

Extended Principal Component Analysis for Spatiotemporal Filtering of Incomplete Heterogeneous GNSS Position Time Series

Kunpu Ji, Yunzhong Shen[✉], Quijie Chen, and Tengfei Feng

Abstract—When ordinary principal component analysis (PCA) is employed to analyze the position time series of a regional global navigation satellite system (GNSS) station network, the GNSS time series are assumed to be homogeneous, and the missing data in the time series must be restored beforehand. To directly process incomplete and heterogeneous GNSS position time series, we develop the extended PCA (EPCA) and weighted EPCA approaches to solving for the missing values based on the best low-rank approximation in the spatiotemporal domain. The proposed approaches are used to process the real GNSS position time series of 24 stations in North China spanning 2011 to 2019 and successfully extract the common mode errors (CMEs). The proposed approaches are compared with modified PCA (MPCA) and weighted MPCA, in which an additional optimization criterion needs to be introduced in the frequency domain. The results show that EPCA can extract more CMEs than MPCA for both the unweighted and weighted cases. Consequently, EPCA outperforms MPCA in reducing noise and improving the accuracy of site velocity estimates. Repeated simulation experiments show that the CMEs extracted by EPCA are closer to the simulated true values than those extracted by MPCA. When the formal errors of the time series are considered, both weighted EPCA and weighted MPCA outperform their unweighted counterparts, and the former outperforms the latter. In addition, EPCA is computationally more efficient than MPCA since fewer unknowns need to be estimated.

Index Terms—Common mode errors (CMEs), global navigation satellite system (GNSS) position time series, principal component analysis (PCA), spatiotemporal filtering.

I. INTRODUCTION

OVER the past three decades, the global navigation satellite system (GNSS) technique has been rapidly developed, and the GNSS position time series of global and regional station networks have provided abundant and high-accuracy information for various geophysical and remote sensing applications, such as crustal deformation monitoring [1], [2], sea-level change estimation [3], [4], [5], terrestrial water storage analysis [6], [7], [8], [9], snow depth inversion [10], [11], [12], and retrieval of atmospheric water vapor [13], [14], [15]. Due

Manuscript received 13 November 2022; revised 27 February 2023 and 12 April 2023; accepted 6 May 2023. Date of publication 18 May 2023; date of current version 6 June 2023. This work was supported by the Natural Science Foundation of China under Grant 41974002, Grant 42274005, and Grant 42192532. (*Corresponding author: Yunzhong Shen.*)

The authors are with the College of Surveying and Geo-Informatics, Tongji University, Shanghai 200092, China (e-mail: kunpuji@tongji.edu.cn; yzshen@tongji.edu.cn; chenqijie2009@163.com; 1911212@tongji.edu.cn).

Digital Object Identifier 10.1109/TGRS.2023.3277460

to various impact factors, GNSS daily solutions inevitably contain site-specific temporal noise and spatially correlated noise, which seriously limits the interpretation of subtle solid Earth deformation [16], [17]. The temporal noise in individual position time series is usually described as a combination of white noise (WN) and power-law noise (PN), and can be estimated with maximum likelihood estimation (MLE) [18], [19], [20], [21], [22], [23], [24], variance component estimation (VCE) [25], [26], [27], and generalized method of wavelet moments with eXogenous inputs (GMWMX) [28]. The spatially correlated noise among stations is related to the scale of networks and has been taken as the main error source in daily GNSS solutions [29], [30], [31], [32]. Such spatially correlated noise is usually referred to as common mode errors (CMEs) that were initially proposed by Wdowinski et al. [33]. The physical sources of CMEs are rather complicated, and the previous studies usually attribute CMEs to a sum of environmental and technique-dependent systematic errors, including imprecise modeling of atmospheric loading, reference frame error, and satellite-related errors (orbits, clocks, or antenna phase center variations) [17], [20], [29], [31], [33], [34], [35], [36]. In fact, the temporal behavior of CMEs is not purely random and exhibits some unmodeled seasonal signals that are related to environmental loadings, such as hydrologic, surface atmosphere, and soil moisture loadings [37], [38], [39]. For this reason, CMEs are also called common mode components (CMCs) in some studies [16], [37], [40], [41]. Due to the existence of CMEs, the amplitudes of temporal noise of GNSS position time series will be overquantified, and as a result, the site velocity and corresponding uncertainty will be misestimated [42], [43], [44]. Therefore, the CMEs should be filtered out from the position time series of a regional network, and this procedure is called spatiotemporal filtering [31], [45], [46].

Many efforts have been devoted to spatiotemporal filtering of the time series of regional GNSS station networks. The stacking method proposed by Wdowinski et al. [33] simply takes the means of each epoch's coordinate components of all stations as the regional CMEs of this epoch and has been successfully applied to process various regional GNSS networks, such as those of Southern California [29], Nevada-California [47], and the Dead Sea Fault [48]. A weighted stacking approach was then developed by Nikolaidis [29], where the formal errors of GNSS daily solutions were

used to construct the weights. Since the stacking approach assumes that the CMEs are spatially homogeneous, it is more suitable for small station networks (spanning less than 600 km) [30], [33]. To overcome this restriction, some modified stacking approaches have been developed by adopting the distances between stations [30] or the correlations between stations [16], [49] as weights. Recently, a robust statistical approach called “CMC Imaging” was proposed by Kreemer and Blewitt [41], where the correlation coefficient between two residual time series was computed via a robust approach.

Although the stacking approach is easy to implement, the assumption of a uniform spatial response (SR) is not realistic. To derive a more reasonable SR for regional networks, some data-driven spatiotemporal filtering approaches have been developed, such as principal component analysis (PCA) [31], [50], independent component analysis (ICA) [36], [42], [46], empirical mode decomposition (EMD) [22], [51], and multichannel singular spectrum analysis (MSSA) [44], [52]. These approaches allow the data themselves to reveal the spatial distribution of the CMEs and, thus, can provide a more rigorous mathematical framework. Dong et al. [31] first introduced PCA to extract the CMEs of the Southern California Integrated GPS Network, in which PCA decomposes the time series into a set of temporally varying modes, and each mode comprises common temporal components and corresponding SRs. Due to its outstanding performance, PCA has been widely applied to process the GNSS time series of regional networks [31], [40], [45], [50].

The PCA approach requires the involved time series to be complete; however, this may not hold for most GNSS position time series due to various reasons, such as equipment failure and outlier elimination. Dong et al. [31] adopted a mixed interpolation scheme to fill the data gaps, but this may not be a good choice for CMEs filtering for the cases of a huge missing percentage and large consecutive missing gaps (even with several years). In addition, after the trend is removed, the GNSS time series presents irregular variation [17]; hence, filling the large data gaps is relatively difficult. Therefore, Shen et al. [45] developed a modified PCA (MPCA) to directly process the incomplete position time series by solving an optimization problem in the frequency domain based on the principle that a time series can be reproduced from its principal components. MPCA is superior to ordinary PCA when the position time series of a regional GNSS station network are not complete [45]. Meanwhile, it will degenerate to ordinary PCA when the GNSS position time series are complete. A similar idea has been extended to singular spectrum analysis (SSA) by Shen et al. [53] for processing incomplete series, which is mathematically proven to be the generalized version of the approach proposed by Schoellhamer [54]. The multidimensional case was later developed by Wang et al. [55] for postprocessing the incomplete gravity recovery and climate experiment (GRACE) monthly gravity field solutions. In addition, the incomplete GNSS position time series are heterogeneous due to different computational strategies and observational conditions. Therefore, Li et al. [56] proposed weighted MPCA, where the weights are constructed with formal errors. However, the weighting

scheme is not comprehensive since the formal errors are only related to the observation error but not necessarily related to the CMEs. For this reason, the modified version of weighted spatiotemporal filtering based on MPCA was developed by Li and Shen [57].

However, an additional optimization criterion in Shen et al. [45] needs to be introduced in the frequency domain to solve all the PCs epoch by epoch, which is computationally inefficient. To achieve better approximation in the spatiotemporal domain and improve computational efficiency, we develop an extended PCA (EPCA) approach for processing incomplete heterogeneous regional GNSS position time series, by which the missing data are computed directly based on the best low-rank approximation in the spatiotemporal domain. The rest of this article is organized as follows: Section II reviews the theory of ordinary PCA, Section III presents the proposed EPCA for spatiotemporal filtering of regional GNSS networks’ data, and Section IV presents a comparison of EPCA and MPCA using real and synthetic time series. Some conclusions and remarks are provided in Section V.

II. ORDINARY PRINCIPAL COMPONENT ANALYSIS

PCA is a widely used technique for dimensionality reduction and extracting the hidden factors that underlie a set of random time series [58], [59], [60], [61]. Given a set of centered time series $\mathbf{x}_i(1 \leq i \leq n)$ with sample length of $m(m > n)$, the data matrix \mathbf{X} is generated by arranging \mathbf{x}_i in column order, i.e., $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$. First, we compute the $n \times n$ covariance matrix \mathbf{B} of \mathbf{X} with [60]

$$\mathbf{B}(i, j) = \frac{1}{m-1} \sum_{k=1}^m \mathbf{X}(k, i)\mathbf{X}(k, j)(1 \leq i, j \leq n) \quad (1)$$

where $\mathbf{B}(i, j)$ and $\mathbf{X}(i, j)$ are the ij th elements of \mathbf{B} and \mathbf{X} . Then, we perform eigenvalue decomposition on \mathbf{B} with

$$\mathbf{B} = \mathbf{V}\Lambda\mathbf{V}^T \quad (2)$$

where \mathbf{V} is an $n \times n$ orthogonal matrix satisfying $\mathbf{V}\mathbf{V}^T = \mathbf{V}^T\mathbf{V} = \mathbf{I}_n$, \mathbf{I}_n is an n dimensional identity matrix, and Λ is a diagonal matrix whose diagonal elements are the eigenvalues $\lambda_i(1 \leq i \leq n)$ sorted in descending order. The principal component matrix \mathbf{A} can be computed with

$$\mathbf{A} = \mathbf{X}\mathbf{V}. \quad (3)$$

The k th column vector of \mathbf{A} represents the k th PC, reading as

$$\mathbf{a}_k = \mathbf{X}\mathbf{v}_k \quad (4)$$

where \mathbf{v}_k is the k th column vector of \mathbf{V} . It is clear from (4) that \mathbf{a}_k is a linear combination of time series $\mathbf{x}_i(1 \leq i \leq n)$ with the elements of \mathbf{v}_k . Mathematically, the eigenvector \mathbf{v}_k is the solution to the optimization problem of finding $\mathbf{v}_k^* \in \mathbb{R}^n$ that maximizes the variance of $\mathbf{X}\mathbf{v}_k^*$ [60], where \mathbb{R}^n denotes n dimensional real space. Furthermore, it is proven that any two different PCs \mathbf{a}_i and $\mathbf{a}_j(i \neq j)$ are uncorrelated, and their variances are the eigenvalues λ_i [60], [61]. In other words, the PCA approach can transform a set of correlated data into

another collection of variables (PCs), which are uncorrelated and ordered by variance so that the first few retain most of the variation present in all of the original variables. The original data \mathbf{X} can be decomposed into several components comprising the PCs and eigenvectors

$$\mathbf{X} = \mathbf{A}\mathbf{V}^T = \sum_{k=1}^n \mathbf{a}_k \mathbf{v}_k^T. \quad (5)$$

In standard PCA terminology, the elements of \mathbf{v}_k are called “PC loadings” and the elements of \mathbf{a}_k are called “PC scores” [60], [62], [63]. In practical applications, we usually reconstruct signals with the first few components

$$\mathbf{S} = \sum_{k=1}^r \mathbf{a}_k \mathbf{v}_k^T \quad (6)$$

where $r < n$ is the number of dominant PCs and \mathbf{S} is the filtered signals. From the point view of optimization, the essence of PCA is to find a matrix $\mathbf{S}^* \in \mathbb{R}^{m \times n}$ with a rank of r that best approximates \mathbf{X} , by minimizing the least-squares norm [58], [60]

$$\min_{\substack{\mathbf{S}^* \\ \text{rank}(\mathbf{S}^*)=r}} : \sum_{i=1}^m \sum_{j=1}^n (\mathbf{X}(i, j) - \mathbf{S}^*(i, j))^2. \quad (7)$$

The solutions can be solved by performing singular value decomposition (SVD) on \mathbf{X} through a hard thresholding constraint.

III. EXTENDED PCA FOR SPATIOTEMPORAL FILTERING

A. Data Preprocessing

For a regional daily station time series network with n stations and m epochs ($m \geq n$), after repairing the offsets and eliminating the outliers, the position time series for each component (i.e., North, East, and Up) is usually modeled as [29], [45], [57]

$$y(t) = a + bt + \sum_{j=1}^2 (A_j \sin 2\pi jt + B_j \cos 2\pi jt) + \varepsilon(t) \quad (8)$$

where $y(t)$ is the coordinate at epoch t ; a and b are the intersect and slope (velocity), respectively; A_j and B_j are annual and semiannual amplitudes, respectively; and $\varepsilon(t)$ is the noise. While there may exist other periodicities in time series, notably, draconitic periods and their overtones [25], [34], [64], [65], [66], they are not estimated here and, thus, remained in the CMEs [41]. With (8), we can filter the deterministic signals and obtain the residuals with the least-squares approach.

After preprocessing the regional time series of all stations, we obtain an $m \times n$ matrix \mathbf{X} for the residual position time series. The columns of \mathbf{X} represent the coordinate components of one station for all epochs, while the rows represent the coordinate components of one epoch for all stations. Since data gaps occur in the original time series [see Fig. 1(a)], the residual matrix \mathbf{X} inevitably contains missing data. Besides, the formal errors measuring the precision of the time series [see Fig. 1(b)], as provided prior information, should be considered in spatiotemporal filtering. To this end, we develop an

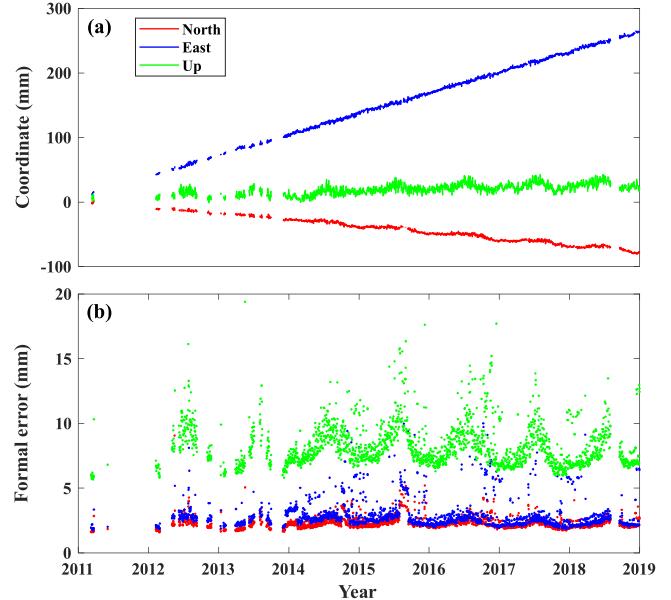


Fig. 1. (a) Position time series and (b) associated formal errors of the SXKL station.

EPCA for spatiotemporal filtering of regional GNSS networks’ data.

B. Unweighted Case

In this section, we focus on the unweighted case where formal errors are not considered and only the data-missing issue is addressed. Since there exist missing values in the datasets \mathbf{X} , the covariance matrix \mathbf{B} can only be computed with the available time series [45]. Let M_i and M_j denote the time series datasets of stations i and j , and m_i and m_j denote the number of epochs; the elements $\mathbf{B}(i, i)$ and $\mathbf{B}(i, j)$ of the covariance matrix \mathbf{B} can be computed with [45]

$$\begin{aligned} \mathbf{B}(i, i) &= \frac{1}{m_i - 1} \sum_{k \in M_i} X(k, i) X(k, i) \\ \mathbf{B}(i, j) &= \frac{1}{m_{ij} - 1} \sum_{k \in M_i \cap M_j} X(k, i) X(k, j) \end{aligned} \quad (9)$$

where m_{ij} denotes the number of epochs of the intersection set $M_i \cap M_j$ and $X(k, j)$ represents the coordinate of the j th station at epoch t_k . According to (6), the CMEs of the j th station at epoch t_i are [31]

$$S(i, j) = \sum_{k=1}^r \mathbf{a}_k(i) \mathbf{v}_k(j) \quad (10)$$

where $\mathbf{a}_k(i)$ is the i th component of \mathbf{a}_k and $\mathbf{v}_k(j)$ is the j th component of \mathbf{v}_k . Similarly, the k th PC can be computed with

$$\mathbf{a}_k(i) = \sum_{j=1}^n X(i, j) \mathbf{v}_k(j). \quad (11)$$

Substituting (11) into (10) yields

$$\begin{aligned} \mathbf{S}(i, j) &= \sum_{k=1}^r \sum_{h=1}^n \mathbf{v}_k(h) \mathbf{v}_k(j) \mathbf{X}(i, h) \\ &= \sum_{h=1}^n \left(\sum_{k=1}^r \mathbf{v}_k(j) \mathbf{v}_k(h) \right) \mathbf{X}(i, h). \end{aligned} \quad (12)$$

Equation (12) establishes the connection between the filtered CMEs \mathbf{S} and datasets \mathbf{X} . Once the datasets \mathbf{X} are prepared beforehand (without missing data), one can derive their CMCS with (12), which are the best approximations to \mathbf{X} among all $\mathbf{S}^* \in \mathbb{R}^{m \times n}$ with the rank of r . Therefore, to keep the best low-rank approximation property, the minimum norm criterion (7) is used to solve for the missing data \mathbf{X}_M in \mathbf{X}

$$\mathbf{X}_M = \operatorname{argmin}_{\mathbf{X}} : \sum_{i=1}^m \sum_{j=1}^n (\mathbf{X}(i, j) - \mathbf{S}(i, j))^2 \quad \text{s.t. Eq (12).} \quad (13)$$

By subtracting $\mathbf{S}(i, j)$ from $\mathbf{X}(i, j)$, we obtain the remaining components

$$\begin{aligned} \mathbf{E}(i, j) &= \left(1 - \sum_{k=1}^r \mathbf{v}_k^2(j) \right) \mathbf{X}(i, j) \\ &\quad - \sum_{h=1, h \neq j}^n \left(\sum_{k=1}^r \mathbf{v}_k(j) \mathbf{v}_k(h) \right) \mathbf{X}(i, h). \end{aligned} \quad (14)$$

Collecting all terms of $\mathbf{E}(i, j)$ for $j = 1, 2, \dots, n$, we have

$$\mathbf{E}(i) = (\mathbf{E}(i, 1) \ \mathbf{E}(i, 2) \ \dots \ \mathbf{E}(i, n))^T = \mathbf{C}\mathbf{X}_i \quad (15)$$

where $\mathbf{X}_i = (\mathbf{X}(i, 1) \ \mathbf{X}(i, 2) \ \dots \ \mathbf{X}(i, n))^T$ and \mathbf{C} is an $n \times n$ coefficient matrix defined as (16), shown at the bottom of the page.

Suppose that the position time series at epoch t_i are available only in the station subset S_i and the number of available stations is c_i . For the convenience of derivation, we separate the right-hand side of (15) into the available and missing data parts

$$\mathbf{E}(i) = \mathbf{C}_{i1}\mathbf{X}_{i1} + \mathbf{C}_{i2}\mathbf{X}_{i2} \quad (17)$$

where $\mathbf{X}_{i1} = \{\mathbf{X}(i, j) \mid j \in S_i\}$ and $\mathbf{X}_{i2} = \{\mathbf{X}(i, j) \mid j \notin S_i\}$ are the available data and missing data for epoch t_i , and \mathbf{C}_{i1} and \mathbf{C}_{i2} are the submatrices of \mathbf{C} , respectively. Therefore, the optimization criterion (13) can be equivalently expressed as

$$\mathbf{X}_M = \operatorname{argmin}_{\mathbf{X}} : \sum_{i=1}^m \mathbf{E}^T(i)\mathbf{E}(i) \quad \text{s.t. Eq (17).} \quad (18)$$

\mathbf{X}_M in criterion (18) is fully composed of $\mathbf{X}_{i2}(i = 1, 2, \dots, m)$, i.e., $\mathbf{X}_M = [\mathbf{X}_{12}^T \ \mathbf{X}_{22}^T \ \dots \ \mathbf{X}_{m2}^T]^T$. To solve for the missing data \mathbf{X}_M , we construct the following objective function:

$$\Phi = \sum_{i=1}^m (\mathbf{E}^T(i)\mathbf{E}(i) + 2\boldsymbol{\mu}_i^T(\mathbf{C}_{i1}\mathbf{X}_{i1} + \mathbf{C}_{i2}\mathbf{X}_{i2} - \mathbf{E}(i))) \quad (19)$$

where $\boldsymbol{\mu}_i$ denotes a vector of n Lagrange multipliers. Differentiating Φ with respect to the unknown parameters \mathbf{X}_{i2} , $\mathbf{E}(t_i)$, and $\boldsymbol{\mu}_i$ and equating them to the null vector, we obtain m sets of equations

$$\left\{ \begin{array}{l} \text{Equation(1)} \\ \text{Equation(2)} \\ \vdots \\ \text{Equation}(m) \end{array} \right. \quad (20)$$

where Equation(i) denotes the equation set for epoch t_i , which is defined as

$$\left\{ \begin{array}{l} \frac{\partial \Phi}{\partial \mathbf{E}_i} = 2(\mathbf{E}_i - \boldsymbol{\mu}_i) = \mathbf{0} \\ \frac{\partial \Phi}{\partial \boldsymbol{\mu}_i} = 2(\mathbf{C}_{i1}\mathbf{X}_{i1} + \mathbf{C}_{i2}\mathbf{X}_{i2} - \mathbf{E}(i)) = \mathbf{0} \\ \frac{\partial \Phi}{\partial \mathbf{X}_{i2}} = 2\mathbf{C}_{i2}^T\boldsymbol{\mu}_i = \mathbf{0}. \end{array} \right. \quad (21)$$

Equations (20) and (21) demonstrate that we can estimate the missing values in \mathbf{X} epoch by epoch since they are mutually independent. After a simple algebraic derivation of (21), we obtain

$$(\mathbf{C}_{i2}^T\mathbf{C}_{i2})\mathbf{X}_{i2} = -\mathbf{C}_{i2}^T\mathbf{C}_{i1}\mathbf{X}_{i1}. \quad (22)$$

The properties of the normal matrix $\mathbf{C}_{i2}^T\mathbf{C}_{i2}$ should be studied before solving for the unknowns \mathbf{X}_{i2} . For the convenience of derivation, we rewrite the coefficient matrix \mathbf{C} in (16) as

$$\mathbf{C} = \mathbf{I}_n - \mathbf{V}_r \mathbf{V}_r^T \quad (23)$$

where $\mathbf{V}_r = (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r)$ with \mathbf{v}_i is the i th column vector of \mathbf{V} . It is easy to prove that \mathbf{C} is symmetric and idempotent; thus, we have

$$\operatorname{Rank}(\mathbf{C}) = \operatorname{Trace}(\mathbf{C}) = n - r \quad (24)$$

where $\operatorname{Rank}(\cdot)$ and $\operatorname{Trace}(\cdot)$ are the operators for computing the rank and trace of a matrix. It is now clear that \mathbf{C} is rank-deficient with the number of rank deficiencies equaling the reconstruction order r . In other words, there are at most $n - r$ linearly uncorrelated variables among the n column

$$\mathbf{C} = \begin{pmatrix} 1 - \sum_{k=1}^r v_k(1)v_k(1) & -\sum_{k=1}^r v_k(1)v_k(2) & \cdots & -\sum_{k=1}^r v_k(1)v_k(n) \\ -\sum_{k=1}^r v_k(2)v_k(1) & 1 - \sum_{k=1}^r v_k(2)v_k(2) & \cdots & -\sum_{k=1}^r v_k(2)v_k(n) \\ \vdots & \vdots & \ddots & \vdots \\ -\sum_{k=1}^r v_k(n)v_k(1) & -\sum_{k=1}^r v_k(n)v_k(2) & \cdots & 1 - \sum_{k=1}^r v_k(n)v_k(n) \end{pmatrix} \quad (16)$$

vectors of \mathbf{C} . Since \mathbf{C}_{i2} is a matrix consisting of $n - c_i$ column vectors drawn from \mathbf{C} , its rank depends on c_i . If $c_i \geq r$, then \mathbf{C}_{i2} is clearly column full rank, i.e., $\text{Rank}(\mathbf{C}_{i2}) = n - c_i$, as all the involved column vectors are linearly uncorrelated; otherwise, \mathbf{C}_{i2} is rank-deficient as it contains at most $n - r < n - c_i$ linearly uncorrelated column vectors. Therefore, the rank of $\mathbf{C}_{i2}^T \mathbf{C}_{i2}$ in (22) is

$$\text{Rank}(\mathbf{C}_{i2}^T \mathbf{C}_{i2}) = \text{Rank}(\mathbf{C}_{i2}) = \begin{cases} n - c_i, & c_i \geq r \\ n - r, & c_i < r. \end{cases} \quad (25)$$

Noting that the number of unknowns in (22) is $n - c_i$, to guarantee a unique solution, the proposed method requires that the number of available data for epoch t_i should not be smaller than r . Therefore, the missing values for epoch t_i are estimated with

$$\mathbf{X}_{i2} = -(\mathbf{C}_{i2}^T \mathbf{C}_{i2})^{-1} (\mathbf{C}_{i2}^T \mathbf{C}_{i1} \mathbf{X}_{i1}). \quad (26)$$

The determination of the reconstruction order r is the key procedure for EPCA. In practical applications, we usually roughly choose the potential sets of reconstruction order by using the F test at a given significance level with the F -values computed with [67]

$$F(1, n - j) = \frac{(n - j)\lambda_j}{\sum_{i=j+1}^n \lambda_i}. \quad (27)$$

Then, the variance contribution and SR distribution of PCs are analyzed to further determine r for estimating the CMEs. In fact, the reconstruction order r serves as the minimum number of dominant PCs used for common components extraction. This means that, if the available station number for an epoch is smaller than r , the so-called “common” property cannot be fulfilled. Some studies have demonstrated that, for spatiotemporal filtering of regional GNSS networks’ data, the reconstruction order r is usually set to one since the first PC contributes the most to the total variance, and only it has the uniform SR [31], [45], [57], [68]. Therefore, EPCA assumes that the datasets for each epoch are not empty. In fact, this requirement also holds for MPCA as the covariance matrix computed with (9) may be nonpositive definite if the datasets at an epoch are missing for all stations.

By fusing (12) and (26), we can compute the CMEs with the available data, i.e.,

$$\mathbf{S}(i, j) = \sum_{k=1}^r \mathbf{v}_k(j) \left(\mathbf{V}_{k1}^T - \mathbf{V}_{k2}^T (\mathbf{C}_{i2}^T \mathbf{C}_{i2})^{-1} (\mathbf{C}_{i2}^T \mathbf{C}_{i1}) \right) \mathbf{X}_{i1} \quad (28)$$

where $\mathbf{V}_{k1} = \{\mathbf{v}_j(k) \mid j \in S_i\}$ and $\mathbf{V}_{k2} = \{\mathbf{v}_j(k) \mid j \notin S_i\}$ are the column vectors composed of the k th row elements of \mathbf{V} related to the available position and missing position. It indicates that EPCA can directly analyze the incomplete time series without interpolation beforehand, in the same manner as MPCA does. Nevertheless, as a byproduct of our approach, the missing data can be interpolated with the estimated signals at missing epochs. It should be emphasized here that the reconstructed components can only reflect the spatial correlation information, but the trends and the temporal correlation information have been removed primarily.

Therefore, if one is then to estimate the velocity using the reconstructed data, the trend terms must be added back.

The core idea of the proposed EPCA approach can be summarized as follows. After substituting (12) into criterion (13), we derive the alternative criterion (18) constraint to (17), by which a Lagrange objective function (19) is constructed to derive (26) for computing the missing data of the incomplete regional GNSS time series, and then, the CMEs (filtered signals) are computed with (28).

C. Weighted Case

Since the GNSS position time series is inhomogeneous, the formal errors of the time series should be considered in spatiotemporal filtering. In this section, we propose two weighted schemes for EPCA. The first is ported to the iterative weighted approach proposed by [57], which is based on MPCA. Recall that the original time series $\mathbf{X}(i, j)$ is composed of signal $\mathbf{S}(i, j)$ and noise $\mathbf{E}(i, j)$, i.e., $\mathbf{X}(i, j) = \mathbf{S}(i, j) + \mathbf{E}(i, j)$. For the available data $\mathbf{X}(i, j), i \in M_j$, the associated formal errors are denoted as $\sigma(i, j), i \in M_j$. To directly use unweighted EPCA for analysis, we should convert the heterogeneous data into homogeneous data with

$$\mathbf{X}_1(i, j) = \mathbf{S}(i, j) + \frac{\sigma_{0j}}{\sigma(i, j)} \mathbf{E}(i, j) \quad (i \in M_j) \quad (29)$$

where σ_{0j} is the standard deviation of the unit weight for the j th station and $\mathbf{X}_1(i, j)$ is the converted time series. It is easy to prove that $\mathbf{X}_1(i, j)$ is homogeneous for each station; thus, it can be processed with the approach presented in Section III-B. Since the true CMEs are impossibly known (otherwise, we do not need to estimate them), we have to replace them with the estimated values; thus, the weighted scheme should be iteratively conducted. For more details, one may refer to [57]. σ_{0j} can be determined based on the principle that the sum of the weights should be equal to the available number

$$\sum_{i \in M_j} \frac{\sigma_{0j}^2}{\sigma^2(i, j)} = m_j. \quad (30)$$

Therefore, we have

$$\sigma_{0j} = \sqrt{\frac{m_j}{\sum_{i \in M_j} 1/\sigma^2(i, j)}}. \quad (31)$$

Another weighted strategy is updating the optimization problem (18) as

$$\begin{aligned} \mathbf{X}_M = \text{argmin} : & \sum_{i=1}^m \sum_{j \in S_i} \mathbf{P}(i, j) \mathbf{E}^2(i, j) \\ & + \sum_{i=1}^m \sum_{j \notin S_i} \mathbf{E}^2(i, j) \quad \text{s.t. Eq.(17)} \end{aligned} \quad (32)$$

where $\mathbf{P}(i, j)$ is the weight of the coordinate at epoch t_i for the j th station, which is defined as $\mathbf{P}(i, j) = \sigma_{0j}^2 / \sigma^2(i, j)$. Therefore, the solutions (26) are updated as

$$\mathbf{X}_{i2} = -(\mathbf{C}_{i2}^T \mathbf{P}(i) \mathbf{C}_{i2})^{-1} (\mathbf{C}_{i2}^T \mathbf{P}(i) \mathbf{C}_{i1} \mathbf{X}_{i1}) \quad (33)$$

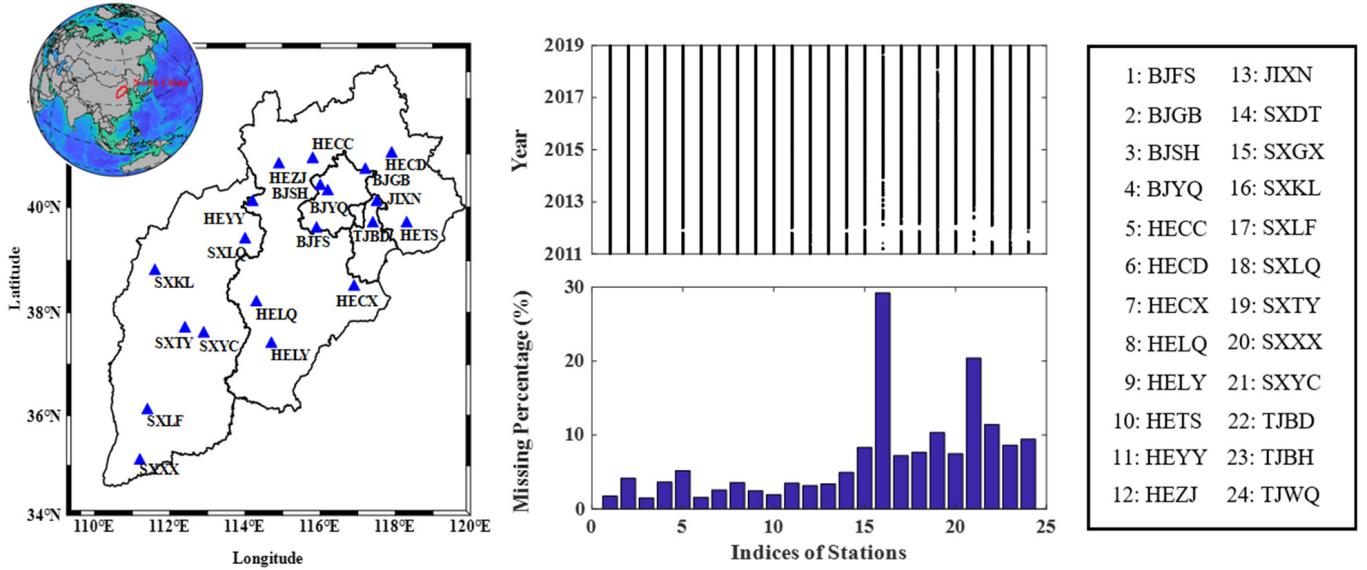


Fig. 2. (left) Geographic locations of 24 stations, (middle) percentage of missing data, and (right) indices of stations.

where $\mathbf{P}(i)$ is a diagonal matrix with the diagonal elements of p_{ij}

$$p_{ij} = \begin{cases} \sigma_{0j}^2 / \sigma^2(i, j), & j \in S_i \\ 1, & j \notin S_i. \end{cases} \quad (34)$$

Both weighted schemes can improve the accuracy of spatiotemporal filtering, which is demonstrated by the real and synthetic time series analysis presented in Section IV. In the following, (29) and (32) are called weighted scheme 1 and weighted scheme 2, respectively.

D. Contributions

The proposed EPCA and weighted EPCA have several advantages over the MPCA [45] and weighted MPCA [57]. First, the EPCA is fully established based on the best low-rank approximation in the spatiotemporal domain, and it theoretically approximates the spatiotemporal observations better than MPCA, where an additional criterion in the frequency domain is introduced to solve the PCs.

Second, the EPCA is theoretically more efficient than the MPCA. The proof is given as follows. The time complexity of an algorithm depends on the most time-consuming part [69], which refers to the solution of (22) for EPCA and the estimation of PC parameters for MPCA. For MPCA, the whole computation procedure should be conducted by m times to obtain the full PC matrix. In each computation, an equation with n unknowns needs to be solved. Therefore, the associated time complexity is $O(m \cdot n^3)$. For EPCA, the whole procedure should also be repeated by m times, but the number of unknowns to be estimated for each computation is $n - c_i$. In other words, the time complexity of EPCA is $O(\sum_{i=1}^m (n - c_i)^3)$, obviously lower than that of MPCA. Therefore, EPCA is theoretically more efficient than MPCA since the former has far fewer unknowns than the latter.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. Data Description

The Crustal Movement Observation Network of China (CMONOC) has provided GNSS position time series for over 20 years. In this study, we choose 24 permanent GNSS stations distributed in North China, of which the geographic locations are presented in Fig. 2 (left). The dataset over the period from 2011 to 2019 can be obtained from the China Earthquake Data Center (CEDC, <https://www.eqdsc.com>). All the raw data were processed with a homogeneous state-of-the-art approach based on GAMIT/GLOBK software. The data availability and the percentages of data missing for the 24 stations are presented in Fig. 2 (middle), and the indices of stations are shown in Fig. 2 (right). The outliers in the GNSS position time series are detected and eliminated by the three times the interquartile range (IQR) criterion [26], [29]; the coordinates with formal errors larger than 10, 10, and 20 mm for North, East, and Up components, respectively, are also removed. The offsets are detected and repaired by the sigseg tool [70] and manual visual inspection.

B. Real Data Analysis

With (9), we can compute the covariance matrix \mathbf{B} with the available data for three coordinate components, and then, the eigenvalues and eigenvectors can be estimated using eigenvalue decomposition of the covariance matrix \mathbf{B} . Fig. 3 presents the histogram of the eigenvalues in the order of PCs. The results show that the first PC represents 30%, 29%, and 31% of the total variance for North, East, and Up coordinate components, respectively, much larger than the other PCs.

The F -values for the first ten leading PCs at a confidence level of 95% are shown in Fig. 4, from which we find that the second PC is significant in all three coordinate components and the third PC is significant only in the East component. To make the comparison more intuitive, we use the normalized amplitude as Dong et al. [31] did, which is a ratio of each

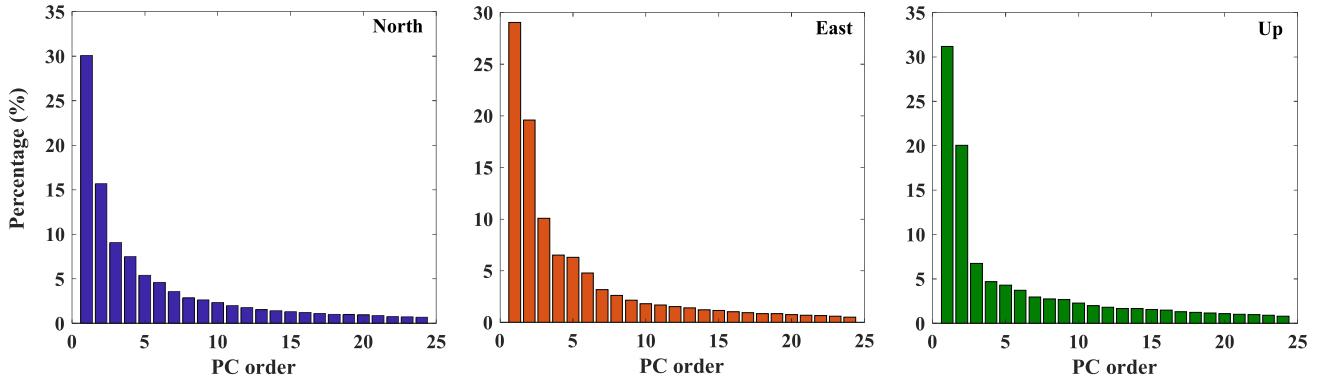


Fig. 3. Histogram of the eigenvalues as a function of PC order for three coordinate components.

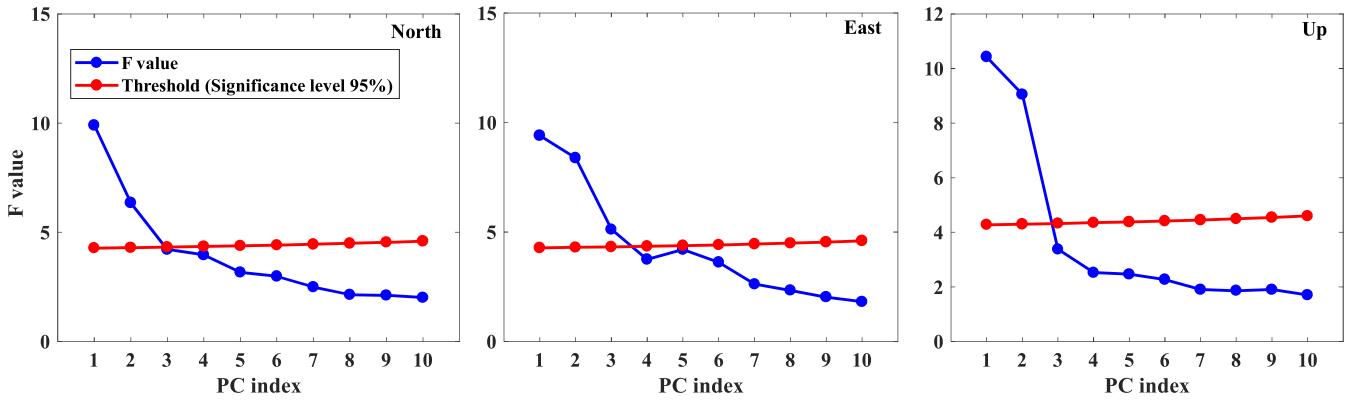


Fig. 4. F values of the first ten leading PCs for three coordinate components.

eigenvector divided by its maximum (absolute) element. The normalized amplitudes of the first three leading PCs for the three coordinate components are presented in the bottom figures of Fig. 5. It is clear that only the first PC has a consistent SR, with means of 46%, 74%, and 80% and minimums of 12%, 19%, and 27% for the North, East, and Up components, respectively. According to the definition of CMEs given by Dong et al. [31], the PC to which most stations (greater than 50%) have a significant normalized response (greater than 25%) should be regarded as one of the CME source signals. Therefore, we compute the CMEs for each station using only the first temporal PC and its eigenvector, which coincides with the previous studies [31], [40], [45], [50]. With (11) and (26), we can compute the temporal PCs with available data, and the first PCs for the three components are presented in the top figure of Fig. 5. The CMEs derived with the first PCs and their associated eigenvectors are shown in Fig. 6.

Since MPCA outperforms the classical interpolation approach and the stacking method, which has been demonstrated by Shen et al. [45], we only compare the results of our EPCA with those of MPCA. The differences in the CMEs estimated by EPCA and MPCA are presented by the green line in Fig. 6. One can clearly see in Fig. 6 that the results of the two methods are in close agreement, and the root mean square (RMS) of the maximum, minimum, and mean differences are only 0.35, 0.04, and 0.16 mm for the North components, 0.25, 0.05, and 0.18 mm for the East components, and 0.70, 0.19, and 0.56 mm for Up components, respectively.

To quantitatively compare the two approaches, we analyze the average errors, noise amplitudes, and velocity estimates, as well as their uncertainties before and after CMEs filtering. Table I presents the average errors of the time series for all 24 stations computed with the following equations:

$$\hat{\sigma}_R(i) = \sqrt{\frac{1}{m_i} \sum_{k \in M_i} X^2(k, i)} \quad (35)$$

$$\hat{\sigma}_F(i) = \sqrt{\frac{1}{m_i} \sum_{k \in M_i} (X(k, i) - S(k, i))^2} \quad (36)$$

where $\hat{\sigma}_R$ and $\hat{\sigma}_F$ are the average errors of the unfiltered and filtered time series, respectively. The results clearly demonstrate that the noise of the time series is greatly reduced after CMEs filtering with both MPCA and EPCA. This indicates that CMEs indeed exist in regional GNSS networks' position time series and should be filtered out for later geophysical studies, for instance, the crustal deformation signal analysis, the assessment of correlated noise, and the estimation of velocity fields [31], [45]. The average error reduction proportions of the filtered time series relative to the unfiltered ones are presented at the top of Fig. 7. With EPCA filtering, the average errors of the time series for the North, East, and Up components are reduced by the means of 19.17%, 22.16%, and 22.17%, higher than those with MPCA filtering by the means of 18.75%, 21.44%, and 21.12%, respectively. For a clear comparison, we present the reduction proportions of EPCA relative to MPCA at the bottom of Fig. 7. The results clearly

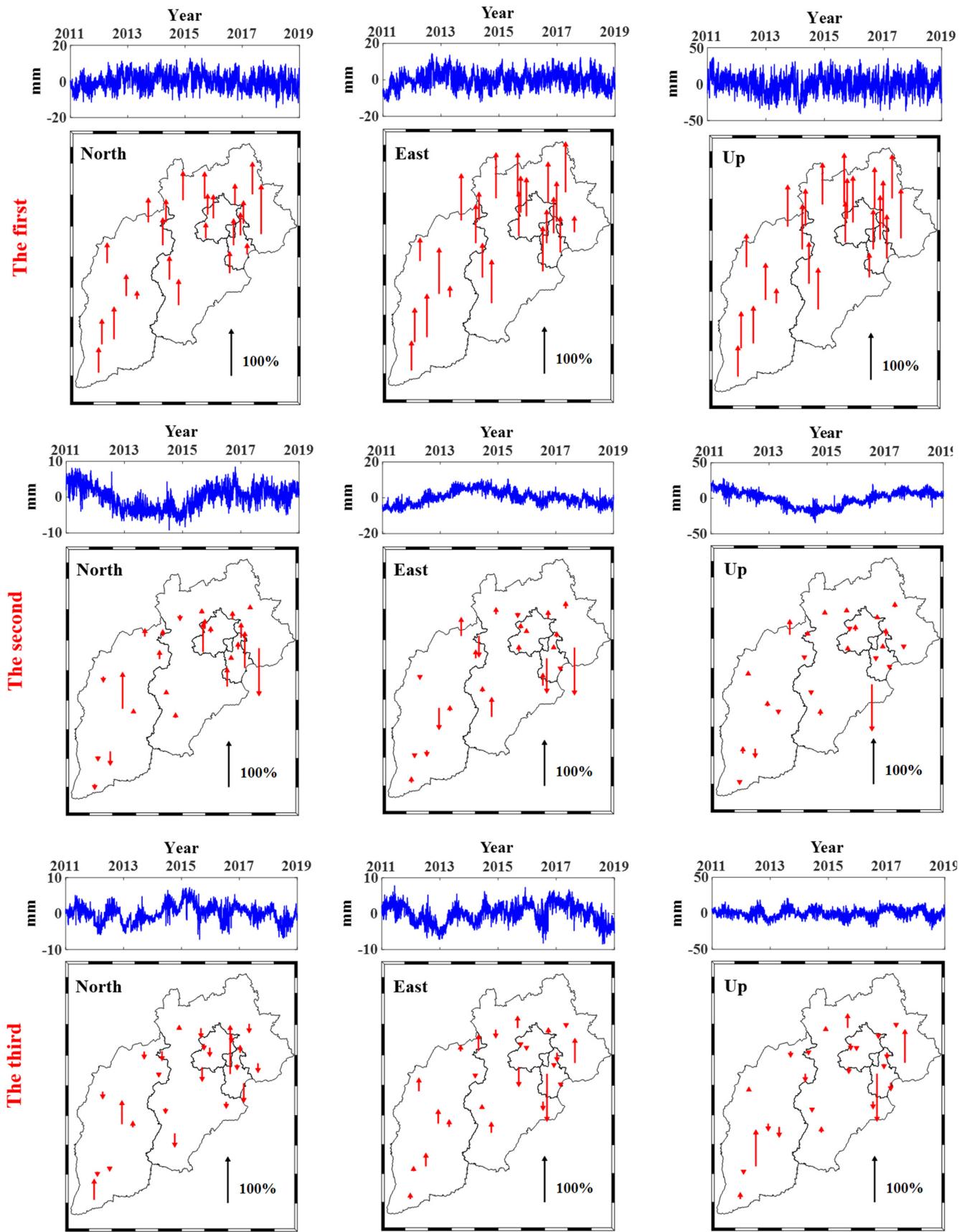


Fig. 5. First three leading PCs and SRs for three coordinate components.

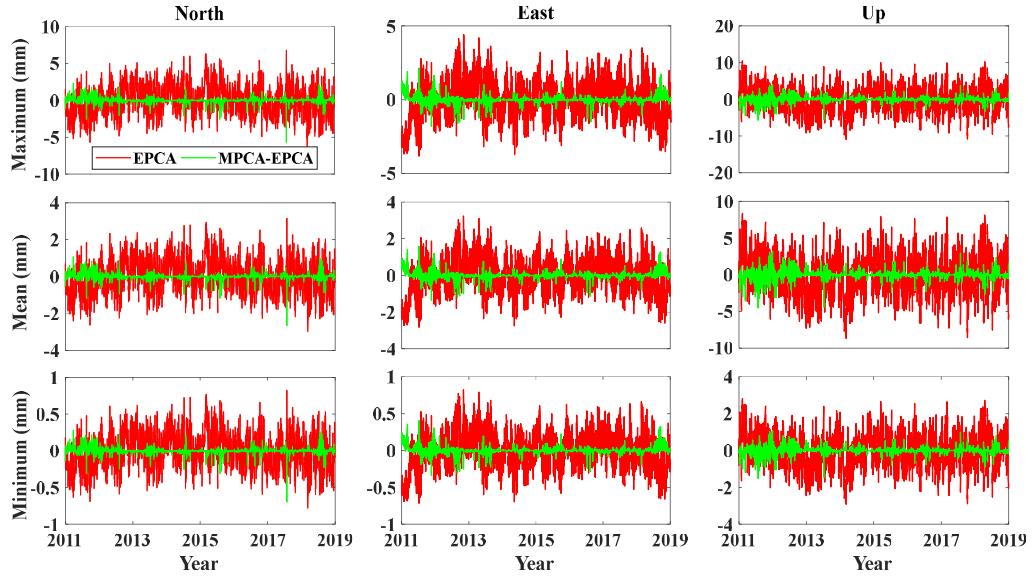


Fig. 6. CMEs extracted by EPCA (red line) and differences between EPCA and MPC-EPCA (green line).

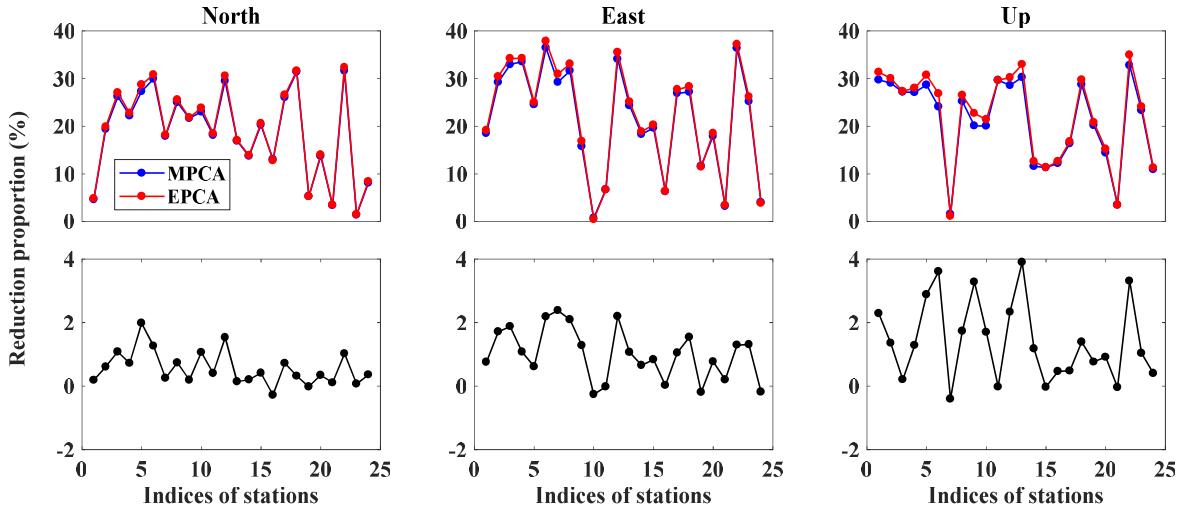


Fig. 7. Reduction proportion of the average errors. (Top) Filtered time series relative to unfiltered series. (Bottom) Filtered time series by EPCA relative to MPC-EPCA.

show that the EPCA outperforms MPC-EPCA on noise reduction, with the highest improvement by 1.98%, 2.38%, and 3.90% for the North, East, and Up components, respectively. Fig. 8 presents the RMS ratios of extracted CMEs over residuals, which further demonstrates that EPCA can extract more CMEs than MPC-EPCA.

In Section III-C, two weighted schemes are presented for EPCA, one of which is taken from Li and Shen [57]. We refer to the weighted MPC-EPCA, weighted EPCA (scheme 1), and weighted EPCA (scheme 2) as MPC-EPCA (W), EPCA (W1), and EPCA (W2), respectively. In what follows, we investigate the impact of considering formal errors on the variation of average errors for EPCA and compare the results with those of the weighted MPC-EPCA [57]. First, we compute the standard deviations of the unit weight for 24 stations with (31), which are presented in the gray bar in Fig. 9. It is obvious that the standard deviations of the vertical components are significantly larger than those of the horizontal components. For the same

coordinate component, the standard deviations of the unit weight are different for different stations, with means of 1.96, 2.00, and 7.26 mm for the North, East, and Up components, respectively. In addition, the standard deviations of the weight factors, measuring the degree of heterogeneity of the time series, are also presented as a blue dot in Fig. 9. We can see that they range from 0.2 to 0.6, with mean values for the North, East, and Up components of 0.28, 0.31, and 0.33, respectively. The results clearly illustrate that formal errors should be considered in spatiotemporal filtering due to the heterogeneity of time series.

Similar to the unweighted case, we compute the weighted average errors of the residual time series after filtering out the CMEs with

$$\hat{\sigma}_W(i) = \sqrt{\frac{1}{m_i} \sum_{k \in M_i} (X(k, i) - S(k, i))^2 P(k, i)}. \quad (37)$$

TABLE I
AVERAGE ERRORS OF POSITION TIME SERIES BEFORE AND AFTER CMEs FILTERING (mm)

Station	Before filtering			After filtering (MPCA)			After filtering (EPCA)		
	North	East	Up	North	East	Up	North	East	Up
BJFS	1.773	2.041	3.866	1.691	1.663	2.716	1.687	1.650	2.654
BJGB	1.259	1.501	3.929	1.015	1.062	2.786	1.009	1.044	2.749
BJSH	1.294	1.350	4.681	0.954	0.905	3.408	0.944	0.888	3.401
BJYQ	1.181	1.263	3.996	0.919	0.839	2.913	0.912	0.830	2.876
HECC	1.422	1.706	4.724	1.034	1.287	3.369	1.013	1.279	3.272
HECD	1.692	1.687	4.447	1.185	1.072	3.375	1.170	1.048	3.253
HECX	1.419	1.691	10.490	1.164	1.196	10.330	1.161	1.168	10.370
HELQ	1.341	1.246	4.348	1.006	0.852	3.251	0.998	0.834	3.195
HELY	1.602	2.183	4.692	1.255	1.838	3.749	1.253	1.815	3.626
HETS	3.083	3.105	5.629	2.373	3.083	4.499	2.347	3.091	4.423
HEYY	1.427	1.913	3.569	1.168	1.785	2.511	1.163	1.785	2.511
HEZJ	1.527	1.556	3.975	1.076	1.026	2.839	1.060	1.003	2.773
JIXN	1.543	1.583	4.420	1.282	1.198	3.082	1.280	1.185	2.962
SXDT	1.774	2.109	5.988	1.531	1.723	5.295	1.528	1.712	5.232
SXGX	2.149	1.973	5.851	1.713	1.585	5.186	1.706	1.572	5.188
SXKL	1.523	1.697	4.723	1.324	1.590	4.145	1.328	1.590	4.126
SXLF	1.444	1.356	4.793	1.068	0.991	4.007	1.060	0.980	3.988
SXLQ	1.425	1.482	4.389	0.977	1.079	3.128	0.974	1.063	3.084
SXTY	2.510	2.618	4.235	2.377	2.313	3.379	2.377	2.317	3.353
SXXX	1.942	1.449	4.216	1.676	1.189	3.607	1.670	1.180	3.574
SXYC	0.925	0.997	3.459	0.894	0.965	3.338	0.893	0.962	3.339
TJBD	1.173	1.241	4.052	0.802	0.790	2.724	0.794	0.780	2.634
TJBH	2.127	1.418	4.732	2.097	1.060	3.627	2.095	1.046	3.590
TJWQ	2.555	3.124	5.981	2.348	2.998	5.327	2.340	3.004	5.306

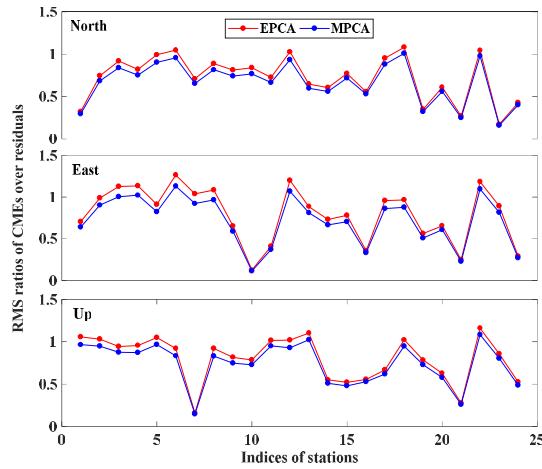


Fig. 8. RMS ratios of extracted CMEs over residuals.

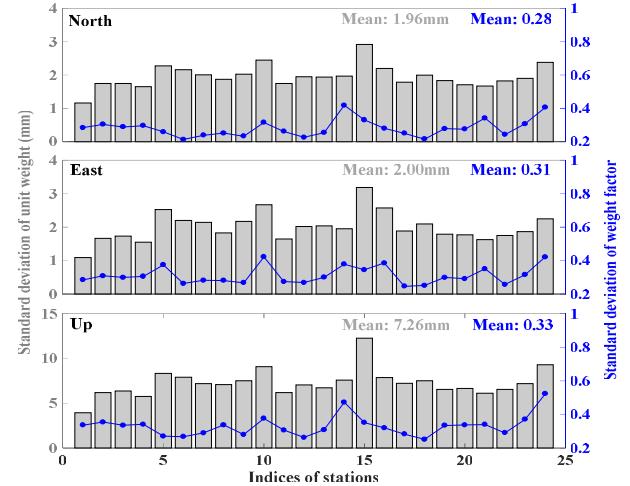


Fig. 9. Standard deviation of unit weight (gray bar) and weight factor (blue line) for each station.

After comparing the average errors of the weighted case with those of the unweighted case, we find that the consideration of formal errors can indeed further reduce the average errors for all stations in all coordinate components, which coincides well with [57]. The reduction proportions of the weighted case relative to the unweighted case for EPCA and MPCA are presented in Fig. 10, where (a) is for MPCA, (b) is for EPCA (weighted scheme 1), and (c) is for EPCA (weighted scheme 2). The reduction proportions are in close agreement for MPCA and EPCA. For MPCA, the maximum

reduction proportions are 9.61% and 10.55% for the North and East components at the HELQ station and 14.33% for the Up component at the BJGB station, while the mean values are 5.79%, 6.65%, and 8.56% for the three coordinate components, respectively.

For EPCA, both of the two weighted schemes effectively improve the accuracy of spatiotemporal filtering. Using weighted scheme 1, the average errors are reduced by the means of 5.66%, 6.58%, and 8.42% for the North, East, and

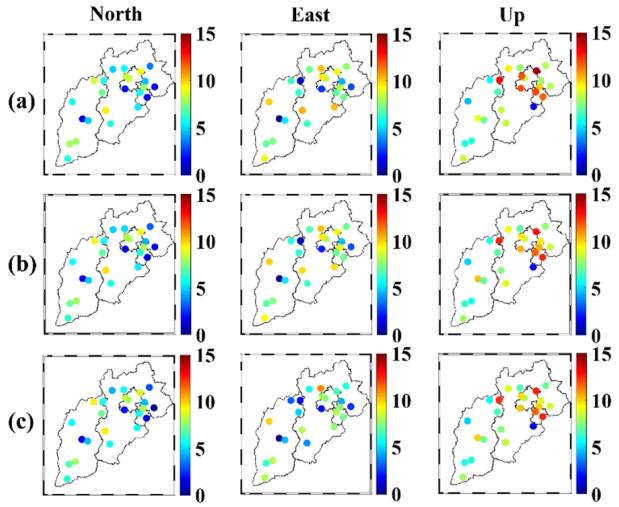


Fig. 10. Average error reduction proportion of the weighted case relative to the unweighted case. (a) MPCAs. (b) EPCA (weighted scheme 1). (c) EPCA (weighted scheme 2).

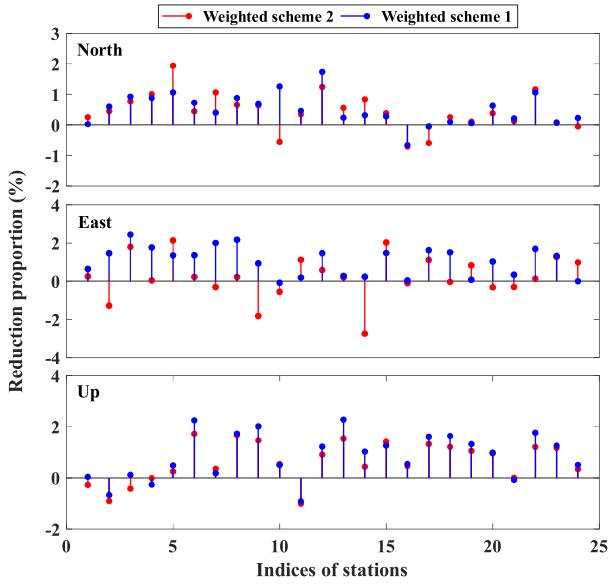


Fig. 11. Average error reduction proportions of EPCA relative to MPCAs for the weighted case.

Up components, while the counterparts for weighted scheme 2 are 5.60%, 5.81%, and 8.21%, respectively. It seems that the vertical component is more sensitive to the weighted case than the horizontal component. To make the comparison clearer, we present the reduction proportions of EPCA relative to MPCAs on weighted average errors in Fig. 11, where the blue and red lines denote weighted schemes 1 and 2 for EPCA, respectively. The results show that EPCA totally outperforms MPCAs even if the formal errors are taken into account. The maximum improvements of EPCA (W1) and EPCA (W2) relative to MPCAs (W) are 1.74% and 1.94% for the North component, 2.45% and 2.14% for the East component, and 2.28% and 1.72% for the Up component, respectively. The mean improvements are 0.51% and 0.45% for the North component, 1.06% and 0.23% for the East component, and 0.87% and 0.64% for the Up component, respectively.

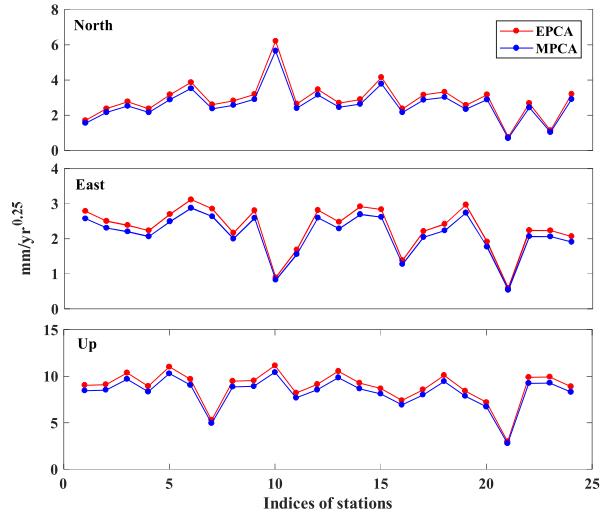


Fig. 12. Estimated FN amplitudes in extracted CMEs.

It is well known that time-correlated (colored) noise exists in the GNSS position time series [19], [20], [25], [50], and the optimal noise model is a combination of WN and PN. In many solutions, the noise can be best characterized by a combination of WN and flicker noise (FN), with a spectral index $\kappa = -1$ [71]. For individual position time series, previous studies have demonstrated that the PN amplitudes are much larger than the WN amplitudes; in other words, the noise is dominated by colored noise [19], [20], [25], [50]. Liu et al. [42] and Serpelloni et al. [50] concluded that the colored noise component is spatially coherent and forms most of the CMEs, while the WN is potentially site-specific and, therefore, mostly unchanged after spatiotemporal filtering. Since the PN amplitudes are dependent on κ , they are not easily comparable. The FN amplitudes, while not perhaps the correct noise model, do allow a degree of comparability [50]. Therefore, to further compare the performance of the two approaches on spatiotemporal filtering, we estimate the FN amplitudes of position time series before and after CMEs filtering using the Hector software [21], which can efficiently assess the noise characteristic of time series in the presence of missing values with the MLE technique. Fig. 12 shows the estimated FN amplitudes in CMEs extracted by EPCA and MPCAs for all 24 stations. It indicates that our EPCA can extract more CMEs than MPCAs as the associated noise amplitude estimates are larger. The results coincide well with the average error analysis. The mean FN amplitudes in the CMEs extracted by MPCAs for the North, East, and Up components are 2.62, 2.12, and 8.29 mm/year^{0.25}, while the counterparts extracted by EPCA are 2.88, 2.30, and 8.86 mm/year^{0.25}, respectively. In other words, the CMEs extracted by EPCA contain more FN than those extracted by MPCAs, with percentages of 9.92%, 7.83%, and 6.88% for North, East, and Up components, respectively. Table II presents the estimated FN amplitudes of the time series before and after CMEs filtering. It shows that the FN of the residual time series is greatly reduced after filtering out the CMEs. The mean FN amplitudes of the unfiltered, MPCAs-filtered, and

TABLE II
FN AMPLITUDES OF TIME SERIES BEFORE AND AFTER CMEs FILTERING ($\text{mm}/\text{yr}^{0.25}$)

Station	Before filtering			After filtering (MPCA)			After filtering (EPCA)		
	North	East	Up	North	East	Up	North	East	Up
BJFS	3.85	3.19	10.25	2.95	2.24	4.10	2.87	2.22	3.91
BJGB	3.10	3.10	10.40	1.87	1.74	4.10	1.82	1.74	4.16
BJSH	3.04	2.66	13.54	1.48	1.40	7.70	1.44	1.34	7.86
BJYQ	2.72	2.44	11.57	1.24	1.08	4.25	1.18	1.04	4.21
HECC	3.79	3.44	12.73	1.87	2.28	5.24	1.80	2.23	5.15
HECD	4.06	3.43	11.25	2.22	1.61	6.15	2.21	1.64	6.08
HECX	3.65	3.22	13.66	1.86	1.79	10.41	1.77	1.76	10.16
HELQ	3.41	2.62	11.08	1.75	1.35	5.32	1.70	1.26	5.03
HELY	3.95	3.67	11.33	2.19	2.45	6.29	2.11	2.43	6.10
HETS	5.25	4.80	14.00	3.68	4.44	8.52	3.74	4.42	8.49
HEYY	3.17	3.07	9.86	1.81	2.33	3.56	1.74	2.26	3.46
HEZJ	3.43	3.23	11.27	1.72	1.75	4.01	1.64	1.65	3.87
JIXN	3.38	3.28	9.88	1.97	1.96	4.28	1.91	1.94	4.43
SXDT	4.13	3.70	14.43	2.86	2.74	9.47	2.80	2.75	9.34
SXGX	5.33	4.28	13.92	3.79	3.11	10.27	3.81	3.02	10.16
SXKL	3.64	3.95	12.70	2.23	2.91	5.70	2.21	2.86	5.46
SXLF	3.60	2.92	11.88	2.17	1.43	7.48	2.15	1.31	7.32
SXLQ	3.69	3.16	11.87	1.38	1.71	4.40	1.34	1.61	4.14
SXTY	5.66	5.84	9.64	4.84	5.17	5.10	4.78	5.17	5.05
SXXX	3.67	2.95	10.44	2.63	2.08	6.54	2.68	2.06	6.36
SXYC	1.83	2.16	7.91	1.72	2.11	6.95	1.72	2.11	6.94
TJBD	3.19	2.64	11.30	1.21	1.27	4.50	1.16	1.26	4.46
TJBH	4.21	2.89	12.37	3.51	1.80	6.81	3.45	1.75	6.76
TJWQ	5.51	5.17	14.10	4.30	4.64	10.14	4.23	4.62	9.96

EPCA-filtered time series are 3.80, 2.39, and 2.34 mm/year^{0.25} for the North component, 3.41, 2.31, and 2.27 mm/year^{0.25} for the East component, and 11.72, 6.30, and 6.20 mm/year^{0.25} for the Up component, respectively. Relative to the unfiltered time series, the mean FN amplitudes of the time series filtered by MPCa and EPCA are reduced by 37.11% and 38.42% for the North component, 32.26% and 33.43% for the East component, and 46.25% and 47.10% for the Up component, respectively. As shown in Table II, the EPCA performs better than the MPCa since the noise amplitudes of the filtered series of the former are smaller than those of the latter for most stations. Since our EPCA method is established based on the best low-rank approximation in the spatiotemporal domain [see (13)], it can theoretically ensure an optimal overall filtering of regional networks' time series, rather than each individual station's time series. This is the reason why EPCA seems to be inferior to MPCa at a few stations. The maximum FN amplitude reductions of EPCA relative to MPCa are 4.84%, 8.39%, and 5.91%, while the means are 2.15%, 2.17%, and 1.70% for the North, East, and Up components, respectively. In addition, the results show that the noise level of the vertical components is significantly larger than that of the horizontal components by a factor of 2~14. This indicates that the precision of the vertical component is lower than that of the horizontal component, which is also verified by the formal errors in Fig. 1(b). The poor precision of the vertical components may be attributed to the geometric structure of the GNSS positioning model [72].

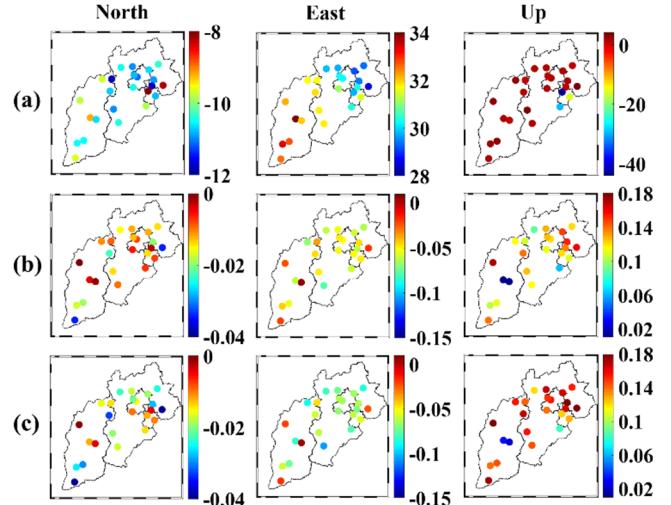


Fig. 13. Velocities of 24 stations (mm/year). (a) Unfiltered. (b) Filtered (MPCA) minus unfiltered. (c) Filtered (EPCA) minus unfiltered.

With the position time series, we can estimate the site velocity. Fig. 13(a) presents the velocity estimates of 24 stations for three components before CMEs filtering. It is clear that the horizontal components show clear linear trends, with mean velocity estimates of -10.33 mm/year for the North component and 31.06 mm/year for the East component, while the vertical components of most stations (except for TJBD, HETS, and HECD) show a negligible velocity with a mean

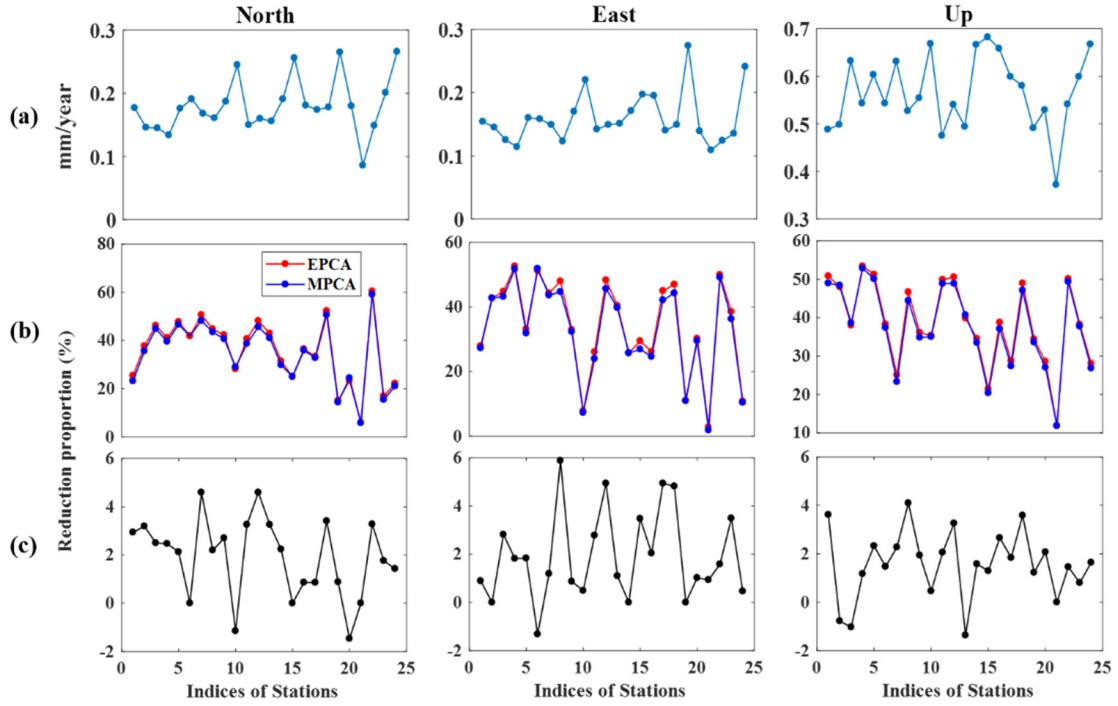


Fig. 14. Uncertainties of velocity estimates. (a) Unfiltered. (b) Reduction proportion of filtered uncertainties relative to unfiltered ones. (c) Reduction proportion of uncertainties filtered by EPCA relative to MPCA.

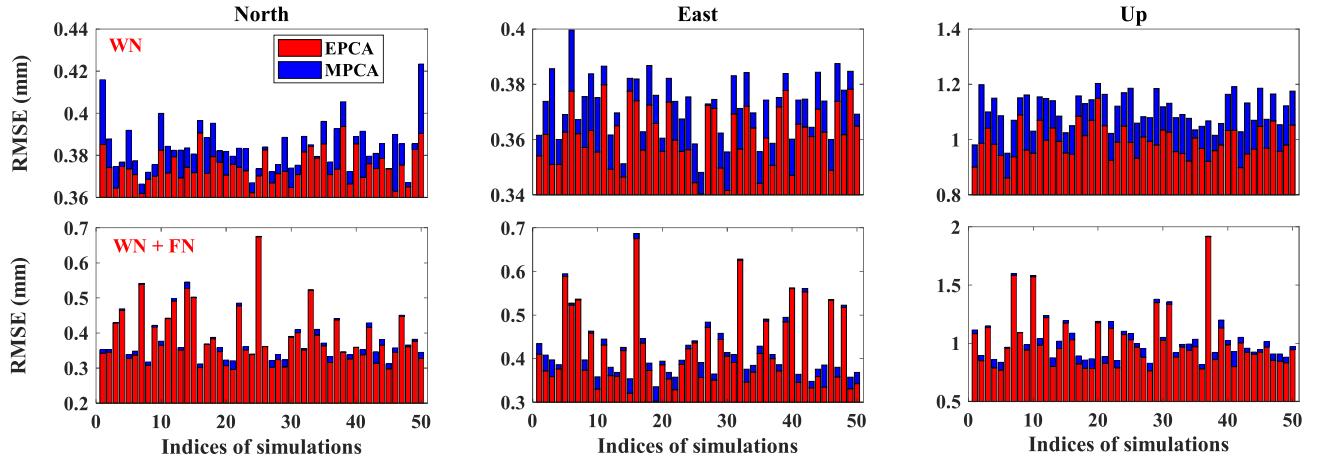


Fig. 15. RMSEs of extracted CMEs for EPCA and MPCA at available epochs.

of 1.67mm/year. Notably, the three stations located in Tianjin and Hebei provinces show significant subsidence, which is related to the overexploitation of groundwater [73], [74]. The velocity differences of 24 stations before and after filtering CMEs are shown in Fig. 13(b) and (c). The results demonstrate that spatiotemporal filtering will change the velocity estimates although the differences are not significant. Furthermore, we find that the uncertainties of the velocity estimates are greatly reduced after the CMEs are filtered, which coincides with previous studies [42], [50]. Fig. 14(a) presents the velocity uncertainties of the unfiltered time series of 24 stations. The uncertainty reduction proportions of filtered data relative to unfiltered data are shown in Fig. 14(b). It shows that EPCA can reduce the velocity uncertainties by 35.8%, 34.0%, and

38.6% for the three coordinate components, which are higher than the reductions produced by MPCA of 34.7%, 32.8%, and 37.6%, respectively. For a clear comparison, we present the uncertainty reduction proportions of EPCA relative to MPCA in Fig. 14(c). The results indicate that EPCA outperforms MPCA in terms of the accuracy improvements of the velocity estimates, with a maximum improvement of 4.60% for the North component at HECX, and 5.88%, and 4.10% for the East and Up components at HELQ, respectively.

C. Synthetic Time Series Analysis

To further validate the performance of the EPCA approach, numerical simulation experiments are conducted, in which the

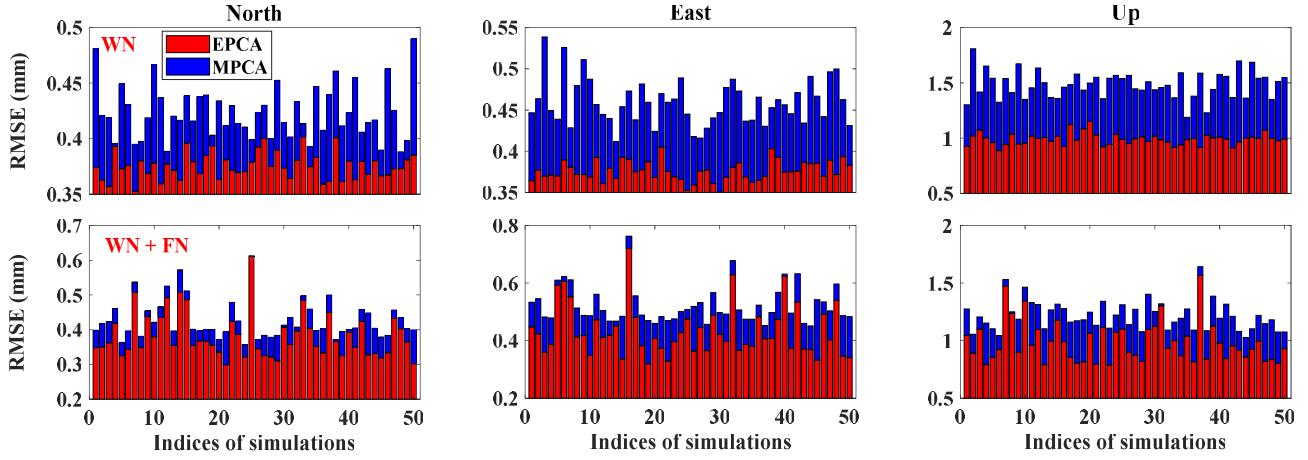


Fig. 16. RMSEs of extracted CMEs for EPCA and MPCa at missing epochs.

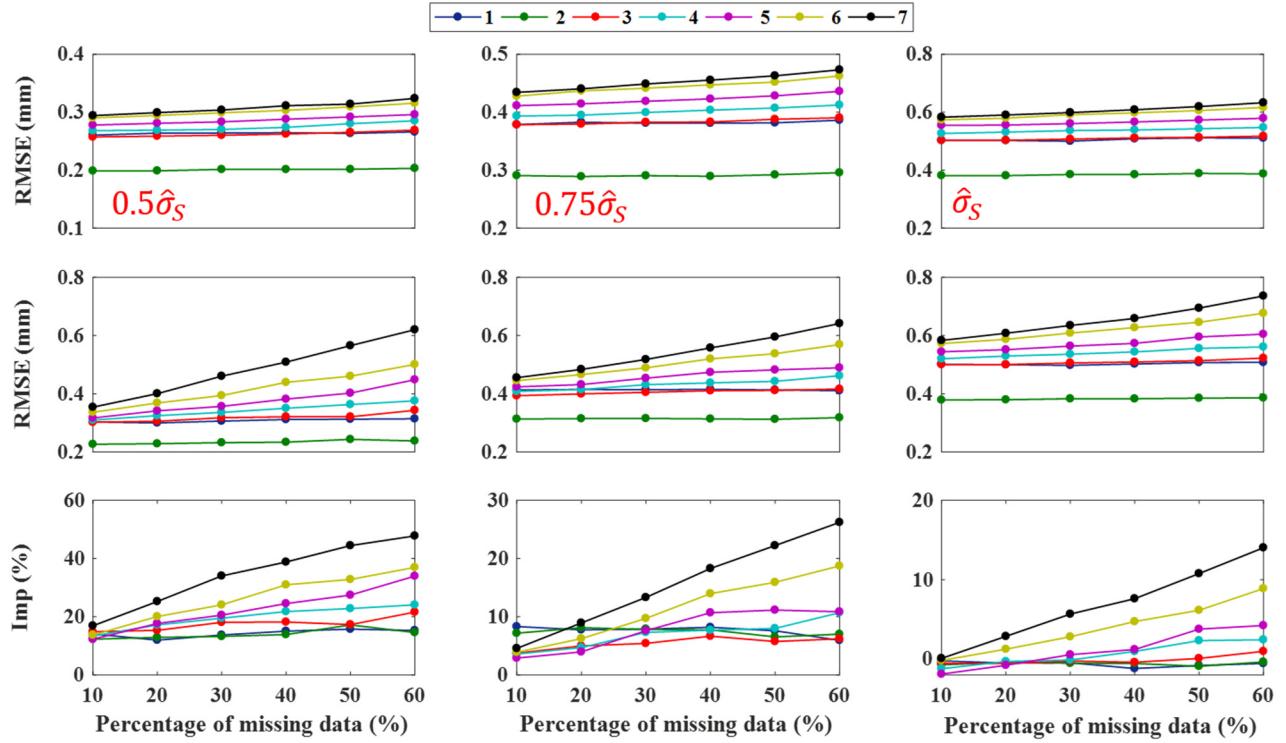


Fig. 17. RMSEs at available epochs. The first and second rows are the RMSEs for EPCA and MPCa, and the third row is the reduction proportion of EPCA relative to MPCa. The legends of numbers indicate the number of stations with data deletion.

extracted CMEs from real data are regarded as the true signals S_0 , and then, the noise e_0 is added to S_0 . Therefore, the synthetic time series are generated by

$$X_0 = S_0 + e_0. \quad (38)$$

The noise e_0 is generated under the assumption of pure WN and WN plus FN, where the noise amplitudes are derived from the real data. To test the two approaches for processing incomplete time series, we delete the complete data according to the missing data positions in the real data. Since the true CMEs are available here, we can compute the RMS errors (RMSEs) of the estimated CMEs at the missing (RMSE_M)

and nonmissing (i.e., available) epochs (RMSE_S) with

$$\text{RMSE}_S = \sqrt{\frac{\sum_{i=1}^m \sum_{j \in S_i} (S(i, j) - S_0(i, j))^2}{\sum_{i=1}^m c_i}} \quad (39)$$

$$\text{RMSE}_M = \sqrt{\frac{\sum_{i=1}^m \sum_{j \notin S_i} (S(i, j) - S_0(i, j))^2}{mn - \sum_{i=1}^m c_i}}.$$

To obtain statistically reliable results, we repeat the simulation experiment 50 times and present the results in Figs. 15 and 16, which demonstrates that EPCA outperforms MPCa as the associated estimates are closer to the true values. Table III presents the average RMSEs of 50 experiments for MPCa and EPCA, where Imp_S and Imp_M represent the

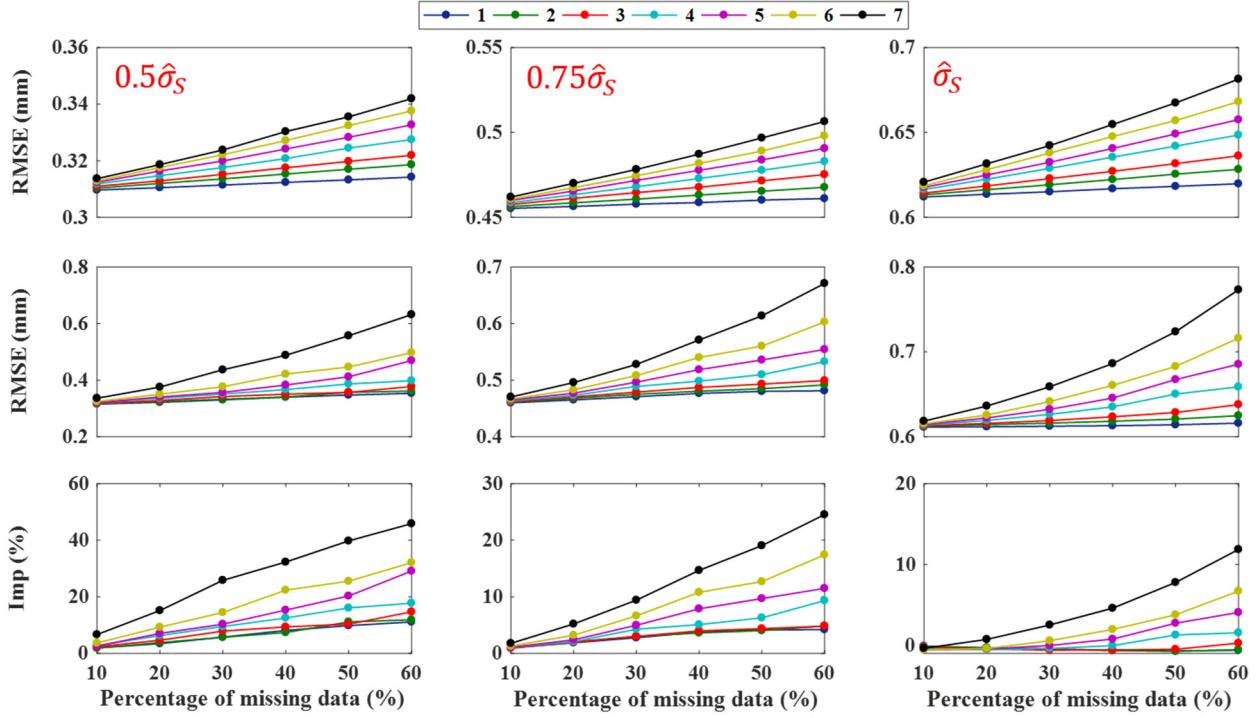


Fig. 18. RMSEs at missing data epochs (the y - and x -axes and meaning of legends are the same as in Fig. 17).

TABLE III
AVERAGE RMSES OF EPCA AND MPCAs

Noise	RMSEs / Imp	North	East	Up
WN	$RMSE_S(EPCA)$	0.3749mm	0.3614mm	0.9916mm
	$RMSE_S(MPCA)$	0.3837mm	0.3722mm	1.1052mm
	$RMSE_M(EPCA)$	0.3760mm	0.3761mm	0.9910mm
	$RMSE_M(MPCA)$	0.4230mm	0.4590mm	1.4844mm
	Imp_S	2.29%	2.90%	10.28%
	Imp_M	11.11%	18.06%	33.24%
WN+FN	$RMSE_S(EPCA)$	0.3829mm	0.4168mm	1.0023mm
	$RMSE_S(MPCA)$	0.3908mm	0.4313mm	1.0459mm
	$RMSE_M(EPCA)$	0.3845mm	0.4349mm	0.9839mm
	$RMSE_M(MPCA)$	0.4237mm	0.5210mm	1.2157mm
	Imp_S	2.02%	3.36%	4.17%
	Imp_M	9.25%	16.53%	19.07%

improvements of EPCA relative to MPCA in terms of the reductions of $RMSE_S$ and $RMSE_M$, respectively. For the pure WN model, Imp_S and Imp_M are 2.29% and 11.11% for the North component, 2.90% and 18.06% for the East component, and 10.28% and 33.24% for the Up component, respectively. For the WN plus FN model, the counterparts are 2.02% and 9.25%, 3.36% and 16.53%, and 4.17% and 19.07% for the North, East, and Up components, respectively. Therefore, EPCA is superior to MPCA in extracting CMEs both for interpolated missing data and available data, and the superiority is more significant for interpolated missing data.

To investigate the impacts of different percentages of missing data on the extracted CMEs with our method, we conduct simulation experiments by deleting different percentages of data at different numbers of stations. The first seven stations

with the highest missing percentages (denoted as SXKL, SXYC, TJBD, SXTY, TJWQ, TJBH, and SXGX) are chosen for deleting data. The percentages of deleted data range from 10% to 60% in 10% increments, and the number of stations for which data are deleted ranges from one to seven. In addition, to investigate how the proposed EPCA method is affected by the noise level, random noise is simulated under a normal distribution with zero mean and variance of $k\hat{\sigma}_s$, where k is the scaling factor (0.5, 0.75, and 1.0) and $\hat{\sigma}_s$ is the RMS of the CMEs. By repeating 50 simulations, we compute the mean RMSEs with (39) and show the results in Figs. 17 and 18, where the first and second rows denote the RMSEs of EPCA and MPCA, respectively, and the third row represents the reduction proportion of EPCA relative to MPCA. It is clear that both the EPCA and MPCA approaches

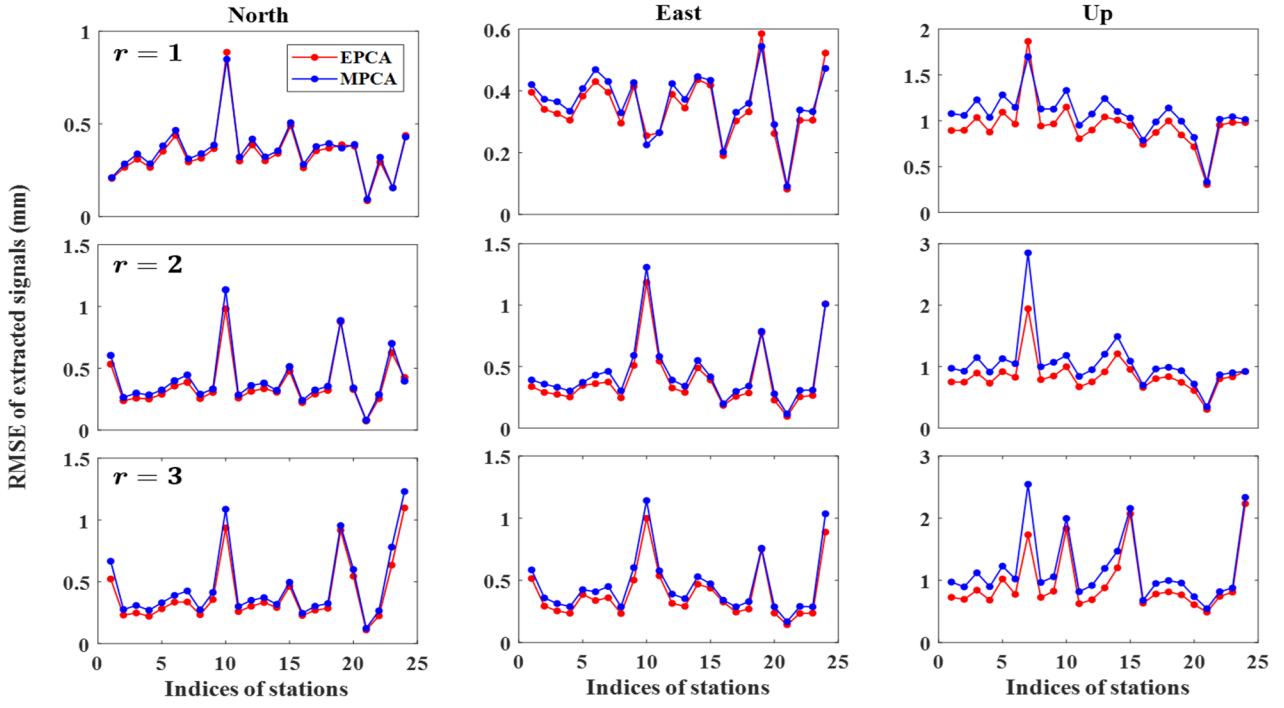


Fig. 19. RMSEs of extracted signals for different cases.

become progressively worse as the percentage of missing data increases. Nevertheless, the superiority of EPCA over MPCA becomes increasingly visible. On the other hand, we find that the lower the noise level is, the larger the improvement of EPCA relative to MPCA.

Although we use the first PC and the associated eigenvectors to estimate CMEs, unknown signals may exist in other modes due to the local effects [31], [57]. Moreover, for a general PCA filtering problem, the reconstruction order r may not always be one. To investigate the impact of different values of r on the performance of EPCA, the true CMEs in simulation experiments are generated from real data with $r = 1 \sim 3$. By the 50 repeated experiments, we compute the average RMSEs of extracted CMEs for each station and present the RMSEs in Fig. 19 and the average RMSEs of the 24 stations in Table IV for different cases. The results show that the superiority of EPCA over MPCA becomes gradually visible as r increases, which agrees with theoretical aspects since the time-domain filtering of EPCA is highly related to r .

To compare the EPCA and MPCA approaches in the weighted case, the random noise e_0 is simulated according to the formal errors [57]. Fig. 20 presents the RMSEs of the estimated CMEs by the two approaches, including the unweighted and weighted cases, which demonstrates that considering formal errors can improve the filtering accuracy as the associated RMSEs are smaller. For MPCA, the average RMSEs of the North, East, and Up components are reduced from 0.46, 0.45, and 1.43 mm for the unweighted case to 0.45, 0.41, and 1.35 mm for the weighted case, with the reductions of 2.17%, 8.89%, and 5.59%, respectively. For EPCA, the improvements of the weighted case over the unweighted case are 6.32%, 13.40%, and 7.41% for weighted scheme 1, and

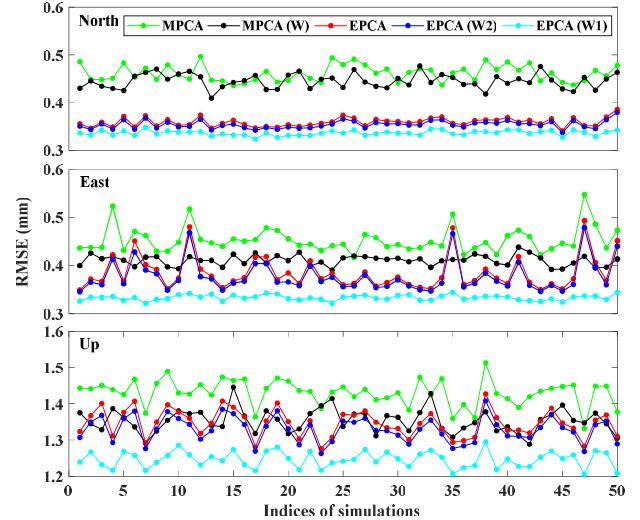


Fig. 20. RMSEs of EPCA and MPCA for the equal weight and weighted cases.

1.45%, 2.05%, and 1.27% for weighted scheme 2, respectively. Therefore, our EPCA outperforms MPCA whether the formal errors are taken into account or not. Moreover, the results indicate that weighted scheme 1 seems to perform better than weighted scheme 2 for EPCA, which coincides with the real data analysis.

To compare the computational efficiency between MPCA and EPCA, we repeat the 50 simulations and measure their computation times in MATLAB 2017 on a Windows 10 system with a 2.8-GHz Intel processor and 8-GB memory. Denoting the computation times of MPCA, MPCA (W), EPCA,

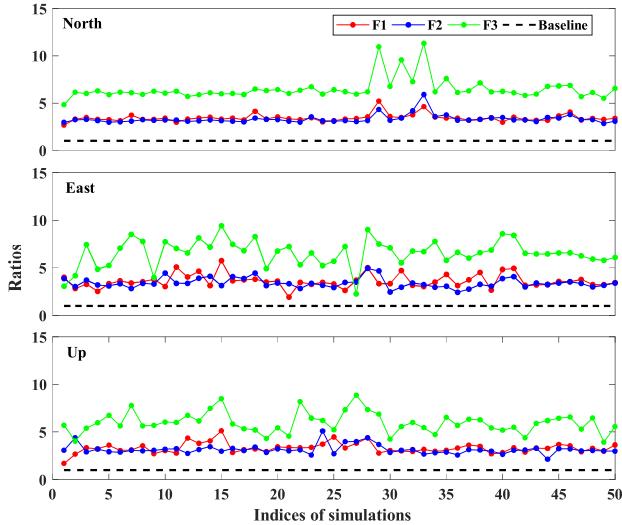


Fig. 21. Ratio of MPCAs to EPCA on computation times.

TABLE IV
AVERAGE RMSES OF EPCA AND MPCAs

		RMSEs of MPCAs (mm)	RMSEs of EPCA (mm)	Imp (%)
North	$r = 1$	0.3574	0.3432	3.97
	$r = 2$	0.4102	0.3741	8.80
	$r = 3$	0.4626	0.4026	12.97
East	$r = 1$	0.3619	0.3450	4.67
	$r = 2$	0.4484	0.3984	11.15
	$r = 3$	0.4572	0.3959	13.41
Up	$r = 1$	1.0684	0.9486	11.21
	$r = 2$	1.0464	0.8511	18.66
	$r = 3$	1.1738	0.9672	17.60

EPCA (W1), and EPCA (W2) as T_M , T_{MW} , T_E , T_{EW1} , and T_{EW2} , respectively, the ratios of EPCA relative to MPCAs, i.e., $F_1 = T_M/T_E$, $F_2 = T_{MW}/T_{EW1}$, and $F_3 = T_{MW}/T_{EW2}$ are computed and presented in Fig. 21, where the black line denotes the baseline with the values of one. The results show that EPCA performs more efficiently than MPCAs regardless of whether the formal errors are considered, which coincides with the theoretical analysis. For the unweighted case, the EPCA can reduce the computation time by an average factor of 3.43, 3.61, and 3.30 for the North, East, and Up components, respectively; for the weighted case, the counterparts are 3.32, 3.40, and 3.15 for weighted scheme 1, and 6.49, 6.56, and 5.97 for weighted scheme 2, respectively.

V. CONCLUSION

This study develops the EPCA approach for extracting CMEs from incomplete and heterogeneous time series. The two most attractive properties of EPCA are that it does not fill data gaps beforehand and is consistent with ordinary PCA in optimal low-rank filtering. The proposed approach is successfully applied to the spatiotemporal filtering of the

regional GNSS position time series of 24 stations located in North China from 2011 to 2019. The results show that, compared to the MPCA and weighted MPCA, which solve an optimization problem in the frequency domain, EPCA can not only obtain better results but is also computationally more efficient since fewer unknowns need to be estimated in EPCA. Specifically, the average errors of time series filtered by EPCA and weighted EPCA are smaller than those filtered by MPCA and weighted MPCA, indicating that EPCA can extract more CMEs. The noise analysis further shows that the FN amplitudes in CMEs extracted by EPCA are larger than those extracted by MPCA, and the counterparts in the filtered time series by EPCA are smaller. In addition, we find that the uncertainties of velocity estimates are greatly reduced after CMEs filtering, but the reduction proportion of EPCA is larger than that of MPCA. In summary, the real data analysis comprehensively demonstrates that EPCA outperforms MPCA on spatiotemporal filtering. The repeated simulations show that the estimated CMEs with our EPCA approach are closer to the simulated true values than MPCA, both under the assumptions of WN and WN plus FN. Moreover, with increases in the percentage of missing data and reconstruction order, the superiority of EPCA over MPCA becomes increasingly obvious.

REFERENCES

- [1] K. R. Franklin and M. Huang, "Revealing crustal deformation and strain rate in Taiwan using InSAR and GNSS," *Geophys. Res. Lett.*, vol. 49, no. 21, Nov. 2022, Art. no. e2022GL101306.
- [2] Y. Pan et al., "GPS imaging of vertical bedrock displacements: Quantification of two-dimensional vertical crustal deformation in China," *J. Geophys. Res., Solid Earth*, vol. 126, no. 4, Apr. 2021, Art. no. e2020JB020951.
- [3] K. M. Larson et al., "Dynamic sea level variation from GNSS: 2020 Shumagin earthquake tsunami resonance and hurricane Laura," *Geophys. Res. Lett.*, vol. 48, no. 4, Feb. 2021, Art. no. e2020GL091378.
- [4] J.-P. Montillet, T. I. Melbourne, and W. M. Szeliga, "GPS vertical land motion corrections to sea-level rise estimates in the Pacific Northwest," *J. Geophys. Res., Oceans*, vol. 123, no. 2, pp. 1196–1212, Feb. 2018.
- [5] A. O. Alothman, M. Bos, R. Fernandes, A. M. Radwan, and M. Rashwan, "Annual sea level variations in the red sea observed using GNSS," *Geophys. J. Int.*, vol. 221, no. 1, pp. 826–834, Dec. 2019.
- [6] G. Carlson, S. Werth, and M. Shirzaei, "Joint inversion of GNSS and GRACE for terrestrial water storage change in California," *J. Geophys. Res., Solid Earth*, vol. 127, no. 3, Mar. 2022, Art. no. e2021JB023135.
- [7] S. Jin and T. Zhang, "Terrestrial water storage anomalies associated with drought in southwestern USA from GPS observations," *Surv. Geophys.*, vol. 37, no. 6, pp. 1139–1156, Nov. 2016.
- [8] Z. Jiang et al., "Insights into hydrological drought characteristics using GNSS-inferred large-scale terrestrial water storage deficits," *Earth Planet. Sci. Lett.*, vol. 578, Jan. 2022, Art. no. 117294.
- [9] Z. Jiang, Y.-J. Hsu, L. Yuan, and D. Huang, "Monitoring time-varying terrestrial water storage changes using daily GNSS measurements in Yunnan, Southwest China," *Remote Sens. Environ.*, vol. 254, Mar. 2021, Art. no. 112249.
- [10] K. Yu, W. Ban, X. Zhang, and X. Yu, "Snow depth estimation based on multipath phase combination of GPS triple-frequency signals," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 9, pp. 5100–5109, Sep. 2015.
- [11] S. Tabibi, F. Geremia-Nievinski, and T. van Dam, "Statistical comparison and combination of GPS, GLONASS, and multi-GNSS multipath reflectometry applied to snow depth retrieval," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3773–3785, Jul. 2017.
- [12] Y. Hu, X. Yuan, W. Liu, J. Wickert, and Z. Jiang, "GNSS-R snow depth inversion based on variational mode decomposition with multi-GNSS constellations," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–12, 2022, Art. no. 2005512.
- [13] F. Alshawaf et al., "Accurate estimation of atmospheric water vapor using GNSS observations and surface meteorological data," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3764–3771, Jul. 2015.

- [14] X. Li et al., "Multi-GNSS meteorology: Real-time retrieving of atmospheric water vapor from BeiDou, galileo, GLONASS, and GPS observations," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 12, pp. 6385–6393, Dec. 2015.
- [15] W. Zhang, S. Zhang, N. Ding, L. Holden, X. Wang, and N. Zheng, "GNSS-RS tomography: Retrieval of tropospheric water vapor fields using GNSS and RS observations," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, 2022, Art. no. 4102313.
- [16] Y. Tian and Z. Shen, "Extracting the regional common-mode component of GPS station position time series from dense continuous network," *J. Geophys. Res., Solid Earth*, vol. 121, no. 2, pp. 1080–1096, Feb. 2016.
- [17] X. He et al., "Review of current GPS methodologies for producing accurate time series and their error sources," *J. Geodyn.*, vol. 106, pp. 12–29, May 2017.
- [18] A. Santamaría-Gómez and J. Ray, "Chameleonic noise in GPS position time series," *J. Geophys. Res., Solid Earth*, vol. 126, no. 3, Mar. 2021, Art. no. e2020JB019541.
- [19] L. Wang, Q. Wu, F. Wu, and X. He, "Noise content assessment in GNSS coordinate time-series with autoregressive and heteroscedastic random errors," *Geophys. J. Int.*, vol. 231, no. 2, pp. 856–876, Jul. 2022.
- [20] S. D. P. Williams, "Error analysis of continuous GPS position time series," *J. Geophys. Res.*, vol. 109, no. B3, 2004, Art. no. B03412s.
- [21] M. S. Bos, R. M. S. Fernandes, S. D. P. Williams, and L. Bastos, "Fast error analysis of continuous GNSS observations with missing data," *J. Geodesy*, vol. 87, no. 4, pp. 351–360, Apr. 2013.
- [22] J.-P. Montillet, P. Tregoning, S. McClusky, and K. Yu, "Extracting white noise statistics in GPS coordinate time series," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 3, pp. 563–567, May 2013.
- [23] X. He, M. S. Bos, J. P. Montillet, and R. M. S. Fernandes, "Investigation of the noise properties at low frequencies in long GNSS time series," *J. Geodesy*, vol. 93, no. 9, pp. 1271–1282, Sep. 2019.
- [24] A. Klos, M. A. Karegar, J. Kusche, and A. Springer, "Quantifying noise in daily GPS height time series: Harmonic function versus GRACE-assimilating modeling approaches," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 4, pp. 627–631, Apr. 2021.
- [25] A. R. Amiri-Simkooei, C. C. J. M. Tiberius, and P. J. G. Teunissen, "Assessment of noise in GPS coordinate time series: Methodology and results," *J. Geophys. Res.*, vol. 112, no. B7, 2007, Art. no. B07413.
- [26] K. Ji and Y. Shen, "A wavelet-based outlier detection and noise component analysis for GNSS position time series," in *Proc. Int. Assoc. Geodesy Symposia*, 2020, pp. 45–53.
- [27] R. Tehranchi, K. Moghtased-Azar, and A. Safari, "Fast approximation algorithm to noise components estimation in long-term GPS coordinate time series," *J. Geodesy*, vol. 95, no. 2, pp. 1–16, Feb. 2021.
- [28] D. A. Cucci, L. Voiril, G. Kermarrec, J.-P. Montillet, and S. Guerrier, "The generalized method of wavelet moments with eXogenous inputs: A fast approach for the analysis of GNSS position time series," *J. Geodesy*, vol. 97, no. 2, p. 14, Feb. 2023.
- [29] R. Nikolaidis, *Observation of Geodetic and Seismic Deformation With the Global Positioning System*. San Diego, CA, USA: Univ. California, 2004.
- [30] B. Márquez-Azúa and C. DeMets, "Crustal velocity field of Mexico from continuous GPS measurements, 1993 to June 2001: Implications for the neotectonics of Mexico," *J. Geophys. Research: Solid Earth*, vol. 108, no. B9, p. 2450, Sep. 2003.
- [31] D. Dong et al., "Spatiotemporal filtering using principal component analysis and Karhunen–Loeve expansion approaches for regional GPS network analysis," *J. Geophys. Res., Solid Earth*, vol. 111, no. B3, Mar. 2006, Art. no. B03405.
- [32] Y. Niu, P. Rebischung, M. Li, N. Wei, C. Shi, and Z. Altamimi, "Temporal spectrum of spatial correlations between GNSS station position time series," *J. Geodesy*, vol. 97, no. 2, p. 12, Feb. 2023.
- [33] S. Wdowinski, Y. Bock, J. Zhang, P. Fang, and J. Genrich, "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., Solid Earth*, vol. 102, no. B8, pp. 18057–18070, Aug. 1997.
- [34] M. A. King and C. S. Watson, "Long GPS coordinate time series: Multipath and geometry effects," *J. Geophys. Res.*, vol. 115, no. B4, 2010, Art. no. B04403.
- [35] K. Zheng, X. Zhang, J. Sang, Y. Zhao, G. Wen, and F. Guo, "Common-mode error and multipath mitigation for subdaily crustal deformation monitoring with high-rate GPS observations," *GPS Solutions*, vol. 25, no. 2, pp. 1–15, Apr. 2021.
- [36] W. Zhou, K. Ding, P. Liu, G. Lan, and Z. Ming, "Spatiotemporal filtering for continuous GPS coordinate time series in mainland China by using independent component analysis," *Remote Sens.*, vol. 14, no. 12, p. 2904, Jun. 2022.
- [37] Z. Zhu, X. Zhou, L. Deng, K. Wang, and B. Zhou, "Quantitative analysis of geophysical sources of common mode component in CMONOC GPS coordinate time series," *Adv. Space Res.*, vol. 60, no. 12, pp. 2896–2909, Dec. 2017.
- [38] W. Tan, J. Chen, D. Dong, W. Qu, and X. Xu, "Analysis of the potential contributors to common mode error in chuanbian region of China," *Remote Sens.*, vol. 12, no. 5, p. 751, Feb. 2020.
- [39] A. Klos, H. Dobslaw, R. Dill, and J. Bogusz, "Identifying the sensitivity of GPS to non-tidal loadings at various time resolutions: Examining vertical displacements from continental Eurasia," *GPS Solutions*, vol. 25, no. 3, pp. 1–17, Jul. 2021.
- [40] P. Yuan, W. Jiang, K. Wang, and N. Sneeuw, "Effects of spatiotemporal filtering on the periodic signals and noise in the GPS position time series of the crustal movement observation network of China," *Remote Sens.*, vol. 10, no. 9, p. 1472, Sep. 2018.
- [41] C. Kreemer and G. Blewitt, "Robust estimation of spatially varying common-mode components in GPS time-series," *J. Geodesy*, vol. 95, no. 1, pp. 1–19, Jan. 2021.
- [42] B. Liu, M. King, and W. Dai, "Common mode error in Antarctic GPS coordinate time-series on its effect on bedrock-uplift estimates," *Geophys. J. Int.*, vol. 214, no. 3, pp. 1652–1664, Sep. 2018.
- [43] W. Jiang, J. Ma, Z. Li, X. Zhou, and B. Zhou, "Effect of removing the common mode errors on linear regression analysis of noise amplitudes in position time series of a regional GPS network a case study of GPS stations in Southern California," *Adv. Space Res.*, vol. 61, no. 10, pp. 2521–2530, May 2018.
- [44] M. Zhou, J. Guo, X. Liu, Y. Shen, and C. Zhao, "Crustal movement derived by GNSS technique considering common mode error with MSSA," *Adv. Space Res.*, vol. 66, no. 8, pp. 1819–1828, Oct. 2020.
- [45] Y. Shen, W. Li, G. Xu, and B. Li, "Spatiotemporal filtering of regional GNSS network's position time series with missing data using principle component analysis," *J. Geodesy*, vol. 88, no. 1, pp. 1–12, Jan. 2014.
- [46] F. Ming, Y. Yang, A. Zeng, and B. Zhao, "Spatiotemporal filtering for regional GPS network in China using independent component analysis," *J. Geodesy*, vol. 91, no. 4, pp. 419–440, Apr. 2017.
- [47] K. D. Smith et al., "Evidence for deep magma injection beneath Lake Tahoe, Nevada-California," *Science*, vol. 305, no. 5688, pp. 1277–1280, Aug. 2004.
- [48] S. Wdowinski et al., "GPS measurements of current crustal movements along the Dead Sea fault," *J. Geophys. Res., Solid Earth*, vol. 109, no. B5, May 2004, Art. no. B05403.
- [49] Y. Tian and Z. Shen, "Correlation weighted stacking filtering of common-mode component in GPS observation network," *Acta Seismol. Sin.*, vol. 33, no. 2, pp. 198–208, 2011.
- [50] E. Serpelloni, C. Faccenna, G. Spada, D. Dong, and S. D. P. Williams, "Vertical GPS ground motion rates in the Euro-Mediterranean region: New evidence of velocity gradients at different spatial scales along the Nubia-Eurasia plate boundary," *J. Geophys. Res., Solid Earth*, vol. 118, no. 11, pp. 6003–6024, Nov. 2013.
- [51] Y. Li, C. Xu, L. Yi, and R. Fang, "A data-driven approach for denoising GNSS position time series," *J. Geodesy*, vol. 92, no. 8, pp. 905–922, Aug. 2018.
- [52] M. Gruszczynska, S. Rosat, A. Klos, M. Gruszczynski, and J. Bogusz, "Multichannel singular spectrum analysis in the estimates of common environmental effects affecting GPS observations," *Pure Appl. Geophys.*, vol. 175, no. 5, pp. 1805–1822, May 2018.
- [53] Y. Shen, F. Peng, and B. Li, "Improved singular spectrum analysis for time series with missing data," *Nonlinear Processes Geophys.*, vol. 22, no. 4, pp. 371–376, Jul. 2015.
- [54] D. H. Schoellhamer, "Singular spectrum analysis for time series with missing data," *Geophys. Res. Lett.*, vol. 28, no. 16, pp. 3187–3190, Aug. 2001.
- [55] F. Wang, Y. Shen, T. Chen, Q. Chen, and W. Li, "Improved multichannel singular spectrum analysis for post-processing GRACE monthly gravity field models," *Geophys. J. Int.*, vol. 223, no. 1, pp. 825–839, May 2020.
- [56] W. Li, Y. Shen, and B. Li, "Weighted spatiotemporal filtering using principal component analysis for analyzing regional GNSS position time series," *Acta Geodaetica et Geophysica*, vol. 50, no. 4, pp. 419–436, Dec. 2015.

- [57] W. Li and Y. Shen, "The consideration of formal errors in spatiotemporal filtering using principal component analysis for regional GNSS position time series," *Remote Sens.*, vol. 10, no. 4, p. 534, Mar. 2018.
- [58] S. Dray, "On the number of principal components: A test of dimensionality based on measurements of similarity between matrices," *Comput. Statist. Data Anal.*, vol. 52, no. 4, pp. 2228–2237, Jan. 2008.
- [59] K. Pearson, "LIII. On lines and planes of closest fit to systems of points in space," *London, Edinburgh, Dublin Phil. Mag. J. Sci.*, vol. 2, no. 11, pp. 559–572, Nov. 1901.
- [60] I. T. Jolliffe and J. Cadima, "Principal component analysis: A review and recent developments," *Phil. Trans. Roy. Soc. A, Math., Phys. Eng. Sci.*, vol. 374, no. 2065, Apr. 2016, Art. no. 20150202.
- [61] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdiscipl. Rev., Comput. Statist.*, vol. 2, no. 4, pp. 433–459, Jul. 2010.
- [62] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics Intell. Lab. Syst.*, vol. 2, nos. 1–3, pp. 37–52, Aug. 1987.
- [63] V. K. Chong, H. F. Fung, and J. R. Stinchcombe, "A note on measuring natural selection on principal component scores," *Evol. Lett.*, vol. 2, no. 4, pp. 272–280, Aug. 2018.
- [64] J. Ray, Z. Altamimi, X. Collilieux, and T. van Dam, "Anomalous harmonics in the spectra of GPS position estimates," *GPS Solutions*, vol. 12, no. 1, pp. 55–64, Jan. 2008.
- [65] A. Santamaría-Gómez, M.-N. Bouin, X. Collilieux, and G. Wöppelmann, "Correlated errors in GPS position time series: Implications for velocity estimates," *J. Geophys. Res.*, vol. 116, no. B1, 2011, Art. no. B01405.
- [66] A. R. Amiri-Simkooei, "On the nature of GPS draconitic year periodic pattern in multivariate position time series," *J. Geophys. Res., Solid Earth*, vol. 118, no. 5, pp. 2500–2511, May 2013.
- [67] K. F. Tiampo, S. Mazzotti, and T. S. James, "Analysis of GPS measurements in eastern Canada using principal component analysis," *Pure Appl. Geophys.*, vol. 169, no. 8, pp. 1483–1506, Aug. 2012.
- [68] U. Kumar, B. F. Chao, and E. T. Y. Chang, "What causes the common-mode error in array GPS displacement fields: Case study for Taiwan in relation to atmospheric mass loading," *Earth Space Sci.*, vol. 7, no. 11, Nov. 2020, Art. no. e2020EA001159.
- [69] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*. Cambridge, MA, USA: MIT Press, 2022.
- [70] A. Vitti, "Sigseg: A tool for the detection of position and velocity discontinuities in geodetic time-series," *GPS Solutions*, vol. 16, no. 3, pp. 405–410, Jul. 2012.
- [71] D. C. Agnew, "The time-domain behavior of power-law noises," *Geophys. Res. Lett.*, vol. 19, no. 4, pp. 333–336, Feb. 1992.
- [72] B. Hofmann-Wellenhof, H. Lichtenegger, and E. Wasle, *GNSS—Global Navigation Satellite Systems: GPS, GLONASS, Galileo, and More*. New York, NY, USA: Springer, 2008.
- [73] W. Feng, M. Zhong, J.-M. Lemoine, R. Biancale, H.-T. Hsu, and J. Xia, "Evaluation of groundwater depletion in north China using the gravity recovery and climate experiment (GRACE) data and ground-based measurements," *Water Resour. Res.*, vol. 49, no. 4, pp. 2110–2118, Apr. 2013.
- [74] T. Feng, Y. Shen, Q. Chen, and F. Wang, "Seasonal driving sources and hydrological-induced secular trend of the vertical displacement in North China," *J. Hydrol., Regional Stud.*, vol. 41, Jun. 2022, Art. no. 101091.



Kunpu Ji received the B.S. degree in surveying engineering from the School of Geomatics Science and Technology, Nanjing Tech University, Nanjing, China, in 2017, and the M.S. degree in geodesy from the College of Surveying and Geo-Informatics, Tongji University, Shanghai, China, in 2020, where he is currently pursuing the Ph.D. degree.

His main research interests include geodetic inversion problems and geodetic time series analysis.



Yunzhong Shen received the Ph.D. degree in geophysical geodesy from the Institute of Geodesy and Geophysics, Chinese Academy of Sciences, Beijing, China, in 2000.

He is currently a Professor with the College of Surveying and Geo-Informatics, Tongji University, Shanghai, China. His research interests include geodetic data processing on satellite positioning and satellite gravimetry. His main research works include the regularized solution to the ill-conditioned inverse problem of recovering a gravitational potential model from satellite-to-satellite tracking data, and variance component and parameter estimation theory in mixed-integer models of multisystem and triple-frequency global navigation satellite system (GNSS) observations.



Qiujie Chen received the B.S. degree in surveying and mapping engineering from the Guangdong University of Technology, Guangzhou, Guangdong, China, in 2007, and the Ph.D. degree in geodesy from Tongji University, Shanghai, China, in 2016.

He is currently a Professor with the College of Surveying and Geo-Informatics, Tongji University. His research interests include a short-arc approach for satellite-based gravity field determination, frequency-dependent noise modeling, non-gravitational acceleration calibration, regularization methodologies, mascon modeling, and simulations for next-generation gravity satellites.



Tengfei Feng received the B.S. and M.S. degrees in surveying and mapping engineering from the Jiangxi University of Science and Technology, Ganzhou, Jiangxi, China, in 2016 and 2019, respectively. He is currently pursuing the Ph.D. degree with the College of Surveying and Geo-Informatics, Tongji University, Shanghai, China.

His research interests mainly include geodetic time series processing and its application on hydrology and surface deformation.