Pure and Applied Geophysics

CrossMark

A Filtering of Incomplete GNSS Position Time Series with Probabilistic Principal Component Analysis

Maciej Gruszczynski,[1] Anna Klos,[1] and Janusz Bogusz[1]

Abstract—For the first time, we introduced the probabilistic principal component analysis (pPCA) regarding the spatio-temporal filtering of Global Navigation Satellite System (GNSS) position time series to estimate and remove Common Mode Error (CME) without the interpolation of missing values. We used data from the International GNSS Service (IGS) stations which contributed to the latest International Terrestrial Reference Frame (ITRF2014). The efficiency of the proposed algorithm was tested on the simulated incomplete time series, then CME was estimated for a set of 25 stations located in Central Europe. The newly applied pPCA was compared with previously used algorithms, which showed that this method is capable of resolving the problem of proper spatio-temporal filtering of GNSS time series characterized by different observation time span. We showed, that filtering can be carried out with pPCA method when there exist two time series in the dataset having less than 100 common epoch of observations. The 1st Principal Component (PC) explained more than 36% of the total variance represented by time series residuals' (series with deterministic model removed), what compared to the other PCs variances (less than 8%) means that common signals are significant in GNSS residuals. A clear improvement in the spectral indices of the power-law noise was noticed for the Up component, which is reflected by an average shift towards white noise from -0.98 to -0.67 (30%). We observed a significant average reduction in the accuracy of stations' velocity estimated for filtered residuals by 35, 28 and 69% for the North, East, and Up components, respectively. CME series were also subjected to analysis in the context of environmental mass loading influences of the filtering results. Subtraction of the environmental loading models from GNSS residuals provides to reduction of the estimated CME variance by 20 and 65% for horizontal and vertical components, respectively.

Key words: Probabilistic principal component analysis, common mode error, GNSS, time series analysis, missing data.

[1] Faculty of Civil Engineering and Geodesy, Military University of Technology, Warsaw, Poland. E-mail: maciej.gruszczynski@wat.edu.pl

1. Motivation and introduction

The advantages of reliable coordinates provided by the globally distributed Global Navigation Satellite System (GNSS) stations have been appreciated by scientists since the early 90s. The position changes of antennae are expressed by coordinates that for decades have been continuously and regularly determined in global reference frame. The GNSS position time series provided by permanent observations have been used primarily to realize and maintain modern kinematic reference frames (Gross et al. 2009) as well as for geophysical studies as a measure of surface displacement or strain (Kreemer et al. 2014; Métivier et al. 2014). The linear trend due to plate tectonics and seasonal signals caused by i.e. environmental mass loading or thermoelastic deformation are very consistent for GNSS position time series recorded by nearby stations. Even the sophisticated time series modelling has not resulted in a total loss of spatial and temporal correlations (Wdowinski et al. 1997, Shen et al. 2013; Bogusz et al. 2015). The so-called Common Mode Error (CME) is the superposition of errors and geophysical phenomena, that similarly affect the coordinates of stations included in the regional networks. It is a spatially correlated type of error, which may be also temporally correlated depending on the temporal structure of the phenomena it absorbs. It is essential to remove, reduce or eliminate the impact of CME in the GNSS networks to improve the accuracy of the estimated velocities (He et al. 2017). The main theoretical contributors to the potential CME sources are (Wdowinski et al. 1997; King et al. 2010):

1. errors in the alignment to the reference frame;
2. errors related to satellites, which are usually observed in small networks as the mismodeling of the satellite: orbits, clocks, or antenna phase center variations;
3. signal emission media effects commonly influencing stations in regional network (troposphere and ionosphere modelling);
4. physical sources of station movements as the mismodeled (or unmodeled) large-scale atmospheric and hydrological effects, as well as small scale crust deformations;
5. errors caused by algorithms, software, and data processing strategies, including ambiguity resolution problem.

To reduce the effect of CME, a number of studies have been preceded by the filtering of the GNSS time series using different methods (Fig. 1). The Common Mode Error term was first introduced by Wdowinski et al. (1997), who described correlated errors in the regional networks of the GNSS stations. The authors used stacking approach assuming that CME is equal to an arithmetic mean of all available residuals at a specified epoch. Nikolaidis (2002) proposed to improve this method and called it a "weighted stacking", indicating that the GPS-derived coordinates with unequal formal errors cannot contribute equally to the final CME estimates. A group of methods known as "stacking" assumes the spatial uniformity of CME that has to be met over a regional
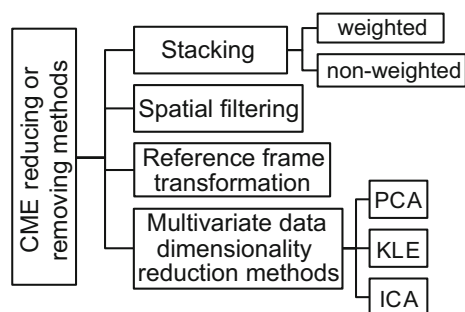

Figure 1
A diagram of various mathematical methods used so far to reduce or remove CME, including weighted and non-weighted stacking, spatial filtering, reference frame transformation and multivariate data dimensionality reduction methods such as Principal Component Analysis (PCA), Karhunen–Loeve Expansion (KLE) and Independent Component Analysis (ICA)

network. This in turn imposes the condition that the estimates of CME at a single epoch are equal for the entire set of stations. In addition, a limit in the maximum size of the GNSS network which can be used to derive the CME estimates is set up. The stacking method can be used for networks with stations as far as 600 km from each other (Wdowinski et al. 1997; Márquez-Azúa and DeMets 2003). For larger networks and stations located up to 2 000 km from each other, various spatial filters can be introduced to differentiate the spatial response of any individual station. The stacking and spatial filtering methods should not be considered as similar approaches either in the sense of the character of the input and output data or mathematical formulas. Spatial filtering means that the CME is extracted with varying spatial responses and is individually and locally fitted to each individual station position time series. Spatial filters take into consideration the length of time series and the distance between stations (Márquez-Azúa and DeMets 2003) or the inter-station distances with correlations between residuals collected for any individual station (Tian and Shen 2011, 2016). It has been shown that CME can be also reduced using a 7-parameter (or 14-parameter) similarity transformation (Ji and Herring 2011; Blewitt et al. 2013). The limitations of aforementioned methods are related to: (1) a limitation of maximal area of network which can be subjected to filtering (mainly in case of stacking), (2) a partial inability to detect stations with strong local effects which will affect the CME estimates or (3) dependence on the conventional weighting procedure. Spatial filters and stacking methods significantly reduce the scattering of the position time series and, as a result, as soon as CME is removed, they improve the precision of velocity estimates. As an example, Wdowinski et al. (1997) reduced the time series root-mean-square (RMS) values for individual stations by more than 50%. Tian and Shen (2016) minimized the RMS values by 20.7, 13.2, and 14.4% on average for the North, East, and Up components, respectively. Despite the fact that the reduction in the scatter of the position time series is important and desirable, it should not be the only main indicator. The preferred filtering method should be able to accurately divide the GNSS residuals into two modes: the spatially

correlated part (CME) and spatially uncorrelated mode (noise). The unwanted effect of smoothing of the GNSS position time series appears when signals that do not correspond to real common effects are removed. It leads to undesired change in the long-term trend (station velocity) and also in the accuracy of this trend being misestimated.

Bearing in mind all of the aforementioned issues, many techniques which reduce the dimensionality of multivariate data have already been implemented to improve the CME filtering. Dong et al. 2006 and Serpelloni et al. 2013 proved, that the Empirical Orthogonal Function (EOF) decomposition provides a more solid numerical framework for the separation of modes than the stacking approach. In addition, this does not assume spatial uniformity of CME as stacking does, but instead employs a uniform temporal function which affects stations across regional network. Dong et al. (2006) were the first to apply the Principal Component Analysis (PCA) and Karhunen–Loeve Expansion (KLE) methods for CME extraction. They are based on different assumptions concerning the construction of the orthonormal vector basis. The former uses the covariance matrix of observations, while the latter applies the correlation matrix of observations. With regards to the fact that the traditional PCA can be applied only for complete data, Shen et al. (2013) proposed the use of a modified PCA (mPCA) to filter the position time series with missing data, which are reproduced from Principal Components (PCs). The PCA approach was further extended by Li et al. (2015), who introduced weighted spatio-temporal filtering. Similarly to weighted stacking, weighted PCA (wPCA) was proposed taking into consideration the individual errors of coordinates. This weighting procedure may cause an unwanted situation when time series with a weak CME response may significantly affect CME value. This may occur, when coordinates from stations affected by strong local effects e.g. local hydrology-induced or station-specific movements are determined by small standard errors. According to earlier publications, the weighting based on errors of observations does not refer to the nature of CME's. The advantages of EOF's for CME filtration have recently been confirmed by Gruszczynski et al. (2016), who showed significant improvement in the accuracy of stations velocities.

The main purpose of this research is to introduce for the first time a probabilistic PCA (pPCA) method for spatio-temporal filtering of GNSS position time series, and to employ it to filter the position time series whilst leaving the missing values without interpolation (Fig. 2a). Although pPCA has previously been employed in various areas of research, e.g. estimation of the EOF's for satellite-derived sea surface temperature (SST) data (Houseago-Stokes and Challenor 2004), a study on the precipitation and absorption squeeze (Andrei and Malandrino 2003), generation of the video textures (Fan and Bouguila 2009), detection of a small target (Cao et al. 2008), investigation of traffic flow volume (Qu et al. 2009), managing self-organizing maps (Lopez-Rubio et al. 2009), detection of outliers (Chen et al. 2009), tracking of the objects (Xiang et al. 2012), speaker recognition (Madikeri 2014), investigation of the nonlinear distributed parameter processes (Qi et al. 2012), nonlinear sensor fault diagnosis (Sharifi and Langari 2017), studying trends of mean temperatures and warm extremes (Moron et al. 2016), denoising of images (Mredhula and Dorairangaswamy 2016) or detection of the rolling element bearing fault (Xiang et al. 2015), according to the best of our knowledge, the pPCA filtering that is readily adapted to the position time series with missing data, has been presented for GNSS residuals (either position or ZTD) for the first time.

We present the pPCA as an alternative approach to spatio-temporal filtering PCA methods proposed by Dong et al. (2006) and by Shen et al. (2013), which will later be referred to as an iterative PCA (iPCA) and modified PCA (mPCA), respectively. Both methods are based on PCA algorithm and characterized by conventional approximates which modify standard PCA to deal with discontinuous time series. In mPCA approach it is assumed that the covariance matrix is initially constructed using all available time series. Gaps are then interpolated by minimizing the weighted quadratic norm of PC unknowns. In iPCA approach it is assumed that residuals can initially be spatially averaged, which means that any missing epoch may be completed using values from other stations that do not have gaps in this specified epoch. However, a problem occurs when there is a gap in the dataset which starts and

Real network stations coordinate availability

(a)

Incidentally occuring data time spans of GNSS time series

(b)   **(c)**   **(d)**
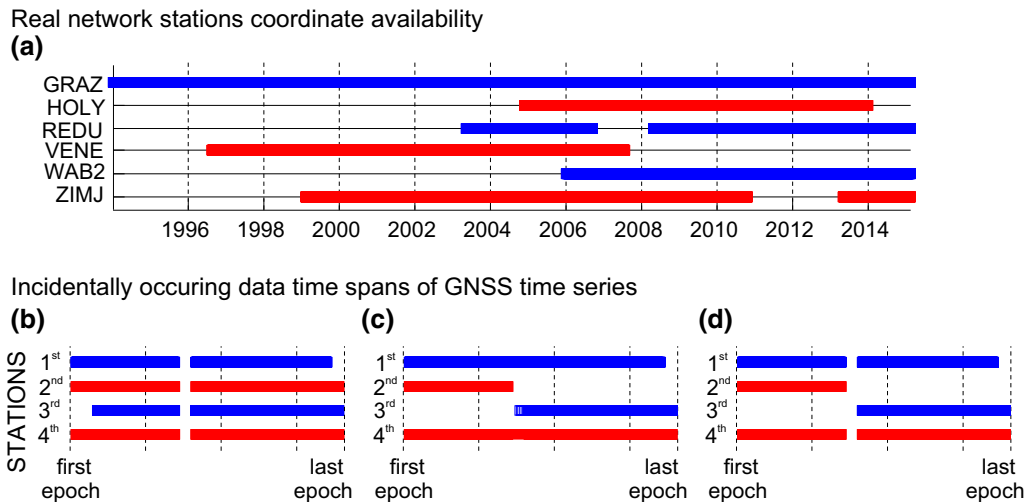

Figure 2
a Availability of observations for six selected stations distributed by the International GNSS Service (IGS); **b–d** graphs show simulated but frequently observed data gaps which may cause the iPCA method (**b**) or mPCA method (**c**) or both of them (**d**) unfunctional

ends at the same time almost for all stations in the network (Fig. 2b). In such a case, there are two options. First, the missing epochs from all series can be deleted, however, some amount of data containing important information for further analysis is removed. Second, during the first stage of interpolation real dependencies in GNSS residuals may be neglected due to the fact, that initial values adopted without a reliable probabilistic model can significantly influence further estimates. The mPCA method fails when any two time series of a network do not have any, or have only a few common epochs of observations (Fig. 2c). In this case, the covariance matrix cannot be set.

Figure 2d shows a theoretical time span of residuals subjected to filtering, where neither iPCA nor mPCA is able to perform orthogonal transformation since a certain gap is present in all data or two series do not have a single common observation. Unlike iPCA and mPCA, the pPCA method which we have introduced in this research, takes into account the probabilistic framework to determine the optimal model for the missing data. Since in pPCA the missing values are considered as latent variables, it is possible to filter even the series shown in Fig. 2d.

In this research, we applied the pPCA method to resolve the problem of a proper spatio-temporal filtering of GNSS position time series when gaps occur at the same time in the regional network and the series do not necessarily have the same observation time span. This method is presented as an alternative to the classic PCA approach and its modifications: mPCA and iPCA. The paper is organized as follows. We start with a set of complete data with a changing amount of simulated gaps to prove the effectiveness of the approach that we employed. Then, we continue with a set of 25 permanent GNSS stations which were included in the latest realization of the International Terrestrial Reference System (ITRS). At the end, we present hard numbers demonstrating the importance of spatio-temporal filtering before uncertainty of linear velocity being determined. It is worth mentioning, that the methodology presented in this research, although applied to the GNSS position time series, is universal and can be successfully adapted to data having spatial relationships gathered by GNSS, as e.g. ZTD (Zenith Total Delay), or any other geodetic instruments such as GRACE (Gravity Recovery and Climate Experiment) or altimetric satellites.

2. Probabilistic Principal Component Analysis

For the network formed by n GNSS stations with a time series spanning m days, before we attempt the spatio-temporal filtering, we are obliged to construct the observation matrix $\mathbf{R}(t_i, r_j)$ ($i = 1, 2,\ldots,m$ and $j = 1, 2,\ldots,n$) for each topocentric component (North, East or Up) separately. Residual time series $\mathbf{r}(t)$ constitute the matrix in such a way that each row corresponds to the epoch of observation t_i, while columns represent each subsequent GNSS coordinate time series r_j from the GNSS stations. To introduce the pPCA procedure, we firstly present the most common derivation of PCA of the matrix $\mathbf{R}$ through eigenvalue decomposition. At this stage, the time series are assumed to be complete. The 4-step basics are given as (Jolliffe 2002):

Step 1: computation of the mean-centered matrix $\mathbf{R}_c$ by subtracting the vector of means of all columns and from each row of $\mathbf{R}$,

Step 2: computation of the covariance matrix $\mathbf{C} \cdot v = \mathbf{R}'_c \cdot \mathbf{R}_c$ which is of n per n-dimension matrix,

Step 3: computation of the eigenvalue decomposition of $\mathbf{C} \cdot v$ given by $\mathbf{C} \cdot v = \mathbf{V} \cdot \mathbf{\Lambda} \cdot \mathbf{V}^{-1}$, where $\mathbf{\Lambda}$ is a matrix with k non-zero diagonal eigenvalues of the covariance matrix and $\mathbf{V}$ is the n per n-dimension matrix with the corresponding eigenvectors in individual columns. The number of eigenvalues may be less than or equal to the number of time series ($n \geq k$), but in most cases with real data, the matrix $\mathbf{C} \cdot v$ is usually of full rank and the number of eigenvectors is equal to the number of the time series ($n = k$),

Step 4: sorting of the eigenvectors and corresponding eigenvalues in a decreasing order. The eigenvalues represent the contribution of each Principal Component mode in the total variance of data. Those principal components are estimated as:

$$a_k(t_i) = \sum_{j=1}^{n} \mathbf{R}(t_i, r_j) \cdot \mathbf{v}_k(r_j) \qquad (1)$$

where $a_k(t_i)$ is the k-th PC of matrix $\mathbf{R}$ and $\mathbf{v}_k(r_j)$ is its corresponding eigenvector (a matching column adapted from $\mathbf{V}$).

A standard PCA approach is applicable only to the complete datasets and any attempt to use this method for data with missing values must be preceded by deleting the rows with missing data, interpolating or modifying PCA algorithm (Ilin and Raiko 2010; Zuccolotto 2012). Real geodetic data are susceptible to incompletion. Since coordinate time series are arranged in the observation matrix by time, any time series that starts later or ends earlier than other stations are also considered as missing. Furthermore, the hardware or software failure or replacement, physical disturbance, data loss or removal of outliers at the pre-analysis contributes to gaps in the data.

We employed a more complex procedure for eigenvalue decomposition in case the data matrix being incomplete. Probabilistic PCA presented here is based on the Expectation–Maximization (EM) algorithm (Roweis 1997; Tipping and Bishop 1999). The regularized EM algorithm has been recently used to interpolate missing values before traditional PCA and ICA were performed for the Chinese regional GNSS network (Ming et al. 2017). In contrast to an interpolation of incomplete time series, the EM algorithm employed in pPCA handles missing values by considering them as additional latent variables. Products of pPCA-based filtering can be interpreted in the same way as results from the traditional PCA, however, the pPCA method stands out by application of a flexible statistical model.

The probabilistic PCA is based on the following latent variable model:

$$\mathbf{t} = \mathbf{W} \cdot \mathbf{x} + \mathbf{\mu} + \mathbf{\varepsilon} \qquad (2)$$

where: $\mathbf{t}$ is a n-dimensional observation vector, $\mathbf{x}$ is a q-dimensional vector of latent variables, $\mathbf{W}$ is a n per q-dimensional transformation matrix, $\mathbf{\mu}$ is the vector mean of $\mathbf{t}$, $\mathbf{\varepsilon}$ is a noise model which compensates for the errors.

In case of filtering of GNSS-derived position residuals, $\mathbf{t}$ can be identified with time series of all available residuals at given epochs, while $\mathbf{x}$ are residuals that are not directly estimated in dataset, e.g. due to the lack of coordinates or as an effect of outliers removal. According to pPCA theorem missing values are rather inferred from other residuals that really exist in time series via the assumption of a spatially correlated CME. $\mathbf{W}$ is the matrix whose columns are composed of the scaled eigenvectors of

sample covariance matrix of residuals, which are necessary to estimate CME.

There is no closed-form analytical solution for **W**, which is the reason for the need to apply an iterative EM algorithm. In this research, we adopted the pPCA theorem proved by Tipping and Bishop (1999), who gave the analytic solution for the model showed in Eq. (2), EM algorithm description and a full characterization of its properties. The Maximum Likelihood (ML) solution for pPCA latent variable model is given by:

$$\mathbf{W}_{\mathrm{ML}} = \mathbf{V}_q \left(\mathbf{\Lambda}_q - \sigma^2 \cdot \mathbf{I}\right)^{1/2} \mathbf{B} \qquad (3)$$

where: **B** is an arbitrary q per q-dimension orthogonal rotation matrix, **I** is an identity matrix, σ^2 is an isotropic variance.

Each of the columns of matrix $\mathbf{V}_q$ (n per q-dimension) is the principal eigenvector of sample covariance matrix of the GNSS residuals, with corresponding eigenvalue in the q per q-dimension diagonal matrix $\mathbf{\Lambda}_q$. Since one of the most important steps of each PCA-based procedure is the decomposition of the covariance matrix into the matrix with eigenvalues and matrix with corresponding eigenvectors, this in case of pPCA the maximization of the likelihood function (Eq. 3) by using EM algorithm is a key issue to obtain the most probable elements of $\mathbf{V}_q$ and $\mathbf{\Lambda}_q$ matrixes. It results in the calculation of principal eigenvectors and eigenvalues necessarily for reliable CME estimation (Eq. 4).

The EM algorithm consists of two main steps: the E-expectation and M-maximization. The parameters of the model given in the Eq. (3) are resolved with the Maximum Likelihood Estimation (MLE) in an iterative manner (Tipping and Bishop 1999) by 3-step procedure:

Step 1 (E-step): calculation of the expected value of the log-likelihood function, given the considered data and the current estimates of the model parameters,

Step 2 (M-step): finding the new parameters by maximizing the log likelihood function using the expected parameters derived in the E-step,

Step 3: repeating Steps 1 and 2 until convergence. For our purposes the convergence criteria was set up as a relative change in the transformation matrix elements less than 10^{-4}.

Using the EM algorithm for finding the principal axes by iteratively maximizing the likelihood function (Eq. 3), the latent variable model defined by Eq. (2) affects mapping from the latent space into principal subspace of the observed data.

One of the most important features related to pPCA method is the fact that the q-number of EOFs to retain, can be specified at the very outset. The reason for limitation of this parameter is the fact that, in case of small value of q in relation to high value of n (number of dimensions—in our case number of GNSS stations) the dimension of **W** transformation matrix is much smaller than the covariance matrix for traditional EOF analysis. This makes pPCA method to be computationally much more efficient and less burdensome for computers. Many papers have focused on the issue of determining the optimum q number of retained EOFs prior to using EM algorithm (e.g. Jolliffe 1972; Houseago-Stokes and Challenor 2004). However, there is no satisfactory and versatile rule. In this research, at the pre-processing stage based on our dataset, we computed the maximum number of principal components which can be retained from pPCA. We found that only the first PC is significant when deterministic model was subtracted prior the pPCA analysis (please see data and methods described in section "GNSS time series"), which is consistent with the considerations of other authors (Dong et al. 2006; Shen et al. 2013). We adopted $q = 3$ value to allow for more variance to be retained. Furthermore, some aspects of computational as well as communication complexity of PCA-based methods were the subject of many analyses (e.g. Roweis 1997; Houseago-Stokes and Challenor 2004; Ilin and Raiko 2010) with leading conclusion that the probabilistic PCA is the most promising PCA approach, especially for large datasets.

Another important advantage of the pPCA method is the ability to interpret its products in the same manner as results of traditional PCA. This allows to adopt the definition and applications of CME estimates (Dong et al. 2006):

$$\mathbf{CME}_j(t_i) = \sum_{k=1}^{p} \mathbf{a}_k(t_i) \cdot \mathbf{v}_k(r_j) \qquad (4)$$

where p is a number of first significant PCs. Following Shen et al. (2013) and Tiampo et al. (2004), we used the Fisher-Snedecor test (Fisher, 1932) for the equality of two variances to decide on the number of significant PCs. CME is removed from the unfiltered residuals $\mathbf{r}(t)$ using arithmetic subtraction, thus obtaining so-called "filtered" residuals. The first few PCs (or just the first PC in some cases) reflect a common source function which affects the regional GNSS network, i.e. CME, and represents the highest contribution to the variance of the GNSS-derived position residuals (Dong et al. 2006).

3. GNSS Time Series

In this research, we employed the daily-sampled GNSS position time series that were produced by the International GNSS Service (Dow et al. 2009) by a network solution referred to as "repro2" to estimate coordinates (Rebischung et al. 2016). Each of the selected stations contributed to the newest realization of the International Terrestrial Reference System (namely ITRF2014, Altamimi et al. 2016). Since it is imperative to investigate the response of the newly adopted pPCA method to the number of missing data, we were obliged to find stations with almost complete time series for reference purposes. Since the distance between stations from the selected network is quite important for this research, a set of 25 stations located in Central Europe were chosen (Fig. 3). The distance between any two stations taken for analysis is shorter than 1870 km which is consistent with the overall assumptions related to applicability of PCA-based filtering methods. Márquez-Azúa and DeMets (2003) found that the spatial correlations of the residual time series are high within 1000 km distance and they gradually decrease to zero for c.a. 6000 km. Similar studies were performed for stations distant at the 10^3 km level, located in China (Li et al. 2015; Shen et al. 2013) or in Australia (Jiang and Zhou 2015).

For the purpose of this study, we used a time span of 2003.2–2015.0 (Fig. 4), when all selected stations were operating. The length of the GNSS coordinate time series is very important whilst the station velocity is expected to be determined with high
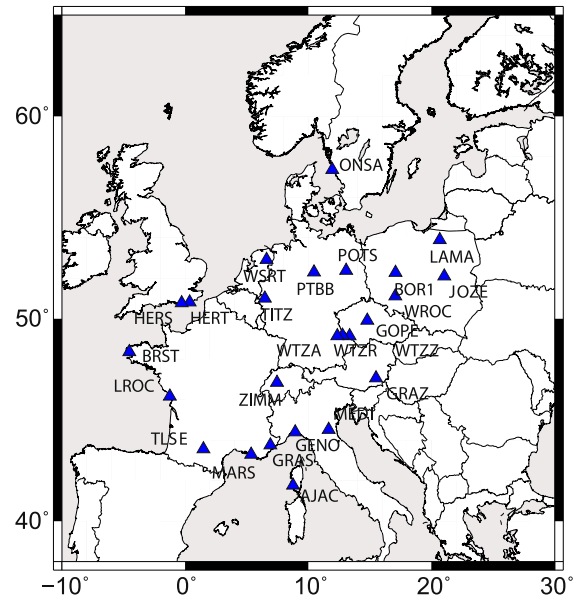

Figure 3
Geographical distribution of 25 stations employed in the analysis

reliability. The data time span commonly assumed by other authors (e.g. Blewitt and Lavallée 2002) as a minimum for reliable velocity estimation is 3 years. However, lately, Klos et al. (2018) argued that this span should be extended to 9 years. Time-dependent improvements in the consistency of GNSS-derived position residuals are explained by avoiding errors caused by mismodeling of seasonal signals and noises. Every incorrectly estimated seasonal signal for regional network of GNSS stations may increase spatially-correlated errors. As a result of studies performed both on GNSS time series and models of surface mass loading deformation, Santamaría-Gómez and Mémin (2015), found that at least 4 years of continuous data is necessary to meet requirements of the accurate modelling of the inter-annual deformations as a step towards reliable estimation of secular velocities. In case of time series included in this analysis, 12.2 years long data time span exceeds assumed minimal levels.

We used epochs of offsets compiled by IERS (International Earth Rotation and Reference System Service) ITRS Product Centre and available at http://itrf.ensg.ign.fr/ITRF_solutions/2014/doc/ITRF2014-soln-gnss.snx to eliminate the influence of discontinuities on final estimates. Any value was considered
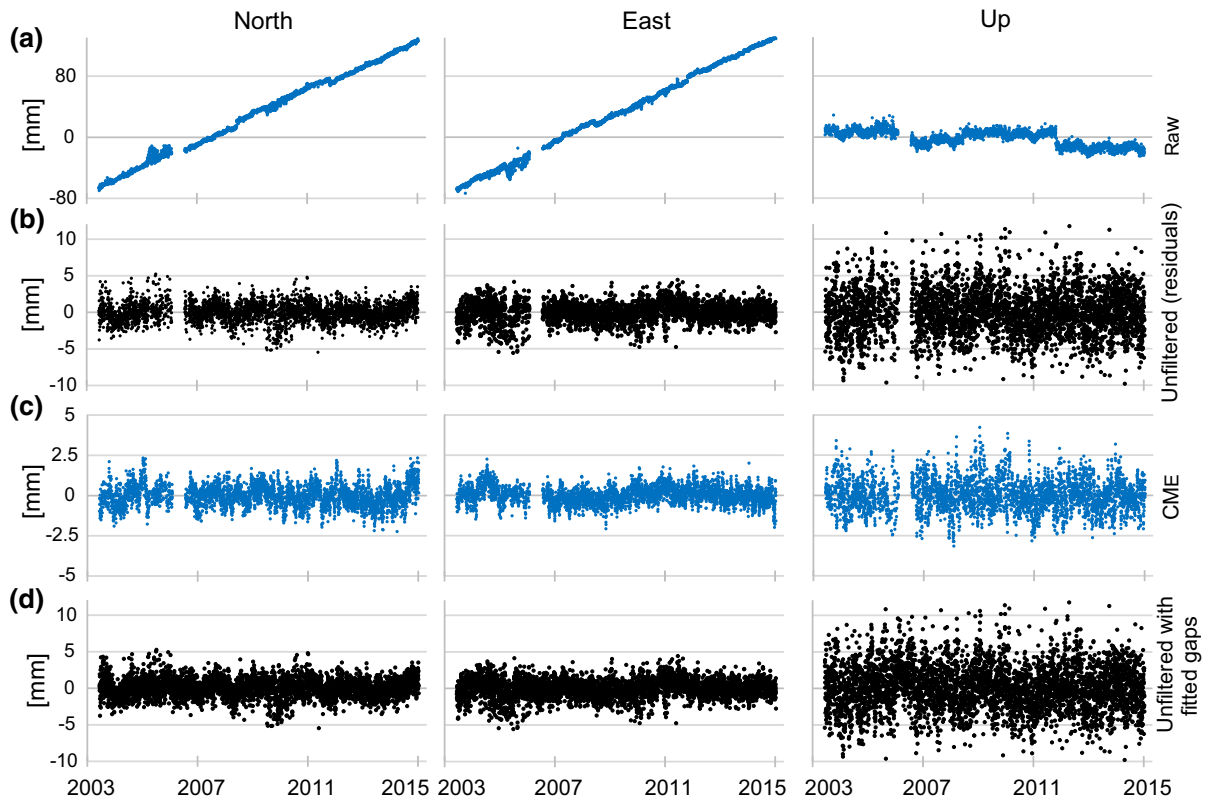
Figure 4

Exemplary GNSS coordinate time series for station BRST (Brest, France). **a** "raw" GNSS coordinate time series. Trend, offsets and seasonal signals can be noticed. **b** "unfiltered" time series subjected to filtering analysis, obtained when model from Eq. (5) was removed. **c** CME estimates based on the pPCA method performed for incomplete time series. **d** Residuals of BRST with gaps being filled. The different scale of vertical axes for (**a**–**d**) should be noticed

as an outlier, when it fell outside 3 times the interquartile range (IQR) below or above the median (Langbein and Bock 2004).

The time series were characterized by 3.8% of gaps on average. Time series form station ONSA (Onsala, Sweden) were the most complete with only 0.5% of missing data, while station BRST (Brest, France) had the greatest amount of missing data: 13.2%.

We modelled each of the topocentric (North, East or Up) position time series (Fig. 4a) $\mathbf{x}(t)$ with a mathematical function that takes the form of:

$$\mathbf{x}(t) = x_0 + v_x \cdot t + \sum_{i=1}^{j} [A_i \cdot \sin(\omega_i \cdot t + \phi_i) + \mathbf{r}(t)]$$

$$(5)$$

where: x_0 represents the initial coordinate at the reference epoch, v_x refers to the linear velocity, $\sum_{i=1}^{j} [A_i \cdot \sin(\omega_i \cdot t + \phi_i)]$ accounts for periodic signals with angular velocities ω_i, $\mathbf{r}(t) = \mathbf{CME}(t) + \varepsilon(t)$ are the residuals being a sum of spatially- and temporally-correlated CME and temporally-correlated noise.

In the following research, the parameters of the deterministic part were estimated with the Maximum Likelihood Estimation (MLE) method according to approach to the deterministic part given by Bogusz and Klos (2016). Unlike the vast majority of described modelling approaches found in literature (Dong et al. 2006; Shen et al. 2013), where only annual and semiannual signals were used, our deterministic model assumes different periodicities: fortnightly,

Chandlerian, tropical and draconitic (see Bogusz and Klos 2016).

CME contains some part of a flicker noise (Klos et al. 2016) with spectral index of -1, which was found to be mostly present in the GNSS position time series (Williams et al. 2004; Bos et al. 2008). The residual time series $\mathbf{r}(t)$ obtained after a deterministic model was subtracted (Fig. 4b), are subjected to further analysis and will be referred to later in this paper as the "unfiltered" time series.

Despite the fact that the time series were detrended and seasonal signal were removed, we still notice the non-zero correlations between residuals $\mathbf{r}(t)$. For the purpose of proving correlation level, we performed two types of tests. Initially, to quantify the level of correlations, we used Lin's concordance correlation coefficient (Lin 1989), as it was recommended for GNSS residuals by Tian and Shen (2016). Lin's concordance coefficient, was introduced to provide a measure of reliability that is based on covariation and correspondence in contrast to commonly used Pearson correlation coefficient, which in turn is a measure of linear covariation between two sets of scores. Differences can be comprehended by geometric interpretation, where two time series are plotted in one scatterplot and best fitted line is imposed. Pearson correlation coefficient specifies how far from the line are data points, whilst Lin's concordance coefficient additionally taking into account the distance to the 45-degree line, which represents perfect agreement. Concordance correlation coefficient ranges from 0 to ± 1 and its interpretation is close to other correlation coefficients, which means, that values near 1 mean perfect concordance and 0 means no correlation. For unfiltered residuals used in this analysis we found that the average Lin's concordance correlation coefficient is equal to median values of 0.39 for horizontal and 0.54 for vertical components, respectively, when the distance between individual stations is less than 500 km (see the red dots in Fig. 11). Correlations are smaller for more distant stations, which is mostly evident for the Up component. Then, we employed the Kaiser–Meyer–Olkin (KMO) index to assess whether we are able to use multivariate analysis to efficiently extract the common signals from a set of stations we employed (Cerny and Kaiser 1977):

$$\text{KMO} = \frac{\sum_j \sum_{k \neq j} \rho_{jk}^2}{\sum_j \sum_{k \neq j} \rho_{jk}^2 + \sum_j \sum_{k \neq j} \hat{\rho}_{jk}^2} \qquad (6)$$

where: $\hat{\rho}_{jk}$ represents a partial correlation, and ρ_{jk} is a correlation coefficient between variables j and k which mean time series from jth and kth stations. Partial correlation is the measure of association between two time series, while controlling or adjusting the effect of one or more additional time series.

The KMO index is a value that describes dataset applied to dimensionality reduction techniques (e.g. pPCA). This index measures the proportion of common variance among the all variables. By definition, the KMO index ranges between 0 and 1. Values close to 1 mean that common signals have a significant variance. For the observation matrix from the real unfiltered residuals of 25 GNSS stations we obtained KMO indices equal to 0.961, 0.966 and 0.988 for the North, East and Up components, respectively.

4. pPCA Filtering of Artificially Incomplete Time Series

In this part of the research, we analyzed and compared iPCA, mPCA and pPCA methods with traditional pre-interpolated PCA approach for the spatio-temporal filtering of the GNSS-derived position time series. Missing values were introduced to real GNSS position time series to simulate the number of gaps we might expect in the observations. In this way, we assessed the ability of each method to deal with incomplete time series and its sensitivity on the number of missing values.

The artificially incomplete residuals were produced in the following manner. First, we used the GNSS position residuals from a set of 25 stations presented in Fig. 3. We fully interpolated them, assuming adequate values of mean and standard deviation of inputted points in such a way that interpolation procedure did not change the variance of the time series. With these assumptions, we obtained time series that imitated 25 unfiltered, fully complete GNSS residuals. An example is shown in Fig. 4d. Then, we randomly chose epochs and removed observations to simulate incomplete data.

We introduced gaps with length from 5 to 40% of the total length of the series with 5% increment. We assumed that the gaps were missing at random (MAR, Little and Rubin 2002) and that the number of stations subjected to introduced gaps is approximately equal to the number of time series which remain complete. Therefore, data gaps were introduced to 13 randomly chosen stations from a set of 25. For the remaining 12 stations, no data were deleted simulating time series as being complete. We were tempted to accept this procedure by two issues. First, to investigate what is the impact of missing data on CME estimates. Second, how much the CME computed for complete time series is affected by values which are missing on other stations.

In case of traditional PCA with interpolated gaps we assumed a white noise model to simulate scatter of residuals to show, how the simple assumption of linear interpolation may bias the CME estimates. The response of each method was then analyzed according to the increased number of missing values. The relative errors of CME were computed based on the unfiltered residuals subjected to introduced gaps and compared to CME estimated for the complete unfiltered residuals. Complete time series were treated as a reference dataset for spatio-temporal filtering. It is worth noting that CME is identical for all methods where time series are complete. The relative error of CME determination was computed as (Shen et al. 2013):

$$\sigma_{CME} = \pm \sqrt{\frac{(\mathbf{CME}_i - \mathbf{CME}_0)^T (\mathbf{CME}_i - \mathbf{CME}_0)}{\mathbf{CME}_i^T - \mathbf{CME}_0)}} \cdot 100$$

(7)

where $\mathbf{CME}_0$ and $\mathbf{CME}_i$ are the vectors of Common Mode Errors computed before data were deleted and after data gaps were interpolated, respectively. When gaps were introduced, we randomly chose stations and epochs to be deleted 100 times and we averaged results of simulations. Figure 5 presents the relative errors for stations where data were deleted, while Fig. 6 includes relative errors of CME for stations where no data were removed. When the interpolation of GNSS-based position residuals has been applied before PCA-based filtering, the CME values were biased reaching values of relative error equal to 10
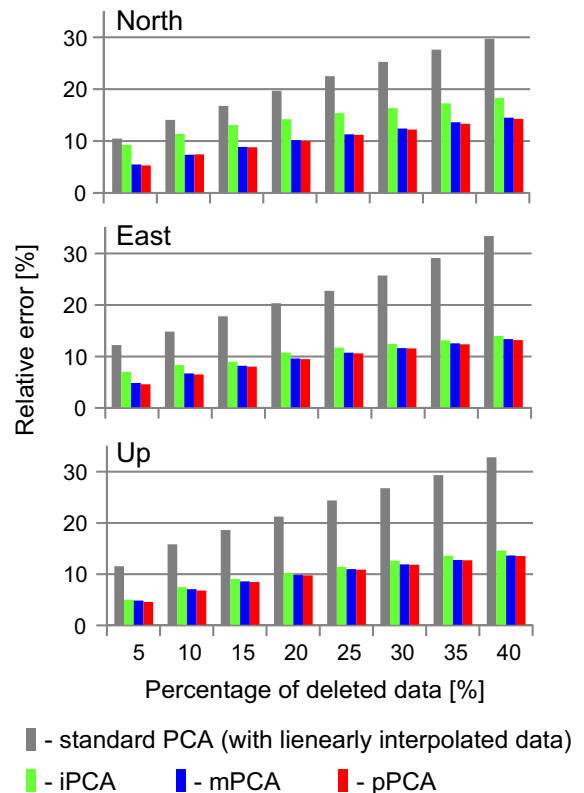

Figure 5

Relative errors of CME (σ_{CME}) where different percentages of data were randomly deleted. CME was computed with the use of pPCA (red bars), iPCA (green), mPCA (blue) and traditional PCA with gaps interpolated (grey). Relative errors relate to the subset of 13 stations that were affected by introduced gaps. The results were averaged for 100 different simulations in every step we performed

and 35%, when 5 and 40% of data were deleted, respectively. The relative errors of CME estimates are similar both for mPCA, iPCA and for pPCA which allows us to conclude that our method can constitute an alternative approach to both methods already mentioned. Only in a few cases the pPCA performance in CME estimation is slightly better than mPCA, but the difference between both of them is not significant and reaches the maximum of 0.1%.

The relative errors of CMEs estimated with pPCA ranged between 5 and 14% for the entire set of stations. The reconstruction of CME computed on the basis of incomplete data is slightly better for the vertical component for which the relative errors of CME are less by 1–3%. This fact may be explained by the higher correlation, which allows to determine

- standard PCA (with lienearly interpolated data)
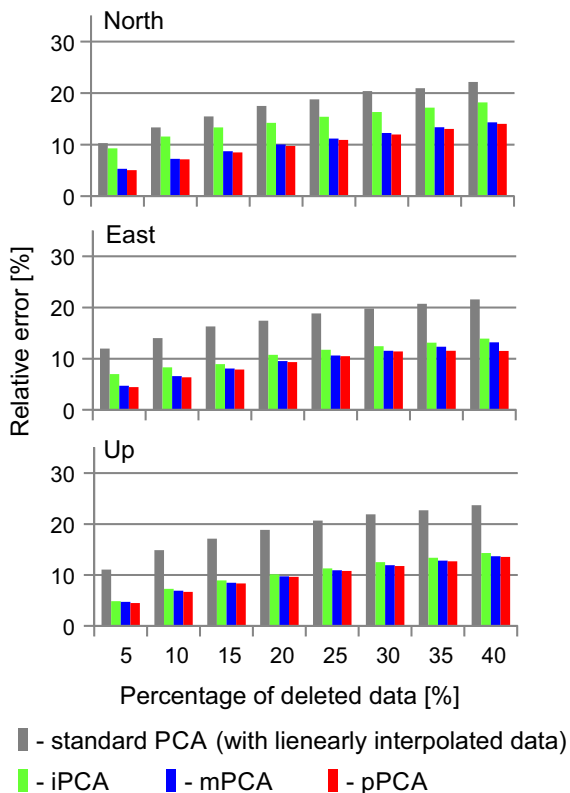
- iPCA   - mPCA   - pPCA

Figure 6

The same as Fig. 5 but the relative errors are computed for the set of 12 stations with complete data for which no values were removed. However, although no data was deleted, the results were biased due to missing data in the remaining 13 time series shown in Fig. 5. The results were averaged for 100 different simulations in every step we performed

more reliable parameters of latent variable model in pPCA. Another important feature, which can be seen in Figs. 5 and 6 is the fact, that compared to mPCA and pPCA, the iPCA method performs worse in the North component than in the East and Up components. Taking into consideration only North component, the differences in relative error of CME estimated with the use of iPCA and pPCA methods are about 4%. It may be due to inhomogeneous spatial response of individual stations to the CME source, which is presented in Fig. 7 and described in the next section.

Similarly to iPCA and mPCA, the relative error of CME reconstructed with the use of pPCA is always less than 14%, even in cases when 40% of residuals were deleted from the 13 stations selected out of 25 (Fig. 5). In the standard PCA approach with

interpolation, the larger the number of missing values, the higher the relative error of CME, rising to 33%. In addition, the CME values were biased for 12 stations, where no data were removed. All relative errors of CME presented in Fig. 6 are non-zero. Despite the fact that each time series derived from 12 stations was not subjected to a deleting procedure, CME estimates were also incorrectly calculated due to the missing values in the remaining time series.

Then, we ran imitation of missing values to simulate a specific case when GNSS coordinate time series have just few common epochs [see stations VENE (Venezia, Italy) and WAB2 (Bern, Switzerland) in Fig. 2a]. As it was mentioned before, every inconsistency in the first and last observed epoch for GNSS time series has to be treated during spatio-temporal filtering as an incompleteness. We simulated a few time series which have relatively small number of common observations. First, we randomly selected 6 stations from a set of 25 complete residuals that we employed and made the time series to end at 2009.12, which means that we purposely deleted data after 2009.12. Later, other 6 stations were randomly selected and we artificially made them to start after 2009.12. Having those two datasets of 6 stations, we gradually added the previously deleted data to prolong the time of common observations in every stage. In each iteration, the remaining set of 13 stations was untouched. For this kind of dataset, spatio-temporal filtering of 25 GNSS residuals was carried out using 4 PCA-based methods. Relative errors of CME were estimated for all stations and the results were averaged. The selection of 12 stations was repeated 20 times independently at every stage to average results for various combinations of stations subjected to data deletion. We also estimated the time which is required to estimate CME for all methods. Table 1 presents relative errors of CME (σ_{CME}, Eq. 7) estimated with the procedure described above.

Based on the data presented in Table 1, we can conclude that pPCA method gives quite consistent results compared to other algorithms. The relative errors of CME averaged for 20 simulations do not exceed 19% in each direction. For analyzed set of stations, the mPCA method requires that the time series have a minimum of 400 common epochs of observations for horizontal components and a
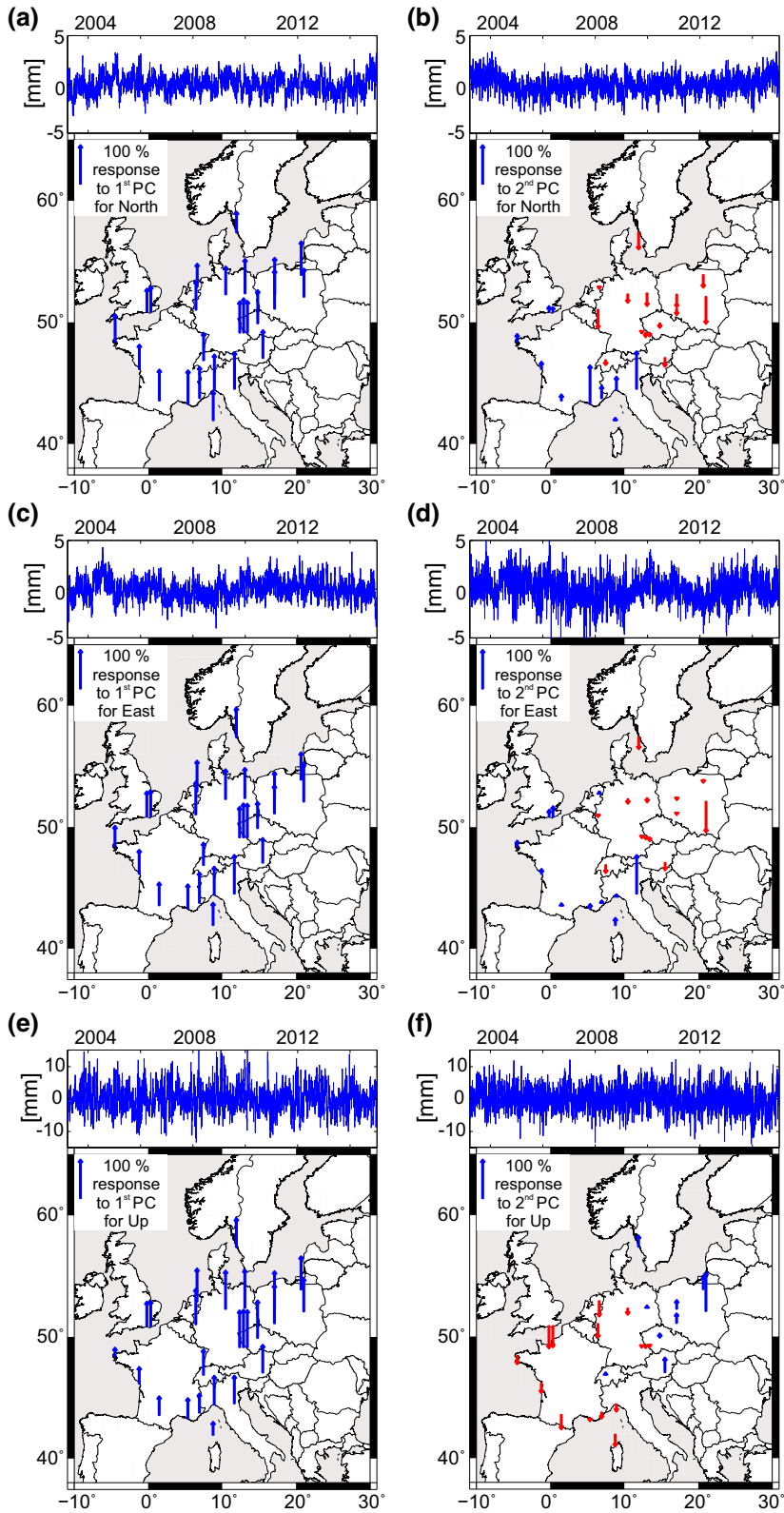
◀Figure 7
Maps of spatial distribution of GNSS residuals responses (normalized eigenvector elements) to the 1st (**a**, **c**, **e**) and 2nd (**b**, **d**, **f**) PC, for North, East and Up components, respectively, and corresponding scaled Principal Components (solid line at the top of each map). Normalized eigenvector elements can be identified with station contribution (positive-blue or negative-red) into the amount of variance in particular PC. Note different scales in horizontal and vertical directions

minimum of 1500 epochs for a vertical one. Otherwise, algorithm is unable to calculate Principal Components, because the covariance matrix estimated at the beginning of the algorithm is not positive semidefinite and some of its eigenvalues are negative. The differences between iPCA and pPCA methods can be seen for horizontal components, where time span of common epochs is shorter than 800 observations. In such a case, the differences in relative error of CME reach 14%. The iPCA, mPCA and pPCA algorithms performed similarly for Up component except in cases in which mPCA was unfunctional.

Standard PCA method is the fastest and also the least complex of those being considered, because residuals are fully interpolated a priori and eigen-decomposition is made only once, but results shown in Table 1 do not give grounds for including this

method for further analysis. Time needed to estimate CME depends on the computing power of the resources. In our case, we conducted each PCA-based filtering method simultaneously using the same HPC-class (High-Performance Computing) resource, therefore, we show relative values of calculation time referred to in pPCA method. This method calculates CME relatively faster than mPCA up to a maximum of 300%, when residuals are loaded with the largest number of missing values. Differences in processing time have decreased almost to zero for these two methods, where residuals have more than 2000 epochs of common observations. Computational time for iPCA method is very similar to pPCA method, i.e. iPCA is up to 20% slower. However, for more complete time series the iPCA seems to be faster than pPCA (up to 50%). Since in our case the number of stations and epochs for daily time series are relatively small, processing time does not seem to be a key factor for defining the superiority of filtering methods to be used for GNSS position time series. However, when long-term hourly (or even more frequent) GNSS time series from network formed by hundreds of stations would have to be employed for filtering procedure, the computational complexity can influence the choice of method.

Table 1

Relative errors of CME averaged for a set of 25 GNSS residuals, in which 12 randomly selected residuals have limited common time span of observations

Number of common epochs	North [%]				East [%]				Up [%]			
	Standard PCA	iPCA	mPCA	pPCA	Standard PCA	iPCA	mPCA	pPCA	Standard PCA	iPCA	mPCA	pPCA
50	99.4	32.6	–	18.6	85.9	22.5	–	18.5	99.9	13.9	–	13.6
100	98.2	27.2	–	18.6	81.6	22.0	–	18.1	91.4	12.9	–	13.2
200	96.1	24.6	–	18.3	77.1	21.6	–	17.3	90.6	12.5	–	12.7
400	69.8	20.0	20.1	18.1	75.5	21.7	30.7	16.7	89.6	12.2	–	13.0
600	56.4	19.7	20.0	17.4	75.2	19.1	20.5	16.5	76.9	12.0	–	12.2
800	55.5	17.9	16.1	17.2	68.8	18.8	17.0	16.5	71.7	11.4	–	12.4
1000	54.6	16.6	15.8	15.6	50.5	17.6	16.7	16.3	61.5	10.9	–	12.0
1500	47.5	15.0	15.6	15.0	45.2	16.6	16.6	16.2	50.3	10.7	10.8	10.9
2000	39.0	15.4	14.2	14.1	37.0	15.1	15.2	14.5	47.7	9.2	9.2	10.4
2500	35.1	14.7	12.5	12.1	26.7	14.4	13.4	13.4	46.7	9.5	9.9	10.2
3000	28.1	13.4	10.3	9.4	22.2	11.3	11.6	10.8	15.8	7.6	7.3	7.8
3500	15.3	12.9	9.1	8.6	15.1	7.9	7.3	6.9	12.5	5.6	5.5	5.5

Number of common epochs for those 12 stations is contained in the first column. It shows how much these 12 stations are overlapped. Number of epochs in the whole dataset is 4314. Symbol '−' means that a certain algorithm was unable to calculate Principal Components

Topocentric component principal Component	North [%]	East [%]	Up [%]
1st	36	36	49
2nd	7	8	7
3rd	6	7	5
4th	5	6	5
5th	4	4	4
6th	4	4	3
7th	4	3	3

On the basis of the results presented in this section, we may conclude that the pPCA method is able to be directly applied to the GNSS position time series with no need to interpolate the data before spatio-temporal filtering. In turn, GNSS time series do not have to start and end at the same epochs, they are not affected by the interpolation procedure. What is more, a gap present in all time series at one (or several) epoch, will facilitate the calculation of CME.

5. pPCA Filtering of Real Time Series

In the following section, we present the results of spatio-temporal filtering performed with pPCA for real dataset consisting of position time series from 25 IGS stations. Residuals are the result of standard pre-processing described previously and they are not subjected to intentional data deleting or interpolating procedure. We employed a set of 25 stations presented in Fig. 3 and used the "unfiltered" residuals $\mathbf{r}(t)$ of their position time series presented in Fig. 4b.

First, we analyzed the eigenvalues and eigenvectors related to each consecutive PCs. Table 2 presents the proportion of variance of all residuals, which is represented by first seven PCs. The eigenvalues can be interpreted as a fraction of the total variance of the residual time series corresponding to each eigenvector. Additionally, the analysis of eigenvalues allowed us to define significant PC (or PCs) which may be interpreted as CME.

The 1st PC explains 36 and 49% of the total variance for horizontal and vertical components,

respectively. Higher order PCs do not contribute to the total variance of residuals higher than 8%. These percentages support the hypothesis that regional phenomena affect the vertical component more than the horizontal components. The first PC is the only one which satisfies the criteria of CME consideration. As indicated by the Fisher-Snedecor test at the 95% confidence level, the variance of residuals which is explained by this mode significantly differs from the variances of the remaining PCs. Therefore, in the following part of the paper, the CME will be calculated using only the 1st PC.

Figure 7 shows scaled PCs obtained through pPCA procedure and their corresponding eigenvectors. Scaled Principal Components are obtained by multiplication of each PC by the normalization factor, which is equal to the maximum response of the network stations to this mode. A procedure to compute the normalized eigenvectors was adopted from Dong et al. (2006). The normalized eigenvector elements refer to the spatial response of individual stations to the CME source if the considered PC can be identified as CME. Those elements may be positive or negative with values between -100 and $+100\%$ (Fig. 7). The theoretical assumption of CME changeability within the considered GNSS network is supported by the results presented in the Fig. 7. The entire set of stations show a positive response to the 1st PC with values higher than 33% for all topocentric components. A minimum response was found for the AJAC (Ajaccio, France) station for Up component (Fig. 7e). The elements of the eigenvector related to 2nd PC are both positive and negative (Fig. 7b, d, f). Such result can be explained by the fact that signals extracted by the 2nd and also by subsequent PCs are due to an uncommon source for that network. They may result from local or regional effects and are unnoticeable for the entire set of stations. The consecutive PCs are characterized by the statistically negligible amount of variance explained by them. Both eigenvalues presented in Table 2 and spatial distribution of station responses shown in Fig. 7 analyzed together allow us to conclude, that 1st PC is the only one that fulfills the CME definition. This has been also confirmed previously by Fisher-Snedecor tests.

Results presented in Fig. 7 show a spatial pattern for the East and Up components found for network station responses to the 1st PC which is identified as CME. For these two components of position, the GNSS residuals responses to the CME are higher for stations situated in Northeastern Europe than for other selected stations. For North component, station responses are more homogeneous. The median value of normalized eigenvector corresponding to 1st PC is equal to 81, 73 and 74% for North, East and Up components, respectively. For the analyzed network, only 5 from 25 stations have relative responses less than 70% to 1st PC in North component. This result is different for Up and East components, where as many as 10 stations have relative responses less than 70% to 1st PC. From 10 stations characterized by the lowest response for CME in Up and East components, 8 of them are located in Southeastern Europe. These are: AJAC (Ajaccio, France), BRST (Brest, France), GRAS (Grasse, France), HERS (Herstmonceux, UK), HERT (Herstmonceux, UK), LROC (La Rochelle, France), MARS (Marseille, France), TLSE (Toulouse, France), see Fig. 3. Spatial pattern, which was found for 1st PC (Fig. 7) is similar to the distribution of power-law noise which was observed earlier by Klos and Bogusz (2017). They showed that vertical components from Central and Northern European stations may be characterized by smaller spectral indices of power-law noise than any other stations in Europe.

The small scale crustal deformations due to surface mass loading are considered to be a very important contributor to the spatially correlated errors in GNSS (Dong et al., 2006; Yuan et al., 2008). To support the explanations of spatial pattern shown in Fig. 7, we analyzed another dataset employing environmental loading models, which are freely available at the École et Observatoire des Sciences de la Terre (EOST) Loading Service (http://loading.u-strasbg.fr/displ_itrf.php). We used the environmental loadings calculated from three different models: ERA (ECMWF Reanalysis) Interim atmospheric model, Modern Era-Retrospective Analysis (MERRA) land hydrological loading and non-tidal ocean loading ECCO2 (Estimation of the Circulation and Climate of the Ocean version 2) model. We averaged the models to be sampled every 24 h to be consistent with GNSS position sampling rate. Then, we summed these models at corresponding epochs to obtain their superposition, which means a joint effect of environmental mass loading on the displacements of the selected ITRF2014 stations. We also limited their time span to be equal to the GNSS residuals (2003.2–2015.0). These time series will be referred to later in this paper as the "environmental loading time series". Subsequently, we submitted them to pPCA procedure to obtain their spatial responses to the 1st and 2nd PCs (Fig. 8) and corresponding scaled PCs.

Comparing spatial distribution of normalized eigenvectors computed for "environmental loading time series" (Fig. 8), to the eigenvectors computed for unfiltered GNSS residuals, a significant similarity can be noticed especially for 1st PC. The level of variance corresponding to the 1st PC reaches 66, 84 and 90% for North, East and Up components, respectively, and differs from each consecutive PC variance. As well as for GNSS residuals, environmental loading time series respond more to the 1st PC in Northeastern Europe with regards to East and Up components. It is worth emphasizing, that we may draw a North–South oriented line separating areas with different responses relating to 2nd PC. More extended investigations of this phenomena does not coincide with the scope of this paper, but we presume, that this effect is related to differences between the influence of the continental and oceanic climate.

As stated previously, environmental loadings, in particular, atmospheric, hydrological and non-tidal oceanic effects, are one of the potential sources of CME in the GNSS coordinate time series. According to the spatial pattern which was presented in Figs. 7 and 8, we estimated how large-scale environmental effects influence the character of CME. For this purpose, we first derived the CME of the "unfiltered" GNSS residual time series using pPCA for a real dataset as described above. Then, to assess the contribution of loading effects to this CME, we derived the CME of the "unfiltered" GNSS residuals adjusted for loading effects. Finally, to make a comparison we calculated CME variances and discussed the results of CME noise analysis together for two dataset.

GNSS-derived position residuals, as well as, other time series of measurements of wide variety of dynamic processes are usually characterized by
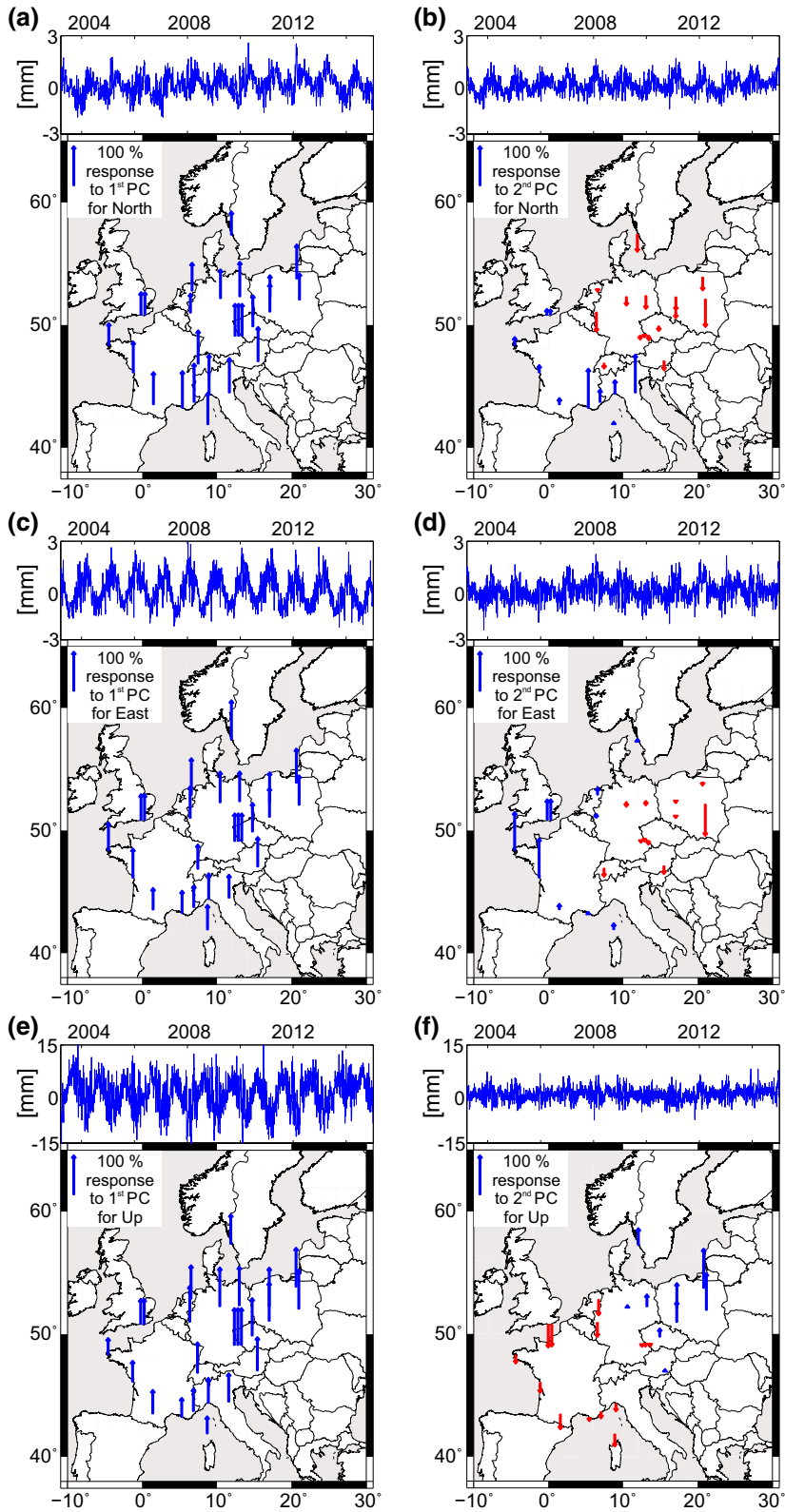
◄Figure 8

Maps of spatial distribution of environmental loading time series responses (normalized eigenvector elements) to the 1st (**a**, **c**, **e**) and 2nd (**b**, **d**, **f**) PC, for North, East and Up component, respectively and corresponding scaled Principal Components (solid line at the top of each map). Loading correction time series are calculated for the same locations as GNSS stations shown in Fig. 3

spectral indices equal to fractional numbers lower than zero (e.g. Langbein and Johnson 1997). In this research, noise analysis was performed with Maximum Likelihood Estimation, which was previously applied in numerous studies describing noise character of GNSS position time series [i.e. Williams et al. (2004), Teferle et al. (2008), Bos et al. (2010) or Klos et al. (2016)]. These researches showed that the noise of GNSS residuals has a character of power-law process with spectral indices varying between -2 (random walk) and 0 (white noise), which are mainly near to -1 (flicker noise). We assumed two different noise models to describe the CME estimates from GNSS "unfiltered" and "filtered" residuals, meaning a combination of power-law and white noise model and autoregressive process. The details of this analysis are described in next paragraphs.

Figure 9 presents the variance of CME, which is identified with 1st PC. CME estimated for "unfiltered" GNSS residuals is characterized by the variances between 0.26 and 1.04 mm^2 for horizontal components and between 0.94 and 17.06 mm^2 for vertical component. The variances of CME were reduced to 0.19–0.87 mm^2 (20% of average reduction) in the horizontal directions and to 0.64–6.87 mm^2 (65%) for vertical direction. A change in CME variances arises from the fact that the environmental loading models remove much of CME variance (Fig. 10), especially with a frequency band between 9 and 12 cpy (cycles per year) mainly affected, which was also noticed before by Gruszczynska et al. (2018). The above described results are consistent with the assertion that GNSS residuals are highly affected by environmental mass loading influences, mostly in the vertical direction.

Within this noise analysis we found that the character of CME is very close to a pure flicker noise for horizontal components, however, it has a character of autoregressive process of first order for vertical component (Fig. 10). Spectral indices we delivered using

MLE analysis computed for CMEs of unfiltered GNSS residuals were equal to − 1.21 and − 1.16 for North and East components, respectively. The contribution of power-law noise was equal to 100.00%, meaning that there is no white noise stored in CME series for horizontal components. Having removed the environmental loadings, spectral indices were equal to − 0.99 and − 0.93 for North and East components, respectively. CME series estimated for Up component are clearly affected by pure autoregressive process of first order (AR(1)), which is flat for low frequencies and stepped when moving to shorter periods. This may indicate that CME is affected by large-scale atmospheric phenomena which also have a character of autoregressive processes (Matyasovszky 2012). Moreover, due to the fact that following stations: BOR1 (Borowa Gora, Poland), GOPE (Pecny, Czech Republic), GRAZ (Graz, Austria), JOZE (Jozefoslaw, Poland), LAMA (Lamkowko, Poland), ONSA (Onsala, Sweden), POTS (Potsdam, Germany), PTBB (Brunswick, Germany), WROC (Wroclaw, Poland), WSRT (Westerbock, Netherlands), WTZA, WTZR, WTZZ (all three in Bad Koetizng, Germany), situated in Central Europe, contributed the most to CME estimates, which was described as the percentage response to 1st PC, the CME we estimated reflects mainly the character of residuals of these stations. A large cut off between 3 and 14 cpy in a power of CME was noticed for CME in Up direction when loading models were removed from series, which causes the CME to resemble a power-law noise. This may indicate that CME in the vertical direction contains environmental effects which affect stations located close to each other.

6. Analysis of GNSS Position Residuals

The GNSS residuals with CME subtracted, which we refer to as "filtered" residuals, were analyzed in terms of the Lin's concordance coefficients between individual pairs of stations in the network under analysis. In addition, the parameters of noise derived from MLE and the uncertainties of velocities estimated for the North, East and Up components using Hector software (Bos et al. 2013) were considered. The values presented in Fig. 11 confirm the benefits
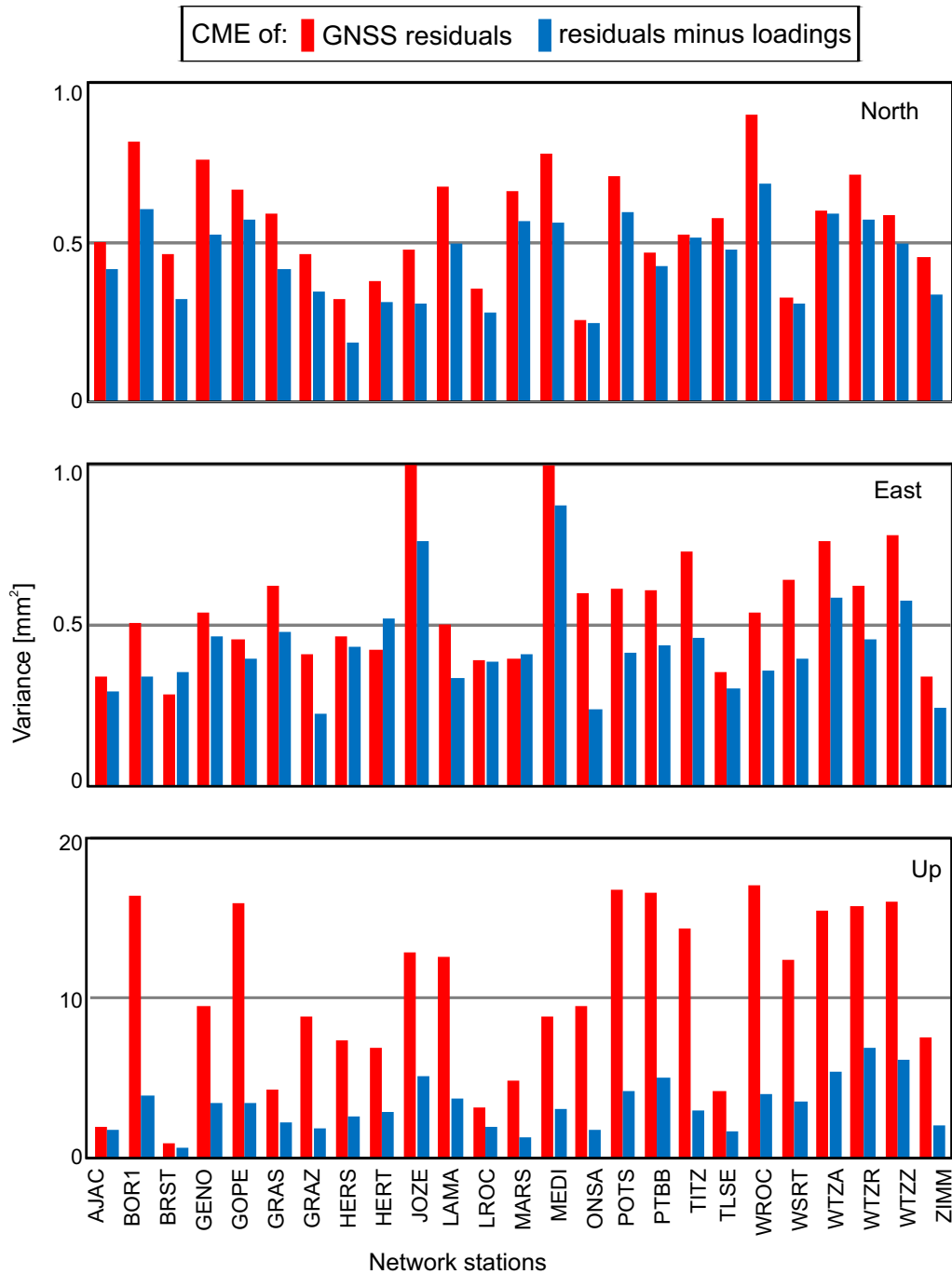
Figure 9

Variance of a CME derived by pPCA filtering method for residuals of 25 stations analyzed in this research. Red bars represent CME estimated for unfiltered GNSS residuals, whilst blue bars show CME of residuals when the superposition of environmental loading models is removed. Note different scales in particular charts
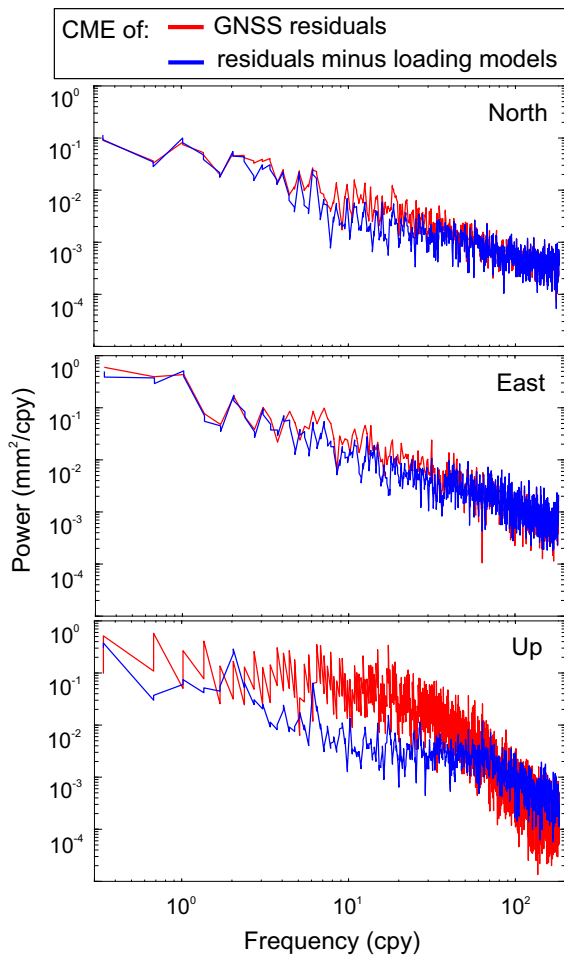
Figure 10
Stacked power spectral densities (PSDs) estimated for CME computed for 25 stations (in red) with a Welch periodogram (Welch 1967), as well as stacked PSDs of CME after environmental loading models were subtracted (in blue)
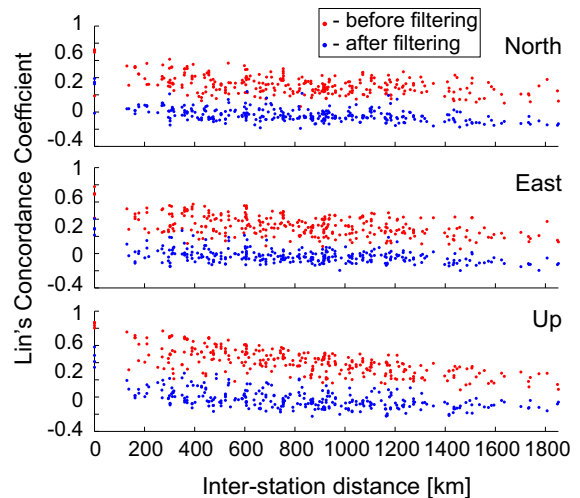

Figure 11
The Lin's concordance correlation coefficient between each pair of 25 stations computed for "unfiltered" (red) and "filtered" (blue) residuals of the position time series is shown as a function of the distance between stations

of spatio-temporal filtering. We showed that pPCA filtering reduces the absolute values of the concordance coefficient on average by 66, 67 and 67% for the North, East and Up components, respectively. The relative reduction is similar for stations in close proximity to each other as well as for stations situated far away, which means that the CME was efficiently reconstructed from 1st PC estimated by the pPCA method.

In the next stage we estimated the variance of residuals, the character of noise and the Bayesian Information Criterion (BIC, Schwarz 1978) which helped to assess the appropriateness of the employed noise model for all stations in the network (Fig. 12).

The changes in noise characteristics for "filtered" and "unfiltered" residuals allowed us to recognize the effect of spatio-temporal filtering on GNSS-derived position residuals.

First, we decided on the preferred noise model to be employed for any individual station. We examined the PSDs of "unfiltered" and "filtered" residuals and found that "unfiltered" residuals in Up direction for stations, BOR1 (Borowa Gora, Poland), GOPE (Pecny, Czech Republic), GRAZ (Graz, Austria), JOZE (Jozefoslaw, Poland), LAMA (Lamkowko, Poland), ONSA (Onsala, Sweden), POTS (Potsdam, Germany), PTBB (Brunswick, Germany), WROC (Wroclaw, Poland), WSRT (Westerbock, Netherlands), WTZA, WTZR, WTZZ (all three in Bad Koetizng, Germany) situated in Central Europe, are affected by pure autoregressive noise model (please see Figures in Supplementary Materials S1). However, when CME is removed from these vertical time series, "filtered" residuals are characterized by pure power-law noise, meaning that we remove the effect that, probably, the atmosphere has on vertical component. On the other hand, we need to be aware of the fact, that we also slightly change the character of other series, which were not affected by atmosphere as much as Central European stations. This makes that, what is noticed from PSDs, these "filtered"
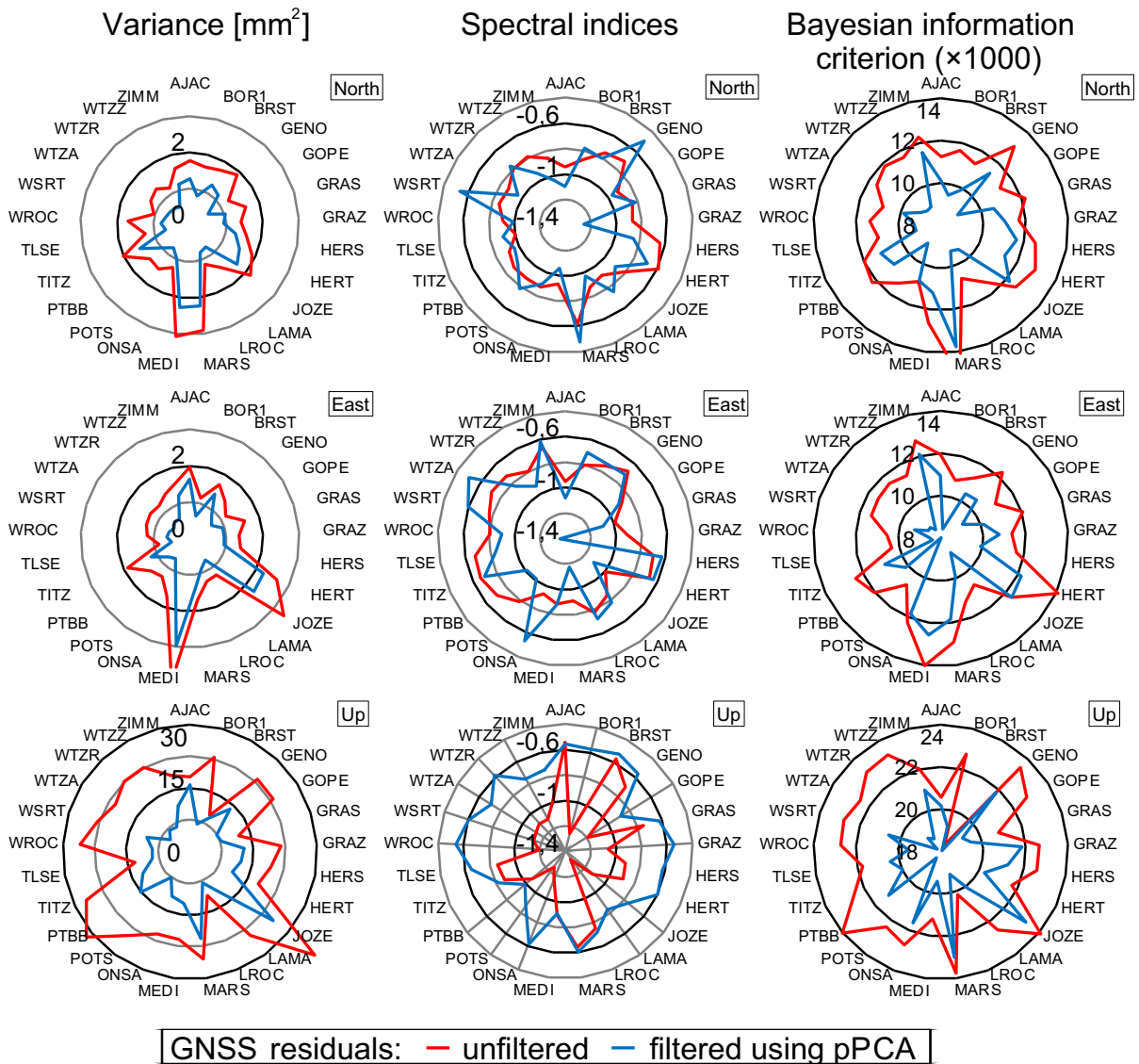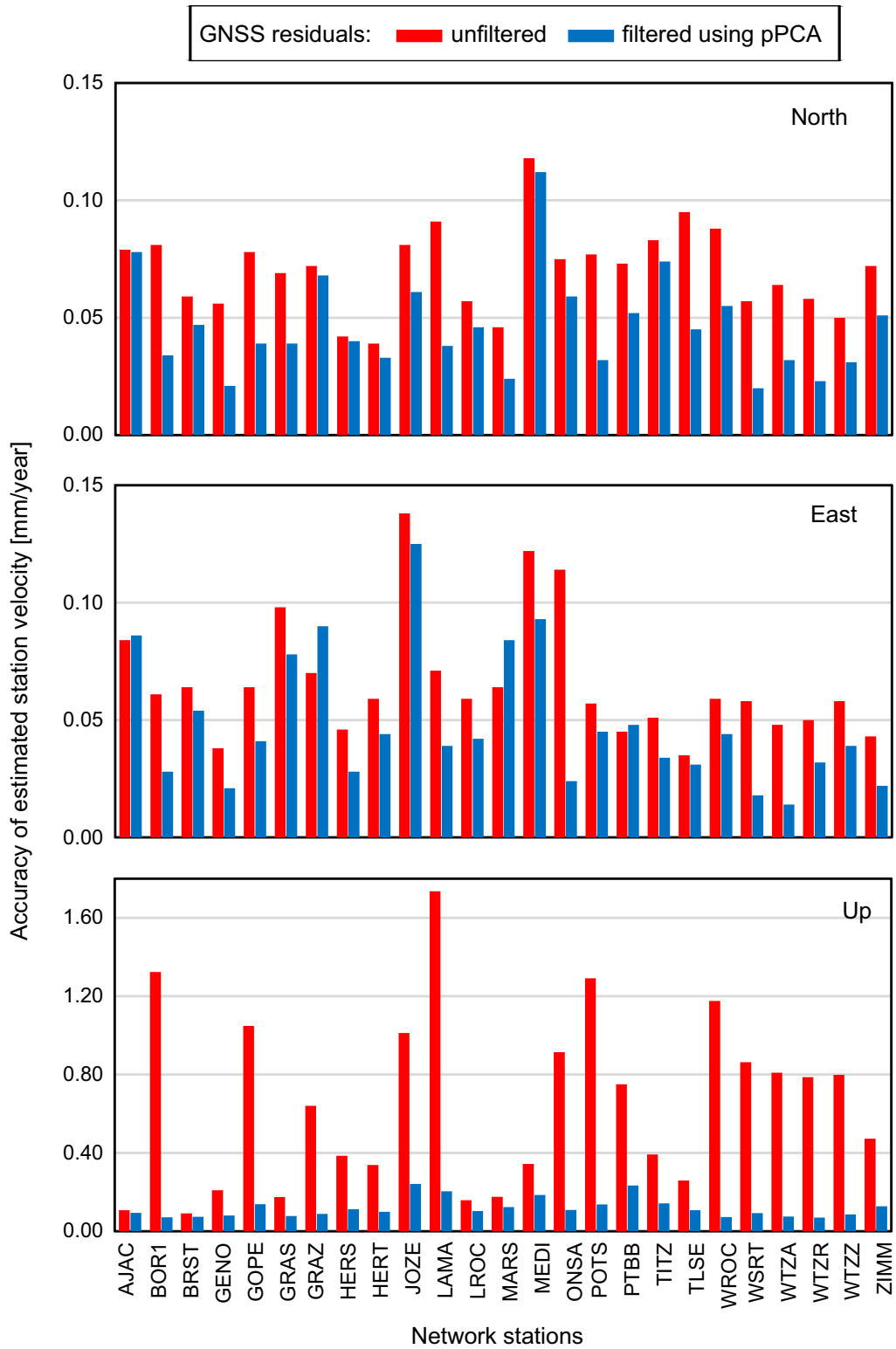
Figure 12
The GNSS residual variances and parameters of noise: spectral indices and BICs estimated for residuals computed before (red) and after (blue) pPCA filtering

residuals are much more affected by white noise than "unfiltered" residuals were, meaning that white noise contributes now more into white plus power-law noise combination.

The variance of "unfiltered" residuals ranged between 1 and 4 mm^2 for the North and East components, whilst it was significantly higher for the Up component and ranged between 10 and 38 mm^2. Spectral indices of power-law noise vary between

-0.6 and -1.0 for the North and East components and between -0.6 and -1.4 for Up component, keeping in mind that for Central European stations, the spectral indices are slightly underestimated because of the portion of AR(1) noise model in residuals. Having filtered the residuals by pPCA, we observed a significant reduction in the variances of between 10 and 74% for all stations with a median decrease estimated at 36, 37 and 46% for the North,

◄Figure 13
Uncertainties in GNSS station velocity determined before spatio-temporal filtering (red bars) and after CME filtering using pPCA (blue bars). Note the different scale on the Up component graph
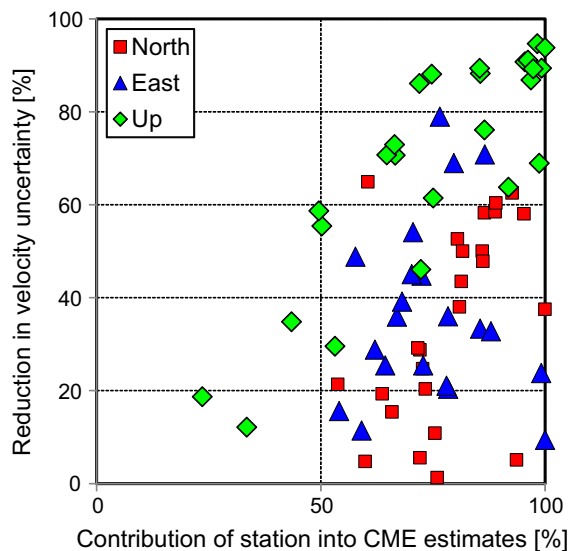

Figure 14
The contribution of each individual station to the CME estimates (normalized eigenvector elements) presented versus reduction in velocity uncertainty [%] when CME was removed. 100% of reduction means that accuracy was reduced to zero

East, and Up components, respectively. A clear improvement in the spectral indices (going towards 0) of the power-law noise was only noticeable for the Up component, but just for stations affected by AR(1) for "unfiltered" residuals, which is reflected by an average shift in the spectral indices towards white noise from − 0.98 to − 0.68 (improvement of almost 30%). This is mainly because a shift between preferred noise models from AR(1) to pure power-law noise was observed. We also estimated the changes in BIC values which confirm the appropriateness of a model to be fitted into certain residuals. We found an improvement in BIC values for all stations and all components after filtration.

Having filtered the CME values, we estimated the uncertainties of the station velocity of the GNSS position time series using the preferred noise model (PL + WN or AR(1) + WN) for each of them.

Figure 13 presents the uncertainties of velocity computed for the "unfiltered" incomplete time series from 25 stations and the uncertainties estimated for the "filtered" ones using the pPCA method. The filtering of the CME leads to a more reliable determination of the GNSS station velocity, especially in the case of the Up component. Prior to filtering, the velocity uncertainties were higher than 1 mm/yr for a few stations, while from Fig. 13 we noticed that after CME was removed by pPCA, the velocity is more precise than 0.2 mm/year.

The largest change of velocity uncertainty equal to 95 and 94% was estimated for the Up component of two Polish stations: BOR1 (Borowa Gora) and WROC (Wroclaw), which is caused by a change of noise model from AR(1) to pure power-law noise. The smallest changes in velocity uncertainty were estimated for the Up component for the AJAC (Ajaccio, France), BRST (Brest, France), GRAS (Grasse, France), LROC (La Rochelle, France), MARS (Marseille, France) and MEDI (Medicina, Italy) stations (Fig. 13).

The changes of velocity uncertainties are significantly different for each topocentric component. The contributions of individual stations to CME estimated for the North vary between 54% for station ONSA (Onsala, Sweden) to 100% for station WROC (Wroclaw, Poland). These contributions have led to a reduction in velocity uncertainty of 21 and 38%, respectively. However, stations with a reduction larger than 38% were also observed. Station WSRT (Westerbork, Netherlands) is one of them, with a maximum reduction of 65%. Both the parameters of noise and velocity accuracy computed before and after filtering, as well as, the reduction of variance in residuals show the importance of GNSS time series filtering.

We noticed a correlation between the contribution of individual stations to CME for the Up component and the reduction in velocity uncertainty. The higher the contribution, the greater the reduction in velocity uncertainty. Pearson correlation coefficients computed for these two variables amount to 46, 32 and 86% for North, East and Up components, respectively (Fig. 14).

7. Discussion and Conclusions

The future of GNSS positioning augmented by continuous measurements provided by permanent stations, will lead to the installation of stations in many new places. Each inequality of operation time span in relation to spatio-temporal analysis should be considered as missing value. This results in the necessity of finding an appropriate method to perform spatio-temporal filtering with no need to limit the series for the same length or to interpolate the gaps. In this study, we proposed probabilistic PCA-based filtering method for the GNSS time series highly affected by missing values or for a situation where stations started and ended operation at different times. We compared the newly applied method with those widely used hitherto: iPCA and mPCA. Moreover, we proved that pPCA gives comparable results but due to its flexible probabilistic model it exceeds in performance both methods, especially in those cases where time series are not characterized by common observational epochs. We compared the traditional PCA filtering approach with the newly employed pPCA and found a few benefits. First, the observations do not have to be interpolated, since pPCA is able to retrieve CME from data with gaps treated in this approach as latent values. Second, the time series may start and end in any epoch, and what is more, they do not have to overlap. This benefit may introduce a fresh perspective of the CME values and may work in any type of network, where the stations do not operate at the same time.

Our analysis of the data from the selected ITRF2014 stations lead us to conclude that CME should not be considered as a uniform signal, homogeneous for all stations. We showed instead that the station spatial responses to the CME may deviate from each other in networks that span up to 1800 km. In case of the considered network, the GNSS stations located in central part of Europe (in Poland, Czech Republic and Germany), contributed the most to the common variability of CME with normalized responses of 87%. The remaining stations contributed 74% on average to CME. The explanation for this phenomenon may simply include the response of stations to environmental loading models, as similar patterns in both GNSS residuals and loading models were observed across the Europe.

It is well known that the vertical component of the GNSS position time series is not determined with the same precision as the horizontal (Wang et al. 2012; Ming et al. 2017). This is due to the principles of satellite navigation systems. The loading processes and spatially-correlated errors have a different effect on vertical component. With this in mind, we noticed a larger reduction in velocity uncertainties in the vertical direction, which is also strictly related to the improvement in the noise characteristics of height component. In addition, the correlation coefficient estimated for pairs of stations decreased much more in the vertical than in horizontal direction. This effect was also confirmed by eigenvalues obtained via the pPCA procedure. These can be interpreted as a percentage of residuals variance represented by each consecutive Principal Components. Since only the 1st PC is identified with CME and eigenvalues corresponding to this PC were equal to 36% of the total variance for horizontal and 49% for vertical direction, we may therefore conclude that CME variance is more significant in Up component. As a result of this, pPCA filtering performs very well especially in Up component. This is very important in the context of increasing expectations regarding to high accuracy of station velocities estimated from the GNSS position time series.

Our results considering environmental loading models are similar to those provided by Zhu et al. (2017) who showed that the RMS for CME estimated for the vertical component is reduced by up to 1.5 mm when loading models are removed. We showed that having removed the environmental loading models, the CME variances are reduced from 10.4 to 3.2 mm^2 on average for vertical components.

The spatial pattern that we noticed in the contribution of individual stations to CME estimations is similar to the spatial dependencies in the amplitudes of power-law noise shown by Klos and Bogusz (2017). The lower the spectral index, the higher the contribution of individual stations to CME. It agrees with the common effect of loadings, which was also investigated using the pPCA method on the basis of the superposition of the environmental models. The stations mostly affected by spatially homogeneous

environmental effects also contributed the most to CME estimates. Following Jiang et al. (2013), stations situated in Central Europe are much more affected by loadings comparing to other parts of Europe. This causes that the vertical displacement we might expect from loading effects are few times higher for stations situated in Central Europe we employed. This dependence was noticed in a form of CME estimated for vertical component, as it resembled the autoregressive noise. When being compared with CMEs for horizontal components, which are of pure power-law character, we may conclude that this CME strictly reflects the atmospheric effect which Central European stations are affected the most. This behavior was also seen for individual PSDs estimated in this research. The autoregressive noise model is preferred over widely employed power-law character for all Central European stations, meaning, that if they contribute the most into CME estimates, this character will also transfer to CME itself. Having removed the CME from "unfiltered" residuals, a power-law noise model became a preferred one for stations affected by autoregressive noise model up till now. So, in other words, we removed the atmospheric effect, which appears in Central European stations and was enough powerful to be transferred to CME estimates. In its turn this brings us a question if the spatial extent of stations should not be limited to the joint environmental impact which loading effects have on position time series. So far, it was stated, that the networks can be as extent as 2 000 km, but then various spatial filters should be employed to differentiate the spatial response of individual stations. Our finding brings here a new light if the environmental loadings impact should not also be taken into consideration.

Hitherto, improvements in the GNSS position time series have resulted in reduction in the scatter of individual time series. Tian and Shen (2016) found an improvement in the scatter of residual time series of 20.7, 13.2, and 14.4% for the North, East and Up components, respectively. Ming et al. (2017) proved that the reduction in scatter when CME was removed was equal to 6.3% for all directions. We estimated the properties of CME using MLE analysis and demonstrated an improvement in colored noise parameters at almost all stations.

In conclusion, according to our analysis we can confidently state, that the newly applied probabilistic Principal Component Analysis is a powerful and efficient tool for the spatio-temporal filtering of any type of geodetic gapped data and not only for the GNSS observations investigated in this paper, being a good alternative for such algorithms as mPCA, iPCA and classical PCA.

REFERENCES

Altamimi, Z., Rebischung, P., Métivier, L., & Collilieux, X. (2016). ITRF2014: a new release of the International Terrestrial Reference Frame modeling nonlinear station motions. *Journal of Geophysical Research: Solid Earth*. https://doi.org/10.1002/2016JB013098.

Andrei, M., & Malandrino, A. (2003). Comparative coreflood studies for precipitation and adsorption squeeze with PPCA as the scales inhibitor. *Petroleum Science and Technology, 21*(7–8), 1295–1315. https://doi.org/10.1081/LFT-120018174.

Blewitt, G., Kreemer, C., Hammond, W. C., & Goldfarb, J. M. (2013). Terrestrial reference frame NA12 for crustal deformation studies in North America. *Journal of Geodynamics, 72,* 11–24. https://doi.org/10.1016/j.jog.2013.08.004.

Blewitt G., Lavallée D. (2002) Effect of annual signals on geodetic velocity, J. geophys. Res.: Solid Earth, vol. 107 (pg. ETG 9-1-ETG 9-11), https://doi.org/10.1029/2001jb000570.

Bogusz, J., Gruszczynski, M., Figurski, M., & Klos, A. (2015). Spatio-temporal filtering for determination of common mode error in regional GNSS networks. Open Geosciences. https://doi.org/10.1515/geo-2015-0021.

Bogusz, J., & Klos, A. (2016). On the significance of periodic signals in noise analysis of GPS station coordinates time series. GPS Solutions, 20(4), 655–664. https://doi.org/10.1007/s10291-015-0478-9.

Bos, M. S., Bastos, L., & Fernandes, R. M. S. (2010). The influence of seasonal signals on the estimation of the tectonic motion in short continuous GPS time-series. Journal of Geodynamics, 49, 205–209. https://doi.org/10.1016/j.jog.2009.10.005.

Bos, M. S., Fernandes, R. M. S., Williams, S. D. P., & Bastos, L. (2008). (2008): fast error analysis of continuous GPS observations. J. Geod, 82, 157–166. https://doi.org/10.1007/s00190-007-0165-x.

Bos, M. S., Fernandes, R. M. S., Williams, S. D. P., & Bastos, L. (2013). Fast error analysis of continuous GNSS observations with missing data. Journal of Geodesy, 87(4), 351–360. https://doi.org/10.1007/s00190-012-0605-0.

Cao, Y., Liu, R. M., & Yang, J. (2008). Infrared small target detection using PPCA. International Journal of Infrared and Millimeter Waves, 29(4), 385–395. https://doi.org/10.1007/s10762-008-9334-0.

Cerny, C. A., & Kaiser, H. F. (1977). A study of a measure of sampling adequacy for factor-analytic correlation matrices. Multivariate Behavioral Research, 12(1), 43–47.

Chen, T., Martin, E., & Montague, G. (2009). Robust probabilistic PCA with missing data and contribution analysis for outlier detection. Computational Statistics & Data Analysis, 53(10), 3706–3716. https://doi.org/10.1016/j.csda.2009.03.014.

Dong, D., Fang, P., Bock, Y., Webb, F., Prawirodirdjo, L., Kedar, S., et al. (2006). Spatio-temporal filtering using principal component analysis and Karhunen-Loeve expansion approaches for regional GPS network analysis. Journal of Geophysical Research, 111, B03405. https://doi.org/10.1029/2005JB003806.

Dow, J. M., Neilan, R. E., & Rizos, C. (2009). The international GNSS service (IGS) in a changing landscape of Global Navigation Satellite Systems. Journal of Geodesy, 83(3–4), 191–198. https://doi.org/10.1007/s00190-008-0300-3. (IGS Special Issue).

Fan, W. T., Bouguila, N. (2009) Generating Video Textures by PPCA and Gaussian Process Dynamical Model. Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications, Proceedings, Book Series: Lecture Notes in Computer Science. Edited by: Bayro Corrochano, E. and Eklundh, JO, 5856:801–808. https://doi.org/10.1007/978-3-642-10268-4_94.

Fisher, R. A. (1932). Inverse probability and the use of likelihood. Proceedings of the Cambridge Philosophical Society., 28(3), 257–261. https://doi.org/10.1017/S0305004100010094.

Gross, R., Beutler, G., Plag, H-P. (2009) Integrated scientific and societal user requirements and functional specifications for the GGOS. In "Global Geodetic Observing System Meeting the Requirements of a Global Society on a Changing Planet in 2020" edited by Hans-Peter Plag and Michael Pearlman, ISBN 978-3-642-02686-7 e-ISBN 978-3-642-02687-4. https://doi.org/10.1007/978-3-642-02687-4, Springer Dordrecht Heidelberg London New York.

Gruszczynska, M., Rosat, S., Klos, A., Gruszczynski, M., & Bogusz, J. (2018). Multichannel singular spectrum analysis in the estimates of common environmental effects affecting GPS observations. Pure and Applied Geophysics. https://doi.org/10.1007/s00024-018-1814-0.

Gruszczynski, M., Klos, A., & Bogusz, J. (2016). Orthogonal transformation in extracting of common mode errors from continuous GPS networks. Acta Geodynamics et Geomaterialia, 13(3), 291–298. https://doi.org/10.13168/AGG.2016.0011.

He, X., Montillet, J.-P., Fernandes, R., Bos, M., Hua, X., Yu, K., et al. (2017). Review of current GPS methodologies for producing accurate time series and their error sources. Journal of Geodynamics, 106, 12–29. https://doi.org/10.1016/j.jog.2017.01.004.

Houseago-Stokes, R. E., & Challenor, P. G. (2004). Using PPCA to estimate EOFs in the presence of missing values. Journal of Atmospheric and Oceanic Technology., 21(9), 1471–1480. https://doi.org/10.1175/1520-0426(2004)021.

Ilin, A., & Raiko, T. (2010). Practical approaches to principal component analysis in the presence of missing values. The Journal of Machine Learning Research, 11, 1957–2000.

Ji, K. H., & Herring, T. A. (2011). Transient signal detection using GPS measurements: transient inflation at Akutan volcano, Alaska, during early 2008. Geophysical Research Letters. https://doi.org/10.1029/2011GL046904.

Jiang, W., Li, Z., van Dam, T., & Ding, W. (2013). Comparative analysis of different environmental loading methods and their impacts on the GPS height time series. Journal of Geodesy, 87, 687–703. https://doi.org/10.1007/s00190-013-0642-3.

Jiang, W. P., & Zhou, X. H. (2015). Effect of the span of Australian GPS coordinate time series in establishing an optimal noise model. Science China: Earth Sciences, 58, 523–539. https://doi.org/10.1007/s11430-014-4996-z.

Jolliffe, I. T. (1972). Discarding variables in a principal component analysis. I: artificial data. Applied Statistics, 21, 160–173. https://doi.org/10.2307/2346488.

Jolliffe, I. T. (2002). Principal Component Analysis. New York: Springer. https://doi.org/10.1007/b98835.

King, M., Altamimi, Z., Boehm, J., Bos, M., Dach, R., Elosegui, P., Fund, F., Hernndez-Pajares, M., Lavallèe, D., Mendes Cerveira, P., Penna, N., Riva, R., Steigenberger, P., van Dam, T., Vittuari, L., Williams, S., Willis, P. (2010) Improved constraints on models of glacial isostatic adjustment: a review of the contribution of ground-based geodetic observations. Surv Geophys 31:465–507. https://doi.org/10.1007/s10712-010-9100-4.

Klos, A., & Bogusz, J. (2017). An evaluation of velocity estimates with a correlated noise: case study of IGS ITRF2014 European stations. Acta Geodynamics et Geomaterialia, 14(3), 255–265. https://doi.org/10.13168/AGG.2017.0009.

Klos, A., Bogusz, J., Figurski, M., & Gruszczynski, M. (2016). Error analysis for European IGS stations. Studia Geophysica et Geodaetica, 60(1), 17–34. https://doi.org/10.1007/s11200-015-0828-7.

Klos, A., Olivares, G., Teferle, F. N., Hunegnaw, A., & Bogusz, J. (2018). On the combined effect of periodic signals and colored noise on velocity uncertainties. GPS Solutions, 22, 1. https://doi.org/10.1007/s10291-017-0674-x.

Kreemer, C., Blewitt, G., & Klein, E. C. (2014). A geodetic plate motion and global strain rate model. Geochemistry, Geophysics, Geosystems, 15, 3849–3889. https://doi.org/10.1002/2014GC005407.

Langbein, J., & Bock, Y. (2004). High-rate real-time GPS network at Parkfield: utility for detecting fault slip and seismic displacements. *Geophysical Research Letters, 31,* 15. https://doi.org/10.1029/2003GL019408.

Langbein, J., & Johnson, H. (1997). Correlated errors in geodetic time series: implications for time-dependent deformation. *Journal of Geophysical Research, 102*(B1), 591–603. https://doi.org/10.1029/96JB02945.

Li, W., Shen, Y., & Li, B. (2015). Weighted spatio-temporal filtering using principal component analysis for analyzing regional GNSS position time series. *Acta Geodaetica et Geophysica, 50*(4), 419–436. https://doi.org/10.1007/s40328-015-0100-1.

Lin, L. I.-K. (1989). A concordance correlation coefficient to evaluate reproducibility Biometrics. *International Biometric Society., 45*(1), 255–268. https://doi.org/10.2307/2532051.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). Hoboken: Wiley. https://doi.org/10.1002/9781119013563.

Lopez-Rubio, E., Ortiz-de-Lazcano-Lobato, J. M., & Lopez-Rodriguez, D. (2009). Probabilistic PCA self-organizing maps. *IEEE Transactions on Neural Networks, 20*(9), 1474–1489. https://doi.org/10.1109/TNN.2009.2025888.

Madikeri, S. R. (2014). A fast and scalable hybrid FA/PPCA-based framework for speaker recognition. *Digital Signal Processing, 32,* 137–145. https://doi.org/10.1016/j.dsp.2014.05.012.

Márquez-Azúa, B., & DeMets, C. (2003). Crustal velocity field of Mexico from continuous GPS measurements. 1993 to June 2001: implications for the neotectonics of Mexico. *Journal of Geophysical Research.* https://doi.org/10.1029/2002JB002241.

Matyasovszky, I. (2012). Spectral analysis of unevenly spaced climatological time series. *Theor Appl Climatol, 111*(3–4), 371–378. https://doi.org/10.1007/s00704-012-0669-z.

Métivier, L., Collilieux, X., Lercier, D., Altamimi, Z., & Beauducel, F. (2014). Global coseismic deformations, GNSS time series analysis, and earthquake scaling laws. *J Geophys Res Solid Earth, 119,* 9095–9109. https://doi.org/10.1002/2014JB011280.

Ming, F., Yang, Y., Zeng, A., & Zhao, B. (2017). Spatiotemporal filtering for regional GPS network in China using independent component analysis. *J Geod, 91*(4), 419–440. https://doi.org/10.1007/s00190-016-0973-y.

Moron, V., Oueslati, B., Pohl, B., Rome, S., & Janicot, S. (2016). Trends of mean temperatures and warm extremes in northern tropical Africa (1961–2014) from observed and PPCA-reconstructed time series. *Journal of Geophysical Research-Atmospheres, 121*(10), 5298–5319. https://doi.org/10.1002/2015JD024303.

Mredhula, L., & Dorairangaswamy, M. (2016). An effective filtering technique for image denoising using probabilistic principal component analysis (PPCA). *Journal of Medical Imaging and Health Informatics, 6*(1), 194–203. https://doi.org/10.1166/jmihi.2016.1602.

Nikolaidis, R. (2002) Observation of geodetic and seismic deformation with the Global Positioning System. Ph.D. thesis. Univ. of Calif., San Diego.

Qi, C. K., Li, H. X., Li, S. Y., Zhao, X. C., & Gao, F. (2012). Probabilistic PCA-based spatiotemporal multimodeling for nonlinear distributed parameter processes. *Industrial and Engineering Chemistry Research, 51*(19), 6811–6822. https://doi.org/10.1021/ie202613t.

Qu, L., Li, L., Zhang, Y., & Hu, J. M. (2009). PPCA-based missing data imputation for traffic flow volume: a systematical approach.

IEEE Transactions on Intelligent Transportation Systems, 10(3), 512–522. https://doi.org/10.1109/TITS.2009.2026312.

Rebischung, P., Altamimi, Z., Ray, J., & Garayt, B. (2016). The IGS contribution to ITRF2014. *J Geod, 90*(7), 611–630. https://doi.org/10.1007/s00190-016-0897-6.

Roweis, S. (1997). EM algorithms for PCA and SPCA. *Advances in Neutral Information Processing Systems, 10,* 626–632.

Santamaría-Gómez, A., & Mémin, A. (2015). Geodetic secular velocity errors due to interannual surface loading deformation. *Geophysical Journal International.* https://doi.org/10.1093/gji/ggv190.

Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*(2), 461–464. https://doi.org/10.1214/aos/1176344136.

Serpelloni, E., Faccenna, C., Spada, G., Dong, D., & Williams, S. D. P. (2013). 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. *Journal of Geophysical Research: Solid Earth, 118,* 6003–6024. https://doi.org/10.1002/2013JB010102.

Sharifi, R., & Langari, R. (2017). Nonlinear sensor fault diagnosis using mixture of probabilistic PCA models. *Mechanical Systems and Signal Processing, 85,* 638–650. https://doi.org/10.1016/j.ymssp.2016.08.028.

Shen, Y., Li, W., Xu, G., & Li, B. (2013). Spatio-temporal filtering of regional GNSS network's position time series with missing data using principle component analysis. *Journal of Geodesy, 88,* 351–360. https://doi.org/10.1007/s00190-013-0663-y.

Teferle, F. N., Williams, S. D. P., Kierulf, H., Bingley, R., & Plag, H.-P. (2008). A continuous GPS coordinate time series analysis strategy for high-accuracy vertical land movements. *Physics and Chemistry of the Earth, 33,* 205–216. https://doi.org/10.1016/j.pce.2006.11.002.

Tiampo, K. F., Rundle, J. B., Klein, W., Ben-Zion, Y., & McGinnis, S. (2004). Using eigenpattern analysis to constrain seasonal signals in southern California. *Pure and Applied Geophysics, 161,* 1991–2003. https://doi.org/10.1007/978-3-0348-7873-9_13.

Tian, Y., & Shen, Z. (2011). Correlation weighted stacking filtering of common-mode component in GPS observation network. *Acta Seismologica Sinica, 33*(2), 198–208. https://doi.org/10.3969/j.issn.0253-37822011.02.007.

Tian, Y., & Shen, Z.-K. (2016). Extracting the regional common-mode component of GPS station position time series from dense continuous network. *Journal of Geophysical Research: Solid Earth, 121*(2), 1080–1096. https://doi.org/10.1002/2015JB012253.

Tipping, M. E., & Bishop, C. M. (1999). Probabilistic principal component analysis. *Journal of the Royal Statistical Society, 61B,* 611–622.

Wang, W., Zhao, B., Wang, Q., & Yang, S. (2012). Noise analysis of continuous GPS coordinate time series for CMONOC. *Advances in Space Research, 49,* 943–956. https://doi.org/10.1016/j.asr.2011.11.032.

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. *Journal of Geophysical Research, 102*(B8), 18057–18070. https://doi.org/10.1029/97JB01378.

Welch, P. D. (1967). The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio Electroacoustics, 15*(2), 70–73.

Wessel, P., Smith, W. H. F., Scharroo, R., Luis, J. F., & Wobbe, F. (2013). Generic mapping tools: improved version released. *Eos, Transactions AGU, 94,* 409–410.

Williams, S. D. P., Bock, Y., Fang, P., Jamason, P., Nikolaidis, R., Prawirodirdjo, L., et al. (2004). Error analysis of continuous GPS position time series. *Journal of Geophysical Research.* https://doi.org/10.1029/2003JB002741.

Xiang, Z.-Y., Cao, T.-Y., Zhang, P., Zhu, T., & Pan, J.-F. (2012). Object tracking using probabilistic principal component analysis based on particle filtering framework. *Advanced Materials Research, 341–342,* 790–797. https://doi.org/10.4028/www.scientific.net/AMR.341-342.790.

Xiang, J., Zhong, Y., & Gao, H. (2015). Rolling element bearing fault detection using PPCA and spectral kurtosis. *Measurement, 751,* 180–191. https://doi.org/10.1016/j.measurement.2015.07.045.

Yuan, L., Ding, X., Chen, W., Guo, Z., Chen, S., Hong, B., et al. (2008). Characteristics of daily position time series from the Hong Kong GPS fiducial network. *Chin J Geophys, 51*(5), 1372–1384. https://doi.org/10.1002/cjg2.1292.

Zhu, Z., Zhou, X., Deng, L., Wang, K., & Zhou, B. (2017). Quantitative analysis of geophysical sources of common mode component in CMONOC GPS coordinate time series. *Advances in Space Research.* https://doi.org/10.1016/j.asr.2017.05.002.

Zuccolotto, P. (2012). Principal component analysis with interval imputed missing values. *AStA Advances in Statistical Analysis, 96*(1), 1–23. https://doi.org/10.1007/s10182-011-0164-3.