J3.8   FINDING INTERESTING CLIMATE PHENOMENA BY EXPLORATORY STATISTICAL TECHNIQUES

Alexander Ilin[*], Harri Valpola, Erkki Oja
Helsinki University of Technology, Helsinki, Finland

1. INTRODUCTION

Finding a meaningful representation of the observed data is a central problem in many research fields. In many applications, collected data contain measurements of a set of variables x_i obtained at different time instances t. These measurements are typically represented as a set of vectors $\mathbf{x}(t) = [\, x_1(t) \, \ldots \, x_i(t) \, \ldots \, x_N(t) \,]^\mathsf{T}$ collected during the observation period $t = 1, ..., T$. In climatology, such datasets may contain observed weather conditions in certain spatial locations during a certain period of time.

A typical task of statistical data analysis is to find important, interesting data characteristics. Often, these interesting properties are captured by a set of vectors $\mathbf{s}(t) = [\, s_1(t) \, \ldots \, s_j(t) \, \ldots \, s_M(t) \,]^\mathsf{T}$ obtained by a linear transformation of the observed data:

$$\mathbf{s}(t) = \mathbf{W_x}\mathbf{x}(t), \qquad (1)$$

where $\mathbf{W_x}$ is the matrix defining the transformation. The matrix $\mathbf{W_x}$ is found such that the vectors $\mathbf{s}(t)$ have suitable properties. In statistical climatology, weather measurements can be analyzed in this way in order to find physically meaningful patterns of climate variability (see, e.g., Richman 1986).

Several classical criteria optimized to find $\mathbf{W_x}$ in (1) include minimum error of reconstructing $\mathbf{x}(t)$ from $\mathbf{s}(t)$, simplicity of the transformation structure, interestingness of the properties in signals $s_j(t)$, independence of signal structures etc. The resulting techniques are principal component analysis (PCA) or empirical orthogonal functions (EOF) (von Storch and Zwiers 1999), factor analysis (Harman 1967), projection pursuit (Jones and Sibson 1987), independent component analysis (ICA) (Hyvärinen et al. 2001) etc. These methods have become popular techniques for *exploratory data analysis* in many applications including climatology (see, e.g., Richman 1986; Aires et al. 2000; Lotsch et al. 2003; Basak et al. 2004).

Recently, we have used a general algorithmic framework called *denoising source separation* (DSS) (Särelä and Valpola 2005) to implement several criteria defining transformation (1) and applied them to analysis of global climate measurements (Ilin et al. 2005, 2006a). In the analysis, we seek uncorrelated components which

---
[*]*Corresponding author address:* Alexander Ilin, P.O. Box 5400, FI-02015 TKK, Espoo, Finland; e-mail: `Alexander.Ilin@tkk.fi`

maximally express some desired, interesting properties. These components are found using an iterative procedure which gradually isolates the desired signal structures.

DSS is a powerful tool for *exploratory analysis* of large spatio-temporal climate datasets. The motivation for seeking a particular type of components can come from general statistical principles (e.g., maximizing non-Gaussianity of components gives the ICA solution), expert knowledge (e.g., some information about the spectral structure of components), or based on some elementary inspection of data (e.g., by observing some regular patterns in them). For example, in the climate data analysis we might be interested in some phenomena that would have prominent variability in a certain timescale or exhibit slow changes. Then, using a properly designed separation algorithm can help extract the components in which this type of structure is most prominent.

In this paper, we overview several separation criteria used to identify model (1), which are implemeted in the common algorithmic framework of DSS. We also report results on applying the corresponding algorithms to analysis of global weather measurements from the NCEP/NCAR reanalysis data.

The paper is organized as follows. Section 2 outlines the general modeling assumptions of the source separation techniques and presents the common algorithmic framework used in this paper for implementing various separation criteria. The analyzed dataset is described in Section 3.

Section 4 presents an algorithm which allows for extraction of components with the cleanest variability in the timescale of interest. In the experiments, we show that the components extracted as the sources with the most prominent interannual variability are clearly related to the well-known El Niño–Southern Oscillation (ENSO) phenomenon.

Section 5 outlines another frequency-based algorithm which tries to find a representation in which components have as distinct power spectra as possible. In the experiments with global climate data, this algorithm is able to obtain a meaningful representation of the slowest climate phenomena as a combination of trends, decadal-interannual (quasi-)oscillations, the annual cycle and components related to slowly changing seasonal variations.

Section 6 presents the analysis which allows for finding

components whose activations (variances) have prominent temporal structure in the timescale of interest. In the experiments on analysis of global temperature measurements, several fast changing components with remarkable slow behavior of their variances have been extracted.

Section 7 outlines an algorithm whose aim is to find groups of dynamically coupled components with the most predictable time course. The potential application of this approach to analysis of climate data may reveal complex climate phenomena with predictable dynamics.

Finally, in Section 8, we discuss the presented techniques and outline some possible directions of future research.

2. DATA ANALYSIS METHOD

2.1 Linear generative model

Many methods related to linear data analysis (1) are formulated in the context of estimating a generative model

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t)\,, \tag{2}$$

where the observations $x_i(t)$ are modeled as linear mixtures of a number of hidden *sources* $s_j(t)$, and the matrix $\mathbf{A} = [a_{ij}]$ contains the mixing coefficients:

$$x_i(t) = \sum_{j=1}^{M} a_{ij} s_j(t), \ i = 1, ..., N\,. \tag{3}$$

The index i runs over the measurement sensors (typically spatial locations), and discretized time t runs over the observation period: $t = 1, ..., T$. The matrix $\mathbf{A}$ is usually called a *mixing matrix* or a *loading matrix* depending on the context.

If we denote the columns of matrix $\mathbf{A}$ by $\mathbf{a}_i$, then (2) can be written as

$$\mathbf{x}(t) = \sum_{j=1}^{M} \mathbf{a}_j s_j(t)\,. \tag{4}$$

In climate data analysis, the time series $s_j(t)$ 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 $s_j(t)$ and the corresponding loading vectors $\mathbf{a}_i$ from the observed data $\mathbf{x}(t)$ alone. This problem is often referred as *blind source separation* (BSS). This problem cannot be solved without extra assumptions or prior knowledge about the unobserved signals or the loading structure. In exploratory settings, this knowledge usually comes from the inspection of the available data, expert knowledge or tested hypotheses.

2.2 Algorithmic framework for estimation

DSS is a general algorithmic framework which identifies the model (3) by assuming that

1. the sources $s_j(t)$ are uncorrelated;

2. the sources $s_j(t)$ maximize some desired properties.

Examples of the desired properties of interest include non-Gaussianity, slowness, prominent dynamic or activation structure etc. DSS can be viewed as an extension of ICA without the strict independence assumption.

Whitening The first, preprocessing step in the algorithmic framework of DSS is called *whitening*. The goal of whitening is to make the covariance structure of the data uniform in such a way that any linear projection of the data has unit variance. The positive effect of such a transformation is that any orthogonal basis in the whitened space defines *uncorrelated* sources. Therefore, this allows for restricting the mixing matrix to be orthogonal afterwards.

Whitening is usually implemented by PCA. Denoting by $\mathbf{X}$ the matrix which contains the observed vectors $\mathbf{x}(t)$ in its columns and assuming that the measurements $x_i(t)$ have been normalized to zero mean, the matrix of sphered data $\mathbf{Y}$ is calculated as

$$\mathbf{Y} = \mathbf{D}^{-1/2}\mathbf{V}^{\mathsf{T}}\mathbf{X}\,, \tag{5}$$

where $\mathbf{D}$ is the diagonal matrix of eigenvalues of the data covariance matrix, defined as $\frac{1}{T}\mathbf{X}\mathbf{X}^{\mathsf{T}}$. The columns of matrix $\mathbf{V}$ are the corresponding eigenvectors. Multiplication by $\mathbf{V}$ corresponds to orthogonal rotation of the data using the directions which maximize the data variance (see, e.g., Diamantaras and Kung 1996) and $\mathbf{D}^{-1/2}$ scales the principal components to unit variance. Each column $\mathbf{y}(t)$ of the matrix $\mathbf{Y}$ is a linear transformation of the observation vector $\mathbf{x}(t)$. It is easy to show that the covariance matrix of the whitened data equals the identity matrix, that is $\frac{1}{T}\mathbf{Y}\mathbf{Y}^{\mathsf{T}} = \mathbf{I}$. The dimensionality of the data can also be reduced at this stage by retaining only the principal components corresponding to the largest eigenvalues in $\mathbf{D}$.

Matrix $\mathbf{Y}$ of whitened data is not unique: Any *orthogonal* rotation of its columns produces a matrix

$$\mathbf{S} = \mathbf{W}\mathbf{Y} \tag{6}$$

that also has unit covariance:

$$\frac{1}{T}\mathbf{S}\mathbf{S}^{\mathsf{T}} = \frac{1}{T}\mathbf{W}\mathbf{Y}\mathbf{Y}^{\mathsf{T}}\mathbf{W}^{\mathsf{T}} = \mathbf{I} \tag{7}$$

Therefore, a set of uncorrelated sources $\mathbf{S}$ can be found by using (6) with the restriction that $\mathbf{W}$ is an orthogonal
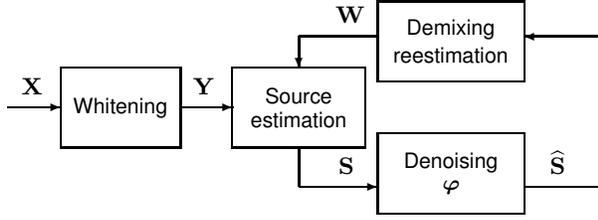
Figure 1: The general iterative procedure of the algorithmic framework of DSS.

matrix. Note that each column $\mathbf{s}(t)$ of the matrix $\mathbf{S}$ is a linearly transformed vector $\mathbf{y}(t)$: $\mathbf{s}(t) = \mathbf{W}\mathbf{y}(t)$. Matrix $\mathbf{W}$ (or the overall transformation matrix $\mathbf{W_x} = \mathbf{W}\mathbf{D}^{-1/2}\mathbf{V}^{\mathsf{T}}$) is often called a *demixing matrix* in the ICA literature.

Iterative procedure based on denoising The rotational ambiguity of the whitened data matrix expressed in (6) can be fixed by using the requirement that the sources maximize some desired properties. In order to optimize these properties, one could maximize a quantity $\mathcal{F}(\mathbf{S})$ that would measure the amount of interesting structure in the source signals $\mathbf{S}$. The measure $\mathcal{F}$ is a function of $\mathbf{W}$ because the sources are defined by demixing through (6). Therefore, $\mathcal{F}$ can be optimized w.r.t. $\mathbf{W}$ using the following gradient-based procedure (see details in Särelä and Valpola 2005; Ilin 2006a):

1. Source estimation using (6) using the current estimate of the demixing matrix $\mathbf{W}$;

2. Updating the source estimates using the gradient-based step:

$$\widehat{\mathbf{S}} = \mathbf{S} + \mu\frac{\partial\mathcal{F}}{\partial\mathbf{S}} = \varphi(\mathbf{S})\,; \tag{8}$$

3. Reestimation of the demixing matrix:

$$\mathbf{W}^{\mathsf{T}} = \mathrm{orth}(\mathbf{Y}\widehat{\mathbf{S}}^{\mathsf{T}})\,, \tag{9}$$

where $\mathrm{orth}(.)$ is an operator giving the orthogonal projection of the matrix $\mathbf{Y}\widehat{\mathbf{S}}^{\mathsf{T}}$ onto the set of orthogonal matrices.

The corresponding iterative procedure is presented in Fig. 1. The iterations continue until the source estimates do not change.

The basic idea of the algorithmic framework of DSS is to design separation algorithms following the general sequence of steps presented in Fig. 1. The separation criterion is introduced in the procedure in the form of a suitably chosen *denoising function* φ. In case the algorithm is derived from an optimized measure $\mathcal{F}$, the corresponding denoising function is given by (8). For many

practical cases, however, it can be easier to construct an update rule

$$\widehat{\mathbf{S}} = \varphi(\mathbf{S}) \tag{10}$$

with a sensible function φ than to derive a gradient-based rule (8) from an objective function. First, the interesting signal structure could be difficult to measure using a simple index $\mathcal{F}$. Second, the derivation of the gradient $\partial\mathcal{F}/\partial\mathbf{S}$ could be cumbersome, especially for complex $\mathcal{F}$. It is also possible that the gradient-based update rule in (8) is not robust as, for example, it can be sensitive to some particular values in $\mathbf{S}$.

In general, the denoising function $\varphi(\mathbf{S})$ should be designed such that it emphasizes the desired (interesting) properties of the signal and removes irrelevant information from $\mathbf{S}$. It can represent a gradient-based update rule or its modification. Sometimes, it is possible to derive an appropriate denoising function from rather heuristic principles.

In DSS terminology, the iterative procedure in Fig. 1 is usually interpreted as extension of the power method for computing the principal components of $\mathbf{Y}$. Without denoising, this procedure is indeed equivalent to the power method, because then Steps 1 and 3 give $\mathbf{w} = \mathrm{orth}(\mathbf{Y}\mathbf{Y}^{\mathsf{T}}\mathbf{w})$. Since $\mathbf{Y}$ is white, all the eigenvalues are equal and the solution without denoising becomes degenerate. Therefore, even slightest changes made by denoising φ can determine the rotation. Since the denoising procedure emphasizes the desired properties of the sources, the algorithm can find the rotation where the properties of interest are maximized.

It should be noted that the presented procedure is very general. The essential part of any specific algorithm implemented in this framework is the denoising procedure. In fact, many existing ICA algorithms fall into the pattern of DSS although they have been derived from other perspectives, typically from a properly chosen cost function (e.g., Hyvärinen and Oja 1997; Barros and Cichocki 2001; Cichocki et al. 2002).

Linear filtering In some cases, the interesting properties of a source signal can be obtained by applying a linear temporal filter. For example, the sources are sometimes known to be cyclic over a certain period of time or to have prominent variability in a certain timescale and a filtering procedure, which passes interesting spectral components and removes all other frequencies, would emphasize this characteristic structure of the sources.

A relevant quantity that measures the amount of interesting structure is the ratio of the variances of a signal after and before filtering:

$$\mathcal{F}(s) = \frac{\mathrm{var}\, s_{\mathrm{f}}}{\mathrm{var}\, s}\,, \tag{11}$$

where s_{f} is the filtered version of the signal s. The mea-

Figure 2: The steps of the separation algorithm in the special case of linear temporal filtering.

sure (11) can be understood as the relative amount of energy contained in the interesting part of the signal and it attains its maximum value of unity if filtering does not change the signal. We use the term *clarity* for this quantity.

The sources with the most prominent characteristics captured by linear filtering could be estimated one after another by maximizing the measure (11) using the iterative procedure outlined in Fig. 1. Then, the denoising procedure simply corresponds to linear temporal filtering. However, in many practical cases, such estimation can be performed in just three steps shown in Fig. 2: the preprocessing step of whitening is followed by filtering and PCA (Särelä and Valpola 2005).

The intuition behind this approach is that filtering on the second step renders the variances of the sphered components different and the covariance matrix of $\widehat{\mathbf{Y}}$ is no more equal to the identity matrix. Note that in many practical situations, this filtering can be done using the same filter as the one used to obtain s_f in (11). Then, PCA can identify the directions which maximize the properties of interest. The eigenvalues obtained from PCA on the third step give the values of the objective function $\mathcal{F}$ for the found sources. Thus, the components are ranked according to the prominence of the desired properties (their clarity values) the same way as the principal components in PCA are ranked according to the amount of variance they explain.

Linear temporal filtering is in practice implemented using a filtering matrix $\mathbf{F}$ operating on the rows of matrix $\mathbf{Y}$: $\widehat{\mathbf{Y}} = \mathbf{YF}$. In the case of frequency-based filtering, this matrix can be designed, for example, based on the orthogonal matrix of the DCT basis. More details about the DSS method were reported by Särelä and Valpola (2005).

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 (2). From (2), (6) and (5), we obtain

$$\mathbf{X} = \mathbf{AS} = \mathbf{AWY} = \mathbf{AWD}^{-1/2}\mathbf{V}^\mathsf{T}\mathbf{X}. \qquad (12)$$

Thus $\mathbf{A}$ should be chosen as the (pseudo)inverse of $\mathbf{WD}^{-1/2}\mathbf{V}^\mathsf{T}$ which is

$$\mathbf{A} = \mathbf{VD}^{1/2}\mathbf{W}^\mathsf{T}. \qquad (13)$$

Since the extracted components $s_i(t)$ are normalized to unit variances, the columns of $\mathbf{A}$ have a meaningful scale.

Note that the signs of the extracted components cannot generally be determined (which is a well-known property of classical ICA) except in some special cases.

3. DATA

In the experiments, we use the proposed techniques to analyze global weather measurements provided by the reanalysis project of the National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP/NCAR) (Kalnay et al. 1996; NCEP data 2004). We analyze measurements of three major atmospheric variables: surface temperature, sea level pressure and precipitation, which are often used for describing global climate phenomena such as ENSO (Trenberth and Caron 2000).

The data represent globally gridded measurements over a long period of time. The spatial grid is regularly spaced over the globe with $2.5° \times 2.5°$ resolution. Although the quality of the data considerably varies in time and throughout the globe, we used the whole period of 1948–2004. Thus, the data is very high-dimensional: more than 10,000 spatial locations by more than 20,000 time instances for each of the three datasets.

The drawback of the reanalysis data is that they are not fully real. The measurements missing in some spatial locations or time instances have been reestimated based on the available data and approximation models. Yet, the data is as close to the real measurements as possible. The main advantage of the analyzed dataset is its regularity which makes the proposed statistical techniques applicable to the analysis of this data.

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. This produced the original data matrix $\mathbf{X}$. The spatial dimensionality of the data was then reduced using the PCA/EOF analysis applied to the weighted data. For each dataset, we retained 100 principal components. This means that for each variable the columns of $\mathbf{Y}$ in (5) have dimension 100, while those of the original $\mathbf{X}$ are over 10,000 dimensional. Yet the principal components explain more than 90% of the total variance, which is due to the high spatial correlation between nearby points on the global grid. The proposed techniques are then applied to the resulting principal components.

4. COMPONENTS WITH PROMINENT VARIABILITY IN CERTAIN TIMESCALE

In the first experiment, we are interested in components which would exhibit prominent slow behavior in the
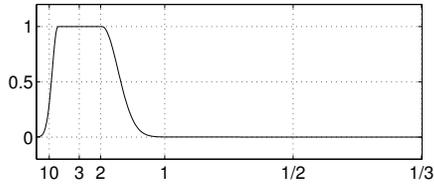
Figure 3: The frequency response of the filter used to emphasize the timescale of the interesting interannual variations. The abscissa is linear in frequency but is labeled in terms of periods, in years.

timescale of interest.

Components with prominent slow behavior can be extracted from the data using DSS with low-pass or band-pass filtering as denoising. This type of denoising is linear and therefore the simple algorithm described in Fig. 2 is applicable here. The extracted slow components are ranked according to their clarity values. That is, the first component contains the least relative amount of other frequencies in its power spectrum.

4.1 ENSO as Component with Most Prominent Interannual Variability

In the experiment described next, we aim at finding components which exhibit prominent variability in the interannual timescale. Therefore, the band-pass filter whose frequency response is shown in Fig. 3 is used as the linear temporal filtering.

Fig. 4 presents the monthly averages of the four components with the most prominent interannual variability extracted from the data combining the three analyzed variables. The time course of the first component (upper curve in Fig. 4) shows striking resemblance with the El Niño index calculated from the sea surface temperature (SST) in the Niño 3 region (presented in Fig. 4). The correlation coefficients between the extracted component and the Niño 3 SST index is 0.9323. Note that the upper components are extracted from climate data consisting of daily measurements from the whole globe, with the only constraint being the emphasis on strong interannual variability.

The spatial patterns corresponding to the two leading components are shown in Fig. 5[*]. The first surface temperature map contains many features traditionally associated with El Niño (Trenberth and Caron 2000). The corresponding sea level pressure map is similar to the classical Southern Oscillation pattern (Trenberth and Caron 2000) and the precipitation map contains many features associated with the ENSO phenomenon. Similar ENSO fea-

---

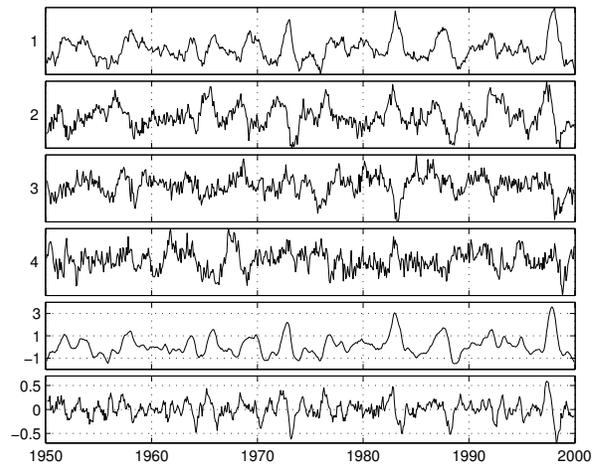[*]The maps are plotted using the mapping toolbox developed by Pawlowicz (2000).

Figure 4: Top four: Monthly averages of the four components with the most prominent interannual variability extracted from the dataset combining surface temperature, sea level pressure and precipitation. Bottom two: The Niño 3 SST index (Niño 3 SST 2004) and its derivative. They bear similarity to the first and second components respectively.
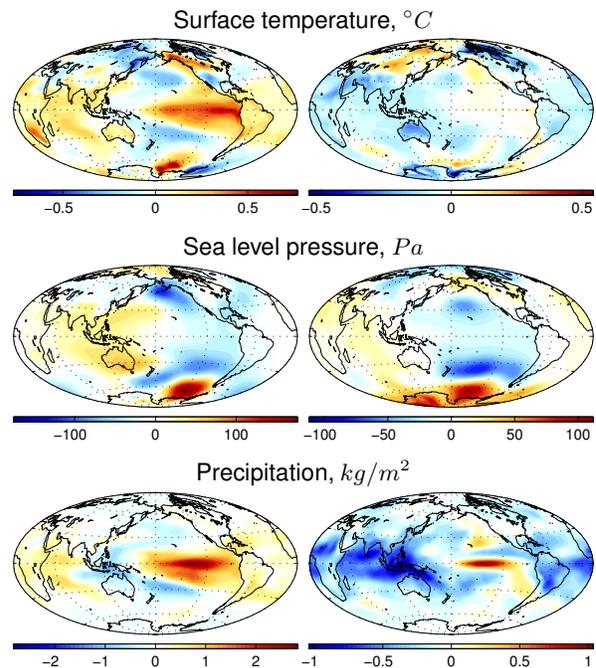

Figure 5: Spatial patterns corresponding to the first (left) and second (right) components with the most prominent interannual variability. The corresponding time courses are the two topmost curves in Fig. 4.

tures can be observed from the regression patterns calculated using the Niño 3 SST index (see Ilin et al. 2006a).

The second extracted component also appears to be related to ENSO and it roughly captures some features of the time derivative of the El Niño index (see Fig. 4). The corresponding precipitation pattern has an interesting localization in the Niño 3 region and mostly negative loadings in the rest of the tropical and subtropical areas. The third and the fourth components show weaker oscillations in the interannual time scale. Similar components also emerge in the results of the spectral separation reported in Section 5.

The same clarity-based analysis was applied to the three datasets separately and the first extracted component was always a good ENSO index. Somewhat surprisingly even the component extracted from sea-level-pressure data resembled more the Niño 3 SST index than Southern Oscillation Index (SOI) although SOI is defined in terms of sea level pressure. More details on the presented experiments can be found in our articles (Ilin et al. 2005, 2006a).

5. COMPONENTS WITH DISTINCT SPECTRAL STRUCTURES

5.1 Spectral separation

The algorithm described in the previous section is useful for extracting components with prominent variability in a certain frequency range. This requires some knowledge about the expected power spectra of the components of interest. Typically, this information does not exist and the prominent spectral characteristics of the sources should be estimated automatically.

In this section, we discuss the algorithm which can be seen as an extension of the previous approach and which achieves signal separation based on the assumption that the sources have *distinct power spectra*. Similarly to the previous approach, the interesting signal properties are emphasized by linear temporal filtering. However, since the sources are expected to have distinct frequency contents, an *individual* filter is applied to each source. The characteristic spectral properties of the sources are not known in advance, and therefore the filters are adjusted to the prominent spectral characteristics of the sources which emerge during the learning procedure. This approach is implemented using the general sequence of steps presented in Fig. 1 in which the denoising function performs temporal filtering using a set of *adaptive* filters.

The corresponding denoising procedure is briefly outlined in the following (see more details in Ilin et al. 2006a; Ilin 2006a). The first step of the denoising procedure is to compute the power spectra of the current source estimates. This gives an idea about the characteristic spec-
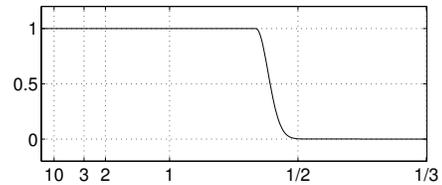

Figure 6: The frequency response of the filter used in the experiments on spectral separation of slow climate components. The abscissa is linear in frequency but is labeled in terms of periods, in years.

tral properties of each source and suggests which frequencies should be emphasized by filtering. The next step is to calculate the individual filters such that the corresponding frequency patterns are made as distinct as possible compared to the other sources. Such a competition procedure naturally requires that all the sources are estimated simultaneously. The competition mechanism is in practice implemented using rather heuristic principles and it is based on partial whitening the power spectra. The final step of the denoising procedure is filtering the source estimates using the filters adapted to the emerging spectral characteristics of the sources. Note that the presented algorithm essentially performs ICA in the frequency domain.

5.2 Frequency-Based Representation of Slow Climate Variability

Steps of the analysis In the experiments, we applied the proposed spectral separation analysis to obtain interesting representation of the slow climate variability. We analyzed the dataset combining three variables: surface temperature, sea level pressure and precipitation.

The first step of the analysis was to identify the subspace of the slowest climate variations using the clarity-based approach described in the previous section. The emphasis was put on the slow variations in the frequency band shown in Fig. 6. Therefore, this filter was used for linear temporal filtering in the three-step procedure described in Fig. 2.

The cleanest components extracted on this stage had prominent patterns both in the time course and in the spatial loadings (see their depiction in Ilin and Valpola 2005). The annual cycle appeared in the two leading components as the cleanest slow source of climate variability. The following components contained mixtures of clearly emerging trends, ENSO oscillations, and other quasi-oscillating slow components. Except for the two annual cycle sources, none of the components had a clear dominant peak in its power spectrum.

The second step of the analysis was to analyze the

cleanest slow components using the algorithm outlined in Section 5.1. This yielded a representation in which rotated components have as distinct spectral properties as possible. We analyzed only 16 cleanest components at this stage as the results were easily interpretable for this number of components. This procedure roughly categorized the 16 resulting components into three groups (also called subspaces): five components with slow trends, six components with prominent interannual variability and five components oscillating with close-to-annual frequencies.

Based on the obtained results, we found it possible to improve the representation within the three subspaces. The subspace of trends was rotated using the clarity-based analysis similarly to Section 4. The filter shown in Fig. 6 was used to specify the interesting frequency range for calculating clarity values. The subspace with prominent interannual variations was rotated by applying again the spectral separation algorithm which makes the spectra as distinct as possible. This somewhat improved the representation within the subspace because the spectral characteristics of the components from the other two subspaces did not affect the competition mechanism on this stage. More details on the steps of the analysis can be found in our journal paper (Ilin et al. 2006a).

Results The time courses of the 16 slow components obtained by the described procedure are shown in Fig. 7. The spatial patterns corresponding to components 1, 10 and 11 are presented in Fig. 7. The loadings corresponding to the rest of the components can be found in the journal article (Ilin et al. 2006a).

The first five sources are the slowest trends. Their power spectra are very similar and therefore their good separation cannot be guaranteed by the spectrum competition procedure. One would naturally require a much longer observation period in order to distinguish differences in the frequency contents of the slowest climate phenomena. Some other criteria such as the spatial localization of the components might give a more meaningful representation in this subspace.

The first component with the constantly increasing time course is most prominent among these sources. This component might be related to global warming as the corresponding surface temperature map has mostly positive values all over the globe (see the corresponding spatial patterns in Fig. 7). The highest temperature loadings of this component are mainly concentrated around the North and South Poles and the sea level pressure map has a clear localization around the South Pole. The precipitation loadings are mostly located in the tropical regions with negative values over the oceans and North Africa and with prominent positive values in the Australian-
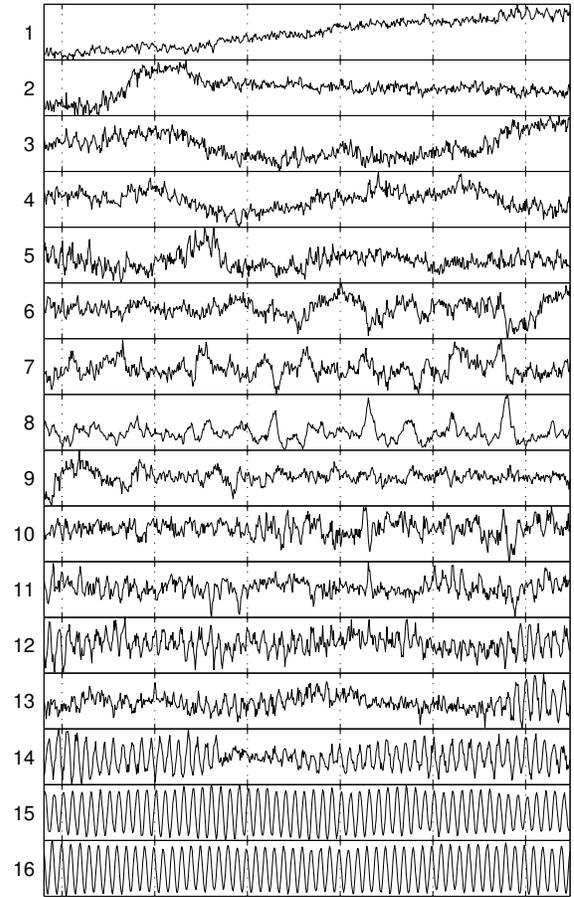

Figure 7: The time course of the components rotated within the subspaces.

Indonesian region, near the Peruvian coast and in South Africa. The other extracted trends might be related to climate phenomena oscillating in the multidecadal time scale such as, for example, the Atlantic Multidecadal Oscillation (Enfield et al. 2001).

The following six components 6–11 exhibit prominent quasi-oscillatory behavior in the interannual timescale. The most prominent sources here are components 7 and 8 which are related to ENSO. These components are very similar to the first two components with the most prominent interannual variability presented in Section 4.1 (see Fig. 4-5). Component 8 is similar to the ENSO index and component 7 bears resemblance with the differential ENSO index. The correlation coefficient of component 8 is 0.90 for the Niño 3 SST index and -0.67 for SOI. The correlation coefficient between component 7 and the differential El Niño is 0.40.

Components 6 and 11 resemble the third and fourth components obtained in the experiments reported in Section 4.1. Component 6 might be related to slowly changing aspects of the ENSO phenomenon as its loadings are
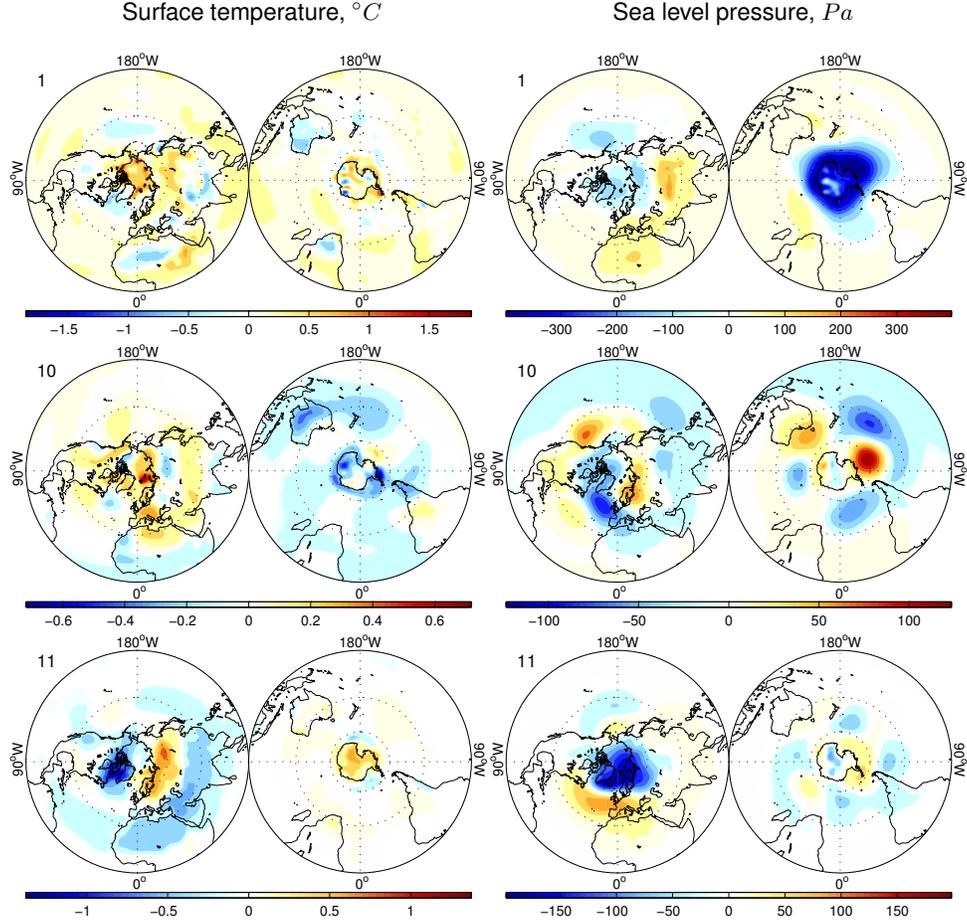
Figure 8: The spatial patterns of components 1, 10 and 11 obtained by frequency-based representation of the 16 most prominent slow components extracted from the data combining surface temperature, sea level pressure and precipitation.

mostly localized in the ENSO regions. Component 11 has quite distinct spatial patterns with a prominent temperature dipole in the Northern Hemisphere and a dominating sea level pressure dipole somewhat resembling the North Atlantic Oscillation or Arctic Oscillation patterns. This component may be related to slowly changing aspects of these phenomena. The correlation coefficient to the Arctic Oscillation index is 0.42.

Component 10 might be related to the interactions of ENSO with the annual cycle. It is well known that ENSO has different effects depending on the time of the year (Trenberth and Caron 2000). This can be modeled by a varying mixing matrix $\mathbf{A} = \mathbf{A}(t)$ whose columns $\mathbf{a}_i$ change throughout the year. The first order approximation yields

$$\mathbf{a}_i(t) = \mathbf{a}_{i,1} + \mathbf{a}_{i,2}s_s(t) + \mathbf{a}_{i,3}s_c(t)\,,$$

where $\mathbf{a}_{i,1}$ are loading vectors of the constant effect, $s_s(t)$

and $s_c(t)$ are the sine and cosine components of the annual oscillations and $\mathbf{a}_{i,2}$, $\mathbf{a}_{i,3}$ are the loading vectors of the seasonally changing effects. This is equivalent to having extra components in model (4):

$$\mathbf{a}_i(t)s_i(t) = \mathbf{a}_{i,1}s_i(t) + \mathbf{a}_{i,2}s_{i,2}(t) + \mathbf{a}_{i,3}s_{i,3}(t)\,,$$

where $s_{i,2}(t) = s_s(t)s_i(t)$ and $s_{i,3}(t) = s_c(t)s_i(t)$ are the annual oscillations modulated (multiplied) by the climate source $s_i(t)$. Component 10 has characteristic spindles happening during El Niño episodes. The frequency of these spindles corresponds to the frequency of the ENSO signal modulated by the annual oscillations.

The last set of the sources are components 12–16 which contain prominent oscillations whose frequencies are close to the frequency of the annual oscillations. The dominating components here is the annual cycle (components 15–16). The rest of the sources resemble the annual oscillations modulated (multiplied) by very slow components. Thus, this set of components might be related

to some phenomena slowly changing the annual cycle. Since the power spectra of these components are quite similar, good separation may not have been achieved here. Some other criteria may be better for finding a more meaningful representation within this subspace.

6. COMPONENTS WITH TEMPORALLY STRUCTURED VARIANCES

The previous sections considered algorithms for extracting prominent components with slowly changing time course. However, interesting slow behavior can be found in fast changing components as well. In our recent conference paper (Ilin et al. 2006b), we introduce an algorithm which seeks fast components with prominent temporal structure of variances. 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 possibility to put emphasis on different timescales.

In order to derive the algorithm, we make a practical assumption that each interesting source $s_j(t)$ in model (4) represents a realization of a zero-mean Gaussian process with changing (non-stationary) variance $v_j(t)$. The idea of the analysis is to extract the components whose variances $v_j(t)$ exhibit the most prominent variations in the timescale of interest. Then, the amount of interesting structure in a signal can be measured by the quantity

$$\mathcal{F} = \frac{1}{2} \log \frac{1}{T} \sum_t v(t) - \frac{1}{T} \sum_t \frac{1}{2} \log v(t) \geq 0, \quad (14)$$

which is always non-negative and attains its minimum value of zero if the process variance $v(t)$ is stationary.

In order to use the proposed measure, one needs to estimate the variances $v(t)$ of a signal at each time instant. The assumption that variances $v(t)$ have prominent variability in the *specific timescale* helps estimate $v(t)$ from one realization of the stochastic process: 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.

The measure $\mathcal{F}$ in (14) is a function of the variances $v(t)$ which are estimated from the source values $s(t)$. Thus, $\mathcal{F}$ is a function of $s(t)$ and can be maximized w.r.t. $s(t)$ using the gradient ascent method explained in Section 2.. The required gradient can be approximated as

$$\frac{\partial \mathcal{F}}{\partial s(t)} \approx \frac{\partial \mathcal{F}}{\partial v(t)} s(t), \quad (15)$$

which yields the denoising function

$$\widehat{s}(t) = g(v(t)) s(t). \quad (16)$$

The values $g(v(t))$ can be termed *masks* as they are applied to the current source estimates to get the new ones.

In our conference paper (Ilin et al. 2006b), we proposed to use a nonlinearity $g(v) = \beta + \tanh(\alpha v)$, with α, β some constants, which can be understood as an approximation of $\partial \mathcal{F}/\partial v$ under some assumptions.

Thus, the denoising procedure $\varphi(\mathbf{S})$ that corresponds to this analysis has the following steps. It starts with estimating the local variances $v(t)$ by filtering the squared signal values using a filter which preserves only the frequencies in the range of interest. Then, the nonlinearity g is applied to the variance estimates in order to calculate the masks $g(v(t))$. In order to emphasize the dominant signal activations, the constant β was chosen such that the minimum values of the masks are put to zero. This does not change the fixed points of the algorithm but speeds up convergence. Finally, the denoised source estimates are calculated by applying the mask to the current source values, as in (16).

6.1 Fast Changing Temperature Components with Prominent Slow Activation Structures

In our conference paper (Ilin et al. 2006b), we present examples of the analysis using the proposed algorithm. We analyze fast surface temperature variations from the NCEP/NCAR reanalysis data. Therefore, the temperature principal components were high-passed preprocessed before the analysis.

When we concentrate on the dominant, *annual* variance variations, two subspaces with different phases of the yearly activations are extracted. The first subspace explains the fast temperature variability in the Northern Hemisphere and has higher activations during Northern Hemisphere (NH) winters. The second subspace corresponds to the fast oscillations in the Southern Hemisphere with higher activations during NH summers. These results are presented in more details in our paper (Ilin et al. 2006b).

In the second experiment, we concentrate on slower, *decadal* timescale of the fast temperature variations. The corresponding filter used for calculation of source activations is shown in Fig. 9. Several components with prominent temporal and spatial structures are extracted. Fig. 10 reproduces the temporal patterns of some of the components found in the data. Note that the activation structure of each component has the annual periodicity. However, the slow structure of the variances emerges very clearly as well.

The temporal patterns of many components is quite remarkable. A salient pattern is, for example, component 4 with an increasing activation level. Some of the other sources (which are not presented here) also appear to have either slowly decreasing or increasing activations.
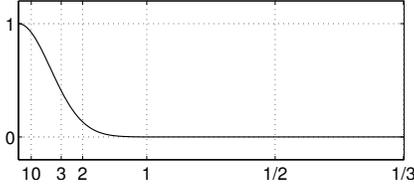
Figure 9: The frequency response of the filter used for finding fast temperature components with slowly structured variances. The abscissa is linear in frequency but is labeled in terms of periods, in years.
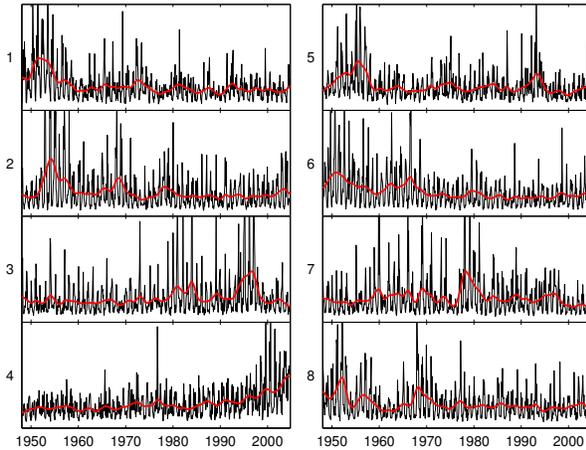

Figure 10: The temporal patterns of the fast changing components extracted from surface temperature measurements. The black curves are the monthly averages of the squared components and the red curves represent activations estimated by low-pass filtering the squared components.

Such components may form a common subspace and they should probably be estimated together by assuming the same activation structure for all of them.

7. COMPONENTS WITH MOST PREDICTABLE TIME COURSE

Frequency-based approach can find a meaningful representation of complex multidimensional data as it can separate different phenomena by the timescales of their prominent variations. This approach is not applicable, however, if the mixed phenomena have similar frequency contents. In this case, a combined time-frequency analysis (see, e.g., Särelä and Valpola 2005) could be useful provided that interesting spectral components of different sources have distinct activation structures. However, the time-frequency analysis is difficult when the observation period is short compared to the timescale of the interesting data variations.

It is also possible that several components are related to the same phenomenon and their separation is not really possible. This might be the case in the climate data. Climate phenomena constantly interact with each other and cannot be independent. The results of the frequency-based analysis of the slow climate variability support this idea. Several components extracted in the experiments reported in Sections 4.1 and 5.2 seem to be related to ENSO. One component gives a good ENSO index, one somewhat resembles a differential ENSO index and one might be related to the seasonal changes of ENSO.

Most probably, different climate phenomena could be described by multidimensional dynamic processes and a meaningful separation criterion would be making the dynamics of different groups of components as decoupled as possible.

The model reported by Ilin (2006b) implements the aforementioned assumptions. There, the sources are decomposed into groups

$$\mathbf{s} = \begin{bmatrix} \mathbf{s}_1^\mathsf{T} & \dots & \mathbf{s}_k^\mathsf{T} & \dots & \mathbf{s}_K^\mathsf{T} \end{bmatrix}^\mathsf{T} . \quad (17)$$

Each group $\mathbf{s}_k$ is assumed to be of known dimensionality and to follow an independent first-order nonlinear dynamic model:

$$\mathbf{s}_k(t) = \mathbf{g}_k(\mathbf{s}_k(t-1)) + \mathbf{m}_k(t) , \quad k = 1, \dots, K , \quad (18)$$

where $\mathbf{g}_k$ is an unknown nonlinear function and $\mathbf{m}_k(t)$ accounts for modeling errors and noise. Assuming separate $\mathbf{g}_k$ in (18) means that the subspaces have decoupled dynamics, that is sources from one subspace do not affect the development of sources from other subspaces.

Without loss of generality, we can retain the assumption that all the sources are mutually uncorrelated and have unit variances. The sources from different subspaces are uncorrelated due to independence and the correlations within the subspaces can always be removed by a linear transformation (whitening). Note that the proposed subspace model can identify the sources only up to linear rotations within the subspaces, which is a known indeterminacy of multidimensional ICA (Cardoso 1998).

Each subspace is estimated so as to minimize the prediction error of the corresponding subspace dynamic model in (18). Hence, the minimized objective function is

$$\mathcal{C} = \frac{1}{2} \sum_t \|\mathbf{s}_k(t) - \mathbf{g}_k(\mathbf{s}_k(t-1))\|^2 . \quad (19)$$

The source values are calculated using the separating structure in (6), and therefore $\mathbf{s}_k(t) = \mathbf{W}_k \mathbf{y}(t)$, where each row of the matrix $\mathbf{W}_k$ defines one source of the k-th subspace. The objective function (19) should be optimized w.r.t. the nonlinear function $\mathbf{g}_k$ and the sources

$s_k(t)$ with the constraint that the demixing matrix is orthogonal. This can be done using the general algorithmic framework outlined in Fig. 1. Therefore, the corresponding denoising procedure alternately updates g_k and $s_k(t)$.

The nonlinearity g_k is updated so as to minimize the cost function (19) keeping the current source estimates $s_k(t)$ fixed. The exact implementation of this step depends on the chosen mathematical model for g_k. For example, one can train a multi-layer perceptron model for g_k using the standard backpropagation procedure with some sort of regularization (see, e.g., Haykin 1999). The update of the sources $s_k(t)$ is done using the gradient descent step, similarly to (8), in which the required gradients depend on the mathematical model for g_k.

8. FUTURE DIRECTIONS

In this paper, we presented several examples how the source separation algorithms can be used for exploratory analysis of global climate measurements. Different separation criteria can be implemented following the same algorithmic framework of DSS by an appropriate choice of operations in the denoising procedure.

We showed how the DSS framework can be tuned to incorporate different separation criteria which proved useful for exploratory analysis of climate data. We used a clarity criterion to extract components with the most prominent interannual variability and a spectral separation criterion to identify slow varying climate phenomena with distinct variability timescales. More results including the time courses of the estimated components and the corresponding spatial patterns can be found online at http://www.cis.hut.fi/alexilin/climate/.

The presented algorithms can be used for finding a physically meaningful representation of the data, for an easier interpretation of the complex climate variability or for making long-term weather forecasts. The meaning of the obtained results still needs to be further investigated, as some of the found components 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 overfitted to the data. In some of the experiments, for example, in the extraction of components with structured variance, some of the results looked like typical overfits. To be sure, the reliability of the results could be tested by cross-validation.

The results of the analysis open up many possible directions for future research. The results on prominent slow climate variability presented in Sections 4–5 suggest that there might be phenomena that could be described by multidimensional processes with complex nonlinear dynamics. This makes the model presented in Section 7 potentially useful in this application. The fact that there are climate phenomena like ENSO which can be observed in different weather variables (such as temperature, air pressure, precipitation) raises the question whether there are other climate phenomena like that. It might be that such phenomena manifest themselves in more complicated ways in the observables and could be extracted using more complex (nonlinear, hierarchical) models.

Other nonlinear effects should also be taken into account because they are known to exist between the state variables. For example, some climate phenomena may affect the fast variations of the weather conditions in certain spatial locations. Also, the most prominent phenomenon in the climate system is the annual cycle and it is quite plausible to assume that climate phenomena may have different effects depending on the time of the year. Then, the combined effect has a nonlinear component as we showed in Section 5.2 for ENSO. Similar nonlinear effects can be expected to be present among all state variables and they could be revealed by dynamic loading matrices.

The results on prominent variance structures reported in Section 6 indicate what kind of features could be found in the fast climate variations when the emphasis is put on different timescales. The presented analysis of the variance structures can be extended in many different ways. For example, it would be interesting to relate the components with prominent variance structures 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 (Valpola et al. 2004).

The presented algorithms can easily be applied to other weather measurements with the possibility to concentrate on various properties of interest, different timescales and spatial localizations. It is also possible that some new interesting properties emerge during such exploratory analysis. This could motivate other types of models and algorithms, and the algorithmic framework used in this work can be a useful tool.

References

Aires, F., A. Chédin, and J.-P. Nadal, 2000: Independent component analysis of multivariate time series: Application to the tropical SST variability. *Journal of Geophysical Research*, **105**, 17437–17455.

Barros, A. K. and A. Cichocki, 2001: Extraction of specific signals with temporal structure. *Neural Computation*, **13**, 1995–2003.

Basak, J., A. Sudarshan, D. Trivedi, and M. S. San-

thanam, 2004: Weather data mining using independent component analysis. *Journal of Machine Learning Research*, **5**, 239–253.

Cardoso, J. F., 1998: Multidimensional independent component analysis. *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '98)*, Seattle, WA, 1941–1944.

Cichocki, A., T. Rutkowski, and K. Siwek, 2002: Blind signal extraction of signals with specified frequency band. *Neural Networks for Signal Processing XII: Proceedings of the 2002 IEEE Signal Processing Society Workshop*, Martigny, Switzerland, 515–524.

Diamantaras, K. I. and S. Y. Kung, 1996: *Principal Component Neural Networks: Theory and Applications*. Adaptive and Learning Systems for Signal Processing, Communications, and Control, John Wiley & Sons.

Enfield, D. B., A. M. Mestas-Nuñez, and P. J. Trimble, 2001: The Atlantic multidecadal oscillation and its relation to rainfall and river flows in the continental U.S. *Geophysical Research Letters*, **28**, 2077–2080.

Harman, H. H., 1967: *Modern Factor Analysis*. University of Chicago Press, 2nd edition.

Haykin, S., 1999: *Neural Networks – A Comprehensive Foundation, 2nd ed.*. Prentice-Hall.

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

Hyvärinen, A. and E. Oja, 1997: A fast fixed-point algorithm for independent component analysis. *Neural Computation*, **9**, 1483–1492.

Ilin, A., 2006a: *Advanced Source Separation Methods with Applications to Spatio-Temporal Datasets*. Ph.D. thesis, Helsinki University of Technology, Finland.

— 2006b: Independent dynamic subspace analysis. *Proceedings of 14th European Symposium on Artificial Neural Networks (ESANN 2006)*, Bruges, Belgium, 345–350.

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

Ilin, A., H. Valpola, and E. Oja, 2005: Semiblind source separation of climate data detects El Niño as the component with the highest interannual variability. *Proceedings of International Joint Conference on Neural Networks (IJCNN2005)*, Montréal, Québec, Canada, 1722–1727.

— 2006a: Exploratory analysis of climate data using source separation methods. *Neural Networks*, **19**, 155–167.

— 2006b: Extraction of climate components with structured variance. *Proceedings of the IEEE World Congress on Computational Intelligence (WCCI 2006)*, Vancouver, BC, Canada, to appear.

Jones, M. and R. Sibson, 1987: What is projection pursuit? *Journal of the Royal Statistical Society, Series A*, **150**, 1–36.

Kalnay, E. and coauthors, 1996: The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, **77**, 437–471.

Lotsch, A., M. A. Friedl, and J. Pinzón, 2003: Spatio-temporal deconvolution of NDVI image sequences using independent component analysis. *IEEE Transactions on Geoscience and Remote Sensing*, **41**, 2938–2942.

NCEP data, 2004: NCEP Reanalysis data provided by the NOAA-CIRES Climate Diagnostics Center, Boulder, Colorado, USA. Available from `http://www.cdc.noaa.gov/`.

Niño 3 SST, 2004: Niño region 3 sea surface temperature index provided by the NCAR Climate Analysis Section from their Website at `http://www.cgd.ucar.edu/cas/`.

Pawlowicz, R., 2000: M_Map: A mapping package for Matlab, available at `http://www2.ocgy.ubc.ca/~rich/map.html`.

Richman, M. B., 1986: Rotation of principal components. *Journal of Climatology*, **6**, 293–335.

Särelä, J. and H. Valpola, 2005: Denoising source separation. *Journal of Machine Learning Research*, **6**, 233–272.

Trenberth, K. E. and J. M. Caron, 2000: The Southern Oscillation revisited: Sea level pressures, surface temperatures, and precipitation. *Journal of Climate*, **13**, 4358–4365.

Valpola, H., M. Harva, and J. Karhunen, 2004: Hierarchical models of variance sources. *Signal Processing*, **84**, 267–282.

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