

Using independent component analysis to separate signals in climate data

Imola K. Fodor and Chandrika Kamath

Lawrence Livermore National Laboratory, 7000 East Avenue, Livermore, CA, USA 94551

ABSTRACT

Global temperature series have contributions from different sources, such as volcanic eruptions and El Niño Southern Oscillation variations. We investigate independent component analysis as a technique to separate unrelated sources present in such series. We first use artificial data, with known independent components, to study the conditions under which ICA can separate the individual sources. We then illustrate the method with climate data from the National Centers for Environmental Prediction.

Keywords: Independent component analysis, climate simulation

1. INTRODUCTION

Observed and simulated global temperature series include the effects of many different sources, such as volcano eruptions and El Niño Southern Oscillation (ENSO) variations. In order to compare the results of different models to each other, and to the observed data, it is necessary to first remove contributions from sources that are not commonly shared across the models considered. Such a separation of sources is also desired in order to assess the effect of human contributions on the global climate.

Atmospheric scientists currently use parametric models and iterative techniques to remove the effects of volcano eruptions and ENSO variations from global temperature trends.¹ Drawbacks of the parametric approach include the non-robustness of the results to the estimated values of the parameters, and the possible lack of fit of the data to the model.

In this paper, we investigate the use of independent component analysis (ICA) as an alternative method for separating independent sources in global temperature series. Instead of fitting parametric models, we let the data guide the estimation, and separate automatically the effects of the underlying sources. We first assess ICA on simple artificial datasets to establish the conditions under which ICA is feasible in our context, then we study its results on climate data from the National Centers for Environmental Predictions.

The rest of this paper is organized as follows. Section 2 provides a brief summary of independent component analysis, Section 3 presents the results of ICA on artificial data, Section 4 describes the results using climate data, and Section 5 summarizes our findings.

2. INDEPENDENT COMPONENT ANALYSIS

This section provides a brief description of ICA.²⁻⁷ Connections of ICA to classical methods such as principal component analysis (PCA), factor analysis (FA), and projection pursuit (PP), are also discussed in the references above.

Assume that $\mathbf{s} = (s_1, \dots, s_p)^T$ is a p -dimensional, unobservable, non-Gaussian random variable whose statistically independent components are mixed using an unknown linear transformation matrix $\mathbf{A}_{k \times p}$ via

$$\mathbf{x} = \mathbf{A}\mathbf{s}, \tag{1}$$

Further author information: (Send correspondence to I.K.F.)

I.K.F.: E-mail: fodor1@llnl.gov, Telephone: 1 925 424 5420, Address: Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, P.O. Box 808, L-561, Livermore, CA, USA 94551

C.K.: E-mail: kamath2@llnl.gov, Telephone: 1 925 423 3768, Address: Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, P.O. Box 808, L-561, Livermore, CA, USA 94551

to produce the observable k -dimensional variable $\mathbf{x} = (x_1, \dots, x_k)^T$, with $k \geq p$. The goal of ICA is to estimate the underlying independent components \mathbf{s} , also referred to as hidden variables or factors, and the mixing matrix \mathbf{A} from realizations of \mathbf{x} .

If the dimensions k and p are equal, then the following inverse relation holds

$$\mathbf{s} = \mathbf{W}\mathbf{x}, \quad (2)$$

where the weight matrix $\mathbf{W}_{p \times k}$ is the inverse of the mixing matrix $\mathbf{A}_{k \times p}$ in Eq. (1).

Statistical independence is a much stronger condition than uncorrelatedness. While the latter involves only the second-order statistics, the former depends on all the higher-order statistics. Formally, the individual components of the random variable $\mathbf{y} = (y_1, \dots, y_n)^T$ are mutually uncorrelated, if for $\forall i \neq j, 1 \leq i, j \leq n$, we have

$$\text{Cov}(y_i, y_j) = E\{(y_i - E(y_i))(y_j - E(y_j))\} = E(y_i y_j) - E(y_i)E(y_j) = 0. \quad (3)$$

In contrast, independence requires that the multivariate probability density function $f(y_1, \dots, y_n)$ factorizes into the product of the univariate densities $f_i(y_i)$, $i = 1, \dots, n$, as

$$f(y_1, \dots, y_n) = f_1(y_1) \dots f_n(y_n). \quad (4)$$

Independence always implies uncorrelatedness, but not vice versa in general. Only if the distribution $f(y_1, \dots, y_n)$ is multivariate normal, are the two equivalent. For Gaussian distributions, the principal components are independent, and thus ICA has no advantage over PCA. In what follows, we will assume that at least one of the hidden components is non-Gaussian.

Since in most applications it is impossible to derive truly independent sources, ICA methods define approximate measures of independence as objective functions (also called contrast, loss, or cost functions), and then search for projections of the observations that optimize those measures. A common measure of independence is the entropy-based mutual information,⁸ which, for uncorrelated variables can be expressed as

$$I(y_1, \dots, y_n) = J(\mathbf{y}) - \sum_{i=1}^n J(y_i), \quad (5)$$

where

$$J(\mathbf{y}) = H(\mathbf{y}_{\text{gauss}}) - H(\mathbf{y}), \quad (6)$$

is the negentropy of \mathbf{y} , $\mathbf{y}_{\text{gauss}}$ is a Gaussian random variable with the same covariance as \mathbf{y} , and

$$H(\mathbf{y}) = - \int f(\mathbf{y}) \log f(\mathbf{y}) d\mathbf{y} \quad (7)$$

is the entropy of \mathbf{y} . For a given covariance matrix, it is well-known that the Gaussian distribution has the maximum entropy. Minimizing the mutual information is roughly equivalent to minimizing the entropy, and therefore amounts to searching for components that are far from Gaussian. If ν denotes a standard normal variable with mean zero and variance one, $\nu \sim \mathcal{N}(0, 1)$, it can be shown⁸ that the negentropy can be approximated by

$$J(y_i) \approx [E\{G(y_i)\} - E\{G(\nu)\}]^2, \quad (8)$$

where $G(\cdot)$ is a non-quadratic function. For the results in this paper, we used the publicly available FastICA algorithm^{2,9} with the contrast function $G(u)$ given by $G(u) = \log \cosh(u)$.

Once the objective function is specified in a computationally tractable form, the independent components are estimated using optimization algorithms such as Newton's method, with the constraint that the terms be orthonormal. The sought components y_i will be estimates of the independent components s_i , and are of the form $y_i = \mathbf{w}_i^T \mathbf{x} = \sum_{j=1}^k w_{ij} x_j$, with weights $\mathbf{w}_i = (w_{i1}, \dots, w_{ik})^T$, for $i = 1, \dots, p$. The corresponding weight matrix in Eq. (2) is given by $\mathbf{W}_{p \times k} = \{w_{ij}\}_{i=1, \dots, p; j=1, \dots, k}$.

-
1. $\mathbf{X}_0 = \mathbf{X} - \bar{\mathbf{X}}$
 2. $\boldsymbol{\Sigma}_{k \times k} = \mathbf{V} \boldsymbol{\Lambda} \mathbf{V}^T$
 3. $\mathbf{R}_{k \times k} = \sqrt{\boldsymbol{\Lambda}^{-1}} \mathbf{V}^T$
 4. $\mathbf{Y}_0 = \mathbf{R} \mathbf{X}_0$, $\text{cov}(\mathbf{Y}_0) = \mathbf{I}_k$
 5. Apply the FastICA code to estimate the p independent components $\hat{\mathbf{S}}_{p \times n}$ and the corresponding mixing matrix $\hat{\mathbf{B}}_{k \times p}$ such that $\hat{\mathbf{Y}}_0 = \hat{\mathbf{B}} \hat{\mathbf{S}}$
 6. $\hat{\mathbf{X}}_0 = \mathbf{R}^{-1} \hat{\mathbf{Y}}_0 = \mathbf{R}^{-1} \hat{\mathbf{B}} \hat{\mathbf{S}}$
 7. $\hat{\mathbf{X}} = \hat{\mathbf{X}}_0 + \bar{\mathbf{X}} = \mathbf{R}^{-1} \hat{\mathbf{B}} \hat{\mathbf{S}} + \bar{\mathbf{X}} = \hat{\mathbf{A}} \hat{\mathbf{S}} + \bar{\mathbf{X}}$, where $\hat{\mathbf{A}}_{k \times p} = \mathbf{R}^{-1} \hat{\mathbf{B}}$
-

Table 1. Pseudo-code for estimating the independent components \mathbf{S} of the original data \mathbf{X} , and for decomposing \mathbf{X} in terms of the estimated components $\hat{\mathbf{S}}$. Quantities with hats involve estimates.

Throughout this paper, we assume that we have n observations, each being a realization of the random variable \mathbf{x} with mean $\boldsymbol{\mu} = \text{E}(\mathbf{x})$ and covariance matrix $\text{Cov}(\mathbf{X}) = \text{E}\{(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T\}$. We denote such an observation matrix by $\mathbf{X} = \{x_{i,j} : 1 \leq i \leq k, 1 \leq j \leq n\}$. In terms of \mathbf{X} , and the corresponding $\mathbf{S} = \{s_{i,j} : 1 \leq i \leq p, 1 \leq j \leq n\}$, the ICA model in Eq. (1) becomes

$$\mathbf{X} = \mathbf{A} \mathbf{S}. \quad (9)$$

Let $\bar{\mathbf{X}}_{k \times 1}$ and $\boldsymbol{\Sigma}_{k \times k}$ denote the sample mean and covariance matrix of the n observations, respectively. Table 1 outlines our procedure to estimate independent components. Since we would like to express the observations in terms of the estimated independent components, we keep track of the pre-processing transformations in steps 1 through 4, and use their inverses in steps 6 and 7 to map back the estimates to the scale of the original input data. Hats above symbols indicate estimated values of the underlying quantities, involving the independent component estimates in step 5 of the algorithm. If $\hat{\mathbf{X}}$ denotes the reconstruction of the original data \mathbf{X} in terms of the estimated independent components $\hat{\mathbf{S}}_{p \times n}$ and corresponding mixing matrix $\hat{\mathbf{B}}_{k \times p}$ in step 5, we can decompose $\hat{\mathbf{X}}$ as

$$\hat{\mathbf{X}} = \hat{\mathbf{A}} \hat{\mathbf{S}} + \bar{\mathbf{X}} \quad (10)$$

$$= \hat{\mathbf{A}} \begin{pmatrix} \hat{s}_1 \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix}_{p \times n} + \hat{\mathbf{A}} \begin{pmatrix} \mathbf{0} \\ \hat{s}_2 \\ \vdots \\ \mathbf{0} \end{pmatrix}_{p \times n} + \cdots + \hat{\mathbf{A}} \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \hat{s}_p \end{pmatrix}_{p \times n} + \bar{\mathbf{X}} \quad (11)$$

$$\equiv \hat{\mathbf{X}}_{\mathbf{S}_1} + \hat{\mathbf{X}}_{\mathbf{S}_2} + \cdots + \hat{\mathbf{X}}_{\mathbf{S}_p} + \bar{\mathbf{X}}, \quad (12)$$

where $\hat{\mathbf{X}}_{\mathbf{S}_j}$, defined above, is of size $k \times n$, and it represents the contribution of the j th estimated independent component to $\hat{\mathbf{X}}$, $j = 1, \dots, p$. The i th row of $\hat{\mathbf{X}}_{\mathbf{S}_j}$, denoted by $\hat{\mathbf{x}}_{i, \mathbf{S}_j}$, gives the contribution of the j th estimated independent component to $\hat{\mathbf{x}}_i$, $i = 1, \dots, k$. Equivalently, we can write Eq. (12) as

$$\hat{\mathbf{X}} = \begin{pmatrix} \hat{\mathbf{x}}_1 \\ \hat{\mathbf{x}}_2 \\ \vdots \\ \hat{\mathbf{x}}_k \end{pmatrix}_{k \times n} = \begin{pmatrix} \hat{\mathbf{x}}_{1, \mathbf{S}_1} + \cdots + \hat{\mathbf{x}}_{1, \mathbf{S}_p} + \bar{\mathbf{x}}_1 \\ \hat{\mathbf{x}}_{2, \mathbf{S}_1} + \cdots + \hat{\mathbf{x}}_{2, \mathbf{S}_p} + \bar{\mathbf{x}}_2 \\ \vdots \\ \hat{\mathbf{x}}_{k, \mathbf{S}_1} + \cdots + \hat{\mathbf{x}}_{k, \mathbf{S}_p} + \bar{\mathbf{x}}_k \end{pmatrix}_{k \times n}. \quad (13)$$

For artificial data, where \mathbf{S} and \mathbf{A} are known, the decomposition in Eq.(12) can be compared with the true decomposition

$$\mathbf{X} = \mathbf{A} \mathbf{S} \quad (14)$$

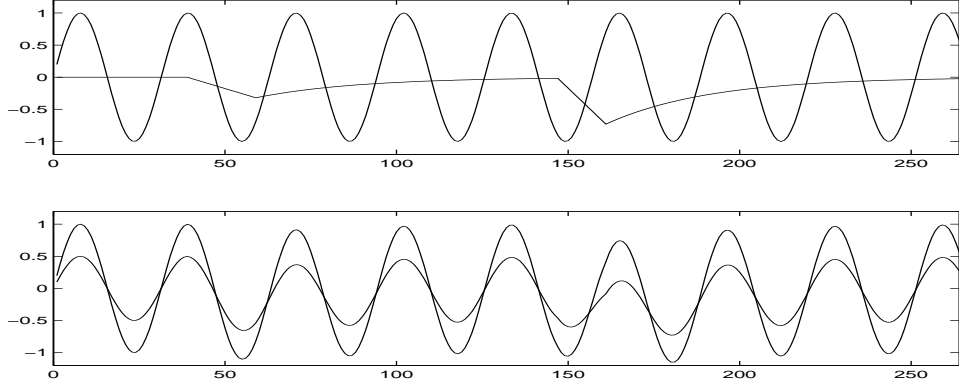


Figure 1. Artificial data. (a) source \mathbf{S} : \mathbf{s}_1 and \mathbf{s}_2 ; (b) observation \mathbf{X} : \mathbf{x}_1 and \mathbf{x}_2 .

$$= \mathbf{A} \begin{pmatrix} \mathbf{s}_1 \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix}_{p \times n} + \mathbf{A} \begin{pmatrix} \mathbf{0} \\ \mathbf{s}_2 \\ \vdots \\ \mathbf{0} \end{pmatrix}_{p \times n} + \cdots + \mathbf{A} \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{s}_p \end{pmatrix}_{p \times n} \quad (15)$$

$$\equiv \mathbf{X}_{\mathbf{S}_1} + \mathbf{X}_{\mathbf{S}_2} + \cdots + \mathbf{X}_{\mathbf{S}_p}. \quad (16)$$

If $\mathbf{x}_{i,\mathbf{S}_j}$ denotes the i th row of $\mathbf{X}_{\mathbf{S}_j}$, $i = 1, \dots, k$, $j = 1, \dots, p$, then, similar to Eq. (13), we have

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_k \end{pmatrix}_{k \times n} = \begin{pmatrix} \mathbf{x}_{1,\mathbf{S}_1} + \cdots + \mathbf{x}_{1,\mathbf{S}_p} \\ \mathbf{x}_{2,\mathbf{S}_1} + \cdots + \mathbf{x}_{2,\mathbf{S}_p} \\ \vdots \\ \mathbf{x}_{k,\mathbf{S}_1} + \cdots + \mathbf{x}_{k,\mathbf{S}_p} \end{pmatrix}_{k \times n}. \quad (17)$$

3. RESULTS WITH DATA FROM A SIMPLE ARTIFICIAL MODEL

In order to understand the use of ICA in separating climate signals, we first experimented with known signals. We considered the simple case of two signals, one a synthetic volcano signal and the other a sinusoid to mimic seasonal variation in the temperature. We modeled the volcano signal using the results in,¹ where the effect of a volcano signal on the global temperature is composed of two parts: a linear cooling from the time of eruption t_{erupt} to a time t_{ramp} , and an exponential recovery thereafter. Thus, a volcano signal v_t can be written as

$$v_t = \begin{cases} \frac{-\Delta T_{max} t}{t_{ramp}}, & t = t_{erupt}, \dots, t_{ramp} \\ -\Delta T_{max} e^{-\frac{t-t_{ramp}}{\tau}}, & t = t_{ramp} + 1, \dots, n, \end{cases} \quad (18)$$

where ΔT_{max} is the maximum cooling effect of the volcano, $t_{ramp} > t_{erupt}$ is the time when the maximum cooling occurs, and τ is an exponential decay time. Following,¹ the volcano signal \mathbf{s}_2 used in our study is the sum of two v_t terms that model the effects of the El Chicón and the Pinutabo major volcanic eruptions of the recent past. The values of the parameters are¹ shown in Table 2. The time period for this data was 264 months, where $t = 0$ corresponds to January 1979, and $t = 263$ corresponds to December 2000. The volcano eruptions occur at months 39 and 147. We chose this particular time period to reflect the climate simulation data we describe in Section 4. The sinusoidal signal \mathbf{s}_1 is simply $\mathbf{s}_1 = \sin\{(1 : n)/5\}$, $n = 264$. The two signals are shown in Fig. 1, panel (a). Next, using the mixing matrix

$$\mathbf{A} = \begin{pmatrix} 1.0 & 0.4 \\ 0.5 & 0.6 \end{pmatrix} \quad (19)$$

we generated the observations \mathbf{X} according to Eq. (9). Panel (b) in Fig. 1 shows the two components of \mathbf{X} , \mathbf{x}_1 and \mathbf{x}_2 . Fig. 2 displays the observation data in terms of the generating sources. The two solid lines in the

	El Chicón	Pinutabo
n	264	264
ΔT_{max}	0.32	0.72
t_{ramp}	20	14
t_{erupt}	39	147
τ	30	30

Table 2. Parameters in Eq. (18) for two major volcano eruptions.

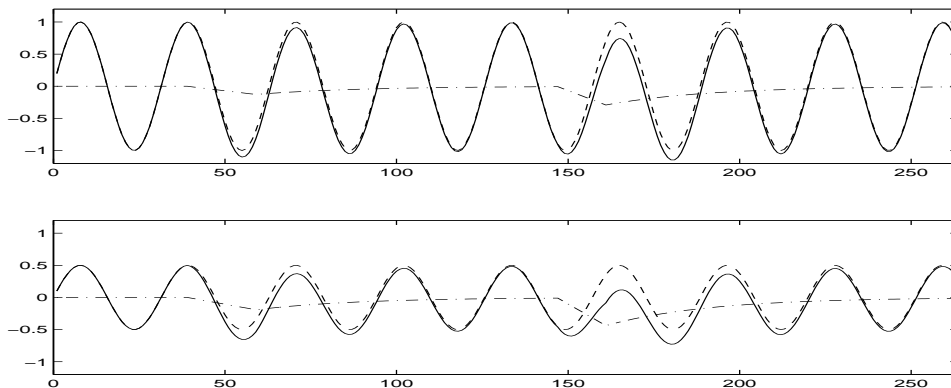


Figure 2. Decomposition of \mathbf{X} in terms of the original signal \mathbf{S} in Fig. 1. (a) $\mathbf{x}_1 = \mathbf{x}_{1,\mathbf{s}_1} + \mathbf{x}_{1,\mathbf{s}_2}$ (solid); $\mathbf{x}_{1,\mathbf{s}_1}$ (dashed); $\mathbf{x}_{1,\mathbf{s}_2}$ (dash-dotted); (b) $\mathbf{x}_2 = \mathbf{x}_{2,\mathbf{s}_1} + \mathbf{x}_{2,\mathbf{s}_2}$ (solid); $\mathbf{x}_{2,\mathbf{s}_1}$ (dashed); $\mathbf{x}_{2,\mathbf{s}_2}$ (dash-dotted).

two panels show the \mathbf{x}_1 and \mathbf{x}_2 vectors of \mathbf{X} , respectively. The dashed lines in the two panels represent the contributions of the first source signal \mathbf{s}_1 to \mathbf{x}_1 and \mathbf{x}_2 , denoted by $\mathbf{x}_{1,\mathbf{s}_1}$ and by $\mathbf{x}_{2,\mathbf{s}_1}$ in Eq. (17), respectively. In this simple case, $\mathbf{x}_{1,\mathbf{s}_1} = 1.0 \times \mathbf{s}_1$ and $\mathbf{x}_{2,\mathbf{s}_1} = 0.5 \times \mathbf{s}_1$. The dash-dotted lines show the contributions of \mathbf{s}_2 to \mathbf{x}_1 and \mathbf{x}_2 , respectively: $\mathbf{x}_{1,\mathbf{s}_2} = 0.4 \times \mathbf{s}_2$ in the upper panel, and $\mathbf{x}_{2,\mathbf{s}_2} = 0.6 \times \mathbf{s}_2$ in the lower panel. In both panels, the solid vectors are the sums of the two other vectors: $\mathbf{x}_1 = \mathbf{x}_{1,\mathbf{s}_1} + \mathbf{x}_{1,\mathbf{s}_2} = 1.0 \times \mathbf{s}_1 + 0.4 \times \mathbf{s}_2$, and $\mathbf{x}_2 = \mathbf{x}_{2,\mathbf{s}_1} + \mathbf{x}_{2,\mathbf{s}_2} = 0.5 \times \mathbf{s}_1 + 0.6 \times \mathbf{s}_2$.

Fig. 3 displays the estimated independent components $\hat{\mathbf{S}}$ obtained after applying the algorithm in Table 1 to \mathbf{X} . The shape of the two original sources, \mathbf{s}_1 and \mathbf{s}_2 displayed in Fig. 1, is clearly visible. However, the scale and the sign of the estimates are arbitrary.

Following the notation in Eq. (13), Fig. 4 presents the decomposition of the estimate $\hat{\mathbf{X}}$ in step 7 of Table 1 in terms of the estimated independent components. The two solid lines in the two panels show the $\hat{\mathbf{x}}_1$ and $\hat{\mathbf{x}}_2$ vectors of $\hat{\mathbf{X}}$, respectively. The dashed lines in the two panels represent the contributions of the first estimated component $\hat{\mathbf{s}}_1$ to $\hat{\mathbf{x}}_1$ and $\hat{\mathbf{x}}_2$, denoted by $\hat{\mathbf{x}}_{1,\mathbf{s}_1}$ and by $\hat{\mathbf{x}}_{2,\mathbf{s}_1}$ in Eq. (13), respectively. The dash-dotted lines show the contributions of $\hat{\mathbf{s}}_2$ to $\hat{\mathbf{x}}_1$ and $\hat{\mathbf{x}}_2$, plus a constant mean correction, respectively: $\hat{\mathbf{x}}_{1,\mathbf{s}_2} + \bar{\mathbf{x}}_1$ in the upper panel, and $\hat{\mathbf{x}}_{2,\mathbf{s}_2} + \bar{\mathbf{x}}_2$ in the lower panel. In both panels, the solid vectors are the sums of the two other vectors. Note that in this case, we knew the true sources, so we knew that we had to add the means to the second terms in order

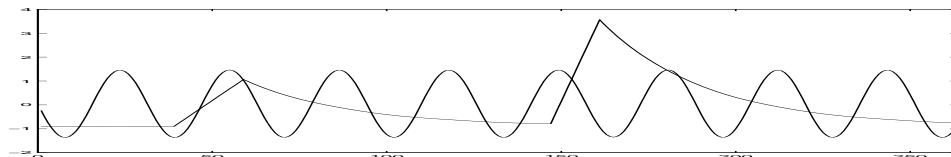


Figure 3. Estimated independent components $\hat{\mathbf{S}}$ for the data in Fig. 2.

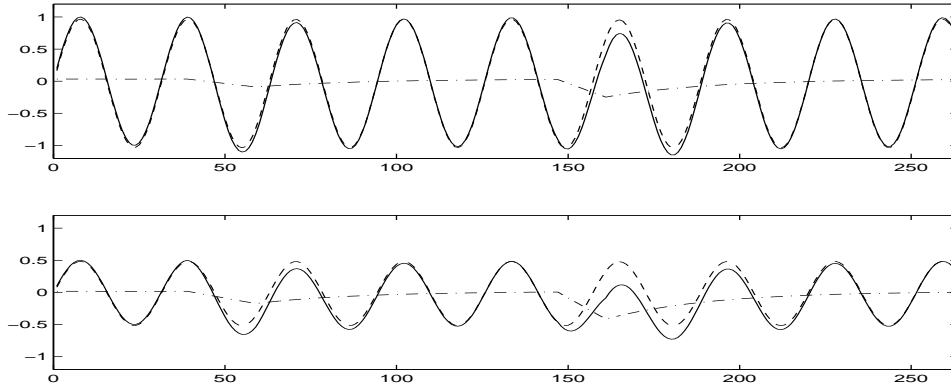


Figure 4. Decomposition of $\hat{\mathbf{X}}$ in terms of the estimated independent components $\hat{\mathbf{S}}$ for the data in Fig. 2. (a) $\hat{\mathbf{x}}_1 = \hat{\mathbf{x}}_{1,s_1} + \hat{\mathbf{x}}_{1,s_2} + \bar{\mathbf{x}}_1$ (solid); $\hat{\mathbf{x}}_{1,s_1}$ (dashed); $\hat{\mathbf{x}}_{1,s_2} + \bar{\mathbf{x}}_1$ (dash-dotted); (b) $\hat{\mathbf{x}}_2 = \hat{\mathbf{x}}_{2,s_1} + \hat{\mathbf{x}}_{2,s_2} + \bar{\mathbf{x}}_2$ (solid); $\hat{\mathbf{x}}_{2,s_1}$ (dashed); $\hat{\mathbf{x}}_{2,s_2} + \bar{\mathbf{x}}_2$ (dash-dotted).

to match the original data. In practical situations, however, the method cannot resolve shifts. Comparing Fig. 4 to Fig. 2, we note the excellent performance of the ICA method in this simple case: it estimated successfully the observed signal in terms of the underlying independent components.

We also experimented with artificial data from more realistic models, and investigated the performance of the FastICA algorithm in the presence of noise and dependencies among the sources.¹² The method was fairly robust to the presence of noise, but it failed when we violated the independence assumption.

4. RESULTS WITH THE CLIMATE DATA

Following our initial study on the use of ICA to separate signals generated by a linear mixing of two synthetic sources that resembled climate data, we tried the techniques on real data from a climate simulation.

4.1. Description of the data

We conducted this study using the monthly mean temperature reanalysis data¹³ from the National Centers for Environmental Prediction (NCEP). The data is on a 144×73 longitude-by-latitude grid, on 17 vertical pressure levels ranging from 10000 hPa close to the surface of the earth to 10 hPa at the highest elevation. It spans 264 months, from January 1979 to December 2000. Figure 5 displays the raw temperatures for January 1979 on the 144×73 latitude by longitude grid at the 1000 hPa pressure level.

Since we expected the ENSO and volcano signals to have a strong latitudinal dependence, we performed our analyses on zonally averaged data. At a given month and level, we first calculated the 73 zonal means over the 144 longitudes. Following standard practices in the atmospheric sciences,¹⁴ we removed the seasonal variation and centered the data as follows. For each month, we replaced the values at each of the $73 \times 17 (= 1241)$ latitude-by-level grid points by subtracting their corresponding monthly means over the entire 22-year period. The resulting time-centered values are referred to as anomalies in the atmospheric sciences literature.

Next, we weighted the data set appropriately to account for the unequal spatial grid sizes and altitude bands. Figure 6 displays the January 1979 anomaly temperatures after the weights have been applied to the time-centered zonally averaged data. Finally, we performed ICA on the $k = 1241$ by $n = 264$ space-by-time dimension anomaly data set $\mathbf{X}_{k \times n}$.

4.2. ICA applied to the full dimensional climate data

Figure 7 shows the first six independent component estimates with the FastICA method applied to the $\mathbf{X}_{k \times n}$ anomaly data. The resulting IC estimates are very localized spatially and are not interpretable scientifically in

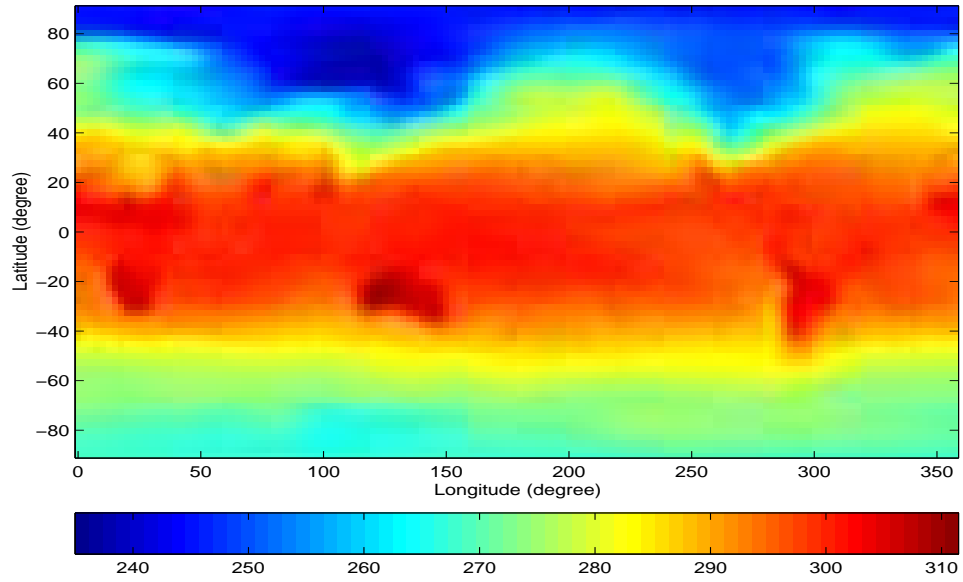


Figure 5. January 1979 raw temperatures (Kelvin) on the 144 longitudes by 73 latitudes spatial grid for pressure level 1000 hPa.

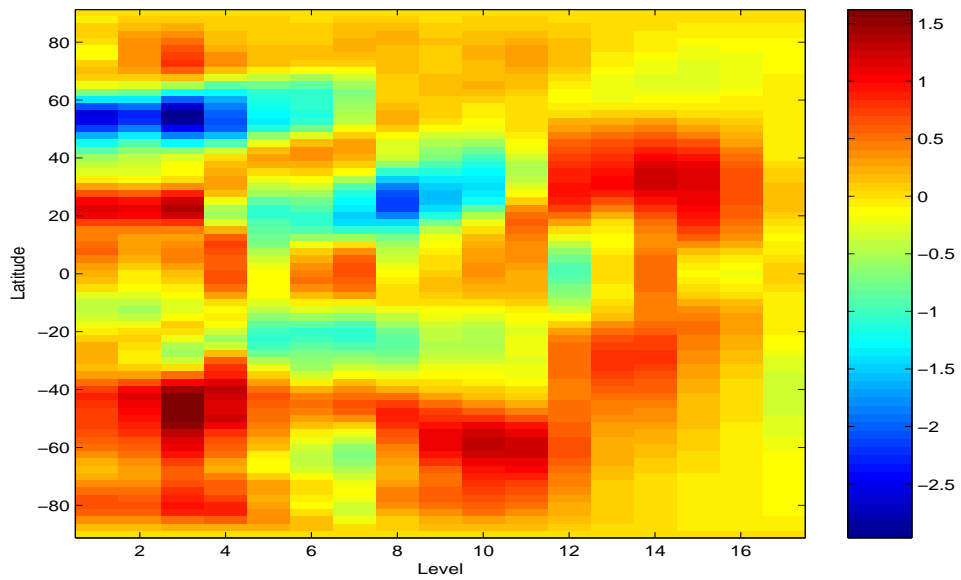


Figure 6. January 1979 zonal anomaly temperatures (Kelvin) on the 17 pressure levels by 73 latitudes grid.

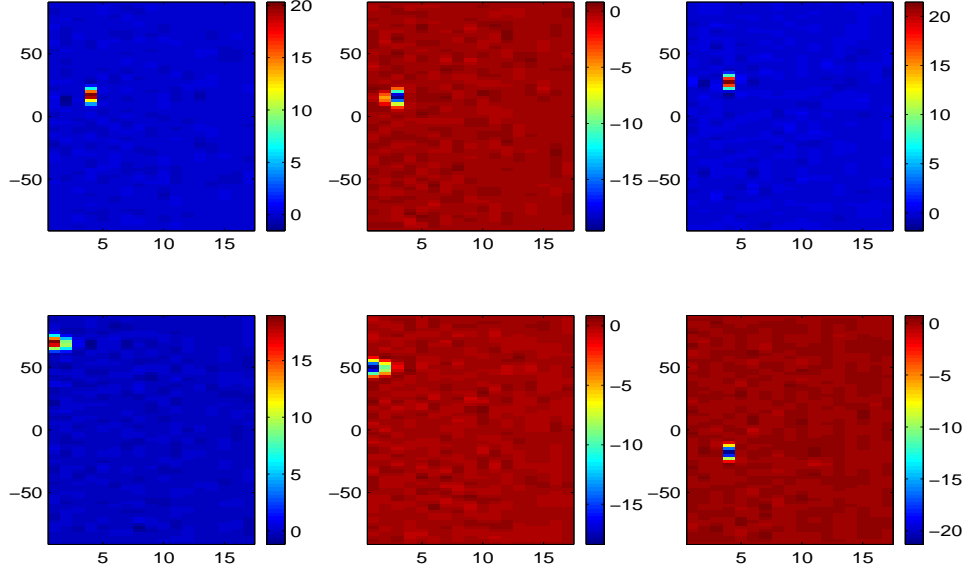


Figure 7. Six independent component bases obtained from the full-dimensional anomaly data.

terms of atmospheric processes. The large spikes are in fact artifacts due to over-learning in the ICA estimation.³ Estimating n independent components along with the mixing matrix from only n samples is statistically impossible since there are more parameters to estimate than available observations. We do have n samples of the p -dimensional signal, but we do not expect to have n truly independent component sources.

4.3. Principal component analysis as a dimension reduction tool

Instead of working with the full high-dimensional anomaly data, we next explored ICA on reduced dimensional representations of the data. Since there are only a limited number of atmospheric processes, before applying ICA we first need to reduce the data to p dimensions, where p is the number of independent components sought. We achieved this dimension reduction by using principal component analysis (PCA).

The PCA technique seeks linear combinations of the data with maximal variance. Given the zero-mean zonal anomaly dataset $\mathbf{X}_{k \times n}$ of dimension of $k = 1241$ spatial grid points and $n = 264$ months, let

$$\boldsymbol{\Sigma}_{k \times k} = \mathbf{X}_{k \times n} \mathbf{X}'_{n \times k} \quad (20)$$

denote the state-space scatter matrix¹⁴ (scaled covariance matrix), where $\mathbf{X}'_{n \times k}$ denotes the transpose of $\mathbf{X}_{k \times n}$. The eigenvalue decomposition of $\boldsymbol{\Sigma}$ provides the eigenvector matrix $\mathbf{E}_{k \times k}$ and the eigenvalue matrix $\mathbf{D}_{k \times k}$ such that

$$\boldsymbol{\Sigma}_{k \times k} = \mathbf{E}_{k \times k} \mathbf{D}_{k \times k} \mathbf{E}'_{k \times k}, \quad (21)$$

where \mathbf{E} is orthonormal, (i.e. $\mathbf{E}'\mathbf{E} = \mathbf{E}\mathbf{E}' = \mathbf{I}_{k \times k}$), and \mathbf{D} is diagonal, $\mathbf{D} = \text{diag}(d_1, \dots, d_k)$ with $d_1 \geq d_2 \dots \geq d_k$. To make the eigenvectors unique, we adopt the convention that the first non-zero term in each of the vectors is positive. The corresponding state-space principal components are

$$\mathbf{A}_{k \times n} = \mathbf{E}_{k \times k} \mathbf{X}_{k \times n}. \quad (22)$$

The p th cumulative sum of the eigenvalues normalized by the total sum of the eigenvalues, defined as

$$\lambda(p) = \frac{\sum_{i=1}^p d_i}{\sum_{i=1}^k d_i}, \quad (23)$$

indicates how much of the variation is explained by the first p , $p \leq k$, components. Table 3 shows the percent of variation explained for selected values of p .

p	$\lambda(p)$	p	$\lambda(p)$
1	0.2033	22	0.9003
2	0.3527	35	0.9501
3	0.4390	50	0.9738
4	0.5057	79	0.9901
5	0.5660	100	0.9946
6	0.6212	150	0.9986
7	0.6626	200	0.9997
8	0.7006	216	0.9999
9	0.7361	234	1.0000
10	0.7645	1241	1.0000

Table 3. Percent of the total variation explained $\lambda(p)$ by the first p principal components.

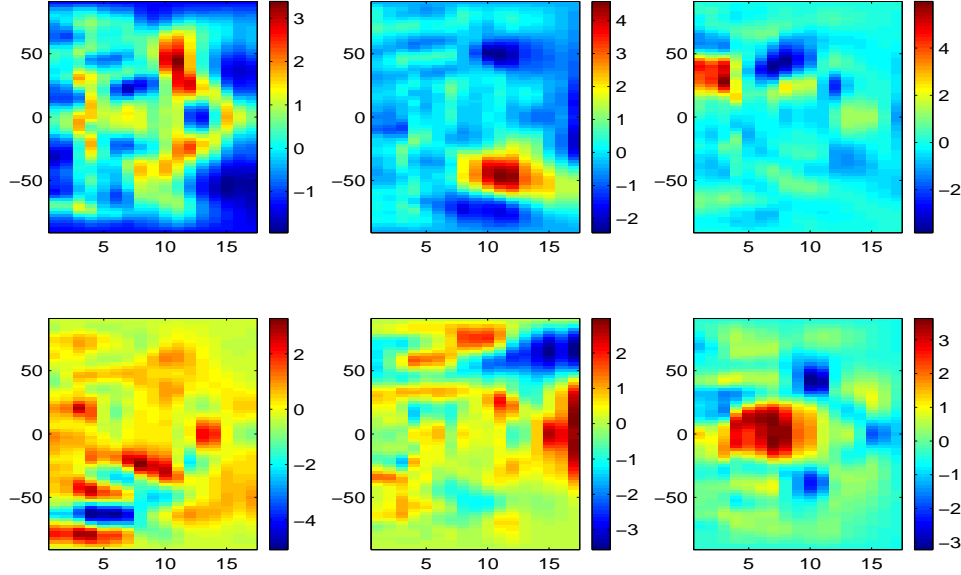


Figure 8. Six independent component bases obtained from the first $p=22$ eigenvectors of the anomaly data scatter matrix.

4.4. ICA applied to the reduced dimensional climate data

The PCA results in Section 4.3 indicate that the first $p = 22$ PCs explain 90% of the variance. In this section, we applied ICA to the anomaly data of reduced dimension, as described by the first $p = 22$ principal components. Since the original anomalies are linear combinations of the PCs, we can easily obtain the IC coefficients corresponding to the original data from the IC coefficients obtained for the PCs. Figure 8 shows six of the p resulting independent basis images when ICA is applied to the first $p = 22$ eigenvectors of Σ . Figure 9 shows the six projections of the anomaly data $\mathbf{X}_{k \times n}$ onto the independent components in Figure 8.

The patterns in the sixth estimated IC basis and corresponding time series suggest characteristics associated with ENSO variations, in particular, the Niño 3.4 index from NCEP. The Niño 3.4 index measures the monthly sea surface temperature deviation from its long-term mean averaged over the Niño 3.4 region (5N-5S, 120-170W). It is a standard index used to monitor ENSO events. If its 5-month running average exceeds +0.4 (-0.4) degrees Celsius, it is considered an El Niño (La Niña) event. If we consider the time series from the last panel in Figure 9 along with the Niño 3.4 index, the correlation between the two series is 0.5788. The two series follow a similar pattern, but they are shifted relative to each other. Lagging the IC time series back a few months increases the value of the correlation coefficient as follows: $\text{corr}(1)=0.6895$, $\text{corr}(2)=0.7410$, $\text{corr}(3)=0.7654$, $\text{corr}(4)=0.7587$, then decreases monotonically further. Figure 10 presents the IC time series shifted back by three months, along

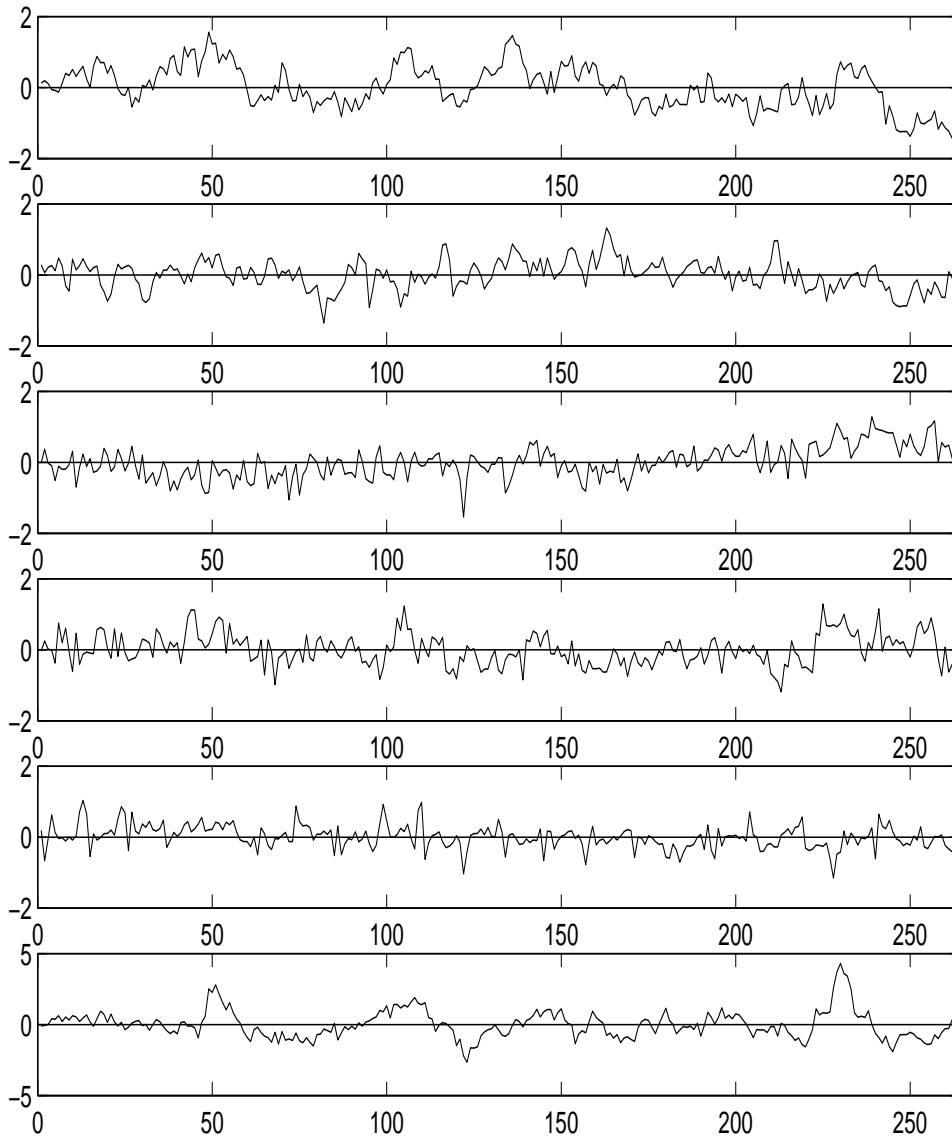


Figure 9. 6 state space IC time series.

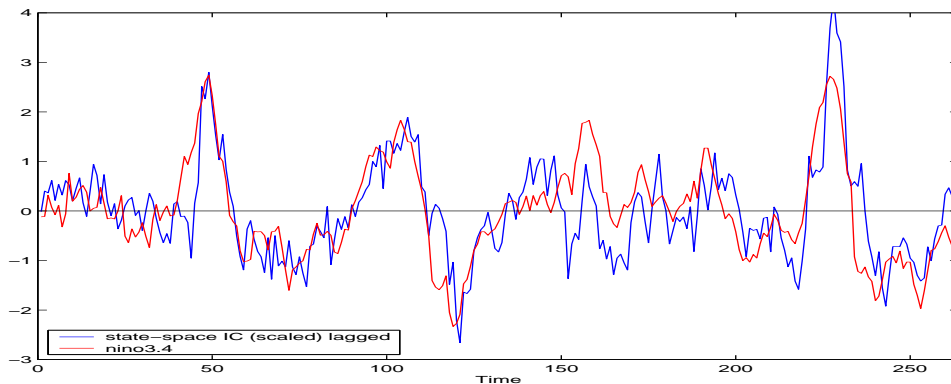


Figure 10. State space IC time series (scaled) with a three month lag, and Niño region 3.4 index. Correlation coefficient is 0.76.

with the Niño region 3.4 index. The excellent match between the two series indicates the success of the ICA approach to automatically separate sources in climate data.

5. SUMMARY

In this paper, we described our initial results with the use of ICA to automatically separate signals in climate data. Our experiments with synthetic data that resembled climate signals indicated that the ICA technique deserved further investigation. While a direct application of ICA to the global temperature data proved problematic, a combination of PCA and ICA was more promising. The simple ICA model used in combination with PCA for dimension reduction resulted in a slightly superior ENSO signal estimate than PCA alone.

Our future plans include interpreting the remaining PC and IC estimates in terms of atmospheric processes; determining the sensitivity of the results to the number of components to estimate; incorporating constraints on the shapes of the sought components (for example, volcano eruptions can only decrease the temperature, but not increase it); investigating possible non-linearities in the mixing process; and providing uncertainty estimates for the estimated components.

ACKNOWLEDGMENTS

We thank Dr. Benjamin D. Santer from the Program for Climate Model Diagnostics and Intercomparison at Lawrence Livermore National Laboratory for introducing us to the problem, for providing the climate data, and for his numerous insights and suggestions.

We would also like to acknowledge the FastICA team at the Laboratory of Information and Computer Science in the Helsinki University of Technology for providing their software in the public domain.

This work was performed under the auspices of the U.S. Department of Energy by University of California Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48, UCRL-JC-151808. This work was partially funded by the SciDAC program in the DOE Office of Advanced Scientific Computing Research.

REFERENCES

1. B. Santer *et al.*, “Accounting for the effects of volcanoes and ENSO in comparisons of modeled and observed temperature trends,” *Journal of Geophysical Research* **106**, pp. 28,033–28,059, November 27 2001.
2. A. Hyvärinen, “Fast and robust fixed-point algorithms for independent component analysis,” *IEEE Transactions on Neural Networks* **10**(3), pp. 626–634, 1999.
3. A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, Series on Adaptive and Learning Systems for Signal Processing, Communications, and Control, Wiley, 2001.

4. T.-W. Lee, *Independent Component Analysis: Theory and Applications*, Kluwer Academic Publishers, 2001.
5. S. Roberts and R. Everson, eds., *Independent Component Analysis: Principles and Practice*, Cambridge University Press, Cambridge, UK, 2001.
6. M. Girolami, ed., *Advances in Independent Component Analysis*, Perspectives in Neural Computing, Springer, 2000.
7. J. Friedman, T. Hastie, and R. Tibshirani, *Elements of Statistical Learning: Prediction, Inference and Data Mining*, Springer, 2001.
8. A. Hyvärinen, "Survey on independent component analysis," *Neural Computing Surveys* **2**, pp. 94–128, 1999. citeseer.nj.nec.com/hyv99survey.html.
9. J. Hurri *et al.* <http://www.cis.hut.fi/projects/ica/fastica/>.
10. J.-F. Cardoso, "High-order contrasts for independent component analysis," *Neural Computation* **11**(1), pp. 157–192, 1999.
11. J.-F. Cardoso. <http://www.tsi.enst.fr/~cardoso/icacentral/>.
12. I. Fodor and C. Kamath, "On the use of independent component analysis to separate meaningful sources in global temperature series," tech. rep., Lawrence Livermore National Laboratory, 2003. In preparation.
13. R. Kistler, E. Kalnay, *et al.*, "The NCEP/NCAR 50-year reanalysis," *Bull. Am. Met. Soc.* **82**, pp. 247–267, 2000.
14. R. Preisendorfer, *Principal Component Analysis in Meteorology and Oceanography*, no. 17 in Developments in Atmospheric Science, Elsevier Science, Amsterdam, 1988.