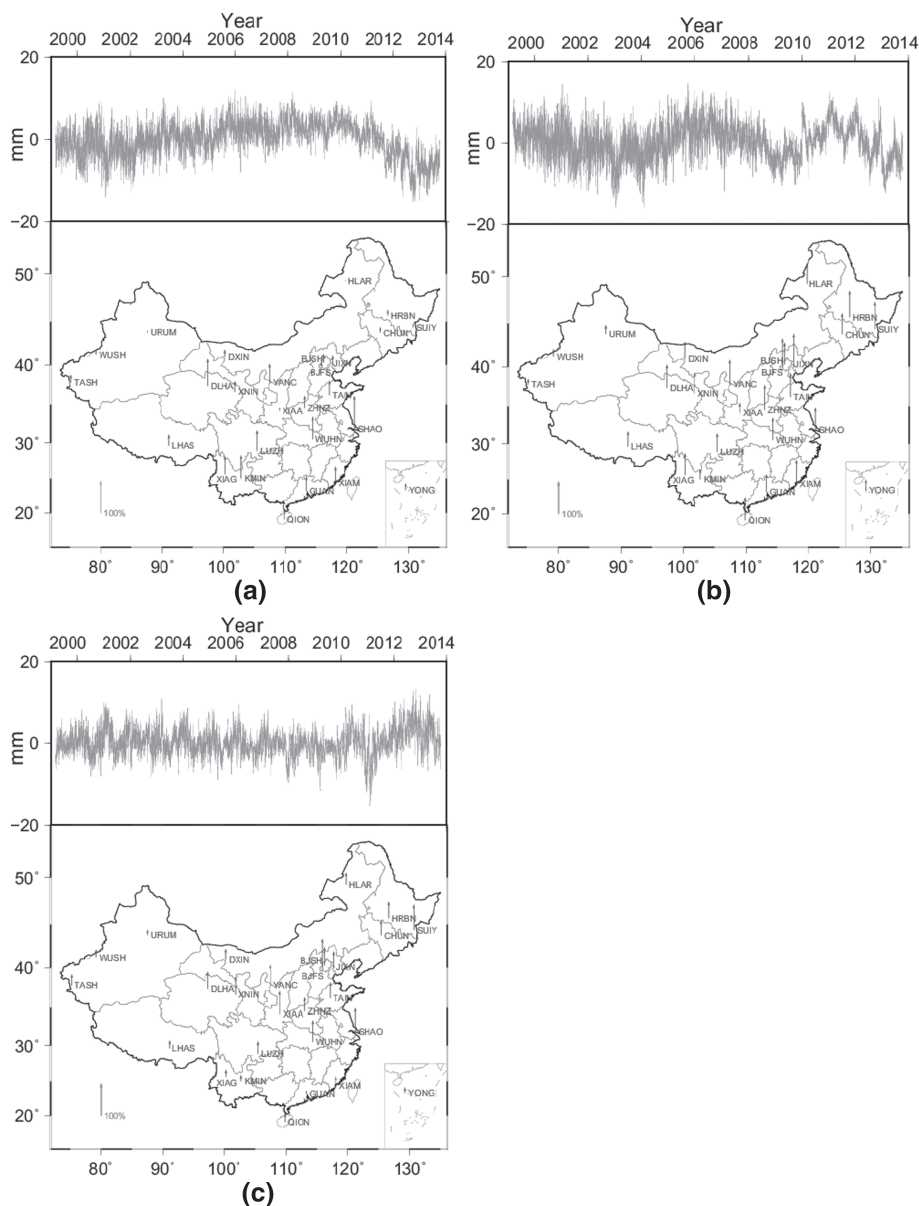


Fig. 6 First PC and Spatial response (a north b east c up)

where, $\hat{s}_p(j, i)$ is computed by (10) and the subscript p means modified PCA. Since modified PCA approach neglects the difference between the variances $\sigma_{(j,i)}^2$, the WSF can estimate the CMEs more precisely than modified PCA. Therefore, the estimate $\hat{\sigma}_w(j)$ will be smaller than $\hat{\sigma}_p(j)$ because the sum of weights is same for WSF and PCA. The results derived by (23) and (24) are shown in Fig. 7 for 27 stations of CMONOC, in which WSF approach outperforms the modified PCA for 21, 19 and 17 out of 27 stations for north, east and up components,

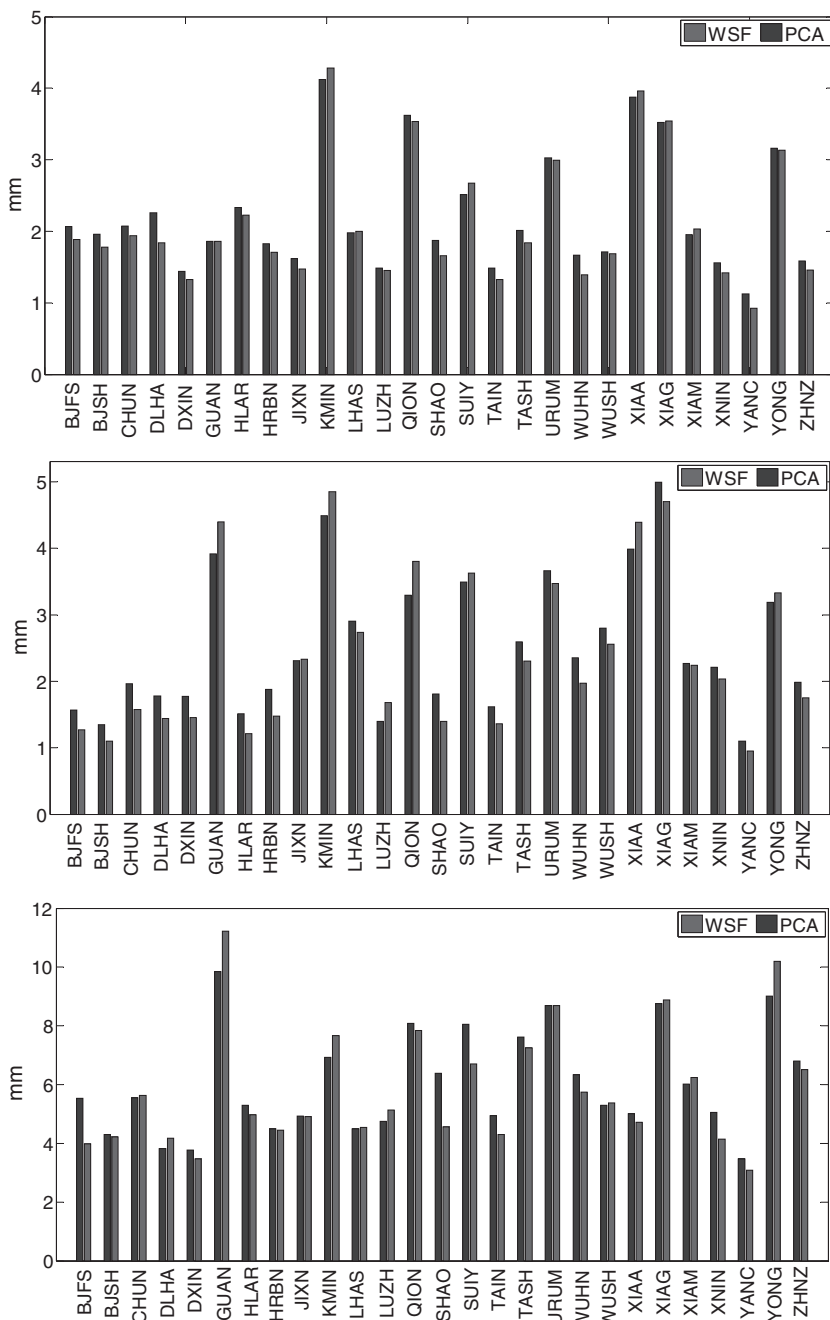


Fig. 7 $\hat{\sigma}_p$ and $\hat{\sigma}_w$ for the 27 stations (north, east, up)

respectively. As for the remaining stations, $\hat{\sigma}_w$ is larger than $\hat{\sigma}_p$. The reason for this should be further investigated. Generally, the mean values of formal standard deviation of unit weight derived from WSF and modified PCA are 2.12, 2.42, 5.88 and 2.21, 2.52, 6.05 respectively

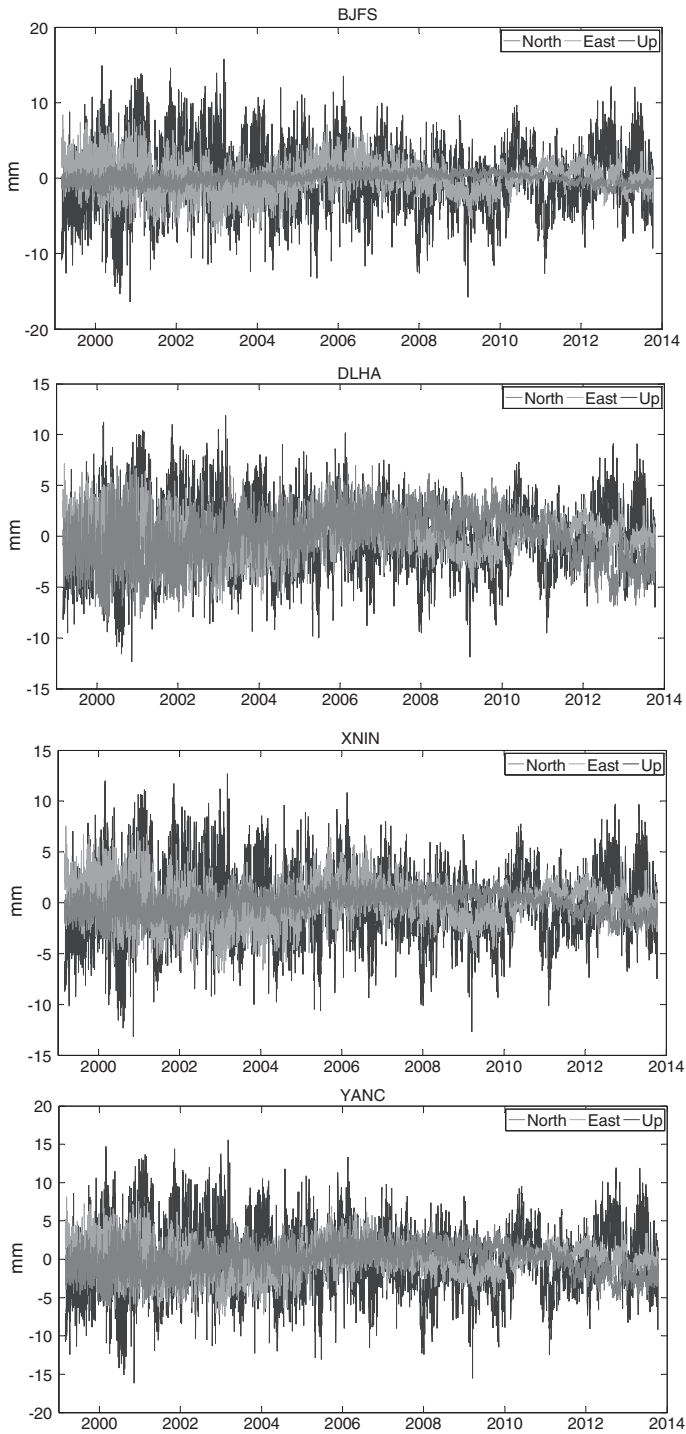


Fig. 8 CMEs of four stations

for north, east and up components. The relative improvements of the mean values of WSF with respect to modified PCA are 4.1, 4.0 and 2.8% for the north, east and up components, respectively.

To further statistically demonstrate the performance of WSF, two simulations are carried out. In the first simulation test, the CMEs of BJFS, DLHA, XNIN, YANC stations with less missing data, which are derived in the real example, are selected for generating simulation data as follows

$$y(j, i) = \bar{s}(j, i) + \varepsilon(j, i), \quad (25)$$

where, $\bar{s}(j, i)$ is the CME of j th station at i th epoch, $\varepsilon(j, i)$ is the random error generated by normal distribution with zero mean and variance $\sigma_{(j,i)}^2$. The real $\sigma_{(j,i)}^2$ of the four stations are already presented in Fig. 2 and the extracted CMEs of the four stations are shown in Fig. 8.

By using the WSF approach we can extract the CMEs ($\hat{s}_w(j, i)$) from the simulated data series $y(j, i)$ and compute the variance of unit weight of recovered CMEs for j th station as

$$\delta_w^2(j) = \frac{1}{m_j} \sum_{i=1}^{m_j} (\hat{s}_w(j, i) - \bar{s}(j, i))^2 P_s(j, i), \quad (j = 1, 2, 3, 4), \quad (26)$$

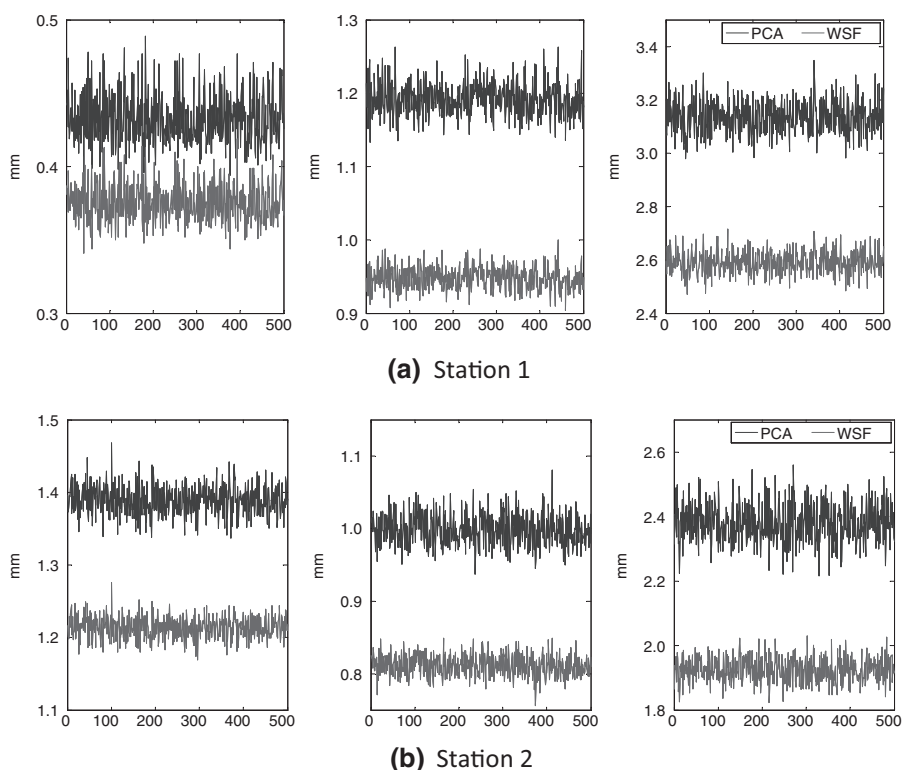


Fig. 9 δ_w and δ_p for 500 experiments (north, east, up)

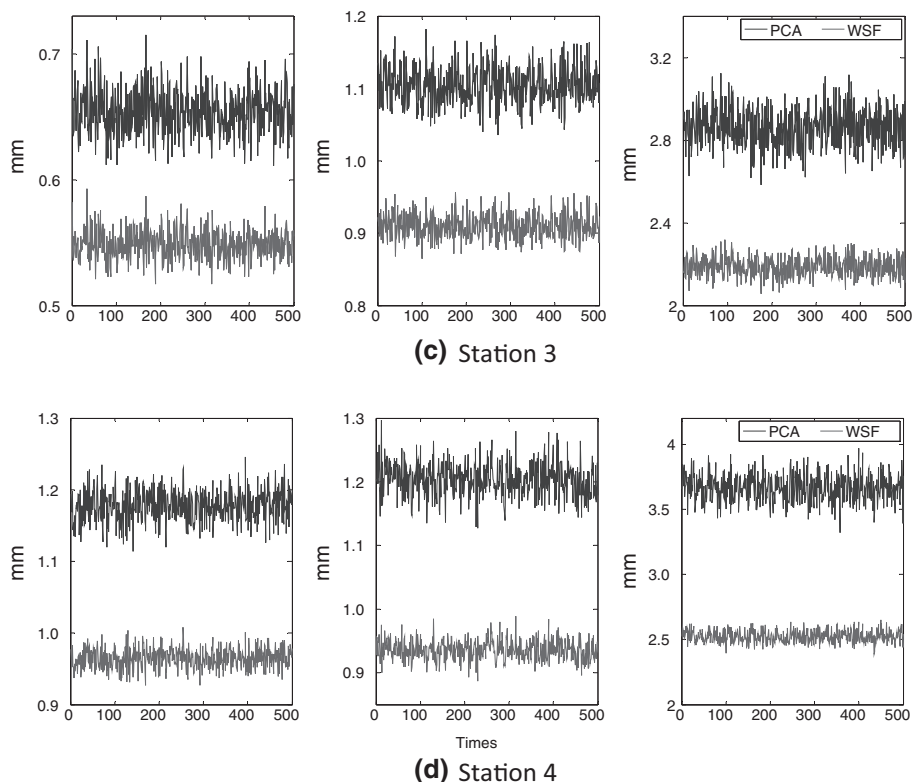


Fig. 9 continued

where, $P_s(j, i)$ is the weight which is also calculated with (20). With unique weights in PCA method, the variance of computed CMEs of j th station can be calculated with

$$\delta_p^2(j) = \frac{1}{m_j} \sum_{i=1}^{m_j} (\hat{s}_p(j, i) - \bar{s}(j, i))^2, \quad (j = 1, 2, 3, 4). \quad (27)$$

We can compute the formal standard deviation of unit weight $\hat{\sigma}_w(j)$ and $\hat{\sigma}_p(j)$ of the simulation data with

$$\hat{\sigma}_w(j) = \sqrt{\frac{1}{m_j} \sum_{i=1}^{m_j} [(y(j, i) - \hat{s}_w(j, i))^2 P_s(j, i)]}, \quad (j = 1, 2, 3, 4), \quad (28)$$

$$\hat{\sigma}_p(j) = \sqrt{\frac{1}{m_j} \sum_{i=1}^{m_j} (y(j, i) - \hat{s}_p(j, i))^2}, \quad (j = 1, 2, 3, 4), \quad (29)$$

where, \hat{s}_w and \hat{s}_p have the same meanings as in (23) and (24). We repeat 500 experiments, the results of δ_w and δ_p are presented in Fig. 9, and the results of $\hat{\sigma}_w$ and $\hat{\sigma}_p$ in Fig. 10. From these two figures, we can find the significant differences between modified PCA and WSF approach. δ_w and $\hat{\sigma}_w$ are smaller than δ_p and $\hat{\sigma}_p$ for all 4 stations therefore CMEs extracted by WSF is more close to their true values and the variance of the residual series filtered by

WSF is lower than that filtered by modified PCA. The mean values of $\bar{\delta}_w$, $\bar{\sigma}_w$, $\bar{\delta}_p$, $\bar{\sigma}_p$ of 500 experiments are shown in Table 1. The relative improvements $\bar{\delta}_{IP}$, $\bar{\sigma}_{IP}$ of $\bar{\delta}_w$, $\bar{\sigma}_w$ with respect to $\bar{\delta}_p$, $\bar{\sigma}_p$ are calculated as

$$\bar{\delta}_{IP} = \frac{\bar{\delta}_p - \bar{\delta}_w}{\bar{\delta}_p} \times 100 \% \quad \bar{\sigma}_{IP} = \frac{\bar{\sigma}_p - \bar{\sigma}_w}{\bar{\sigma}_p} \times 100 \% \quad (30)$$

It is obvious in Table. 1 that all $\bar{\delta}_{IP}$ and $\bar{\sigma}_{IP}$ are positive and over 11% and 5%, which means that the performance of WSF is better than modified PCA.

Since local effects exist in the real GNSS time series, they are considered in the second simulations. The PCs from 2 to 5 derived from the real time series are treated as local effects in this simulation. These orders of PC are simulated to the stations from 1 to 4, respectively. Since strong local effects will sway the extracted CMEs, it is necessary to scale the PCs used for local effects in order to keep the CME as dominated component. Then the time series in this simulation are generated with

$$y(j, i) = \bar{s}(j, i) + q(j, i) \times scale + \varepsilon(j, i), \quad (31)$$

where, the *scale* is constant, and the others are same as those in (25). The scale should be less than 0.5; otherwise the local effects are over strong (Dong et al. 2006). We choose the scales of 0.1, 0.2 and 0.3 to generate the time series and do the same comparisons as we did in the first simulation, respectively. When the scales are 0.1 or 0.2, all $\bar{\delta}_{IP}$ and $\bar{\sigma}_{IP}$ are positive,

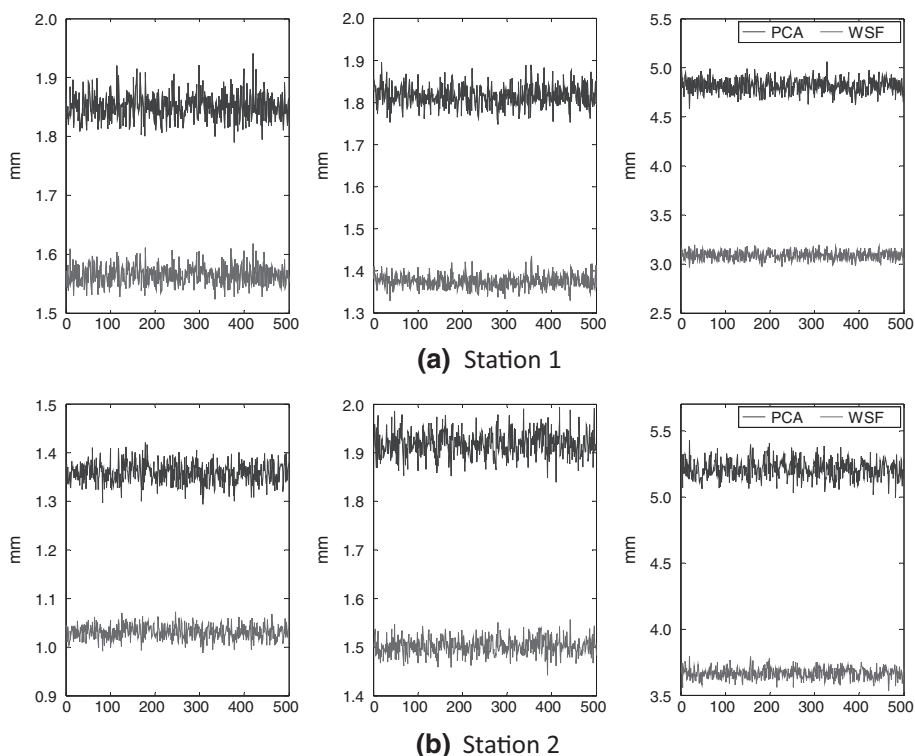


Fig. 10 $\hat{\sigma}_w$ and $\hat{\sigma}_p$ for 500 experiments (*north, east, up*)

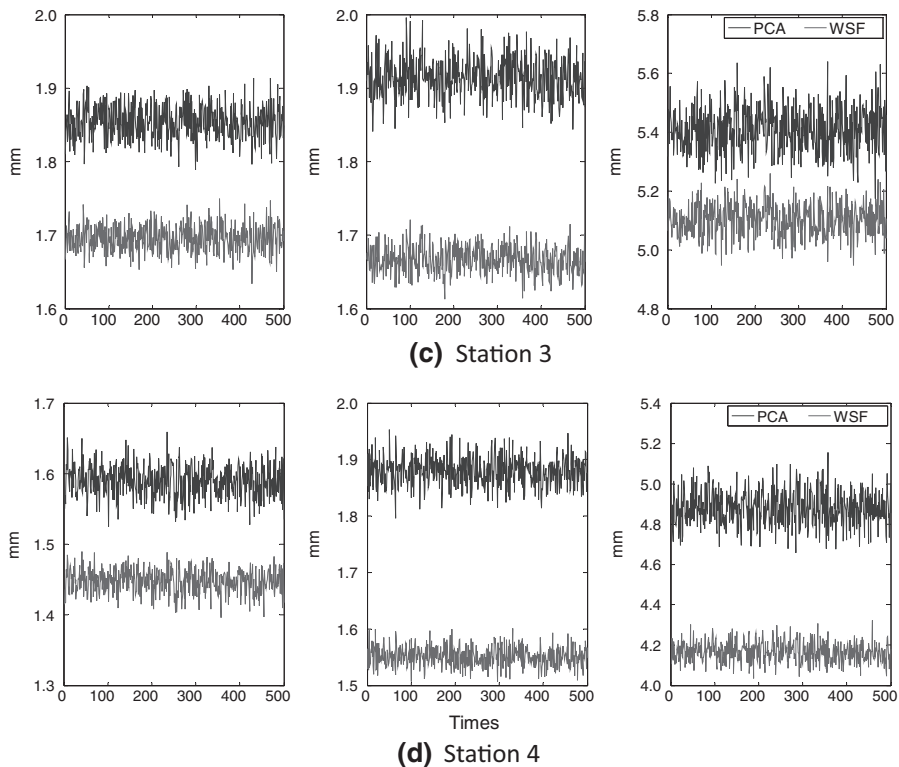


Fig. 10 continued

which means our WSF performs better than PCA in all stations. When it is 0.3, three $\bar{\delta}_{IP}$ and one $\bar{\sigma}_{IP}$ values are negative, as shown in Table 2 for the results of 500 experiments. Thus, as the local effects get larger, one cannot guarantee that the WSF can get better results than the modified PCA. Therefore, the significant local effects in the real time series analysis is probably a good explanation why the performance of WSF is worse than modified PCA at 6, 8 and 10 of 27 stations for the north, east and up components, respectively. Anyway, our WSF approach in general outperforms modified PCA on average in both real and simulated series if the local effects are not too large. Since the stations with large local effects can be determined by using Karhunen-Loeve expansion approach (Dong et al. 2006), the impacts of large local effects can be decreased by enlarging the formal errors of the stations.

5 Conclusions

This paper develops a PCA-based weighted spatiotemporal filtering for extracting the CMEs from the GNSS position time series with the consideration of non-uniform formal error and missing data. For the real GNSS time series of 27 stations of CMONOC network, our WSF can provide better results than modified PCA at 21, 19 and 17 stations for north, east and up components, respectively. The mean values of the formal standard deviation of unit weight derived from WSF and modified PCA are 2.12, 2.42, 5.88 and 2.21, 2.52, 6.05 for north,

Table 1 Mean values of $\bar{\delta}_w$, $\bar{\sigma}_w$, $\bar{\delta}_p$, $\bar{\sigma}_p$ and improvement ($\bar{\delta}_{IP}$, $\bar{\sigma}_{IP}$) of series without local effects derived from modified PCA and WSF

	Station1			Station2			Station3			Station4		
	North	East	Up	North	East	Up	North	East	Up	North	East	Up
$\bar{\delta}_p$	0.43	1.19	3.14	1.39	0.99	2.38	0.65	1.10	2.86	1.18	1.20	3.66
$\bar{\delta}_w$	0.38	0.95	2.59	1.21	0.81	1.93	0.55	0.91	2.19	0.96	0.94	2.52
$\bar{\delta}_{IP}$	11.63	20.17	17.52	12.95	18.18	18.91	15.38	17.27	23.43	18.64	21.67	31.15
$\bar{\sigma}_p$	1.85	1.82	4.82	1.36	1.92	5.21	1.85	1.91	5.41	1.59	1.88	4.88
$\bar{\sigma}_w$	1.57	1.38	3.09	1.03	1.50	3.67	1.70	1.67	5.10	1.45	1.55	4.17
$\bar{\sigma}_{IP}$	15.13	24.18	35.89	24.26	21.88	29.56	8.11	12.56	5.73	8.81	17.55	14.55

Table 2 Mean values of $\bar{\delta}_w$, $\bar{\sigma}_w$, $\bar{\delta}_p$, $\bar{\sigma}_p$ and improvement ($\bar{\delta}_{IP}$, $\bar{\sigma}_{IP}$) of series with local effects (scale = 0.3) derived from modified PCA and WSF

	Station1			Station2			Station3			Station4		
	North	East	Up	North	East	Up	North	East	Up	North	East	Up
$\bar{\delta}_p$	0.65	1.96	3.88	1.72	1.22	2.60	0.76	1.26	3.05	1.34	1.37	3.97
$\bar{\delta}_w$	0.44	2.19	4.07	1.73	1.14	2.32	0.64	1.14	2.40	1.16	1.20	2.83
$\bar{\delta}_{IP}$	32.31	−11.73	−4.90	−0.58	6.56	10.77	15.79	9.52	21.31	13.43	12.41	28.72
$\bar{\sigma}_p$	2.62	2.20	5.40	1.60	2.50	5.93	2.23	2.44	6.14	1.90	2.32	5.52
$\bar{\sigma}_w$	2.32	1.52	3.38	1.16	2.26	4.41	2.09	2.41	6.15	1.82	2.19	5.35
$\bar{\sigma}_{IP}$	11.45	30.91	37.41	27.50	9.60	25.63	6.27	1.23	−0.16	4.21	5.60	3.08

east and up components, respectively; the corresponding relative improvements are 4.1, 4.0 and 2.8%. In the two simulations, with small local effects, our WSF performs better than the modified PCA for all coordinate components of all stations. When the local effects get larger, the WSF can still provide better results than the modified PCA at most of the position time series. Generally, by considering the formal errors, the WSF approach can achieve better performance than the modified PCA in regional filtering.

Acknowledgments This work was mainly sponsored by Natural Science Foundation of China (Projects: 41274035, 41474017; 41374031). It is also supported by the Foundation of State Key Laboratory of Geodesy and Earth's Dynamics (Projects: SKLGED2013-6-2-Z and SKLGED2014-2-4-E). The reviewers and editors are acknowledged for their constructive comments, and Dr. Ling Yang from the University of New South Wales is appreciated for correcting the English grammar.

References

- Amiri-Simkooei AR (2009) Noise in multivariate GPS position time-series. *J Geod* 83:175–187
- Dong D, Fang P, Bock Y, Webb F, Prawirodirdjo L, Kedar S, Jamason P (2006) Spatiotemporal filtering using principal component analysis and Karhunen-Loeve expansion approaches for regional GPS network analysis. *J Geophys Res* 111:3405–3421
- Ji KH, Herring T (2013) A method for detecting transient signals in GPS position time-series: smoothing and principal component analysis. *Geophys J Int* 193:171–186

- Ji L, Wang Q, Wang S (2014) Present-day 3D deformation field of Northeast China, observed by GPS and leveling. *Geod Geodyn* 5(3):34–40
- Langbein J, Johnson H (1997) Correlated errors in geodetic time series: implications for time-dependent deformation. *J Geophys Res* 102(B1):591–603
- Langbein J (2004) Noise in two-color electronic distance meter measurements revisited. *J Geophys Res* 109:B04406. doi:10.1029/2003JB002819
- Lee S-W, Kouba J, Schutz B, Kim DH, Lee YJ (2013) Monitoring precipitable water vapor in real-time using global navigation satellite system. *J Geod* 87:923–934
- Mao A, Harrison CGA, Dixon TH (1999) Noise in GPS coordinate time series. *J Geophys Res* 104(B2):2797–2816
- Nikolaidis R (2002) Observation of geodetic and seismic deformation with the global positioning system, Ph.D Thesis, University of California, San Diego
- Nilsson T, Elgered J (2008) Long-term trends in the atmospheric water vapor content estimated from ground-based GPS data. *J Geophys Res* 113:D19101. doi:10.1029/2008JD010110
- Shen Y, Li W (2014) Spatiotemporal signal and noise analysis of GPS position time series of the permanent stations in China. In: Exertier P (ed) *Earth on the edge: science for a sustainable planet*. Springer, Heidelberg, pp 231–237
- Shen Y, Li W, Xu G, Li B (2014) Spatiotemporal filtering of regional GNSS network's position time series with missing data using principle component analysis. *J Geod* 88:1–12
- Tian Y, Shen ZK (2011) Correlation weighted stacking filtering of common-mode component in GPS observation network. *Acta Seismol Sin* 33(2):198–208
- Tian Y, Gang L, Zhang J (2013) Seasonal GPS positioning errors due to water content variations in atmosphere. AGU fall meeting, San Francisco 9–13 December
- Wdowinski S, Bock Y, Zhang J, Fang P, Genrich J (1997) Southern California permanent GPS geodetic array: spatial filtering of daily positions for estimating coseismic and postseismic displacements induced by the 1992 Landers earthquake. *J Geophys Res* 102:18057–18070
- 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* 109:B03412. doi:10.1029/2003JB002741
- Yuan L, Ding X, Chen W, Guo Z, Chen S, Hong B, Zhou J (2008) Characteristics of daily position time series from the Hong Kong GPS fiducial network. *Chin J Geophys* 51(5):1372–1384
- 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. *J Geophys Res* 102(B8):18035–18055