JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 105, NO. D13, PAGES 17,437–17,455, JULY 16, 2000

Independent component analysis of multivariate time series: Application to the tropical SST variability

Filipe Aires[1] and Alain Chédin
Laboratoire de Météorologie Dynamique du CNRS, École Polytechnique, France

Jean-Pierre Nadal
Laboratoire de Physique Statistique, École Normale Supérieure de Paris

Abstract. With the aim of identifying the physical causes of variability of a given dynamical system, the geophysical community has made an extensive use of classical component extraction techniques such as principal component analysis (PCA) or rotational techniques (RT). We introduce a recently developed algorithm based on information theory: independent component analysis (ICA). This new technique presents two major advantages over classical methods. First, it aims at extracting statistically independent components where classical techniques search for decorrelated components (i.e., a weaker constraint). Second, the linear hypothesis for the mixture of components is not required. In this paper, after having briefly summarized the essentials of classical techniques, we present the new method in the context of geophysical time series analysis. We then illustrate the ICA algorithm by applying it to the study of the variability of the tropical sea surface temperature (SST), with a particular emphasis on the analysis of the links between El Niño Southern Oscillation (ENSO) and Atlantic SST variability. The new algorithm appears to be particularly efficient in describing the complexity of the phenomena and their various sources of variability in space and time.

1. Introduction

A time series is a collection of observations of a dynamical system made sequentially in time. This work is concerned with the identification of the physical causes of the variability of a given dynamical system. In many cases, observed time series are well represented as a mixture, linear or nonlinear, of different statistically independent components (or sources, or factors). An important goal of statistics is then to retrieve these components from the observed data. Our approach is based on statistics computed from a set of time series samples generated by the dynamical system and is aimed at extracting the internal regularities of a given time series. Such kind of a statistical method is called a component extraction technique.

There are three main statistical component extraction methods: principal component analysis (PCA), singular value decomposition (SVD), and factor analysis (FA). Other related methods are the rotational techniques (RT); they rotate, orthogonally or obliquely, the PCA, SVD, or FA projection axes to facilitate the physical interpretation of the extracted components. Each method relies, implicitly or explicitly, on some assumptions about the structure of the observations. In particular, two major hypotheses are made in the PCA, SVD, or FA techniques such as (1) the linearity assumption, which assumed that the observed time series are linear mixtures of components; and (2) the second-order statistics hypothesis, which assumed that the stochastic variables of the problem have Gaussian probability distribution functions. Moreover, the RT relies on empirical optimization criteria.

The independent component analysis (ICA) method, presented in this paper, is based on information theory. It was recently developed in the context of signal processing and neural coding modeling. We argue that the ICA approach may overcome the main pitfalls of the standard techniques of time series analysis. The two major advances of the ICA approach are as follows: (1) the hypothesis of linearity of the mixture model is not required, and (2) the fact that it extracts statistically independent components, even if these components have non-Gaussian probability distribution functions, making use of higher-order statistics, whereas the PCA, SVD, or FA approaches make use of second-order statistics only.

The ICA paradigm is directly related to the notion of redundancy reduction proposed by the biologist *Barlow*

[1]Now at Columbia University/Nasa Goddard Institute for Space Studies, New York, USA.

Copyright 2000 by the American Geophysical Union.

Paper number 2000JD900152.
0148-0227/00/2000JD900152$09.00

[1960] in the 1960s, as a criterion that might be at the basis of neural coding. The relevance of sensory coding has been studied in several models during the last 10 years [*Atick*, 1992]. The terminology "ICA" comes from the domain of signal processing, where this idea of representing a signal in terms of its independent components was rediscovered in the 1980s [*Jutten and Herault*, 1991]. Since then, ICA has been the subject of numerous studies at the frontier between signal processing and neural coding theory.

In the particular case where the signal has a linear structure (i.e., a linear superposition of independent components), ICA is also named blind source separation (BSS) and is of interest for many signal processing applications [*Bar-Ness*, 1982; *Jutten and Herault*, 1991]. There exists a wide variety of algorithms that can perform BSS [*Jutten and Herault*, 1991; *Comon*, 1994; *Pham et al.*, 1992; *Delfosse and Loubaton*, 1995; *Cardoso*, 1989; *Amari et al.*, 1996; *Bell and Sejnowski*, 1995; *Nadal and Parga*, 1997] and some of them can be adapted to ICA for nonlinear mixtures [*Nadal and Parga*, 1994, 1997; *Storck and Deco*, 1997].

ICA searches for statistically independent components in a data set by minimizing an objective cost function. Different equivalent cost functions, called "contrasts" [*Comon*, 1994], can be chosen, all based on a model describing the statistical nature of the data. In the case of a linear superposition of independent components, ICA shares some common features with RT: both rotate the PCA axes according to some criteria. In RT the criterion of rotation is an empirical additional constraint still based on second-order statistics. The rotation in ICA can also be seen as the consequence of an additional constraint but based on higher-order statistics [*Comon*, 1994; *Nadal and Parga*, 1997]. The ICA paradigm is thus fundamentally different from the other classical approaches.

To our knowledge, ICA has not yet been used for component extraction in time series analysis, although related studies in the context of time series prediction exist [*Weigend and Gershenfeld*, 1994; *Storck and Deco*, 1997].

As already mentioned, various kinds of models, learning algorithms, and quality criteria for statistical independence can be used in ICA. For illustrative purposes we will focus on the infomax approach to ICA [*Nadal and Parga*, 1994], from which simple algorithms have been derived [*Bell and Sejnowski*, 1995] and can be easily adapted to nonlinear data mixture structure [*Storck and Deco*, 1997; *Nadal and Parga*, 1997]. However, we will restrict the numerical application to the linear case assuming that the data have a linear structure. The nonlinear case will be the subject of a forthcoming publication.

The numerical application presented in this work is concerned with the particular case of geophysical time series analysis. This kind of application is characterized by the fact that data have temporal and spatial dimensions, which allow the problem to be processed in various ways. As a consequence, the terminology must be clarified: it is important to distinguish the statistical techniques (PCA, SVD, FA, RT, and finally ICA) from the various ways of dealing with the data (e.g., analysis in the space, time, frequency domains).

The paper is organized as follows: we present, in section 2, general decorrelation solutions with the two most classical techniques of statistical component extraction: PCA and RT. The ICA paradigm is introduced in section 3 and in section 3.3, a particular algorithmic implementation of ICA is presented for the linear case. An application to geophysical data series is discussed in section 4 using the time series of tropical sea surface temperature variability.

2. Classical Component Extraction Techniques

We present here an overview of the classical techniques of component extraction. These methods rely on the hypothesis that the observations $\boldsymbol{x}$ are linear mixtures of uncorrelated components $\boldsymbol{\sigma}$ (See Notation section at the end of the article). Here after the vectors and the matrices are indicated in bold face type. The aim of the standard techniques is to infer uncorrelated components $\boldsymbol{h}$ that are estimators of $\boldsymbol{\sigma}$.

2.1. Linear Hypothesis

Let $\mathcal{D}$ be a set of M observations (data) of dimension N: $\mathcal{D} = \{\boldsymbol{x}^j \in \mathbb{R}^N \; ; \; j = 1, \; \ldots, \; M\}$. The multidimensional data $\boldsymbol{x} \in \mathcal{D}$ is a sample (or observation), like a time series or a geophysical field. We suppose here that data are centered and normalized. So the observation $\boldsymbol{x}$ is replaced by

$$\frac{(\boldsymbol{x} - < \boldsymbol{x} >)}{\sigma_x}, \tag{1}$$

where σ_x is the standard deviation of $\boldsymbol{x}$ and the angle brackets represent the statistical expectation. Let $\boldsymbol{O}$ be the $N \times M$ matrix with the set of M samples $\boldsymbol{x} \in \mathcal{D}$ in columns.

It is assumed that the observation $\boldsymbol{x}$ (or response) is the result of the mixture of Q different components $\boldsymbol{\sigma} = (\sigma_i \; ; \; i = 1, \; \ldots, \; Q)$. In classical component extraction techniques the number of components Q, not known in practice, has to be specified, and the observation $\boldsymbol{x}$ is assumed to be a linear mixture of the components $\boldsymbol{\sigma}$:

$$\boldsymbol{x} = \boldsymbol{A} \cdot \boldsymbol{\sigma} + \boldsymbol{\varepsilon}, \tag{2}$$

where $\boldsymbol{A}$ is the $N \times Q$ mixture matrix and the dot is the scalar product.

Given a set of observations $\boldsymbol{x}$, we want to determine the components $\boldsymbol{\sigma}$. The inversion of the mixture model (2) is generally an ill-posed problem. Two strategies are possible: (1) estimating directly the generalized inverse $\boldsymbol{A}^\dagger$ of the mixture matrix $\boldsymbol{A}$ by a $Q \times M$ matrix $\boldsymbol{J}$, this is the approach adopted in PCA, SVD, and ICA

techniques; (2) estimating $\boldsymbol{A}$ by a matrix $\hat{\boldsymbol{A}}$ and then inverting the estimated model by $\boldsymbol{J} = \hat{\boldsymbol{A}}^{\dagger}$, this is the approach of FA.

Taking one or the other strategies, the noise term ε is neglected and we search for an estimator $\boldsymbol{h}$ of $\boldsymbol{\sigma}$ with

$$\boldsymbol{h} = \boldsymbol{J} \cdot \boldsymbol{x}. \tag{3}$$

The Q rows of $\boldsymbol{J}$ are called filters because they filter the observation $\boldsymbol{x}$ to obtain the estimator $\boldsymbol{h} = \boldsymbol{J} \cdot \boldsymbol{x}$ of $\boldsymbol{\sigma}$. The Q columns $\{\boldsymbol{g}_j \in \mathbb{R}^N \; ; \; j = 1, \; \ldots, \; Q\}$ of the estimated matrix $\boldsymbol{J}^{\dagger}$ are called base functions because the observations $\boldsymbol{x}$ can be decomposed as

$$\boldsymbol{x} \simeq h_1 \cdot \boldsymbol{g}_1 + h_2 \cdot \boldsymbol{g}_2 + \cdots + h_Q \cdot \boldsymbol{g}_Q \tag{4}$$

If $Q = N$ and if all base functions $\boldsymbol{g}_i$ are orthogonal, the latter equation is an equality. The components $\{h_j \; ; \; j = 1, \; \ldots, \; Q\}$ are the projections of the observation $\boldsymbol{x}$ on the base functions.

2.2. Decorrelation of the Components

The goal of classical component extraction techniques is to infer uncorrelated components $\{h_j \; ; \; j = 1, \; \ldots, Q\}$ from the observations $\boldsymbol{x}$. Extracting uncorrelated components implies the diagonalization of the $Q \times Q$ covariance matrix $< \boldsymbol{h} \cdot \boldsymbol{h}^t >$ (the superscript t represents the transpose operation). The scaling factor is undetermined, but generally the data are normalized. So the covariance matrix becomes the identity matrix: $< \boldsymbol{h} \cdot \boldsymbol{h}^t >= \boldsymbol{I}_{Q \times Q}$, where $\boldsymbol{I}_{Q \times Q}$ is the $Q \times Q$ identity matrix.

$$\Leftrightarrow \boldsymbol{J} \cdot < \boldsymbol{x} \cdot \boldsymbol{x}^t > \cdot \boldsymbol{J}^t = \boldsymbol{I}_{Q \times Q} \tag{5}$$

$$\Leftrightarrow \boldsymbol{J}^t \cdot \boldsymbol{J} = < \boldsymbol{x} \cdot \boldsymbol{x}^t >^{-1} = \boldsymbol{C}_{xx}{}^{-1}, \tag{6}$$

where $\boldsymbol{C}_{xx} =< \boldsymbol{x} \cdot \boldsymbol{x}^t >= \frac{1}{M} \boldsymbol{O} \cdot \boldsymbol{O}^t$ is the $N \times N$ covariance matrix of $\boldsymbol{x} \in \mathcal{D}$. This means that $\boldsymbol{J}$ has the form

$$\boldsymbol{J} = \boldsymbol{V} \cdot \boldsymbol{C}_{xx}{}^{-1/2}, \tag{7}$$

where $\boldsymbol{V}$ is any $Q \times N$ matrix with $\boldsymbol{V}^t \cdot \boldsymbol{V} = \boldsymbol{I}_{N \times N}$.

Let $\boldsymbol{\Sigma}$ be the diagonal matrix of the decreasing eigenvalues of $\boldsymbol{C}_{xx}$ and let $\boldsymbol{E}$ be the matrix with the associated normalized eigenvectors of $\boldsymbol{C}_{xx}$ in the columns. By definition we have $\boldsymbol{C}_{xx}{}^{-1/2} = \boldsymbol{E} \cdot \boldsymbol{\Sigma}^{-1/2} \cdot \boldsymbol{E}^t$. Expression (7) can then be written

$$\boldsymbol{J} = \boldsymbol{V} \cdot \boldsymbol{E} \cdot \boldsymbol{\Sigma}^{-1/2} \cdot \boldsymbol{E}^t. \tag{8}$$

In the no-noise condition ($\varepsilon = \mathbf{0}$ in equation (2)), and if $Q \leq N$, only Q eigenvalues are strictly positive in $\boldsymbol{\Sigma}$. Let $\tilde{\boldsymbol{\Sigma}}$ be the $Q \times Q$ diagonal matrix with the strictly positive eigenvalues of $\boldsymbol{\Sigma}$ and let $\tilde{\boldsymbol{E}}$ be the $N \times Q$ matrix of the associated eigenvectors. Expression (8) can be replaced by

$$\boldsymbol{J} = \boldsymbol{V} \cdot \tilde{\boldsymbol{E}} \cdot \tilde{\boldsymbol{\Sigma}}^{-1/2} \cdot \tilde{\boldsymbol{E}}^t = \boldsymbol{\Theta} \cdot \boldsymbol{J}_0, \tag{9}$$

where $\boldsymbol{\Theta} = \boldsymbol{V} \cdot \tilde{\boldsymbol{E}}$ is an undetermined $Q \times Q$ matrix, such as $\boldsymbol{\Theta}^t \cdot \boldsymbol{\Theta} = \boldsymbol{I}_{Q \times Q}$ and $\boldsymbol{J}_0 = \tilde{\boldsymbol{\Sigma}}^{-1/2} \cdot \tilde{\boldsymbol{E}}^t$, a $Q \times N$ matrix.

The solution of the decorrelation problem is then reduced to the determination of a $Q \times Q$ matrix $\boldsymbol{\Theta}$ (with $\boldsymbol{\Theta}^t \cdot \boldsymbol{\Theta} = \boldsymbol{I}_{Q \times Q}$), provided the data from $\mathcal{D}$ are premultiplied by $\boldsymbol{J}_0$ (this step is called prewhitening). With the undetermination in $\boldsymbol{\Theta}$ an infinity of solutions exists which solves the decorrelation of components: the criterion of diagonalization of the covariance $< \boldsymbol{h} \cdot \boldsymbol{h}^t >$ (or the correlation) matrix leads to an infinite number of alternative final solutions $\boldsymbol{J}$.

Some solutions for $\boldsymbol{J}$ are presented in the next sections. However, all decorrelation solutions make the Gaussian hypothesis for the components $\boldsymbol{\sigma}$ in equation (2). This hypothesis is not always valid in real world applications (see the tropical sea surface temperature application in section 4, for example): we will see that ICA does not make this assumption.

2.3. Principal Component Analysis

Principal component analysis [*Morrison*, 1976], also called Hotteling principal components or the Kharunen-Loeve algorithm, leads to the determination of a particular solution of (9) and is very frequently used in many areas.

In the PCA approach, the noise term $\boldsymbol{\varepsilon}$ is not explicitly taken into account in the model of equation (2). It is assumed that the variance produced by noise is eliminated when deleting small eigenvalues of $\boldsymbol{C}_{xx}$ in the diagonal of $\tilde{\boldsymbol{\Sigma}}$.

The PCA solution is obtained by expressing $\boldsymbol{\Theta}$, in (9), as $I_{Q \times Q}$. Therefore the solution matrix becomes $\boldsymbol{J} = \boldsymbol{J}_0 = \tilde{\boldsymbol{\Sigma}}^{-1/2} \cdot \tilde{\boldsymbol{E}}^t$. This solution corresponds to the decorrelation solution where the first component describes the axis along which the variance is highest (in the full data space). The second component describes the axis (in the data subspace) along which the remaining variance is highest, *etc.* This technique is often preferred in comparison to other component extraction methods due to its mathematical simplicity. This simplicity implies a number of very useful properties (such as space/time equivalence [*Von Storch and Hannoschock*, 1984] in geophysical applications).

There are many examples of geophysical applications of the PCA. For example, *Lorentz* [1951] analyzed the seasonal and irregular variations of the Northern Hemisphere sea level pressure profile. However, as *Richman* [1986] has shown, the orthogonality of base functions has not always a physical significance. So this technique should only be applied if an a priori information is given concerning the orthogonality of the physical phenomena under study.

The consequences of this orthogonality constraint on unsuitable problems are that the structure of the basis function is global in the domain under study and often appears to be the same, similar to a spectrum Fourier analysis: the first base function possesses one node,

the second possesses two opposite nodes, the third base function possesses two opposite nodes with a different orientation, the fourth possesses four opposite nodes, *etc.* Another problem for PCA solutions is the constraint for the successive extracted components to explain the maximum remaining variance: this may lead to the mixing problem of physical phenomena in the extracted components [*Kim et al.*, 1999]; that is, physical phenomena, without links between us, are gathered into one extracted component. So if this approach is useful as a tool for compressing information, it can lead to misinterpretation for physical analysis.

2.4. Rotation Techniques

PCA, SVD, or FA techniques extract the uncorrelated components $\boldsymbol{h}$ by projecting the observations $\boldsymbol{x}$ on the rows of the matrix $\boldsymbol{J}$. Therefore the rows of $\boldsymbol{J}$ can be seen as a set of axes describing a subspace of the space of observations. In the RT approach, the axes of a previous decorrelation solution (such as PCA) are rotated. For instance, rotating the axes of the PCA matrix $\boldsymbol{J}_0$ leads to the equation

$$\boldsymbol{J} = \boldsymbol{\Theta} \cdot \boldsymbol{J}_0, \tag{10}$$

where $\boldsymbol{\Theta}$ is a rotation matrix.

The goal of rotational techniques is to overcome the undeterminacy of decorrelation solutions using additional information than that of second-order statistics. This additional information takes the form of a localization constraint: the criteria for rotation are derived from the idea of the so-called "simple structure" [*Morrison*, 1976]. Quantities are defined in order to determine the distribution of the variance of the data on each axis with the aim of localizing the basis functions in its domain (space, time, *etc.*) and then facilitating their interpretation. This type of analysis yields physically meaningful results and is not only a pure data reduction technique.

There are various criteria for rotation [*Richman*, 1986]. *Richman* cites many possible ways of orthogonal rotation criteria (such as quartimax, varimax, transvarimax) and of oblique rotation criteria (quartimin, biquartimin, oblimax). The number of possible criteria is a real drawback. Minimizing these quantities can lead to the rotation of axes in an orthogonal way (the axes stay orthogonal) or in an oblique way (the orthogonality of axes is not preserved).

Rotation techniques have been used in many works [*Horel*, 1981; *Barnston and Livezey*, 1987; *Eder et al.*, 1999] in order to improve the interpretation of extracted components: the result is that the extracted components are more geographically localized where the PCA technique searches for global components. However, the user of this technique may expect meeting problems in the case where the physical components are not localyzed.

So concerning the localization or the nonlocalization of physical components, it appears that neither the PCA nor the rotational techniques are adapted for any kind of application. The use of PCA or rotational techniques is then difficult if no a priori information is available on the physical phenomena at the origin of the variability of observed data. These assumptions, even without physical meaning, are partly reponsible for the sensitivity of extracted components to the geographical range under study. We give, in Table 1, the different hypothesis made by the classical and the ICA techniques.

3. Independent Component Analysis

We suppose now that the observations $\boldsymbol{x}$ are a mixture, linear or not, of statistically independent components $\boldsymbol{\sigma}$:

$$\boldsymbol{x} = \mathcal{A}(\boldsymbol{\sigma}). \tag{11}$$

The goal of the ICA paradigm is to find an inverse mapping $\Phi : \boldsymbol{x} \to \boldsymbol{h}$, where $\boldsymbol{h}$ is an estimator of $\boldsymbol{\sigma}$, specifying that the $\{h_i\ ;\ i = 1, \ \ldots, \ Q\}$ are statistically independent. So the estimator $\boldsymbol{h}$ is defined as a deterministic (linear or nonlinear) function of the observation

$$h_i = \Phi_i(\boldsymbol{W}_i, \boldsymbol{x})\ ;\ \ i = 1, \ \ldots, \ Q, \tag{12}$$

where $\{\boldsymbol{W}_i\ ;\ \ i = 1, \ \ldots, \ Q\}$ denotes the set of functional parameters. The number Q of independent components, not known in advance in most cases, has to be specified. A first initial guess is to take Q equal to the number N of inputs (dimension of $\boldsymbol{x}$); we will return to that point later on.

The mapping Φ will be searched among a given family of mappings, the simplest example being the linear mappings. In fact, any given family of mappings represents some prior hypotheses on the data structure (it is an implicit mixture model for the data): the case of a linear mixture of components is detailed below. The mapping is parameterized by the functional parameters

Table 1. Theoretical Hypothesis Made by PCA, RT, and ICA Techniques

	PCA	RT	ICA
Mixture model	linear	linear	linear or nonlinear
Component distributions	Gaussian	Gaussian	no assumption
Domain of basis functions	global	local	no assumption
Mutual constraint of basis functions	orthogonality	orthogonality or nonorthogonality	no assumption

$\{\boldsymbol{W}_i\ ;\ i = 1, \ \ldots, \ Q\}$ in such a way that some gradient descent algorithm onto a chosen cost function is possible.

3.1. An Objective Cost Function for Statistical Independence

3.1.1. Minimizing the redundancy. The statistical independence of Q variables $\{h_i\ ;\ i = 1, \ \ldots, \ Q\}$ is realized when the probability distribution function P_h of $\boldsymbol{h}$ is factorized as

$$P_h(\boldsymbol{h}) = P_1(h_1) \cdot P_2(h_2) \ldots P_Q(h_Q), \qquad (13)$$

where P_i denotes the marginal probability distribution functions of h_i. Statistical independence is stronger a constraint than decorrelation, which is based on second-order statistics only. It is only for Gaussian distributions that decorrelation is equivalent to factorization, but data may show also non-Gaussian statistics. An interesting example concerning the non-Gaussian character of the El-Nino phenomenon is given by *Burgers and Stephenson* [1999]. The goal of ICA is to search for $\boldsymbol{h}$, a linear or nonlinear mixture of the data $\boldsymbol{x}$, so that the probability distribution function of $\boldsymbol{h}$ satisfies as much as possible the above factorization (13). One has thus to define objective cost functions that quantify the quality of the solution $\boldsymbol{h}$ in terms of factorization. Several choices are possible [*Comon*, 1994]: here we focus on criteria derived from information theory [*Comon*, 1994; *Nadal and Parga*, 1994].

Information theory [*Blahut*, 1988] gives us tools for quantifying the statistical dependency between random variables. The fundamental quantity is the mutual information between variables. Given two random variables h_1 and h_2, the mutual information $I(h_1, h_2)$ is defined as the Kullback divergence between the joint distribution $P_{12}(h_1, h_2)$ and the factorized distribution $P_1(h_1) \cdot P_2(h_2)$:

$$I(h_1, h_2) = H(h_1) - H(h_1/h_2), \qquad (14)$$

$$= \int_{-\infty}^{+\infty} dh_1 dh_2 P_{12}(h_1, h_2) \log(\frac{P_{12}(h_1, h_2)}{P_1(h_1) P_2(h_2)}), \qquad (15)$$

where $H(\cdot)$ is the entropy of one variable and $H(h_1/h_2)$ is the entropy of h_1 given h_2. The mutual information $I(h_1, h_2)$ is a positive quantity and is null if and only if h_1 and h_2 are statistically independent; that is, $P_{12}(h_1, h_2) = P_1(h_1) \cdot P_2(h_2)$ (except possibly on a zero measure set). The mutual information depends on all higher-order statistics of the variables. This definition extends naturally to an arbitrary number Q of variables. The redundancy $\mathcal{R}(\boldsymbol{h})$ between the Q components of $\boldsymbol{h}$ is precisely defined as the mutual information between these Q variables:

$$\mathcal{R}(\boldsymbol{h}) = \int_{-\infty}^{+\infty} \prod_{i=1}^{Q} dh_i \ P_h(\boldsymbol{h}) \ \log \frac{P_h(\boldsymbol{h})}{\prod_{i=1}^{Q} P_i(h_i)}. \qquad (16)$$

The goal of ICA is thus to find a linear or nonlinear mapping $\Phi : \boldsymbol{x} \rightarrow \boldsymbol{h}$ which minimizes redundancy (16). However, computing redundancy (16) directly is not generally possible in practice, particularly for $\boldsymbol{h}$ in high-dimension spaces. Fortunately, one can build alternative cost functions leading to redundancy reduction and which can be used in a gradient descent scheme. The following general result allows one to build an efficient cost function that can be used with many different models Φ (linear or not).

3.1.2. Infomax approach. As stated above, the reduction of redundancy is based on higher-order statistics. So to obtain statistical independence, the manipulation of higher moments (like $< {h_i}^3 >, < {h_i}^4 >, < {h_i}^5 >, \ \ldots$) is required. Applying nonlinearities f_i on the h_i's allows one to pick up these higher-order moments because the Taylor expansion uses powers of the h_i values. So we consider a nonlinear transformation (postfiltering step) on the estimator $\boldsymbol{h} = \Phi(\boldsymbol{x})$:

$$y_i = f_i(h_i); \ \ i = 1, \ \ldots, \ Q, \qquad (17)$$

where the $f_i(\cdot)$ values are nonlinear, bounded, invertible functions of a single variable. The redundancy of equation (16) is zero if and only if the redundancy is zero for the variables $\boldsymbol{y} = (y_i\ ;\ i = 1, \ \ldots, \ Q)$. So the minimization of redundancy $\mathcal{R}(\boldsymbol{h})$ of equation (16) is equivalent to the minimization of $\mathcal{R}(\boldsymbol{y})$.

Redundancy (16) characterizes the statistical dependence of the h_i values or the y_i values but is not a priori related to the dependency between these variables and the observation $\boldsymbol{x}$. Let us now consider the mutual information between the observation and the filtered components $\boldsymbol{y}$.

Technically, to define $I(\boldsymbol{x}, \boldsymbol{y}) = H(\boldsymbol{y}) - H(\boldsymbol{y}|\boldsymbol{x})$ one has to introduce some additive noise in the definition of the output $\boldsymbol{y}$ (equation (17)). The variable $\boldsymbol{y}$ is a deterministic (linear or nonlinear) function of the data; see equations (12) and (17). The equivocation $H(\boldsymbol{y}|\boldsymbol{x})$ is a constant (it depends on the noise distribution only). Leaving out this constant, one can then take the limit of vanishing noise. So in our case, the mutual information $I(\boldsymbol{x}, \boldsymbol{y})$ is equal, up to a constant, to the output entropy $H(\boldsymbol{y})$ given by

$$H(\boldsymbol{y}) = - < ln[P_y(\boldsymbol{y})] > = - \int_{-\infty}^{+\infty} d\boldsymbol{y} \cdot P_y(\boldsymbol{y}) \cdot ln[P_y(\boldsymbol{y})] \qquad (18)$$

Nadal and Parga [1994] have proven that the maximization of the mutual information $I(\boldsymbol{x}, \boldsymbol{y})$ (hence of the output entropy $H(\boldsymbol{y})$) with respect to the adaptation of both the functional parameters $\{\boldsymbol{W}_i\ ;\ i = 1, \ \ldots, \ Q\}$ and the functions $\{f_i(\cdot)\ ;\ i = 1, \ \ldots, \ Q\}$ leads to the minimization of redundancies $\mathcal{R}(\boldsymbol{y})$ and $\mathcal{R}(\boldsymbol{h})$. Finally, the entropy $H(\boldsymbol{y}) = H\left(f_i(\Phi_i(\boldsymbol{W}_i, \boldsymbol{x}))\right); \ \ i = 1, \ \ldots, \ Q)$ of equation (18) can be maximized by gradient ascent with respect to the $\boldsymbol{W}_i$ values and the functions f_i. Full redundancy reduction is, however, possible only if the mapping family (the Φ_i values) is well chosen; that is, if

it is properly related to the data structure. In the other case, a consequence could be that the solution of the minimization of statistical dependence does not imply decorrelation. This is now illustrated in the case of a linear mixture.

3.2. Linear Case

3.2.1. Linear mixtures and perceptron architectures. Let us now make the hypothesis that the observations $\boldsymbol{x}$ are a linear mixture of independent components $\{\sigma_i\ ;\ i = 1, \ldots, Q\}$, with unknown probability distribution function $\{\rho_i(\sigma_i)\ ;\ i = 1, \ldots, Q\}$:

$$\boldsymbol{x} = \boldsymbol{A} \cdot \boldsymbol{\sigma}, \tag{19}$$

where $\boldsymbol{A}$ is the unknown mixture matrix, assumed to be invertible. The goal of ICA is thus to compute an estimate $\boldsymbol{J}$ of the inverse of $\boldsymbol{A}$, so $\boldsymbol{h} = \boldsymbol{J} \cdot \boldsymbol{x}$ is an estimator of $\boldsymbol{\sigma}$, with an underterminacy on scale and permutation (e.g., remark of section 2.1). The proper model associated with (19) is thus a linear one and the functional parameters $\boldsymbol{W}_i$ of equation (12) constitute the rows of the matrix $\boldsymbol{J}$,

$$h_i = \Phi_i(\boldsymbol{W}_i, \boldsymbol{x}) = \boldsymbol{J}_i \cdot \boldsymbol{x} = \sum_{j=1}^{N} \boldsymbol{J}_{ij} x_j, \tag{20}$$

where $\boldsymbol{W}_i = \boldsymbol{J}_i$ is the vector whose components are given by the ith row of the matrix $\boldsymbol{J}$. To make use of the result briefly presented above, we consider the processing obtained by applying a transfer function to each h_i, leading to the global model

$$F(\boldsymbol{W}, \cdot) : \boldsymbol{x} \to \boldsymbol{y} = f(\boldsymbol{h}) = f(\boldsymbol{J} \cdot \boldsymbol{x}). \tag{21}$$

By analogy with neural network modeling, the parameters $\boldsymbol{J}$ are called the "couplings" or "synaptic weights," and the f_i values "transfer functions." This model (21) is called a perceptron in neural network modeling [*Hertz et al.*, 1992], with synaptic couplings $\boldsymbol{J}$ and transfer functions f_i (Figure 1). For a nonlinear model for the mixture of components, we will use a multi layer perceptron.

The general result given by *Nadal and Parga* [1994] gives, in that particular case, that maximizing the mutual information $I(\boldsymbol{x}, \boldsymbol{y})$ with respect to the coupling matrix $\boldsymbol{J}$ and the transfer functions leads to redundancy reduction: at the optimum, one has $\boldsymbol{J} = \boldsymbol{A}^{-1}$ and in addition,

$$\frac{df_i(h_i)}{dh_i} = \rho_i(h_i)\ ;\ i = 1, \ldots, Q. \tag{22}$$

This means that in the optimal case, the transfer functions would be related to the probability distribution of the components. At the optimum, each output unit y_i has a uniform distribution, corresponding to the maximization of entropy (Figure 2) [*Nadal and Parga*, 1994].

3.2.2. Infomax principle in the linear case. We now write the entropy of equation (18) in a convenient form for algorithmic purposes. As in the work of *Bell and Sejnowski* [1995], we make in the integral in equation (18) the change of variable $\boldsymbol{y} \to \boldsymbol{h}$ which is well defined since we consider invertible transfer functions and the invertible matrix $\boldsymbol{J}$. One has

$$P_y(\boldsymbol{y}) = \frac{P_h(\boldsymbol{h})}{|\boldsymbol{G}|}, \tag{23}$$

where $\boldsymbol{G}$ is the $Q \times Q$ Jacobian matrix equal to

$$\boldsymbol{G}_{ik} = \frac{\partial y_i}{\partial x_k} = \frac{df_i}{dh_i} \boldsymbol{J}_{ik}. \tag{24}$$

The entropy then reads $H(\boldsymbol{y}) = - < ln(P_h(\boldsymbol{h})) > + < ln(|\boldsymbol{G}|) >$; that is,

$$H(\boldsymbol{y}) = - < ln(P_h(\boldsymbol{h})) > + < \ln(|\frac{df_i}{dh_i}|) > + \ln |det \boldsymbol{J}|. \tag{25}$$

It is now easy to perform some gradient ascent on $H(\boldsymbol{y})$ (or a gradient descent on $-H(\boldsymbol{y})$). The first term in equation (25) is the entropy of the input data, hence a constant. The other terms depend on the couplings and the transfer functions. To maximize $H(\boldsymbol{y})$, one may restrict the transfer functions to a parametrized family of functions. A particular scheme is presented by *Pham et al.* [1992]. Another strategy, followed in the work of *Bell and Sejnowski* [1995], is to expect that full optimization of the transfer functions is not necessary for performing ICA. The transfer functions in (25) are necessary in order to generate terms depending on higher-order statistics. Choosing an adequate fixed transfer function might then be enough for obtaining redundancy reduction. For example, if one expects the component distribution to have a unique maximum,

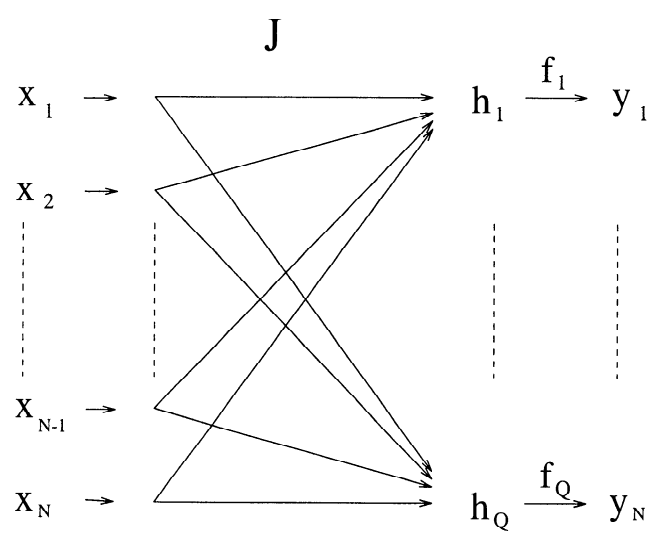

Figure 1. Component extraction model, the perceptron architecture: $\boldsymbol{x}$ is the observation and $\boldsymbol{h}$ is the extracted component vector.

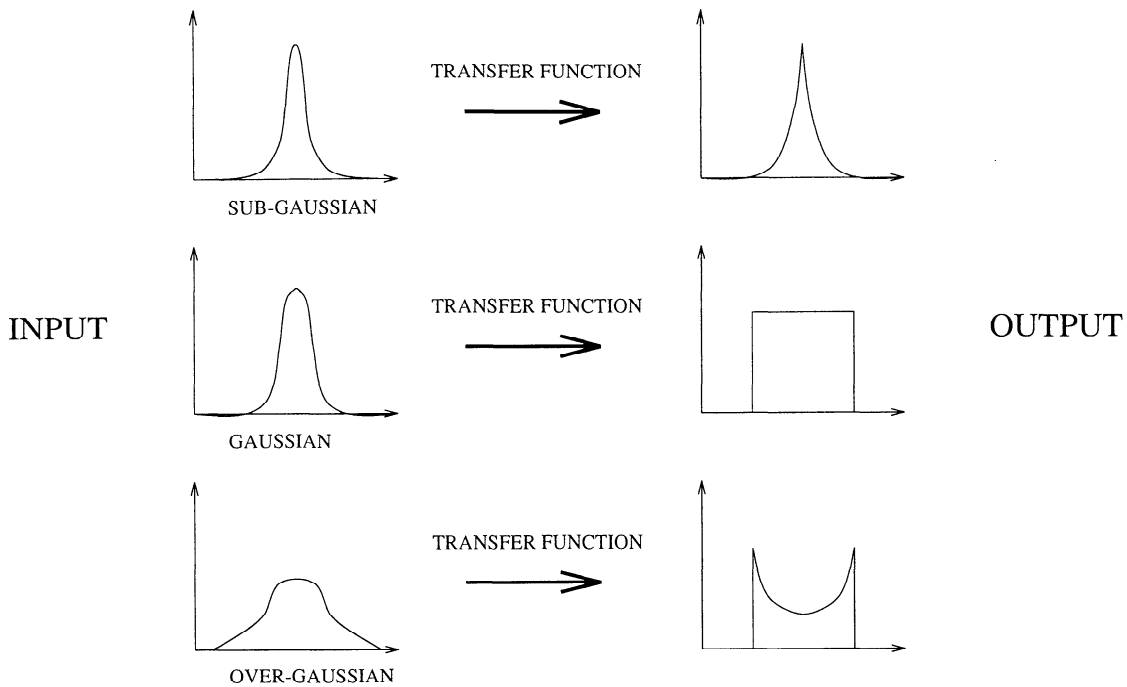

Figure 2. Probability density functions in the output layer of the component extraction model, before (component $\boldsymbol{h}$) and after (variable $\boldsymbol{y}$) the transfer function f.

one can try transfer functions which are monotonic (so that their derivatives have the required shape). In doing so, there is no longer any guaranty that the source separation will be obtained. However, promising results have been obtained [*Bell and Sejnowski*, 1995], and this strategy can be easily improved by introducing some partial adaptation of the transfer functions.

3.3. Algorithmic Implementation of ICA in the Linear Case

The algorithm (See web page http://www.cnl.salk.edu /~tewon/ica_cnl.html from the Computational Neuroscience Laborabory of Terry Sejnowski at The Salk Institute for links to recent literature, softwares and demos concerning the ICA paradigm) we have chosen to implement in the linear case relies on the work of *Bell and Sejnowski* [1995]. They have developed an algorithm that has the advantage of being extremely simple; we present the natural gradient version [*Amari*, 1998] of it in section 3.3.1. We have modified this algorithm in section 3.3.2: a preprocessing step is added in order to limit the number of extracted components [*Nadal et al.*, 2000].

3.3.1. A simple learning algorithm. Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be, respectively, the input and the output of the neural network described in section 3.2.1. Every transfer function is taken as the logistic function:

$$f_i(h) = (1 + e^{-h})^{-1}. \tag{26}$$

The quality criterion used to estimate the weights $\boldsymbol{J}$ of the neural model is, as explained in section 3.1.2, the output entropy $H(\boldsymbol{y})$. We have seen in the later section that the minimization of $-H(\boldsymbol{y})$ is equivalent to the minimization of

$$- < ln(|\boldsymbol{G}|) >= - < \ln(|\frac{df_i}{dh_i}|) > - \ln|det\boldsymbol{J}|. \tag{27}$$

To avoid local minima during the minimization of the criterion, stochastic steepest descent is used. The learning step is made for only one sample chosen iteratively and stochastically in $\mathcal{D}$

$$\Delta\boldsymbol{J} \propto \frac{\partial ln(|\boldsymbol{G}|)}{\partial \boldsymbol{J}}. \tag{28}$$

If the transfer function is the logistic function (26), the later equation becomes [*Bell and Sejnowski*, 1995]:

$$\Delta\boldsymbol{J} \propto [\boldsymbol{J}^t]^{-1} + \bar{\boldsymbol{y}} \cdot \boldsymbol{x}^t, \tag{29}$$

where

$$\bar{y}_i = \frac{\partial}{\partial y_i}\frac{\partial y_i}{\partial h_i} = \frac{\partial}{\partial h_i} ln(\frac{\partial y_i}{\partial h_i}) = 1 - 2y_i. \tag{30}$$

This learning step requires the inversion of the matrix $\boldsymbol{J}^t$: it is expensive in time computation and may be unstable. *Amari* [1998] shows that one should use the "natural gradient" (instead of the absolute gradient). This is a more robust learning step that consists in multiplying the absolute gradient by $\boldsymbol{J}^t \cdot \boldsymbol{J}$:

$$\Delta\boldsymbol{J} \propto ([\boldsymbol{J}^t]^{-1} + \bar{\boldsymbol{y}} \cdot \boldsymbol{x}^t) \cdot (\boldsymbol{J}^t \cdot \boldsymbol{J}) = \boldsymbol{J} + \bar{\boldsymbol{y}} \cdot \boldsymbol{h}^t \cdot \boldsymbol{J}. \tag{31}$$

For the weights of the bias term,

$$\Delta \boldsymbol{J}_0 \propto \bar{\boldsymbol{y}} \cdot (\boldsymbol{J}^t \cdot \boldsymbol{J}). \tag{32}$$

In terms of individual weight, this corresponds to

$$\begin{cases} \Delta J_{ik} \propto J_{ik} + \bar{y}_i \cdot \sum_l J_{lk} \cdot h_l \\ \Delta J_{i0} \propto \bar{y}_i \cdot \sum_k \sum_l J_{lk} \cdot J_{li} \end{cases} \tag{33}$$

3.3.2. Preprocessing of the data and initialization of the weights. The ICA solution could be initialized randomly [*Bell and Sejnowski*, 1995], but it is a good strategy to begin the learning process with a "reasonable" initial solution: learning is faster and the risk for the learning algorithm to be trapped in a local minimum of the cost function is reduced.

A natural approach is to use as initial condition a linear projection that decorrelates observations. This allows for the extraction of second-order statistics. This preprocessing step is called "data whitening" and is obtained by performing PCA or the SVD (theoretically, in this context, equivalent to the PCA but more stable numerically). In addition, it is possible to use PCA or SVD for reducing the dimension of the input space. *Nadal et al.* [2000] demonstrated that the ICA solution is not perturbed by this preprocessing step. We have seen in section 2.2 that a decorrelation solution has the following form (equation (9)): $\boldsymbol{J} = \boldsymbol{\Theta} \cdot \boldsymbol{J}_0$, where $\boldsymbol{J}_0$ is the $Q \times N$ matrix of the PCA (or the SVD) solution. So it is possible to preprocess the data in $\mathcal{D}$ as

$$\tilde{\boldsymbol{x}} \leftarrow \boldsymbol{J}_0 \cdot (\boldsymbol{x} - < \boldsymbol{x} >). \tag{34}$$

The new database is constituted by samples $\tilde{\boldsymbol{x}}$ with uncorrelated coordinates and dimension Q. Then, the ICA learning technique computes a $Q \times Q$ particular matrix solution for $\boldsymbol{\Theta}$: $\boldsymbol{J}_{\text{ica}}$. The global solution, taking into account the prewhitening step, is the $Q \times N$ matrix $\boldsymbol{J} = \boldsymbol{J}_{\text{ica}} \cdot \boldsymbol{J}_0$. The Q ICA base functions are the columns of the inverse $N \times Q$ matrix $\boldsymbol{J}^{-1}$. This matrix is computed using the decomposition $\boldsymbol{J}^{-1} = \boldsymbol{J}_0{}^{-1} \cdot \boldsymbol{J}_{\text{ica}}{}^{-1}$, where $\boldsymbol{J}_0{}^{-1} = \boldsymbol{J}_0{}^t$ and $\boldsymbol{J}_{\text{ica}}{}^{-1}$ is estimated by the SVD technique instead of the Jacobi algorithm for stability considerations.

The two major advantages of this preprocessing step are (1) the ICA learning process estimates a $Q \times Q$ matrix $\boldsymbol{\Theta}$ instead of a $N \times N$ matrix and (2) the initial solution $\boldsymbol{\Theta} = \boldsymbol{I}_{Q \times Q}$ corresponds to the PCA or the SVD solution.

It is important to note that the initial PCA solution is changed by the ICA learning step only if the statistical independence criterion considers that the modification is necessary. For the matrix $\boldsymbol{J}_0$ the SVD solution is preferred to the PCA solution for stability considerations. The estimation of $\boldsymbol{\Theta}$ by the ICA algorithm ($\boldsymbol{\Theta} = \boldsymbol{J}$ in previous notations) may be seen as a rotation of the SVD initial solution $\boldsymbol{J}_0$.

4. Application of ICA for Analysis of Tropical SST

We present in this study, for evaluating ICA applicability in geophysical data, an application concerning a real and complex problem. Between the large variety of geophysical variables, the sea surface temperature (SST) offers a good opportunity: it is the result of many climatological processes involving the ocean, the atmosphere, and their coupling [*Chen et al.*, 1997]. So SST appears to be a complex mixture reflecting variabilities in a large variety of timescales acting in different geographical areas [*Barnett*, 1991]. Moreover, long global time series of SST observations are available.

The study of tropical ocean SST variability is of particular interest because it is governed by several important and complex nonperiodic phenomena like the El Nino Southern Oscillation (ENSO), which represents about 30% of the total tropical variance and is the major source of SST variability in the tropics. The variance in the Indian and the Atlantic Oceans is about half as large [*Bottomley et al.*, 1990].

While the Pacific ENSO was intensively studied [*Rasmusson and Carpenter*, 1982], the tropical Atlantic has received less attention. Strong links have been pointed out between climatic phenomena in the tropical Atlantic and various climate-related disasters occurring around the tropical Atlantic Basin. The droughts of Northeast Brazil, for example, are closely related to warm SST anomalies in the tropical Atlantic Ocean [*Enfield and Mayer*, 1997]. These problems are presently of increasing interest for climatologists.

Many studies suggest that important relationships exist between northern and southern equatorial Atlantic SSTs [*Chu*, 1984], between Pacific and Atlantic SSTs [*Enfield and Mayer*, 1997], and between Indian and Pacific SSTs [*Klein et al.*, 1999]. We will focus our analysis on these teleconnections.

4.1. Data Sources

4.1.1. GISST2.2 Database. The GISST2.2 (global sea-ice and sea surface temperature) database of the UKMO (United Kingdom Meteorological Office) [*Parker et al.*, 1995] contains a global gridded monthly climatology of sea surface temperature. This SST database was created by mixing satellite observations and *in situ* measurements. The spatial resolution is $1° \times 1°$. We have selected in this database the better consistant period that ranges from 1961 to 1994. GISST2.2 is a robust and well-known SST database and has already been used in many studies [*e.g., Torrence and Webster*, 1998].

We have chosen to study tropical SST variability (20°S-20°N in latitude; 0°-360° in longitude) during the period 1961-1994 = 34 years. We use gridded monthly means, so the temporal dimension is T = 34 years $\times$ 12 months = 408. With $1° \times 1°$ surface

resolution and for the region under study, there exist 11,275 over sea pixels. This resolution could be artificial in several regions since most of this resolution merely derives from interpolation. However, the results of our analysis are independent of this point, the pixels being processed independently. So the spatial dimension is $M = 11,275$.

4.1.2. Preprocessing of data. The database is supposed to be centered and normalized as explained in section 2.1. In general, geophysical data are primarily affected by seasonal variability. The study of other (interannual or intraseasonal) variabilities is rendered easier by the removal of seasonal variability. For that reason the database is also deseasonalized by removing from each of the 11,275 pixels under study its corresponding season. Making this, we made the classical assumption that the forcing of the seasonal cycle has additive consequences on SST. The season-removing process transforms the rough data into anomalies. This is the only filtering procedure used in this study. So the following spectral analysis of the extracted component (see sections 4.2.2 and 4.3) is not conditioned to find a predetermined kind of periodicities; in other words, no assumptions are made about the periodicities we are interested in.

4.1.3. Non-Gaussianity of the SST. A random variable is characterized by all its statistical moments: the first moment is the mean, the second moment is the variance, the third moment is the skewness, the fourth moment is the kurtosis [*Press et al.*, 1992]. For Gaussian variables, moments higher than 2 are zero. When data have zero-mean, the "skewness" $\text{skew}(X) = \frac{<X^3>}{\sigma^3}$ and the "kurtosis" $\text{kurt}(X) = \frac{<X^4>}{\sigma^4} - 3$. These "normalized" moments are often used to test a departure from the Gaussian behavior (Figure 3). The skewness measures the symmetry of the probability distribution function: when the skewness is positive, rare events have more probabilities to be high than low, and the reverse is true when the skewness is negative. The kurtosis is a measure of the sharpness of the distribution: a negative kurtosis indicates that the distribution have no sharp central peak and have large tails, when a positive kurtosis indicates that the distribution has a sharp central peak.

We have analyzed the first four moments of the GISST database to analyze the Gaussianity of the SST. In Plate 1, we have represented the mean, the standard deviation, the skewness, and the kurtosis of the SST using all months. In Plate 2, this analysis is made only for the months of May to test the time variability of these statistics. For skewness we have represented only the absolute values higher than 0.2: the standard deviation of the skewness estimate for a Gaussian distribution is equal to $\sqrt{15/N} = 0.19$ [*Press et al.*, 1992]. For kurtosis we have represented only the absolute values higher than 0.5: the standard deviation of the kurtosis estimate for a Gaussian distribution is equal to $\sqrt{96/N} = 0.48$. The coherence on local regions gives also a confidence of the robustness of these statistical moment estimates. Plate 1 clearly shows the non-Gaussian behavior of the SST (i.e., nonzero skewness and kurtosis) over large areas of the ocean. In particular, as previously commented by *Burgers et al.* [1999], the upwelling region of the eastern Pacific, which is strongly affected by ENSO, has a positive skewness (warmer SSTs more probable than colder) and a positive kurtosis (distribution more peaked than the Gaussian distribution). Recently, *Timmermann* [1999] uses the skewness to caracterize the statistical changes of ENSO in a CO_2 doubling experiment. Burgers shows also that the kurtosis is negative in the warm-pool region; probably because of nonlinear radiative/convective greenhouse feedbacks causing SST saturation at 30°C. These local effects demonstrate the non-Gaussian character of the ENSO phenomenon.

We also see in Plate 2 that the tropical Atlantic variability of the SST is governed, in particular in May, by non-Gaussian processes in a dipole (north/south) spa-

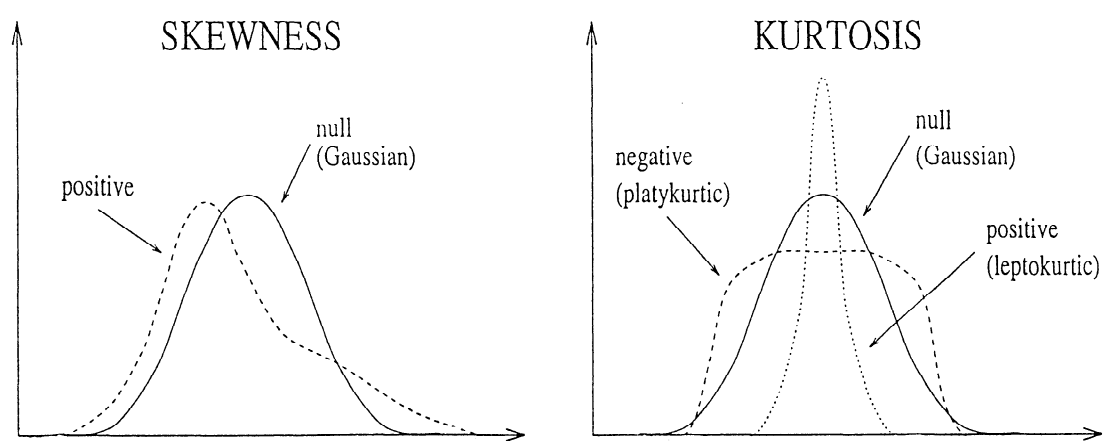

Figure 3. Measure of non-Gaussian behavior of distributions by (a) their third moment (i.e., skewness) and (b) fourth moment (i.e., kurtosis).

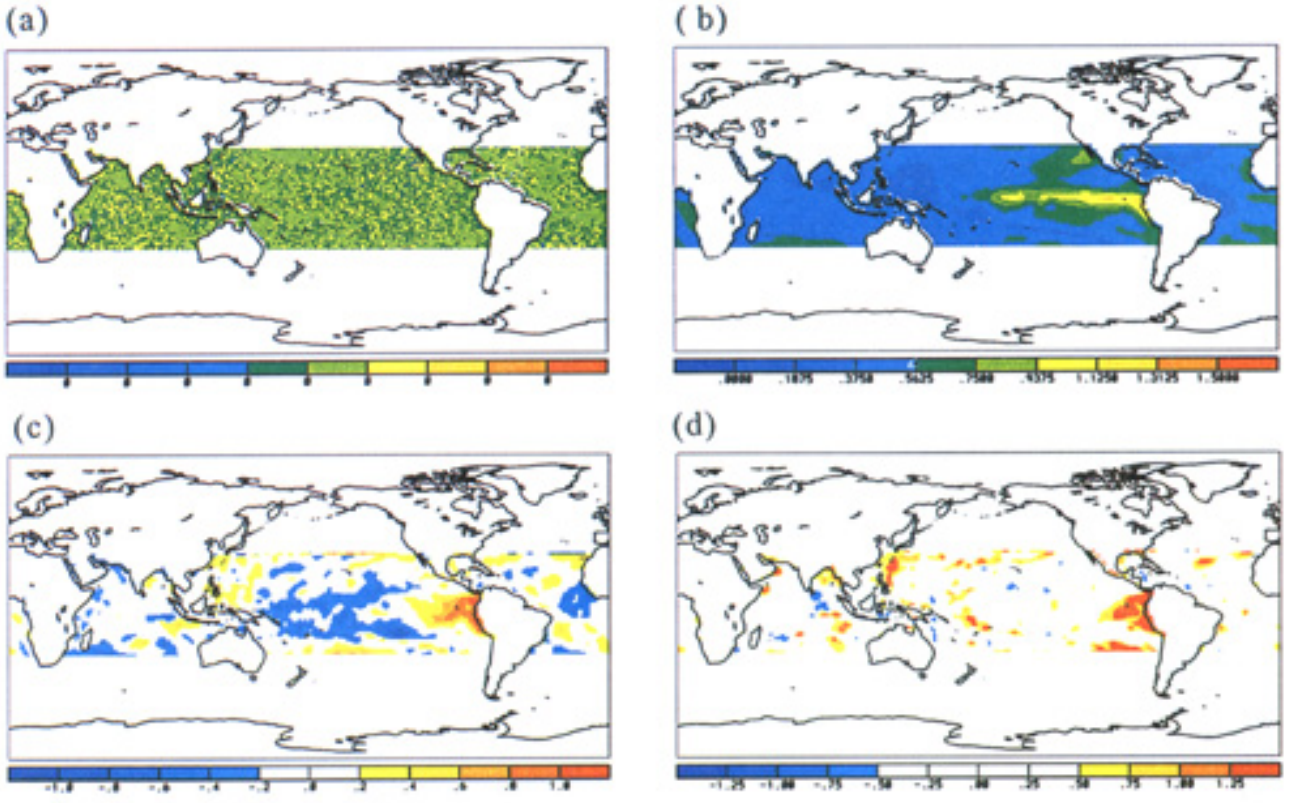

Plate 1. First four moments of the tropical sea surface temperature (SST) anomalies for all months: (a) average, (b) standard deviation, (c) skewness, and (d) kurtosis.

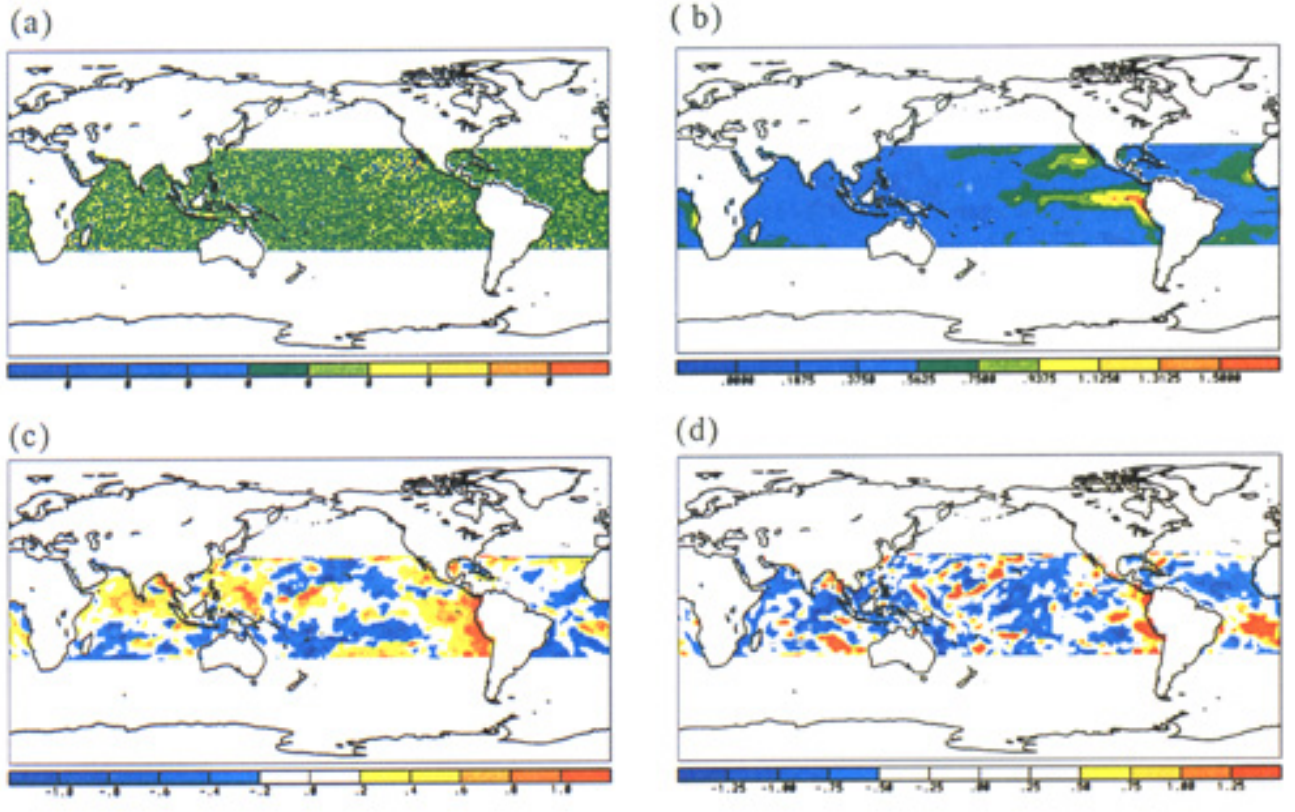

Plate 2. First four moments of the tropical SST anomalies for May months: (a) average, (b) standard deviation, (c) skewness, and (d) kurtosis.

tial structure suggesting that the well-known Atlantic dipole phenomenon [*Chu*, 1984] could also be related to a non-Gaussian behavior.

Therefore we may conclude that the physical processes at the origin of climatological variability of tropical SST are governed, at least partly, by non-Gaussian variables. The use of the ICA approach is then required because it does not make the Gaussian hypothesis and is well adapted for the analysis of higher-order distributions, contrary to classical techniques.

4.2. Methodology

We consider here that the database $\mathcal{D}$ is composed of a geophysical time series $\Upsilon = \{X_s{}^t;\ s = 1, \ \ldots, \ S;\ t = 1, \ \ldots, \ T\}$, where s is the spatial index (in a regular grid or in an irregularly spaced network of measurement stations), and t is the time index. We want to determine the estimator $\boldsymbol{h}$ of the components $\boldsymbol{\sigma}$ causing the variability in the Υ time series.

We have seen, in sections 2 and 3, different kinds of techniques to extract components from an observed signal (PCA, RT, ICA). However, there are many approaches to apply these techniques in Υ. Each approach emphasizes one particular aspect of the data: temporal, spatial, or frequency. The best strategy to adopt also depends on the available data set because statistical techniques require a sufficient number of samples.

4.2.1. Geophysical model adopted: analysis in time. Observations $\boldsymbol{x}$ have to be extracted from the Υ database to compose the inference set $\mathcal{D} = \{\boldsymbol{x}^j \in \mathbb{R}^N\ ;\ j = 1, \ \ldots, \ M\}$. Analysis in time is particularly well adapted here, the temporal dimension (408) of the database $\mathcal{D}$ being much smaller than its spatial dimension (11,275). So this analysis concerns the study of the temporal correlations in Υ. In that case the samples $\boldsymbol{x}(j) \in \mathcal{D}$ are chosen as $\boldsymbol{x}(j) = (X_j{}^t\ ;\ t = 1, \ \ldots, \ T)$, describing the temporal evolution of one variable in a given geographical location j. The number M of samples is the spatial dimension S of the time series Υ. The dimension N of the samples is the temporal dimension T in Υ. So using the database $\mathcal{D}$ can be written as

$$\mathcal{D} = \{\boldsymbol{x}(j) \in \mathbb{R}^{408}\ ;\ j = 1, \ \ldots, \ 11,275\}. \quad (35)$$

The ICA technique is applied to analyze the database $\mathcal{D}$: it computes a $Q \times T$ filter-matrix $\boldsymbol{J}$. Each row of $\boldsymbol{J}$ is a filter: the scalar product of a time series by the ith filter estimates the component h_i (equation (3)). The temporal base functions $\{\boldsymbol{g}_i(t)\ ;\ i = 1, \ \ldots, \ Q\}$ are the columns of the inverse $T \times Q$ matrix $\boldsymbol{J}^{-1}$. This analysis results in the decomposition of one time series $\boldsymbol{x}(j)$ in the pixel j:

$$\boldsymbol{x}(j) \simeq h_1(j) \cdot \boldsymbol{g}_1 + h_2(j) \cdot \boldsymbol{g}_2 + \cdots + h_Q(j) \cdot \boldsymbol{g}_Q, \quad (36)$$

where each of the Q base functions $\boldsymbol{g}_i$ describes a canonical dynamical behavior (or behavior "prototype," or "cluster" [*Richman*, 1981]) of the variable. Determination of these prototypes uses the statistical independence criterion between estimated components h_i.

It is important to know that a complex physical phenomenon has generally different dynamical behaviors acting in different geographical locations. So to be properly described, such phenomena must be decomposed into various prototypes describing all the variabilities in time. This is called the splitting of a given climatological phenomenon into several extracted components. The temporal base function of one component can be more intense, for example, before or after the maximum phase of the phenomenon, and this component will be used differently in each pixel locations.

On the other hand, some physical phenomena could be mutually linked, so one component could gather these several physical phenomena. The set of Q prototypes extracted by the analysis technique aims at summarizing, as well as possible, all kinds of temporal variations present in the data set. Extracting these prototypes is the key to a better physical understanding of the dynamical behavior of the phenomenon, and important conclusions could be drawn from this meaningful representation.

The term $h_i(j)$ is the component score of the ith base function $\boldsymbol{g}_i$ in the observation $\boldsymbol{x}(j)$. The h_i values are supposed to be mutually statistically independent. These components are better "state variables" for prediction (their physical meaning helps at defining a prediction model, and their statistical use for prediction is easier through their independence. The projection of each pixel time series onto the ith base function $\boldsymbol{g}_i$ gives us the ith component "score map" $\{h_i(j)\ ;\ j = 1, \ \ldots, \ S\}$. A component "score map" indicates the geographical location of the pixels following the dynamical behavior of the corresponding base function $\boldsymbol{g}_i$. They have no direct links with anomalies.

4.2.2. Interpretation of the extracted components. Interpretation of the extracted components goes through the analysis of the temporal base functions $\boldsymbol{g}_i$ and of their geographical localization brought by the component score maps.

Two important tools are also available. The first one is the spectral analysis of the base functions $\{\boldsymbol{g}_i(t)\ ;\ i = 1, \ \ldots, \ Q\}$: based on the spectral density estimate, a smoothing version of the periodogram given by a Fourier analysis [*Eder et al.*, 1999]. This analysis allows characterizing its different time periodicities via the spectral peaks observed. Here periodicities higher than 7 years have been removed on figures because, for statistical significance considerations, these periodicities cannot be retrieved from the spectral analysis of time series of only 34 years.

The second tool for the interpretation of the components is the level of correlation between each base function and well-known climatological indices. We give, in Table 2, a list of indices (see the NOAA Climate Prediction Center web site: http://www.cpc.noaa.gov/data) characterizing the three climatological phenomena that are of interest for our application: the ENSO (El Niño / Southern Oscillation), the NA index (North Atlantic

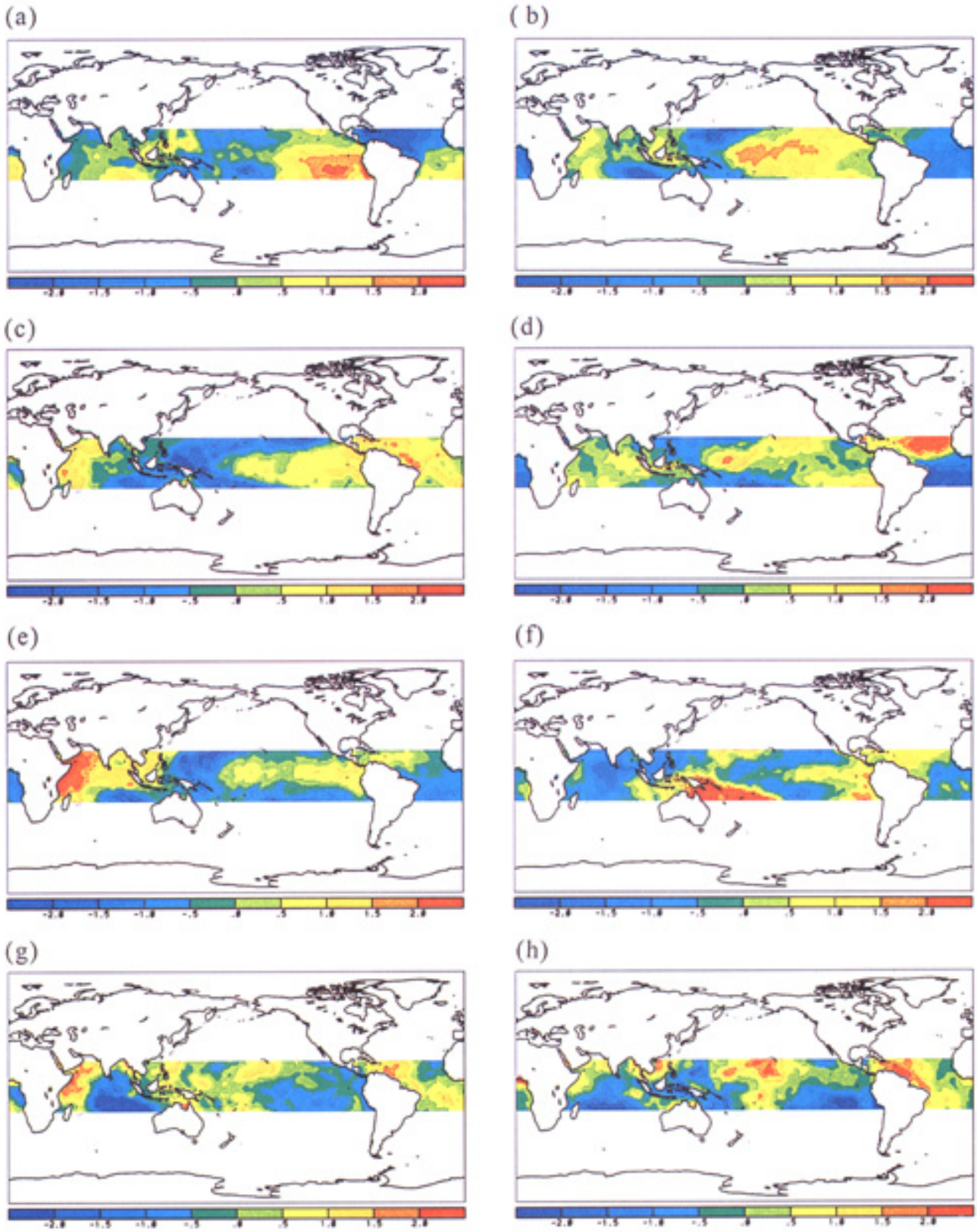

Plate 3. Geographical amplitude h_i of extracted components: ICA component score maps.

Table 2. Climatological Indices Used in This Study

Acronym	Name	Region
SOI/SLP	Southern Oscillation Index / sea level pressure	Tahiti-Darwin
NIÑO 3	El Niño	(5°N-5°S;150°W-90°W)
HFSST = Ship6	Hawaï-Fiji / SST	Pacific
NA	North Atlantic index	(5°-20°N;60°-30°W)
SA	South Atlantic index	(0°-20°S;30°W-10°E)

index), and the SA index (South Atlantic index), as defined by the NOAA. See, for example, *Ropelewski and Jones* [1987] for a detailed study of the Southern Oscillation Index (SOI). The correlation between two time series can be lagged: a positive or a negative lag is introduced in one of the two time series before the computation of the correlation. This technique allows analysing delayed effects between physical phenomena.

4.3. ICA Results

Prewhitening of data is performed, as explained in section 3.3.2, by projecting the data on the 10 first PCA components to reduce the dimensionality [*Nadal et al.*, 2000]. Reducing the dimension of data to 10 seems to be a good compromise between the number of ICA components extracted and the variance explanation: this prewhitening step keeps a sizeable part, 55.9%, of the total variance in data (Figure 4).

We present in Figure 5 the resulting ICA base functions, and the associated component score maps (section 4.2.1) are presented in Plate 3. Among the 10 components extracted by the ICA, eight may be quite easily interpreted on the basis of our present knowledge of tropical SST variability.

4.3.1. Component ICA 1. The link of this component with the tropical North Atlantic variability is well illustrated with an important node in the score map (Plate 3 a) in this region. Furthermore, the correlation of the temporal base function with the NA index is -0.59. The spectral peaks (Figure 5 b) of the temporal base function of this component are 5.3, 2.3 years, and 1 year. The percent variance explained by this component is 8.1%.

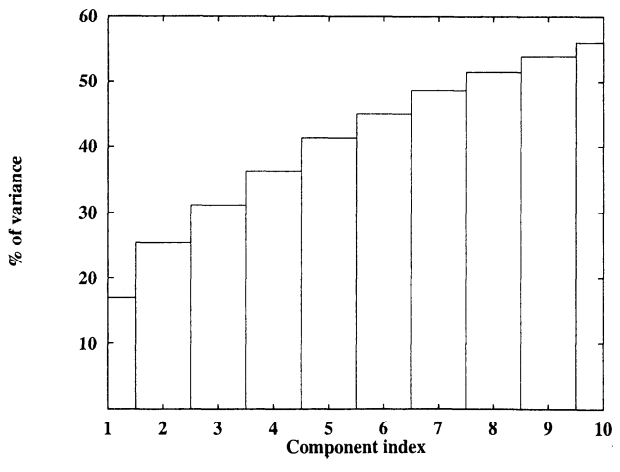

Figure 4. Cumulative percentage of variance of the 10 first singulat value decomposition (SVD) components of the tropical SST anomalies.

The interesting thing is that an opposite node also exists in the tropical Southeast Pacific (i.e., upwelling region, over Equator and Peru), that is a region strongly affected by ENSO (with anomalous wet conditions). This teleconnection is confirmed by the work of *Delecluse et al.* [1994], suggesting that the Pacific ENSO provides one possible source of external forcing for the interannual SST variability in the eastern equatorial Atlantic.

4.3.2. Component ICA 2. The corresponding score map (Plate 3 b) shows the classical ENSO pattern in the equatorial Pacific [*Wallace et al.*, 1998]. The variance explained by this component represents 4.7%. Spectral peaks of the base function associated to this component are 4.2 and 1.8 years (Figure 5 d). The 4.2 year spectral peak results from the quasi-quadrennial cycle of the ENSO. The 1.8 year spectral peak illustrates the tendency for a quasi-biennial rhythm [*Rasmusson et al.*, 1990]. The correlation of the temporal base function with the Ship6 index is 0.50.

4.3.3. Component ICA 3. The component score map (Plate 3 c) clearly shows the link of this component with the ENSO phenomenon. The variance explained represents 3.6% of the total variance. Spectral peaks of the temporal base function of this component are 4, 2 years, and 1 year (Figure 5 f). The periodicity of 1 year corresponds to the tendency of ENSO for the phase locking with the annual cycle. The link with the equatorial Atlantic variability is real, and the correlation of the temporal base function with the NA index is -0.44. This component, as ICA component 1, indicates the correlation between the equatorial Atlantic variability and the ENSO phenomenon [*Delecluse et al.*, 1994].

4.3.4. Component ICA 4. The structure of the component score map (Plate 3 d) shows an interhemispheric "dipole" structure in the tropical Atlantic. This "dipole" pattern is a well known mode of variability in equatorial Atlantic SST [*Servain et al.*, 1999]. The percent of the explained variance is 6%. Spectral peaks of the associated temporal base function are 2.1 years

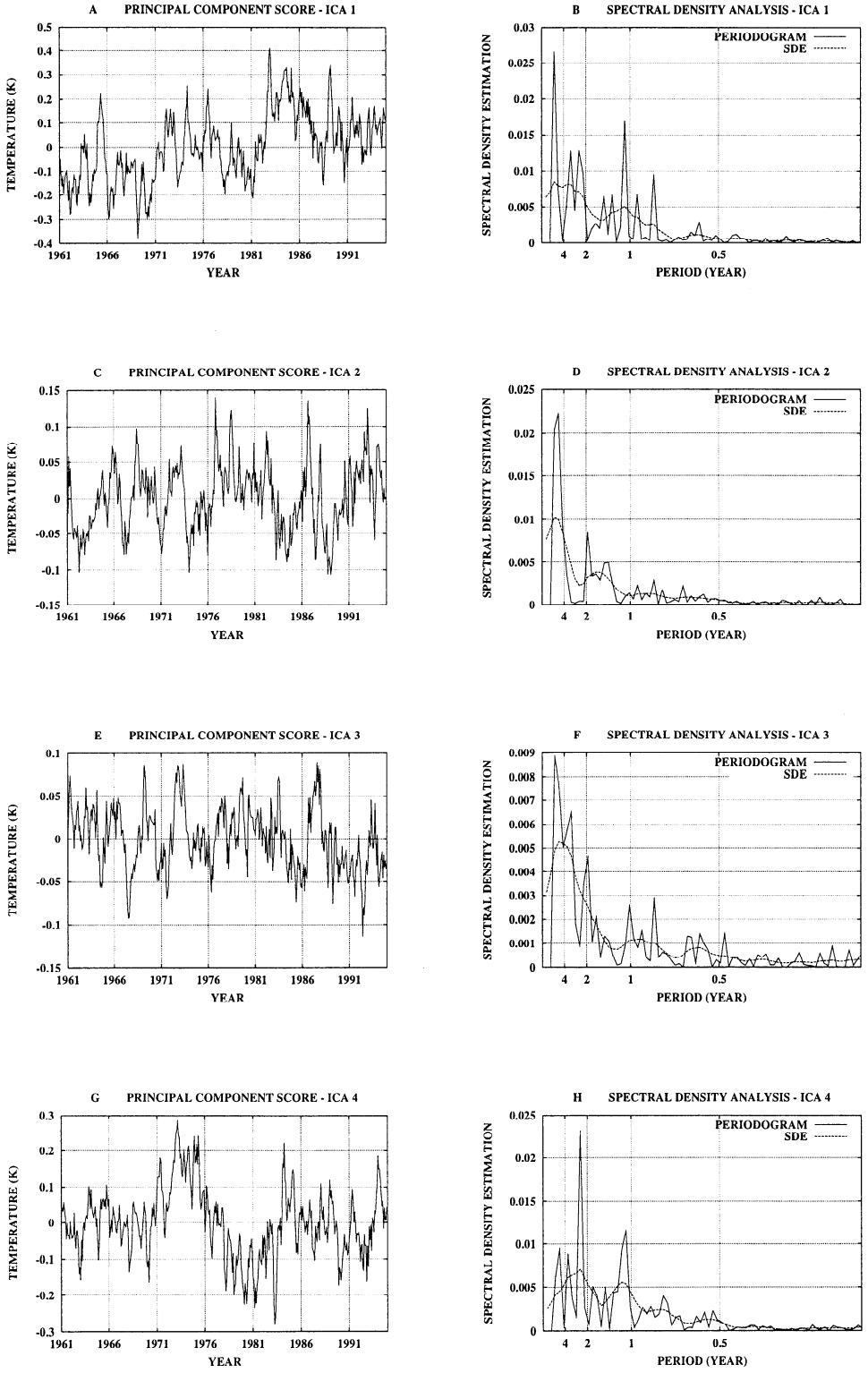

Figure 5a. Independent component analysis (ICA) components: $N^0 1$ to $N^0 4$: temporal base functions g_i (left) and corresponding frequency spectra (right)

and 1 year (Figure 5 h). Also observed in the associated temporal base function (Figure 5 g), a periodicity greater than 10 years, the spectral analysis does not allow identifying such a low frequency due to the shortness of the time series. The correlation of the temporal base function with the NA index is -0.60 and the correlation with the SA index is 0.41.

This interhemispheric SST gradient is well known as the principal source of variability in the tropical Atlantic Ocean. Its preferred periodicities are quasi-

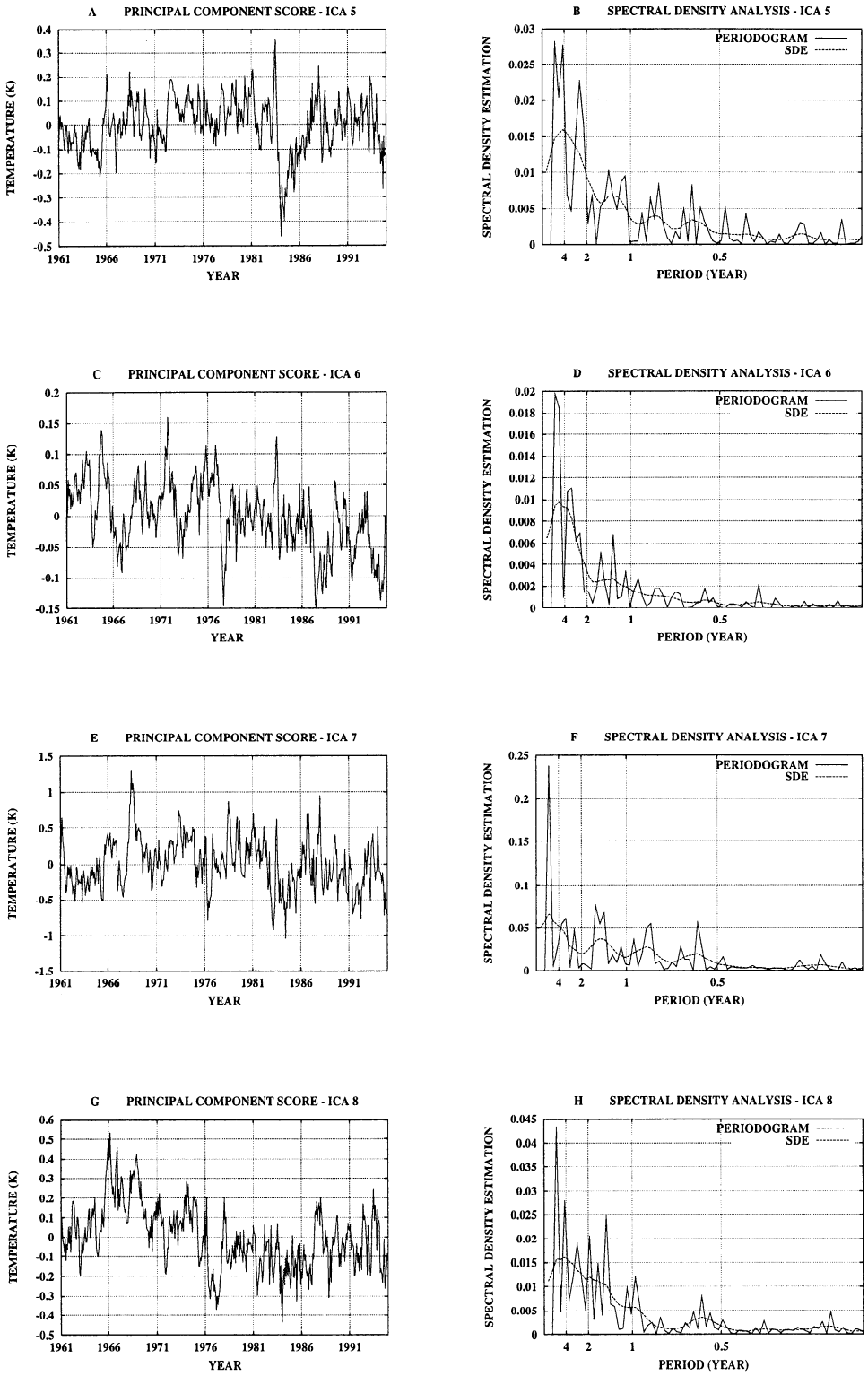

Figure 5b. Independent component analysis (ICA) components: $N^0 5$ to $N^0 8$: temporal base functions g_i (left) and corresponding frequency spectra (right).

biennial and around 13 years [*Chu*, 1984]. The physical origin of the "dipole" is not yet understood. It has, however, been shown that it has links with the extra tropical Atlantic and particularly with the NAO [*Namias*, 1972]. Another link was found with the ENSO phenomenon [*Enfield and Mayer*, 1997; *Delecluse et al.*, 1994]. The structure of the component score map exhibits a pattern in the Pacific Ocean close to the classical ENSO pattern (given, for example, by the score map of component ICA 2).

This component is very important because it seems to be linked with various climatic disasters around the tropical Atlantic basin. In particular, the Northeast Brazil precipitation has shown a high correlation with this "dipole" SST anomaly [*Harzallah et al.*, 1996].

4.3.5. Component ICA 5. The link of this component with the ENSO is shown by the component score map (Plate 3 e). Furthermore, spectral peaks (Figure 5 j) of the temporal base function of this component are 4.2, 2.13, and 1.3 years with also some lower periodicities. We recognize the periodicities of the ENSO phenomenon. The explained variance reprensents 5.8%.

4.3.6. Component ICA 6. The component score map (Plate 3 f) of this component is clearly linked with the ENSO phenomenon. This hypothesis is confirmed by the 4.2 years spectral peak (Figure 5 l) of the temporal base function. This component represents another temporal variability associated to the dynamical process of ENSO. For example, the correlation is high (-0.48) between its temporal base function (Plate 3 a) and the Ship6 index when we introduce a lag of -8 months. The component score map also brings into evidence a link with the Indian Ocean. This component represents 4.3% of the total variance.

4.3.7. Component ICA 7. The component score map (Plate 3 g) shows that it affects essentially the Indian Ocean. A dipole structure is present in the Indian Ocean with a southeast/northwest gradient. This pattern is close to the September SST anomaly found in the work of *Terray* [1995, Figure 7]. Correlations of the temporal base function with those of other components ICA 5 (0.57), ICA 2 (0.51), and ICA 3 (-0.47), each being linked with ENSO, indicate that this component is related to the same ENSO phenomenon. This analysis is confirmed by the observation of SST anomalies (not shown). Spectral peaks of the temporal base function of this component are 4.2 and 1.3 years and 10 and 7 months (Figure 5 n). This component is a very important one because it represents 8.8% of the total variance.

4.3.8. Component ICA 8. The component score map (Plate 3 h) appears to be close to the score map of the component ICA 1. The variance explained associated to this component represents 7.2%. Spectral peaks of the temporal base function are 4.2, 2 years, and 1 year and also 7 months (Figure 5 p). These spectral peaks are also close to those of the component ICA 1. So this component is probably related to the same phenomenon but with a different temporal variability.

4.4. Discussion

Contrary to PCA results (not shown), where at best two or three components may be interpreted, ICA extracts eight quite easily interpretable components. We think that the PCA constraint of maximization of explained variance in the fisrt PCA components has a negative effect on solution: the linear/Gaussian/orthogonal PCA technique artificially mix variabilities that are not physically connected. This is the well-known mixing problem [*Kim et al.*, 1999]. On the other hand, statistical independence criterion used in ICA appears to extract individual components characterizing the dynamical complexity of each physical phenomena. Each basis function is being more a kind of physically plausible prototype that would be extracted, in a totally different approach, by a unsupervised classification approach, like clustering or Kohonen map techniques.

As shown in section 4.1.3, the ENSO phenomenon is highly non-Gaussian. So the PCA or related techniques are not well adapted for the analysis of this physical process, and the use of the ICA in this context appears interesting. ICA describes the variability of the ENSO phenomenon with several ENSO-related ICA components describing particular temporal behavior acting in different geophysical locations and exhibing teleconnections between oceans. This is the well-known splitting of one complex climatological phenomenon, with links on various oceans, into several statistically independent components that are prototypes of a dynamical behavior. Extracted ICA temporal base functions are prototypes that are used by the analysis to describe all the temporal variabilities of the ENSO phenomenon in all geographical regions.

If the component ICA 2 is the component with the geographical aspect, the closest to classical ENSO pattern, components 3, 5, and 6 (for the Pacific variabilities), components 1 and 8 (for relationships with the Atlantic variabilities), and component 7 (for the Indian Ocean variabilities) also display an ENSO information concerning a different dynamical behavior in other geographical locations. Spectral peaks associated to these extracted temporal base functions (quasi-quadrennal, quasi-biennial, and annual periodicities) are those associated to ENSO phenomenon [*Zhang et al.*, 1997]. We see that our analysis is then particularly robust since no filtering of data was used before the analysis.

The fact that a schematic sequence of component was not found by the analysis shows the aperiodicity of the ENSO phenomenon. Then ICA confirms that the quasi-periodic conception of the ENSO cycle is too simple and that, as commented by *Wallace et al.* [1998], the canonical conception of the ENSO scheme is nothing more than an ideal case. The ICA technique shows that the ENSO phenomenon cannot be reduced to a simple dynamical behavior (like an index), in all regions over the tropical Pacific [*Wallace et al.*, 1998]: there exist many ENSO components describing its space-time variabilities and the teleconnections between the oceans.

More generally, we could say that the ICA technique, which has no constraint either for globalization or for localization of base functions, allows the determination of interesting teleconnections. Previous studies on tropical SST using the rotational technique [*Kawamura*, 1994] or MSSA approach [*Moron et al.*, 1998] does not found these teleconnections. Our global ICA analysis gathers relationships between the ENSO and the Indian

and Atlantic Ocean variabilities confirming the findings of many previous specific approaches. In particular, we observe that teleconnections exist between the ENSO/Pacific and the tropical Atlantic (components ICA 1, 3, and 4) and between the ENSO/Pacific and the Indian Ocean variability (component 7).

5. Conclusion

We have introduced a new component extraction technique, based on information theory, for geophysical time series analysis. This technique aims at extracting statistically independent components, a constraint stronger and more pertinent than the decorrelation at the basis of classical techniques (PCA, SVD, FA). It can also be extended to a nonlinear mixture model.

This new technique has not the limitations of classical techniques: orthogonality of components extracted, globalization or localization inadequacy of PCA or RT, mixing of modes due to the constraint of maximum variance explained by extracted components, inadequacy or subjectivity of standard statistical quality criteria, *etc.* Assumptions at the basis of ICA and classical techniques (PCA, RT) have been summarized in Table 1, which shows that only ICA does not make artificial assumptions introducing misunderstanding on the analysis.

To illustrate ICA capability to analyze complex climatological phenomena, this technique has been applied to the tropical SST. Results obtained with ICA gathered results of many other specific studies using a large variety of techniques (time, space, frequency emphasis) and experimental conditions. So ICA appears to be a new tool particularly well adapted for the analysis of geophysical variabilities, with important teleconnections. The ICA technique clearly shows that important relationships exist between the northern and the southern equatorial Atlantic SSTs via the "dipole" pattern, between the Pacific and the Atlantic SSTs via the ENSO phenomenon, and the Pacific and Indian Oceans. These teleconnections were not found by rotational techniques or MSSA approaches. More work remains to be done to analyze in detail the results of ICA. However, our study has already shown that the potentialities of the ICA technique are very high compared to those of the classical techniques (like PCA or RT). It allows the geographical and temporal analysis of the physical components that constitute the geophysical time series.

Depending on the availability of a priori information about the basis functions at the origin of the variability in observations, the ICA technique can be used in other ways that analysis in time. An observation $\boldsymbol{x}$ can be a field describing the geographical variability of one parameter for a given date t. This is often referred to as EOF analysis [*von Storch and Frankignoul*, 1998]. The EOF analysis could be sophisticated in various ways: complex EOF if patterns evolve, extended EOF if many variables are linked, periodically extended EOF or cyclostationary EOF if periodicity assumptions could be made, *etc* [*Kim et al.*, 1999]. It is also possible to privilege the frequency aspect with the SSA (singular spectrum analysis) approach [*Broomhead and King*, 1986; *Vautard and Ghil*, 1989; *Vautard et al.*, 1992; *Plaut et al.*, 1995] or to analyze the space/time structure with the MSSA (multi-channel singular spectrum qnalysis) approach [*Plaut and Vautard*, 1994; *Vautard et al.*, 1996; *Zhang et al.*, 1997]. The interested reader is referred to *Richman* [1986] for more details concerning the different approaches.

This work opens numerous future projects: (1) develop a nonlinear mixture model with its associated learning algorithm [*Burel*, 1992]; (2) study the adaptation of the transfer functions (section 3.2.1); (3) analyze the behavior of a set of more variables to understand the physical processes inducing the variability on observations, for example, by adding the atmospheric water vapor for the study of the ENSO and its teleconnections; (4) use other spatiotemporal approaches (complex EOF, extended EOF, periodically extended EOF, cyclostationary EOF, *etc* [*Kim et al.*, 1999]) with the ICA technique; and (5) apply the results of ICA analysis to test ENSO predictions.

Appendix A: Notation

Matrices and vectors

- $\boldsymbol{A}$ mixing matrix.
- $\boldsymbol{\varepsilon}$ noise.
- $\boldsymbol{\sigma}$ components to extract.
- $\boldsymbol{h}$ estimator of components $\boldsymbol{\sigma}$.
- $\boldsymbol{x}$ a sample of the database, input of the neural network.
- $\tilde{\boldsymbol{x}}$ a prewhitened data.
- $\boldsymbol{y}$ output of the neural network.
- $\boldsymbol{O}$ $N \times M$ data matrix.
- $\boldsymbol{C}_{xx}$ covariance matrix of data ($=< \boldsymbol{x} \cdot \boldsymbol{x}^t >$).
- $\boldsymbol{W}$ couplings, synaptic weights, estimate of $\boldsymbol{A}^{-1}$.
- $\boldsymbol{J}_0$ preprocessing matrix.
- $\boldsymbol{V}$ matrix of eigenvectors of $\boldsymbol{C}_{xx}$.
- $\boldsymbol{\Sigma}$ eigenvalues of $\boldsymbol{C}_{xx}$.
- $\boldsymbol{\Lambda}$ an arbitrary diagonal matrix with nonzero element (arbitrary scales and signs).
- $\boldsymbol{\Pi}$ an arbitrary permutation matrix.
- $\boldsymbol{\Theta}$ a rotation matrix.
- $\boldsymbol{G}$ $Q \times Q$ Jacobian matrix $\boldsymbol{G}_{ij} = \frac{\partial y_i}{\partial x_j} = \frac{df_i}{dh_i}\boldsymbol{W}_{ij}$.

Scalars

- S dimension of the geographical field.
- Q number of components.
- N dimension of inputs $\boldsymbol{x}$.
- M number of samples in $\mathcal{D}$.
- T dimension of the time index.
- q time delay.

Sets

- $\mathcal{D}$ general data set.
- Υ geophysical data set.

Functions

- g_i base function.
- f_i, f transfer function.
- ρ_i probability distribution function of components σ_i.
- P_h probability distribution functions ($P_1 = P_{h_1}$).
- $\mathcal{R}$ redundancy.
- I information.
- H entropy.
- C correlation.
- F neural network mapping.
- Φ_i output coordinates of neural mapping.
- μ correlation function.

Symbols

- $\hat{}$ used for estimated variables.
- $\boldsymbol{A}^t$ transpose of matrix $\boldsymbol{A}$.
- $\boldsymbol{A}^\dagger$ generalized inverse of matrix $\boldsymbol{A}$.

Acknowledgments. The authors would like to thank Noëlle Scott, Augustin Vintzileos, Robert Vautard, and Ulrike Burkhardt for many fruitful discussions. They also would like to thank the UKMO (United Kingdom Meteorological Office) for providing us with the GISST2.2 database.

References

Amari, S.-I., Natural gradient works efficiently in learning, *Neural Comput.*, **10** (2), 251–276, 1998.

Amari, S. I., A. Cichocki, and H. H. Yang, A new learning algorithm for blind signal separation, in *Advances in Neural Information Processing Systems*, vol. 8, edited by D.S. Touretzky, M.C. Mozer, and M.E. Hasselmo, MIT Press, Cambridge, Mass., 1996.

Atick, J. J., Could information theory provide an ecological theory of sensory processing, *Network Comput. Neural Syst.*, **3**, 213–251, 1992.

Barlow, H. B., The coding of sensory messages, in *Current Problems in Animal Behaviour*, Edited by W.H. Thorpe, and O.L. Zangwill, pp. 331–360, Cambridge Univ. Press, New York, 1960.

Bar-Ness Y., J.W. Carlin, and M.L. Steinberg, Bootstrapping Adaptive Cross Pol Cancelers for Satellite Communication, Paper presented at the *Int. Conference on Comm.*, Conference Proceedings paper No: 4F.5, Philadelphia, PA, June 13-17, 1982.

Barnett, T., The interaction of multiple timescales in the tropical climate system, *J. Clim.*, **4**, 269–285, 1991.

Barnston, A., and R. Livezey, Classification, seasonality and persistence of low-frequency atmospheric circulation patterns, *Mon. Weather Rev.*, **115**, 1083–1126, 1987.

Bell, A. J., and T. J. Sejnowski, An information-maximization approach to blind separation and blind deconvolution, *Neural Comp.*, **7** (6), 1004–1034, 1995.

Blahut, R. E., *Principles and Practice of Information Theory*, Addison-Wesley-Longman, Reading, Mass., 1988.

Bottomley, M., C.K. Folland, J. Hsiung, R.E. Newell, and D.E. Parker, *Global Ocean Surface Temperature Atlas (GOSTA)*, Her Majesty's Stn. Off., Norwich, England, 1990.

Broomhead, D., and G. King, Extracting qualitative dynamics from experimental data, *Physica D*, **20**, 217–236, 1986.

Burel, G., Blind source separation of sources: A nonlinear algorithm, *Neural Networks*, **5**,937–947, 1992.

Burgers, G., and D.B. Stephenson, The "normality" of El Niño, *Geophy. Res. Lett.*, **26** (8), 1027–1030, 1999.

Cardoso, J.-F., Source separation using higher-order moments, Int. Conf. Acoust. Speech Signal Process., **89**, 2109–2112, 1989.

Chen, D., S. Zebiak, M. Cane, and A. Busalacchi, Initialization and predictability of a coupled ENSO forecast model, *Mon. Weather Rev.*, **124**, 773–788, 1997.

Chu, P., Time and space variability of rainfall and surface circulation in the northeast Brazil-tropical atlantic sector, *J. Meteor. Soc. Jpn.*, **62**, 362–370, 1984.

Comon, P., Independent Component Analysis, a new concept?, *Signal Process.*, **36**, 287–314, 1994.

Delecluse, P., J. Servain, C. Levy, K. Arpe, and L. Bengtsson, On the connection between the 1984 Atlantic warm event and the 1982-83 ENSO, *Tellus*, **4**, ser. A, 448–464, 1994.

Delfosse, N., and P. Loubaton, Adaptive blind separation of independent sources: A deflation approach, *Signal Process.*, **45**, 59–83, 1995.

Eder, B., S. LeDuc, and J. Sickles, A rotated principal component analysis of TOMS data for 1980-1992, *J, Geophys. Res.*, **104**, 3691–3709, 1999.

Enfield, D., and D. Mayer, Tropical Atlantic Sea Surface Temperature variability and its relation to El Niño - Southern Oscillation, *J. Geophys. Res.*, **102**, 929–945, 1997.

Harzallah, A., J. Rocha De Aragao, and R. Sadourny, Interannual rainfall variability in north-east Brazil: Observation and model simulation, *Int. J. Climatol.*, **16**, 861–878, 1996.

Hertz, J., A. Krogh, and R. Palmer, *Introduction to the Theory of Neural Computation*, Cambridge Univ. Press, New York, 1992.

Horel, J., A rotated principal component analysis of the interannual variability of the northern hemisphere 500 mb height field, *Mon. Weather Rev.*, **109**, 2080–2092, 1981.

Jutten, C., and J. Herault, Blind separation of sources, part I, An adaptive algorithm based on neuromimetic architecture, *Signal Process.*, **24**, 1–10, 1991.

Kawamura, R., A rotated EOF analysis of global sea surface temperature variability with interannual and interdecadal scales, *Int. J. Phys. Oceanogr.*, **24**, 707–715, 1994.

Kim, K.-Y., and Q. Wu, A comparison study of EOF techniques: Analysis of nonstationary data with periodic statistics, *J. Clim.*, **12**, 185–199, 1999.

Klein, S.A., B.J. Soden, and N.-C. Lau, Remote sea surface temperature variations during ENSO: Evidence for a tropical atmospheric bridge, *J. Clim.*, **12** (4), 917–932, 1999.

Lorenz, E., Seasonal and irregular variations of the northern hemisphere sea-level pressure profile, *J. Meteorol.*, **8**, 52–59, 1951.

Moron, V., R. Vautard, and M. Ghil, Trends, interdecadal and interannual oscillations in global sea-surface temperatures, *Clim. Dyn.*, **14**, 545–569, 1998.

Morrison, D., *Multivariate Statistical Methods*, McGraw-Hill, New York, 1976.

Nadal, J.-P., and N. Parga, Nonlinear neurons in the low-noise limit: A factorial code maximizes information transfer, *Network Comput. Neural Syst.*, **5**, 565–581, 1994.

Nadal, J.-P., and N. Parga, Redundancy reduction and independent component analysis: Conditions on cumulants and adaptive approaches, *Neural Comput.*, **9** (7), 1421–1456, 1997.

Nadal, J.-P., E. Korutcheva, and F. Aires, Blind source separation in the presence of weak sources, *Neural Networks*, in press, 2000.

Namias, J., Influence of the northern hemisphere general circulation on drought in northern Brazil, *Tellus*, **24**, 336–343, 1972.

Parker, D., M. Jackson, and E. Horton, The 1961-1990

GISST2.2 sea surface temperature and sea-ice climatology, *Tech. Rep.*, 63, Clim. Res., Hadley Cent., UK, 1995.

Pham, D.-T., P. Garrat, and C. Jutten, Separation of a mixture of independent sources through a maximum likelihood approach, *Proc. EUSIPCO*, 771–774, 1992.

Plaut, G., and R. Vautard, Spells of low-frequency oscillations and weather regimes in the northern hemisphere, *J. Atmos. Sci.*, 51(2), 210–236, 1994.

Plaut, G., M. Ghil, and R. Vautard, Interannual and interdecadal variability in 335 years of central England temperature, *Science*, **268**, 710–713, 1995.

Press, W.H.P., S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, *Numerical Recipes in Fortran*, Cambridge Univ. Press, New York, 1992.

Rasmusson, E., and T. Carpenter, Variations in tropical sea surface temperature and surface wind fields associated with the Southern Oscillation/El Niño, *Mon. Weather Rev.*, **110**, 354–384, 1982.

Rasmusson, E., X. Wang, and C. Ropelewski, The biennial component of ENSO variability, *J. Mar. Syst.*, **1**, 71–96, 1990.

Richman, M., Obliquely rotated principal components: An improved meteorological map typing technique?, *J. Appl. Meteorol.*, **20**, 1145–1159, 1981.

Richman, M., Rotation of principal components, *J. Climatol.*, **6**, 293–335, 1986.

Ropelewski, C., and P. Jones, An extension of the Tahiti-Darwin southern oscillation index, *Mon. Weather Rev.*, **115**, 2161–2165, 1986.

Servain, J., I. Wainer, J.P. McCreary, and A. Dessier, Relationship between the equatorial and meridional modes of climatic variability in the tropical Atlantic, *Geophys. Res. Lett.*, **26** (4), 485–488, 1999.

Storck, J., and G. Deco, Nonlinear independent component analysis and multivariate time series analysis, *Physica D*, **108**, 335–349, 1997.

Terray, P., Space-Time structure of monsoon interannual variability, *J. Clim.*, **8** (11), 2595–2619, 1995.

Timmermann, A., Detecting the nonstationary response of ENSO to greenhouse warming, *J. Atmos. Sci.*, **56**, 2313–2325, 1999.

Torrence, C., and P. Webster, The annual cycle of persistence in the El Niño/Southern Oscillation, *Q. J. R. Meteorol. Soc.*, **124**, 1985–2004, 1998.

Vautard, R., and M. Ghil, Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series, *Physica D*, **35**, 395–424, 1989.

Vautard, R., P. Yiou, and M. Ghil, Singular-spectrum analysis: A toolkit for short, noisy chaotic signals, *Physica D*, **58**, 95–126, 1992.

Vautard, R., C. Pires, and G. Plaut, Long-range atmospheric predictability using space-time principal components, *Mon. Weather Rev.*, **124** (2), 288–307, 1996.

Von Storch, H., and G. Hannoschock, Comments on empirical orthogonal function-analysis of wind vectors over the tropical Pacific region, *Bull. Am. Meteorol. Soc.*, **65** (2), 162–162, 1984.

Von Storch, H., and C. Frankignoul, Empirical modal decomposition in coastal oceanography, Chap. 16, in *The Sea*, vol. 10, edited by Kenneth H. Brink and Allan R. Robinson, pp. 419–455, John Wiley, New York, 1998.

Wallace, J., E. Rasmusson, T. Mitchell, V. Kousky, E. Sarachik, and H. von Storch, On the structure and evolution of ENSO-related climate variability in the tropical pacific: Lessons from TOGA, *J. Geophys. Res.*, **103**, 14,241–14,259, 1998.

Weigend, A. S., and N. A. Gershenfeld (Eds.), *Time Series Prediction*, Addison-Wesley-Longman, Reading, Mass., 1994.

Zhang, X., J. Corte-Real, and X. Wang, Low-frequency oscillations in the northern hemisphere, *Theor. Appl. Climatol.*, **57**, 125–133, 1997.

Filipe Aires, NASA Goddard Institute for Space Studies, 2880 Broadway, New-York, NY 10025. (filipe@sharki.giss.nasa.gov).

Alain Chédin, CNRS Laboratoire de Météorologie Dynamique, École Polytechnique, 91128 Palaiseau Cedex, France.

J.-P. Nadal, Laboratoire de Physique Statistique, École Normale Supérieure de Paris/CNRS, 75231 Paris Cedex 05, France.

(Received August 26, 1999; revised February 10, 2000; accepted February 22, 2000.)