2006 International Joint Conference on Neural Networks
Sheraton Vancouver Wall Centre Hotel, Vancouver, BC, Canada
July 16-21, 2006

Extraction of Components with Structured Variance

Alexander Ilin, Harri Valpola, and Erkki Oja, *Fellow, IEEE*

***Abstract*— We present a method for exploratory data analysis of large spatiotemporal data sets such as global long-time climate measurements, extending our previous work on semiblind source separation of climate data. The method seeks fast changing components whose variances exhibit slow behavior with specific temporal structure. The algorithm is developed in the framework of denoising source separation. It finds sources iteratively and alternates between estimating the variance structure of extracted sources and using the structure to find new source estimates. The performance of the algorithm is first demonstrated on a simple example of a semiblind source separation problem with artificially generated signals. Then, the proposed technique is applied to the global surface temperature measurements coming from the NCEP/NCAR reanalysis project. Fast changing temperature components whose variances have prominent annual and decadal structures are extracted. The extracted annual components reflect higher temperature variability over the continents during winters. The components with slower changing variances might correspond to some interesting weather phenomena characterized by slowly changing temperature variability in specific regions.**

I. Introduction

In statistical climatology, weather measurements are often analyzed in order to find consistent weather patterns which correspond to meaningful climate phenomena. One of the classical tools in this problem is principal component analysis (PCA) better known in climatology as empirical orthogonal functions (EOF) [1]. PCA is a powerful technique which can be used for reducing the dimensionality of highly multidimensional climate data and therefore it is often used as a preprocessing step. Rotated PCA is a more suitable technique for isolating meaningful modes of weather variations. The basic idea is to rotate the principal components using the concept of "simple structure" and thus obtain a representation which is more localized in space or in time [2].

Independent component analysis (ICA) is a recently introduced technique which can be used for the rotation of principal components. The only criterion used by ICA is the assumption of the statistical independence of the components. Even though ICA can sometimes give a meaningful representation of weather data [3], [4], [5], the statistical independence is quite a restrictive assumption which can often lead to naive solutions.

Recently, we have used a data analysis technique called *denoising source separation* (DSS) to the analysis of climate data [6], [7], [8]. The general DSS framework was proposed in [9] and it can be seen as generalization of ICA with relaxed independence assumption. Instead, DSS seeks uncorrelated components which maximally express some desired, interesting properties. The general DSS algorithm is an iterative procedure which gradually isolates components with the desired structure. The motivation for extracting a particular type of properties can come from general statistical principles (e.g. maximizing non-Gaussianity of components gives the ICA solution), expert knowledge (e.g. information about periodicity of components), or based on some elementary inspection of data (e.g. by observing some regular patterns in it). Thus, DSS presents a powerful tool for *exploratory analysis* of multivariate data.

In our previous works, we concentrated on slow climate oscillations. In [6], [8], we showed that optimization of the criterion that we termed clarity helps find the sources exhibiting the most prominent periodicity in a specific timescale. In the experiments, the components with the most prominent interannual oscillations were clearly related to the well-known El Niño–Southern Oscillation (ENSO) phenomenon. Later, we extended the analysis to a more general case where slow components were separated by their frequency contents [7], [8]. The sources found using the frequency-based criterion give a meaningful representation of the slow climate variability as combination of trends, interannual oscillations, the annual cycle and slowly changing seasonal variations.

A relevant model is independent dynamic subspace analysis [10] in which several components are assumed to share common dynamics. The algorithm extracts groups of components so as to minimize the prediction error of their common dynamic model. This allows for finding the groups of components with the most predictable time course.

In this work, we extend our exploratory approach to the analysis of fast changing climate phenomena. In particular, we design an algorithm which seeks fast components with prominent temporal structure of variances. The proposed algorithm is again based on the general DSS procedure in which the variance structure of the sources is isolated using a properly designed filter. The motivation of the proposed analysis comes from the inspection of the global weather measurements and the observation that fast weather variations have distinct yearly structure. This raises the question whether there are similar variations on slower timescales. The aim of the algorithm is to capture such prominent slow variability of the variances with the possibilty to put emphasis on different timescales of variance oscillations.

The paper is organized as follows: In Section II, we present the general modeling assumptions and give a short

Alexander Ilin is with the Lab. of Computer and Information Science, Helsinki University of Technology, P.O. Box 5400, FI-02015 TKK, Espoo, Finland (email: Alexander.Ilin@tkk.fi).

Harri Valpola is with the Lab. of Computational Engineering, Helsinki University of Technology, P.O. Box 9203, FI-02015 TKK, Espoo, Finland (email: Harri.Valpola@tkk.fi)

Erkki Oja is with the Lab. of Computer and Information Science, Helsinki University of Technology, P.O. Box 5400, FI-02015 TKK, Espoo, Finland (email: Erkki.Oja@tkk.fi).

0-7803-9490-9/06/$20.00/©2006 IEEE

introduction to the general DSS concept. In Section III, we explain how the method is tuned to extract components with the desired variance structure. A simple example presented there shows how the proposed algorithm works on artificially generated data. Section IV describes the climate measurement data analyzed in this paper and Section V presents the results of the proposed analysis on the climate data. Finally, we discuss the results and possible future directions in Section VI.

II. Method

A. Source separation methods

Our analysis is based on the linear mixing model when the multivariate measurements $x_j(t)$ are assumed to be linear combinations of some hidden components $s_i(t)$ (also called sources, factors or latent variables):

$$x_j(t) = \sum_{i=1}^{N} a_{ji} s_i(t),\ j = 1, ..., M\,. \tag{1}$$

The index j runs over the measurement sensors (typically spatial locations), and discretized time t runs over the observation period: $t = 1, ..., T$. This can be expressed in matrix formulation as

$$\mathbf{X} = \mathbf{AS}\,, \tag{2}$$

where $\mathbf{X}$ denotes the matrix of observations (the sensor index j denotes the rows and the time index t denotes the columns), the matrix of sources $\mathbf{S}$ is defined likewise and the coefficients a_{ji} make up the mixing matrix $\mathbf{A}$.

If we denote the columns of matrix $\mathbf{A}$ by $\mathbf{a}_i$ and the observation vector at time t by $\mathbf{x}(t)$, then the model can be written as

$$\mathbf{x}(t) = \sum_{i=1}^{N} \mathbf{a}_i s_i(t)\,. \tag{3}$$

In climate data analysis, the time series $s_i(t)$ usually correspond to the time-varying states of the climate system, and the loading vectors $\mathbf{a}_i$ are the spatial maps showing the typical weather patterns corresponding to the components.

The goal of the analysis is to estimate the unknown components $s_i(t)$ and the corresponding loading vectors $\mathbf{a}_i$ from the observed data $\mathbf{X}$. With minimum a priori assumptions about the sources, the problem is called *blind source separation* (BSS). A popular method for solving the BSS problem is *independent component analysis* in which the only assumption used for the source separation is the statistical independence of the sources [11].

When some prior information exists about what is interesting in the data (e.g. the general shape of the time curves of the sources or their frequency contents), the source separation problem is sometimes called *semiblind*. For example, in the climate data we might be interested in some phenomena that would be cyclic over a specific period or would exhibit slow changes [8]. Then, exploiting the prior knowledge may significantly help in finding a useful representation of the data.

Denoising source separation (DSS) [9] is a general algorithmic framework which can incorporate the prior knowledge or preferences in order to identify the model in Eq. (1). The sources $s_i(t)$ estimated by DSS are generally assumed 1) to be uncorrelated and 2) to have some structure known from the available prior information. Typically, maximizing the structure of components makes them more independent. Thus, DSS provides a general framework for designing semiblind source separation algorithms which can be seen as generalization of ICA.

DSS is based on an estimation procedure in which the prior knowledge (and hence the separation criterion) is expressed in the form of a *denoising function*. Here, we will give a brief exposition of the DSS framework, more details including rigorous derivations and analysis were reported in [9].

B. The DSS algorithmic framework

1) Whitening: The requirement that the sources are uncorrelated is assured in DSS by using a preprocessing step called whitening [11]. Whitening is a liner transformation usually performed by PCA with normalization of the principal components to unit variances. Here, we denote the matrix of whitened data by $\mathbf{Y}$. It has the same dimensions as $\mathbf{X}$ unless dimensionality reduction is combined with PCA.

The matrix of whitened data has uniform covariance structure, which means that any matrix

$$\mathbf{S} = \mathbf{WY}\,, \tag{4}$$

produced by an orthogonal matrix $\mathbf{W}$, has unit covariance. Thus, any orthogonal basis in the whitened space defines uncorrelated sources of unit variance. Therefore, the sources can be estimated by Eq. (4) with the restriction that the demixing matrix $\mathbf{W}$ is orthogonal.

2) Iterative procedure: The optimal demixing matrix $\mathbf{W}$ is found so as to maximize the desired structure of components $\mathbf{S}$, that is by using the second DSS requirement. This requirement is introduced in the algorithm in the form of the denoising function. The purpose of denoising is to emphasize the desired structure in the current source estimates, which assures gradual maximization of the interesting properties.

The general algorithmic framework of DSS is the iterative procedure presented in Fig. 1. Whitening is followed by three successive steps:

1: Source estimation: $\mathbf{S} = \mathbf{WY}$,

2: Applying the denoising function: $\mathbf{S}_{\text{new}} = \mathbf{f}(\mathbf{S})$,

3: Reestimation of $\mathbf{W} = \mathrm{orth}(\mathbf{S}_{\text{new}}\mathbf{Y}^T)$.

The iterations continue until the source estimates do not change. The components can be extracted either simultaneously (symmetric approach) or one after another (deflation approach). In the symmetric approach, the operator $\mathrm{orth}(.)$ in Step 3 gives the orthogonal projection of the matrix $\mathbf{S}_{\text{new}}\mathbf{Y}^T$ onto the set of orthogonal matrices. The basis of the deflation approach is the Gram-Schmidt orthogonalization method when the vector defining the currently found component is made orthogonal to the previously found vectors in $\mathbf{W}$ (see, e.g. [11]).

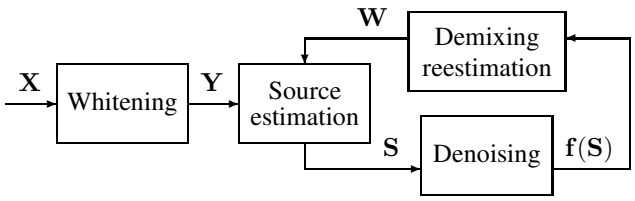

Fig. 1. The steps of the DSS algorithmic framework.

Without denoising, the iterative procedure is equivalent to the power method for computing the principal components of $\mathbf{Y}$. Since $\mathbf{Y}$ is white, all the eigenvalues are equal and the solution without denoising becomes degenerate. Therefore, even slightest changes made by denoising can determine the DSS solution. Since the denoising procedure emphasizes the desired properties of the sources, DSS can gradually isolate the desired structure of the sources.

The described procedure can be significantly simplified if the denoising function $\mathbf{f}$ can be implemented by a linear temporal filter, operating on the rows of its matrix argument: $\mathbf{f}(\mathbf{S}) = \mathbf{SF}$, with $\mathbf{F}$ the filtering matrix. Then, the separation can be preformed in just three steps: whitening, linear denoising (filtering) and PCA on the denoised data [9]. Such simple denoising was used in our previous works on exploratory climate data analysis [6], [8] but it is out of the scope of this paper.

3) Spatial patterns: In the applications, we are interested not only in the sources (rows of matrix $\mathbf{S}$), but also in the matrix $\mathbf{A}$ in Eq. (2). The i-th column of $\mathbf{A}$ is a spatial map showing how the effect of the i-th source is distributed over the sensor array. The mixing matrix can be calculated from the whitening and demixing matrices [8]. Since the extracted components s_i are normalized to unit variances, the columns of $\mathbf{A}$ have a meaningful scale.

Note that since we are now looking for components with a certain variance structure, the signs of the extracted components are not determined. Such ambiguity arises when $\mathbf{f}(\mathbf{S}) = -\mathbf{f}(-\mathbf{S})$. The ambiguity of the solution is even higher for subspace models such as *independent subspace analysis* [12], [13], [10]. There, the sources are decomposed into groups and the sources within a group are generally assumed dependent while components from different groups are mutually independent. Such models can be estimated only up to orthogonal rotations of sources within the groups.

A subspace of sources can be visualized by the observation variance explained by its components. From Eq. (1) and the condition that the sources s_i are mutually uncorrelated and have unit variance, it follows that the variance of the observations equals

$$\operatorname{var}\{x_j\} = \sum_{i=1}^{N} a_{ji}^2 \operatorname{var}\{s_i\} = \sum_{i=1}^{N} a_{ji}^2 \,. \tag{5}$$

Thus, the variances explained by the sources from one subspace $\{s_i | i \in I\}$ is

$$\operatorname{var}_I\{\mathbf{x}\} = \sum_{i \in I} \mathbf{a}_i^2 \,, \tag{6}$$

where $\mathbf{a}_i^2$ denotes the vector of the squared elements of the mixing vector $\mathbf{a}_i$. The quantity in Eq. (6) is a vector whose dimensionality equals the number of sensors and therefore it can be represented as a spatial pattern showing the effect on the observation variance in different spatial locations.

III. Isolating components with structured variance

A. Measuring structure

In this work, the goal of the analysis is to extract fast changing components whose variance (or activations) would have prominent temporal structure. This analysis was motivated from the inspection of the global climate measurements from the NCEP/NCAR reanalysis project [14], [15]. In particular, it is easy to observe that the intensity of fast weather oscillations depends on the time of the year. This variability is so prominent that any other structure in the variance is not immediately visible but it could be there.

An assumption that we make in our analysis is that the interesting sources have non-stationary variances, that is their level of activation changes with time. Moreover, the variances of the sources have prominent temporal structure in a specific timescale chosen for investigation. The analysis can be done for different timescales of variance oscillations and the found interesting components would generally be different. It is important to be able to neglect annual variability in the variance because it is so dominant that it would mask any other, weaker phenomena.

Let us now derive the optimization criterion which follows from the desired source structure. We regard here the source values $\{s(t) \mid t = 1, \ldots, T\}$ as a realization of a stochastic process $\{s_t\}$ consisting of random variables s_t. Note that the difference in notations: $s(t)$ denotes the sample from a random variable s_t. We assume that variables s_t are Gaussian, with zero mean and changing variances $v(t)$. We can define the mean variance of $\{s_t\}$ as

$$\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} v(t) \,. \tag{7}$$

We propose to measure the amount of structure in each source using the difference $\mathcal{F} = H'(\nu) - H'(s)$, where $H'(s)$ denotes the (differential) entropy rate of $\{s_t\}$ and $H'(\nu)$ is the entropy rate of a Gaussian process $\{\nu_t\}$ with i.i.d. zero-mean variables ν_t whose variances $\mathrm{E}\{\nu_t^2\}$ are stationary and equal the mean variance of $\{s_t\}$ defined in Eq. (7). The Gaussian process with stationary variances has the highest entropy rate among all the processes with the same mean variance. Therefore, $\mathcal{F}$ is a good measure of non-stationarity, it is always nonnegative and it attains its minimum value of zero if and only if $\{s_t\}$ is a Gaussian process with stationary variances. The proposed measure resembles negentropy [11] which is often used as as measure of non-Gaussianity of a random variable.

Now we note that the assumption that variances $v(t)$ have prominent variability in the known timescale helps estimate $v(t)$ from one realisation of the stochastic process. Then,

given a realisation of length T, the quantity in Eq. (7) can be estimated as $\frac{1}{T}\sum_t v(t)$. The Gaussian variables s_t are assumed independent given $v(t)$ and therefore the entropy rate of $\{s_t\}$ can be estimated as

$$H'(s) = \frac{1}{T}\sum_t H(s_t) = \frac{1}{T}\sum_t \frac{1}{2}\log 2\pi e v(t)\,, \qquad (8)$$

where $H(s_t)$ denotes the entropy of s_t. This yields

$$\mathcal{F} = \frac{1}{2}\log\frac{1}{T}\sum_t v(t) - \frac{1}{T}\sum_t \frac{1}{2}\log v(t) \geq 0\,. \qquad (9)$$

In practice, whitening makes $\frac{1}{T}\sum_t s(t)^2 = 1$ for any source estimate, which allows for the assumption that

$$\frac{1}{T}\sum_t v(t) = 1\,. \qquad (10)$$

This simplifies Eq. (9) to

$$\mathcal{F}_1 = -\frac{1}{T}\sum_t \frac{1}{2}\log v(t)\,. \qquad (11)$$

The statistic $\mathcal{F}$ is a good measure of structure, which is related to non-stationarity of variances and has some connection to non-Gaussianity. The latter can be seen by noting from Eq. (10) that the variances $v(t)$ fluctuate around unity and therefore we can use the approximation $\log(1+\epsilon) \approx \epsilon - \frac{1}{2}\epsilon^2$. The yields the quantity

$$\mathcal{F}_2 \propto \frac{1}{T}\sum_t v^2(t) - 1 \qquad (12)$$

which measures the magnitude of the variance fluctuations around the mean variance. For a process with stationary and unit variance, $\mathcal{F}_2$ equals zero. Now note that if the local variance $v(t)$ is approximated by $s^2(t)$, Eq. (12) gives the fourth moment of the random variable s. Such higher-order moments are often used for measuring non-Gaussianity [11].

In order to use the proposed measure, we have to be able to estimate the variances $v(t)$ of a signal in each time instant. This is usually done by estimating local sample variances because the variance is assumed to change slowly. We, however, want to concentrate on a *specific timescale* of variance oscillations and therefore we assume that the variance can be estimated in practice by filtering the squared signal values $s^2(t)$ such that only the interesting frequencies are preserved.

B. The algorithm

The criteria in Eq. (11) and (12) are functions of the variances $v(t)$ which are estimated from the sources $s(t)$. Thus, $\mathcal{F}_1$ and $\mathcal{F}_2$ are functions of $s(t)$ and can be maximized w.r.t. $s(t)$ by the gradient ascent method:

$$s_{\text{new}}(t) = s(t) + \mu\frac{\partial\mathcal{F}}{\partial s(t)}\,, \qquad (13)$$

where μ is the step size. Note that the orthogonality constraint for the demixing matrix $\mathbf{W}$ makes it possible to modify the update rule in Eq. (13) by adding a term $\beta s(t)$, with β some constant, without changing the fixed points of the algorithm (see details in [9]). Thus, the update rule

$$s_{\text{new}}(t) = f(s(t)) \propto \beta s(t) + \frac{\partial\mathcal{F}}{\partial s(t)} \qquad (14)$$

optimize the same criterion as the rule in Eq. (13). In the DSS terminology, Eq. (14) describes the denoising function.

The gradients of $\mathcal{F}_1$ and $\mathcal{F}_2$ can be shown to yield the denoising function

$$s_{\text{new}}(t) = g(v(t))s(t)\,, \qquad (15)$$

where the nonlinearity g is given by

$$\text{for } \mathcal{F}_1: \quad g(v) \propto \beta - 1/v\,, \qquad (16)$$

$$\text{for } \mathcal{F}_2: \quad g(v) \propto \beta + v\,, \qquad (17)$$

and β is a constant which can be chosen arbitrarily. Note that we use the term *mask* for the values of $g(v(t))$ as they are applied to the current source estimates to get the new ones.

However, neither of the two nonlinearities is robust. The nonlinearity in Eq. (17) behaves nicely for small values of v but it gives too much weight to large v. This makes the algorithm very sensitive to outliers and very often results in overfitting [16]. Note that $\mathcal{F}_2$ is related to higher-order moments which often suffer from this problem. In contrast, the nonlinearity in Eq. (16) saturates for large v but it is sensitive to small v where the gradient approaches infinity.

Therefore, we propose to use the denoising function in Eq. (15) with a monotonic nonlinearity $g(v)$ which starts from zero and saturates for large values, for example,

$$g(v) = \tanh(v)\,. \qquad (18)$$

It is possible to show that g in Eq. (18) approximates the nonlinearity in Eq. (16) with a specific choice of β and some constraints on the smallest possible values of $v(t)$.

The proposed denoising procedure consists of several steps:

1: Estimate source variances by filtering the squared signal such that only the interesting frequencies are preserved:

$$\{v_i(t)\,|\,\forall t\} = \text{filter}_f\{s_i^2(t)\,|\,\forall t\}\,. \qquad (19)$$

2: Use the variance values as the masks if the measure $\mathcal{F}_2$ is optimized (this follows from Eq. (17) with $\beta = 0$):

$$m_i(t) = v_i(t) \qquad (20)$$

Otherwise, apply the nonlinearity g, e.g. from Eq. (18):

$$m_i(t) = g(v_i(t))\,. \qquad (21)$$

3: We shift the masks such that their minimum values are put to zero. This does not change the fixed points of the algorithm but speeds up convergence.

$$m_i(t) = m_i(t) - \min_t m_i(t)\,. \qquad (22)$$

4: Compute the new source estimates:

$$s_{i,\text{new}}(t) = m_i(t)s_i(t)\,. \qquad (23)$$

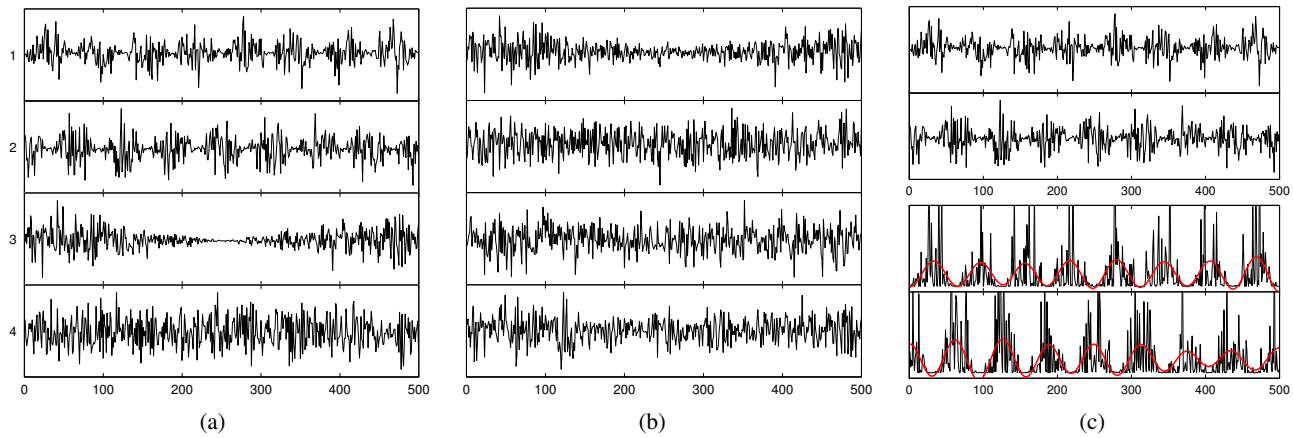

Fig. 2. An artificial example of source separation based on the temporal structure of variances. (a) - Artificially generated sources with structured variances. (b) - Linear mixtures of sources. (c) - Top: The sources extracted by the proposed algorithm with the emphasis on variance oscillations corresponding to eight bursts during the observation period. Bottom: The squares of the found components (black) and their filtered versions $v(t)$ (red).

The proposed algorithm can also be used for subspace analysis where several sources are assumed to share the same variance structure. In this case, the subspace activation can be estimated on Step 1 by taking the average of the squared sources from the same subspace:

$$\{v(t) \,|\, \forall t\} = \mathrm{filter}_f \left\{ \frac{1}{K} \sum_i^K s_i^2(t) \,\middle|\, \forall t \right\} . \qquad (24)$$

Then, the same mask calculated from $v(t)$ is applied to each component from the corresponding K-dimensional subspace.

C. Artificial example

We demonstrate the proposed algorithm on a simple artificial example. The sources in Fig. 2a have prominent variance structures. The variances of the first two sources change periodically with eight activation bursts during the observation period. The phase of the activations is different for the two sources. The third component has slowly changing variance and the variance of the fourth component is stationary. Fig. 2b presents linear mixtures of the sources generated using a randomly chosen mixing matrix. The oscillating variance structure is almost invisible in the observations, it is somewhat possible to capture it only in observation 4.

The algorithm was set to extract two sources whose variances have prominent variability in the timescale corresponding to the activation periodicity of sources 1 and 2. The motivation for investigating this timescale may come, for example, from the visual inspection of observation 4 or from the periodicity of some known phenomena (such as the annual cycle in climatology). The masks are calculated here using Eq. (20) and the sources are estimated one after another by using deflation.

The extracted sources are shown in Fig. 2c to reconstruct the original sources 1 and 2. The prominent activation structure in the timescale of interest is very clear from Fig. 2c. The results show that the proposed algorithm is able to extract the components which are most structured temporally in a chosen timescale. In the presented example, it would also be possible to extract the original component 3 if the emphasis were put on the slowest variance oscillations.

IV. Climate data and preprocessing method

We apply the proposed DSS-based algorithm to surface temperature measurements provided by the reanalysis project of the National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP/NCAR) [14], [15]. The data represent globally gridded measurements daily over a long period of time. The spatial grid is regularly spaced over the globe with $2.5^\circ \times 2.5^\circ$ resolution.

The reanalysis data is not fully real because the missing measurements have been reestimated based on the available data and approximation models. Yet, the data is as close to the real measurements as possible. Although the quality of the data is different, we used the whole period of 1948–2004. Thus, the data contain more than 10,000 spatial locations and about 20,000 time instances.

To preprocess the data, the long-term mean was removed and the data points were weighted to diminish the effect of a denser sampling grid around the poles: each data point was multiplied by a weight proportional to the square root of the corresponding area of its location. The spatial dimensionality of the data was then reduced using the PCA/EOF analysis applied to the weighted data. We retained 100 principal components which explain more than 90% of the total variance, which is due to the high spatial correlation between nearby points on the global grid.

In this work, we are interested in fast changing climate phenomena. Thus, the principal components were further high-passed filtered to retain only high-frequencies. The cut-off frequency of the filter was set such that only the oscillations with a period shorter than one month are passed.

V. Experiments on climate data

A. Fast components with annually structured variance

The aim of the first experiment is to extract fast changing climate components whose variances have prominent *annual*

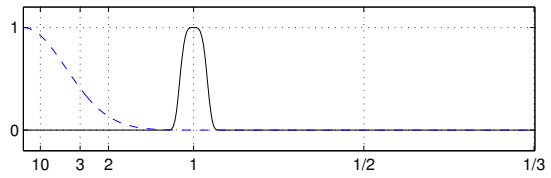

Fig. 3. The frequency response of the filters used for finding components with annually (solid) and slowly (dashed) structured variances from daily measurements. The abscissa is linear in frequency but is labeled in terms of periods, in years.

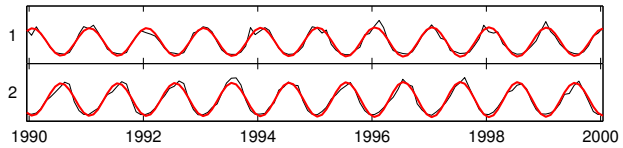

Fig. 4. The activation structure of the two subspaces with annually structured variances for a fragment of data. The upper and lower curves correspond to the subspaces with 47 components and 33 components, respectively. The black curves are the monthly averages of the squared values of all the sources from the same subspace and the red curves are the filtered versions of the black curves.

structures. Therefore, a filter which passes frequencies within a narrow band around the annual frequency (see Fig. 3) was used in the denoising procedure. The masks were computed using Eq. (20).

As the first step, we did exploratory analysis by applying the proposed algorithm several times in order to extract as many annually structured components as possible. Deflation was used to extract distinct components. It turned out that it is possible to find many sources with annually structured variances. For the considered 100-dimensional data, more than eighty found components had very clear annual activation structure. Another observation was that the found components can be categorized into two groups with the same phase of the annual activation. One group contained about 45-50 components while the other group had about 30-35 sources. The rest of the components had relatively weak annual patterns.

These results suggest that there are two subspaces of components with annually structured variances. Thus, we applied the algorithm again in order to extract two subspaces of sources sharing the same variance structure. Eq. (24) was used for the variance estimation in the denoising procedure. The dimensionalities of the subspaces were set to 47 and 33.

Fig. 4 shows a fragment of the variance temporal structures of the two extracted subspaces. The black curves are the monthly averages of the squared values $s_i^2(t)$ for all the sources from the same subspace and the red curves represent their filtered versions. The annual periodic structure of the variances emerges very clearly. The phase of the activation is different for the two subspaces. The first subspace corresponds to the higher temperature variability during Northern Hemisphere (NH) winters and the peaks of the activations for the second subspace are during NH summers.

The spatial patterns corresponding to the two subspaces,

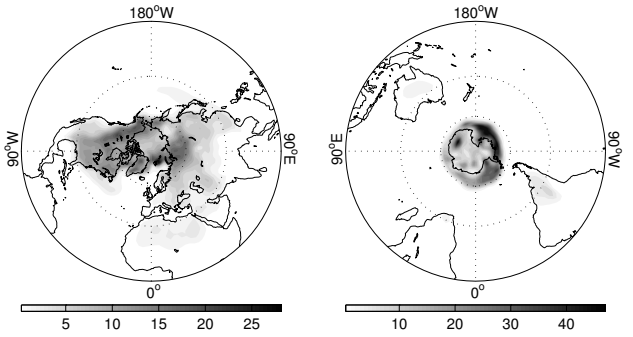

Fig. 5. The observation variance explained by the components of the subspaces with annually structured variances. The pattern shown on the left corresponds to the first subspace (upper curves in Fig. 4) and the pattern shown on the right corresponds to the second subspace (lower curves in Fig. 4).

calculated according to Eq. (6), are shown in Fig. 5 [17]. The group of components with the activations during NH winters has a clear localization in NH with much weaker loadings in the Southern Hemisphere (SH) (the SH pattern is not shown for this subspace). The loadings of the other subspace are, in contrast, mostly localized in SH. The two subspaces capture most of the fast variability of the observations. Thus, when concentrated on the annual variance oscillations, the proposed algorithm can separate two subspaces which reflect higher temperature variability during winters in the Northern and Southern Hemispheres.

B. Fast components with slowly changing variance

The goal of the second experiment is to find prominent components whose variance is structured in the very slow timescale. Therefore, the variance estimation in the denoising procedure includes a filter which emphasizes variance oscillations in the slow, decadal timescale (see its frequency response in Fig. 3). The components were extracted one by one using the deflation approach and the masks were computed using Eq. (21). The extracted components were ordered according to the amount of structure estimated by the measure in Eq. (12). The leading components are therefore considered most prominent for the used separation criterion.

The temporal structures corresponding to the twenty leading components are shown in Fig. 6. Note that the dominant loadings of several components are mostly located either in NH or SH. For such components, we use labels N1-N5 and S1-S5, respectively. The components with prominent patterns in both hemispheres are labeled as B1-B10. It can be seen from Fig. 6 that the annual variability is the dominant variance structure of the extracted components. It is especially clear for the components mostly localized in one of the hemispheres. However, slow variance oscillations also emerge very clearly. In many sources, several successive years have either increased or decreased level of activation.

The temporal structure of many components is quite remarkable. A salient pattern is, for example, component S2 with the constantly increasing activation level. Increasing

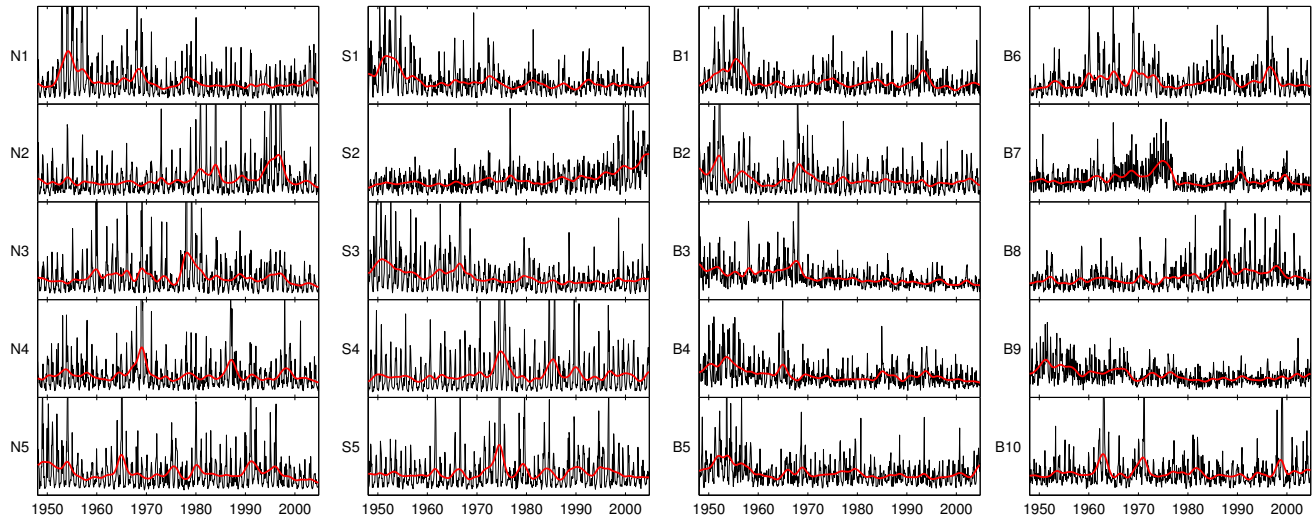

Fig. 6. The monthly averages of the squared components (black) and the variances estimated by low-pass filtering the squared components (red).

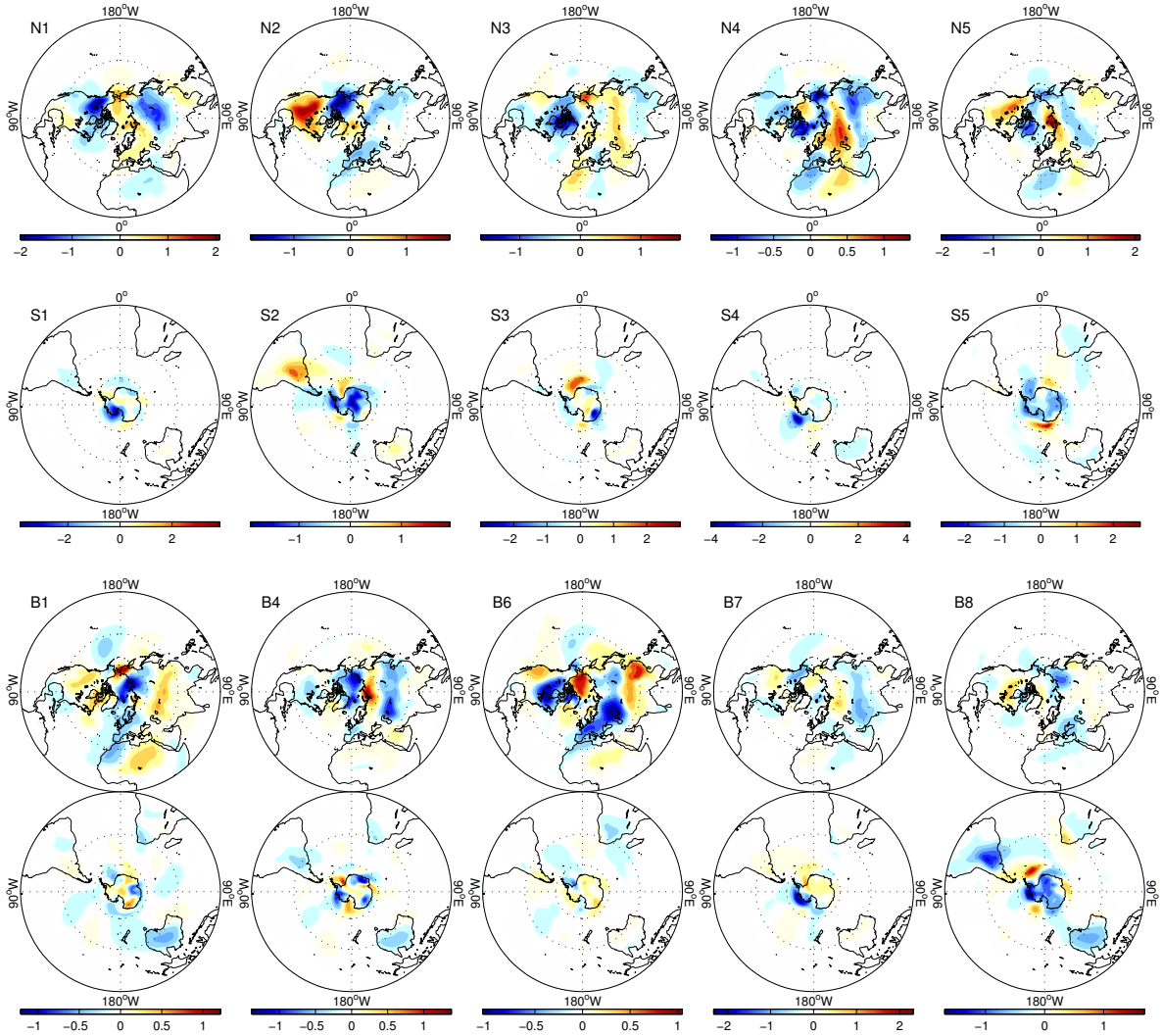

Fig. 7. The spatial patterns of some components with prominent loadings in NH (N1-N5, top row), in SH (S1-S5, second row) and in both hemispheres (B1, B4, B6-B8, bottom rows). The scale of the maps is in degrees centigrade.

variability can also be noticed in B8. On the contrary, some of the sources (e.g. S3, B3, B4, B5, B9) appear to have slowly decreasing activations. Such components may form a common subspace and they should probably be estimated together using the subspace version of the proposed algorithm. Another prominent temporal pattern is B7 with a remarkably abrupt change in the mid-seventies.

The temperature maps corresponding to the extracted components (due to the space limitation, we present only some of them in Fig. 7) also have several prominent patterns, for example, a dipole structure in North America in component B2. The dominant loadings of the aforementioned groups (S3, B3, B4, B5, B9) and (S2, B8) are mostly located in the same regions, which is another indication of the similarities within these groups.

VI. Discussion

The paper presents a new extension of our previous works [6], [7], [8] on applying the DSS framework to the exploratory analysis of climate data. In this work, we proposed an algorithm which seeks components with temporally structured variances. The optimized criterion has some connection to higher-order statistics and can be used for independent component analysis and blind source separation (BSS) for different types of data. It can be proven rigorously that the proposed algorithm can identify the original signals mixed in the data, provided that the signals have non-stationary variances and they are mutually independent.

An advantage of the proposed approach is that it is possible to concentrate on different timescales of data variations by changing the filter used in variance estimation. For solving the BSS problem, the emphasis on a properly chosen timescale can improve the separation results, especially for noisy data when other separation criteria cannot provide reliable components. In the exploratory analysis of data, the method allows for finding different interesting phenomena in the same dataset by concentrating on different timescales.

In this paper, the proposed algorithm was used to analyze fast oscillations of global surface temperatures. When we concentrated on the dominant, annual variance oscillations, two subspaces with different phases of the yearly activations were extracted. The first subspace explains the fast temperature variability in the Northern Hemisphere and has higher activations during NH winters. The second subspace corresponds to the fast oscillations in the Southern Hemisphere with higher activations during NH summers.

In the second experiment, we concentrated on the slower, decadal timescale of the fast temperature oscillations. Several components with prominent temporal and spatial structures were extracted. The meaning of the found components needs to be further investigated, some of them may correspond to significant climate phenomena while others may reflect some artifacts produced during the data acquisition. A third alternative would be that the components may have been ovefitted to the data, but we do not believe so because they do not have the structure typical for overfitted solutions [16]. In contrast, without using the nonlinearity g in Eq. (21), some of the results looked like typical overfits. To be sure, the reliability of the presented results could be tested by cross-validation.

The approach presented in this paper can be extended in many different ways. For example, it would be interesting to relate the components presented here to the known climate phenomena visible as specific projections of global weather data. It would also be possible to use more information for more robust variance estimation. The additional information could be in the form of other components extracted from climate data or a hierarchical variance model as in [18].

References

[1] H. von Storch and W. Zwiers, *Statistical Analysis in Climate Research*. Cambridge, U.K.: Cambridge University Press, 1999.

[2] M. B. Richman, "Rotation of principal components," *Journal of Climatology*, vol. 6, pp. 293–335, 1986.

[3] F. Aires, A. Chédin, and J.-P. Nadal, "Independent component analysis of multivariate time series: Application to the tropical SST variability," *Journal of Geophysical Research*, vol. 105, no. D13, pp. 17,437–17,455, 2000.

[4] A. Lotsch, M. A. Friedl, and J. Pinzón, "Spatio-temporal deconvolution of NDVI image sequences using independent component analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 12, pp. 2938–2942, 2003.

[5] J. Basak, A. Sudarshan, D. Trivedi, and M. S. Santhanam, "Weather data mining using independent component analysis," *Journal of Machine Learning Research*, vol. 5, pp. 239–253, 2004.

[6] A. Ilin, H. Valpola, and E. Oja, "Semiblind source separation of climate data detects El Niño as the component with the highest interannual variability," in *Proceedings of International Joint Conference on Neural Networks (IJCNN'2005)*, Montréal, Québec, Canada, 2005, pp. 1722–1727.

[7] A. Ilin and H. Valpola, "Frequency-based separation of climate signals," in *Proceedings of the 9th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD'2005)*, Porto, Portugal, 2005, pp. 519–526.

[8] A. Ilin, H. Valpola, and E. Oja, "Exploratory analysis of climate data using source separation methods," *Neural Networks*, vol. 19, no. 2, pp. 155–167, 2006.

[9] J. Särelä and H. Valpola, "Denoising source separation," *Journal of Machine Learning Research*, vol. 6, pp. 233–272, 2005.

[10] A. Ilin, "Independent dynamics subspace analysis," in *Proc. of the 14th European Symposium on Artificial Neural Networks (ESANN 2006)*, Bruges, Belgium, April 2006, to appear.

[11] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. John Wiley, 2001.

[12] A. Hyvärinen and P. O. Hoyer, "Emergence of phase and shift invariant features by decomposition of natural images into independent feafture subspaces," *Neural Computation*, vol. 12, no. 7, pp. 1705 – 1720, 2000.

[13] J. F. Cardoso, "Multidimension independent component analysis," in *Proc. ICASSP'98*, Seattle, WA, 1998.

[14] E. Kalnay and coauthors, "The NCEP/NCAR 40-year reanalysis project," *Bulletin of the American Meteorological Society*, vol. 77, pp. 437–471, 1996.

[15] NCEP data, "NCEP Reanalysis data provided by the NOAA-CIRES Climate Diagnostics Center, Boulder, Colorado, USA," November 2004. [Online]. Available: http://www.cdc.noaa.gov/

[16] A. Hyvärinen, J. Särelä, and R. Vigário, "Spikes and bumps: Artefacts generated by independent component analysis with insufficient sample size," in *Proceedings of International Workshop on Independent Component Analysis and Blind Signal Separation (ICA'99)*, Aussois, France, 1999, pp. 425–429.

[17] R. Pawlowicz, "M_Map: A mapping package for Matlab," 2000. [Online]. Available: http://www2.ocgy.ubc.ca/∼rich/map.html

[18] H. Valpola, M. Harva, and J. Karhunen, "Hierarchical models of variance sources," *Signal Processing*, vol. 84, no. 2, pp. 267–282, 2004.