

Analyse de séries temporelles géophysiques et théorie de l'information : l'analyse en composantes indépendantes

Analysis of geophysical time series and information theory: Independent Component Analysis

Filipe Aires^{a*}, Alain Chédin^a, Jean-Pierre Nadal^b

^a Laboratoire de météorologie dynamique du CNRS, École polytechnique, 91128 Palaiseau cedex, France**

^b Laboratoire de physique statistique de l'ENS, 24, rue Lhomond, 75231 Paris cedex 05, France***

(Reçu le 27 novembre 1998, accepté après révision le 15 mars 1999)

Abstract — With the aim of identifying the physical causes of variability of a given dynamic system, the geophysical community has made extensive use of the statistical component extraction techniques. We introduce here a recently developed algorithm based on information theory: the Independent Component Analysis. This 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). Secondly, the linear hypothesis for the mixture of components is not required. This new method is presented in the context of geophysical time series analysis. The ICA algorithm is applied to the study of the variability of the tropical Sea Surface Temperature, with a particular emphasis on the analysis of the links between El Niño/Southern Oscillation and the Atlantic SST variability. (© Académie des sciences / Elsevier, Paris.)

Independent Component Analysis (ICA) / neural networks / time series analysis / sea surface temperature / El-Niño

Résumé — Dans le but d'identifier les causes physiques de la variabilité d'un système dynamique, la communauté géophysique utilise de façon intensive les techniques statistiques d'extraction de composantes. Un algorithme récemment développé, fondé sur la théorie de l'information, est introduit dans ce travail : l'analyse en composantes indépendantes (ACI). Cette technique présente deux avantages majeurs sur les techniques classiques. Premièrement, elle a pour but d'extraire des composantes statistiquement indépendantes, là où les techniques classiques cherchent uniquement la décorrélation. Deuxièmement, l'hypothèse linéaire pour le mélange des composantes n'est pas requise. Cette nouvelle technique est présentée dans le contexte de l'analyse de séries temporelles géophysiques. L'algorithme ACI est appliqué à l'étude de la variabilité de la température de surface de l'océan (TSO) tropical, avec une attention particulière pour l'analyse des liens entre le phénomène *El Niño/Southern Oscillation* (Enso) et la variabilité de la TSO Atlantique. (© Académie des sciences / Elsevier, Paris.)

analyse en composantes indépendantes (ACI) / réseaux de neurones / analyse de séries temporelles / température de l'océan / El Niño

Note présentée par Robert Dautray.

* Correspondance et tirés à part.
aires@ara01.polytechnique.fr

** Unité mixte CNRS 8589, associée à l'ENS, à l'École polytechnique et à l'université Paris-6.

*** Laboratoire associé au CNRS, à l'ENS et aux universités Paris-6 et Paris-7.

Abridged version

This work concerns the identification of the physical causes of variability of a given dynamic system. In many cases, observed time series are well represented as a mixture, linear or nonlinear, of different statistically independent components. This physical model is called the mixture model. An important goal of statistics is then to retrieve these components from the observed data. Such statistical methods are called component extraction techniques.

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 of Neural Coding modelling. We argue that the ICA approach is a particularly promising technique which may overcome the main pitfalls of the standard techniques (such as the Principal Component Analysis (PCA), the Singular Value Decomposition (SVD) and the Factor Analysis (FA), etc.) for time series analysis. The two major advances of the ICA approach over the classical techniques are: first, the hypothesis of linearity of the mixture model is not required, and second, the fact that it extracts statistically independent components, even if these components have non-Gaussian probability distribution function, making use of higher order statistics – whereas the PCA, SVD or FA approaches make use of second order statistics only.

The ICA technique aims at extracting statistically independent components, a constraint stronger than the decorrelation at the basis of classical techniques and can also be used with a non-linear model. This model (linear or non-linear) is represented by an artificial neural network, like the perceptron (*figure 1*). The statistical independence can be expressed as the minimization of an objective cost function. Different equivalent cost functions (Comon, 1994) can be chosen, all based on the physical mixture model of the observations and on higher order statistics (Comon, 1994; Nadal and Parga, 1997). We focus here on the *infomax* approach to ICA (Nadal and Parga, 1994), from which simple algorithms have been derived (Bell and Sejnowski, 1995): the *infomax* approach reduces, in the case of the perceptron model, to the minimization of the output entropy of the neural network. Furthermore, this approach can be easily adapted to nonlinear data mixture structure (Nadal and Parga, 1997). However, the numerical

application presented here is restricted to the linear case – that is, assuming that the data mixture had a linear structure.

To our knowledge, ICA has not yet been used for component extraction in time series analysis. The numerical application presented in this work concerns the case of geophysical time series. This kind of application is characterized by the fact that data have temporal and spatial dimensions, which allows various kinds of processing (period analysis in time, in space, in frequency, etc.). The ICA technique can be integrated in all these classical processings for geophysical time series analysis. The ICA technique allows one to carry out the geographical and temporal analysis of the physical components that constitute the geophysical time series. Here, the ICA technique is applied to the analysis of the tropical Sea Surface Temperature (SST).

Comparing the SVD and the ICA techniques has led us to the following conclusions. The SVD components that can be physically interpreted are included in the results of the ICA. Furthermore, the first SVD components (essentially the ENSO component) are decomposed into more ICA components that are geographically and temporally better located and then more easily interpretable. Thus, the information given by the ICA appears to be more complete.

To analyse the ICA solution, the score maps (i.e. the maps giving the amplitude of extracted components in each pixel under study) and the temporal base functions have to be interpreted. We use spectral analysis to extract the frequencies in the base functions and we study their correlations with indexes identifying climatological phenomena, as the El-Niño/Southern Oscillation (ENSO), the North Atlantic Oscillation (NAO), South Atlantic Oscillation (SAO), etc.

The ICA technique clearly shows (*figure 2*) that important relationships exist between the northern and the southern equatorial Atlantic SSTs via a “dipole” pattern, between the Pacific and the Atlantic SSTs via the ENSO phenomenon, and between tropical and extra-tropical Atlantic SSTs via correlations with the NAO or the SAO (*figure 2d*). More work remains to be done to analyze in details the results of the ICA. But our study has already shown that the potential of the ICA technique is very high compared to that of the classical techniques (like PCA, SVD, FA, etc.).

1. Introduction

Dans de nombreux cas, une série temporelle peut être représentée par un mélange, linéaire ou non, de composantes statistiquement indépendantes. Ce modèle physique est appelé modèle de mélange. Un des buts majeurs des statistiques est de retrouver ces composantes à partir de la donnée observée. Ce type de technique est appelé *extraction de composantes*. L'analyse en composantes indépendantes (ACI), présentée dans cet article, est basée sur la théorie de l'information. Elle a été récemment

développée dans les contextes du traitement du signal et de la modélisation des réseaux de neurones. Cette technique évite les principaux écueils des techniques classiques que sont l'analyse en composantes principales (ACP), la décomposition en valeurs singulières (DVS), l'analyse factorielle (AF), etc. Les deux avancées de l'approche ACI sur les techniques classiques sont, premièrement, que l'hypothèse de linéarité du mélange n'est pas requise et, deuxièmement, que l'ACI extrait des composantes statistiquement indépendantes (i.e. une contrainte plus forte que la décorrélation).

La recherche de composantes statistiquement indépendantes dans l'ACI est exprimée par la minimisation d'une fonction de coût objective. Plusieurs fonctions de coût (Comon, 1994) peuvent être utilisées, toutes étant basées sur le modèle de mélange des composantes et sur les statistiques d'ordres supérieurs (Comon, 1994 ; Nadal et Parga, 1997). Nous nous intéresserons dans ce travail à l'approche *infomax* (Nadal et Parga, 1994), à partir de laquelle des algorithmes simples ont été établis (Bell et Sejnowski, 1995). Par ailleurs, cette approche *infomax* est également pertinente pour le cas d'un mélange non linéaire (Nadal et Parga, 1994, 1997). Toutefois, nous nous limiterons, dans notre application numérique, au cas linéaire.

À notre connaissance, l'ACI n'a jamais été utilisée pour l'analyse de séries temporelles. L'application numérique présentée dans ce travail concerne le cas particulier de l'analyse de séries temporelles géophysiques. Ici, la technique de l'ACI est appliquée à l'analyse de la température de surface de l'océan tropical.

2. L'analyse en composantes indépendantes

2.1. L'analyse de série temporelle géophysique

On dispose d'une base de données \mathcal{D} possédant M observations $\mathbf{x}(j) = (x_j^t; t = 1, \dots, N)$, avec $j \in \{1, \dots, M\}$ (dans la suite du texte, les vecteurs et matrices seront notés en gras). M est la dimension spatiale de la série et N sa dimension temporelle. On suppose que les séries temporelles $\mathbf{x}(j)$ sont un mélange, linéaire ou non, de composantes $\sigma = \{\sigma_i; i = 1, \dots, Q\}$ statistiquement indépendantes :

$$\mathbf{x}(j) = \mathcal{A}(\sigma(j)) \quad (1)$$

où \mathcal{A} est une fonction de mélange inconnue, supposée inversible.

Le but de l'ACI est de trouver une application $\Phi : \mathbf{x} \rightarrow \mathbf{h}$, où \mathbf{h} est un estimateur de σ , tel que les $\{h_i; i = 1, \dots, Q\}$ soient statistiquement indépendants. L'estimateur \mathbf{h} est défini comme une fonction déterministe (linéaire ou non) de l'observation :

$$h_i = \Phi_i(\mathbf{W}_i, \mathbf{x}); \quad i = 1, \dots, Q \quad (2)$$

avec $\{\mathbf{W}_i; i = 1, \dots, Q\}$ l'ensemble des paramètres de Φ . Le nombre Q de composantes doit être spécifié. L'application Φ est recherchée dans une famille donnée d'applications — l'exemple le plus simple étant celui des applications linéaires. En fait, chaque famille d'applications représente une information a priori sur la structure des données observées. L'application Φ est paramétrée par $\{\mathbf{W}_i; i = 1, \dots, Q\}$; les \mathbf{W}_i seront déterminés par un algorithme de descente de gradient sur une fonction de coût donnée.

2.2. Une fonction de coût objective pour l'indépendance statistique

L'indépendance statistique de Q variables $\{h_i; i = 1, \dots, Q\}$ est réalisée lorsque la distribution de probabilité (d.p.) $P_{\mathbf{h}}$ de \mathbf{h} est factorisée : $P_{\mathbf{h}}(\mathbf{h}) = P_1(h_1) P_2(h_2) \dots P_Q(h_Q)$, où P_i est la d.p. de h_i . L'indépendance statistique est une contrainte plus forte que la décorrélation : la première est basée sur les statistiques d'ordres supérieurs, la deuxième sur les statistiques d'ordre 2 uniquement. Ce n'est que pour des distributions gaussiennes qu'on a l'équivalence, mais dans beaucoup d'applications, le cas Gaussien est exclu. Le but de l'ACI est l'estimation de \mathbf{h} telle que la d.p. de \mathbf{h} satisfasse au mieux la factorisation précédente. Une fonction de coût objective doit alors être définie pour quantifier la qualité de la solution \mathbf{h} . Plusieurs choix sont possibles ; nous utilisons dans ce travail la théorie de l'information (Comon, 1994 ; Nadal et Parga, 1994).

La théorie de l'information apporte des outils pour quantifier la dépendance statistique de variables stochastiques. La quantité fondamentale est l'*information mutuelle*. Étant donné deux variables stochastiques h_1 et h_2 , l'information mutuelle $I(h_1, h_2)$ est définie par la divergence de Kullback entre la distribution jointe $P_{12}(h_1, h_2)$ et la distribution factorisée $P_1(h_1) P_2(h_2)$:

$$I(h_1, h_2) = H(h_1) - H(h_1/h_2) \\ = \int_{-\infty}^{+\infty} dh_1 dh_2 P_{12}(h_1, h_2) \log \left(\frac{P_{12}(h_1, h_2)}{P_1(h_1)P_2(h_2)} \right) \quad (3)$$

où $H(\cdot)$ est l'entropie d'une variable et $H(h_1/h_2)$ est l'entropie conditionnelle de h_1 sachant h_2 . L'information mutuelle $I(h_1, h_2)$ est une quantité positive, qui n'est nulle que si et seulement si h_1 et h_2 sont statistiquement indépendants. Elle dépend des statistiques de tous ordres. Cette définition s'étend naturellement à un nombre Q de variables : la *redondance* $\mathcal{R}(\mathbf{h})$ entre les Q composantes de \mathbf{h} est définie comme l'information mutuelle entre ces Q variables :

$$\mathcal{R}(\mathbf{h}) = \int_{-\infty}^{+\infty} \prod_{i=1}^Q dh_i P_{\mathbf{h}}(\mathbf{h}) \log \frac{P_{\mathbf{h}}(\mathbf{h})}{\prod_{i=1}^Q P_i(h_i)} \quad (4)$$

Le but de l'ACI est donc de trouver une application $\Phi : \mathbf{x} \rightarrow \mathbf{h}$ qui minimise la redondance (4). Toutefois, le calcul direct de la redondance est généralement impossible en pratique, particulièrement pour des \mathbf{h} sur des espaces de grandes dimensions. Des fonctions de coût exprimant la redondance et utilisables algorithmiquement doivent donc être définies.

Atteindre l'indépendance statistique des h_i suppose de manipuler les moments d'ordres supérieurs, tels que $\langle h_i^3 \rangle$, $\langle h_i^4 \rangle$, $\langle h_i^5 \rangle$, etc., où $\langle \rangle$ est l'espérance mathématique. Appliquer une non-linéarité f_i sur les composantes recherchées h_i permet de s'affranchir du calcul de ces

moments. En effet, le développement de Taylor des f_i fait intervenir des puissances de h_i . On considère donc une transformation de l'estimateur $\mathbf{h} = \Phi(\mathbf{W}, \mathbf{x}) : \{y_i = f_i(h_i) ; i = 1, \dots, Q\}$, où les $f_i(\cdot)$ sont des fonctions d'une variable, non linéaires, bornées et inversibles. La redondance de l'équation (4) est nulle, si et seulement si la redondance des variables $\mathbf{y} = (y_i ; i = 1, \dots, Q)$ est nulle. Ainsi, la minimisation de la redondance $\mathcal{R}(\mathbf{h})$ est équivalente à la minimisation de $\mathcal{R}(\mathbf{y})$. La redondance $\mathcal{R}(\mathbf{h})$ caractérise la dépendance des h_i ou des y_i mais elle n'est pas reliée a priori à la dépendance entre ces variables et les observations \mathbf{x} . En effet, l'estimateur \mathbf{h} est une fonction déterministe (linéaire ou non linéaire) de l'observation. Ainsi, l'information mutuelle $I(\mathbf{x}, \mathbf{y})$ est égale, à une constante près, à l'entropie $H(\mathbf{y})$:

$$H(\mathbf{y}) = - \langle \ln [P_{\mathbf{y}}(\mathbf{y})] \rangle = - \int_{-\infty}^{+\infty} d\mathbf{y} P_{\mathbf{y}}(\mathbf{y}) \ln [P_{\mathbf{y}}(\mathbf{y})] \quad (5)$$

Nadal et Parga (1994) ont montré que la minimisation de l'information mutuelle $I(\mathbf{x}, \mathbf{y}) = H(\mathbf{y}) + \text{cste}$ par rapport à l'ajustement des paramètres $\{\mathbf{W}_i ; i = 1, \dots, Q\}$ de l'estimateur et par rapport aux fonctions $\{f_i(\cdot) ; i = 1, \dots, Q\}$ implique la minimisation des redondances $\mathcal{R}(\mathbf{y})$ et $\mathcal{R}(\mathbf{h})$. Ainsi, l'entropie $H(\mathbf{y}) = H(f_i(\Phi_i(\mathbf{W}_i, \mathbf{x})) ; i = 1, \dots, Q)$ de l'équation (5) peut être minimisée par une descente de gradient par rapport aux \mathbf{W}_i et aux fonctions f_i .

2.3. Le cas linéaire

On fait l'hypothèse que l'observation \mathbf{x} est une combinaison linéaire de composantes indépendantes $\{\sigma_i ; i = 1, \dots, Q\}$ avec des d.p. inconnues $\{\rho_i(\sigma_i) ; i = 1, \dots, Q\}$. L'expression (1) se simplifie en : $\mathbf{x} = \mathbf{A} \cdot \boldsymbol{\sigma}$, où \mathbf{A} est la matrice de mélange (inconnue), supposée inversible. On peut reprendre cette équation :

$$\mathbf{x}(j) = \sigma_1(j) \cdot \mathbf{g}_1 + \sigma_2(j) \cdot \mathbf{g}_2 + \dots + \sigma_Q(j) \cdot \mathbf{g}_Q \quad (6)$$

où les Q fonctions de bases \mathbf{g}_i (i.e. i^{e} colonne de la matrice \mathbf{A}) sont des séries temporelles décrivant un comportement dynamique donné. Ce comportement dynamique peut être la trace d'un phénomène physique sur une certaine localisation géographique.

Le but de l'ACI est alors d'estimer l'inverse de \mathbf{A} par une matrice \mathbf{J} . Ainsi, $\mathbf{h} = \mathbf{J} \cdot \mathbf{x}$ est une estimation de $\boldsymbol{\sigma}$ (avec une indétermination sur l'amplitude et l'ordre des coordonnées de $\boldsymbol{\sigma}$). Les paramètres \mathbf{W}_i de l'estimateur de l'équation (2) sont les lignes de la matrice \mathbf{J} ,

$$h_i = \Phi_i(\mathbf{W}_i, \mathbf{x}) = \mathbf{J}_i \cdot \mathbf{x} = \sum_{t=1}^N J_{it} x_t \quad (7)$$

où $\mathbf{W}_i = \mathbf{J}_i$ est le filtre dont les coordonnées sont données par la i^{e} ligne de la matrice \mathbf{J} . Les fonctions de base $\{\mathbf{g}_i(t) ; i = 1, \dots, Q\}$ sont les colonnes de la matrice inverse \mathbf{J}^{-1} .

On considère maintenant l'application d'une fonction de transfert sur chaque composante h_i :

$$F(\mathbf{W}, \cdot) : \mathbf{x} \rightarrow \mathbf{y} = f(\mathbf{h}) = f(\mathbf{J} \cdot \mathbf{x}) \quad (8)$$

On reconnaît en ce modèle un réseau de neurones : les paramètres \mathbf{J} sont les « poids synaptiques » et les f_i sont les « fonctions de transfert » (voir figure 1). Ce modèle (8) est appelé *perceptron*. Les entrées de ce réseau de neurones sont les observations \mathbf{x} , les h_i sont les composantes extraites et les sorties y_i sont les transformations non linéaires des composantes extraites. Le résultat principal, donné dans Nadal et Parga (1994) est que, dans le cas particulier du perceptron, la minimisation de l'information mutuelle $I(\mathbf{x}, \mathbf{y})$, par rapport aux paramètres \mathbf{J} et aux fonctions de transfert f_i , entraîne la réduction de la redondance : à l'optimum, $\mathbf{J} = \mathbf{A}^{-1}$ et $df_i(h_i) / dh_i = \rho_i(h_i) ; i = 1, \dots, Q$. Cela signifie que les fonctions de transfert sont directement liées aux d.p. des composantes. À l'optimum, les unités de sorties y_i auront une d.p. uniforme, ce qui correspond bien à la maximisation de l'entropie de sortie.

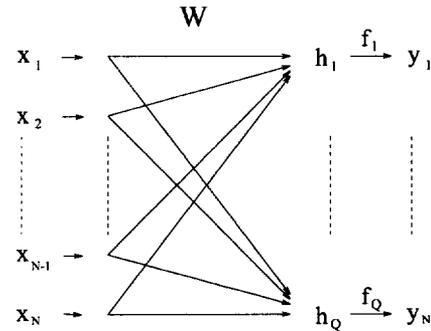


Figure 1. Architecture neuronale du perceptron.

Neural architecture of the perceptron.

On réécrit maintenant l'entropie de l'équation (5) sous une forme convenant à une implémentation algorithmique (Bell et Sejnowski, 1995). Pour cela, dans l'intégration de l'équation (5), on fait le changement de variable $\mathbf{h} \rightarrow \mathbf{y}$, qui est bien défini, puisque les fonctions de transfert et la matrice \mathbf{J} sont inversibles. On a : $P_{\mathbf{y}}(\mathbf{y}) = P_{\mathbf{h}}(\mathbf{h}) / |\mathbf{G}|$, où $|\mathbf{G}|$ est le déterminant de la matrice jacobienne $Q \times Q$ égale à : $\mathbf{G}_{it} = \partial y_i / \partial x_t = (df_i / dh_i) J_{it}$. L'entropie devient alors : $H(\mathbf{y}) = - \langle \ln (P_{\mathbf{h}}(\mathbf{h})) \rangle + \langle \ln (|\mathbf{G}|) \rangle$, c'est-à-dire :

$$H(\mathbf{y}) = - \langle \ln (P_{\mathbf{h}}(\mathbf{h})) \rangle + \langle \ln \left(\left| \frac{df_i}{dh_i} \right| \right) \rangle + \ln |\det \mathbf{J}| \quad (9)$$

Il est maintenant aisé de faire une descente de gradient sur $-H(\mathbf{y})$. Le premier terme du membre de droite de l'équation (9) est l'entropie des entrées (une constante). Les autres termes dépendent de \mathbf{J} et des fonctions de transfert. Pour minimiser $-H(\mathbf{y})$, on suppose dans notre approche que l'on peut réduire la redondance sans optimisation complète des fonctions de transfert. La fonction

de transfert reste nécessaire pour pouvoir atteindre des statistiques d'ordres supérieurs, mais choisir une unique fonction de transfert adéquate reste suffisant pour la réduction de la redondance. La fonction logistique $f_i(h) = (1 + e^{-h})^{-1}$ est souvent choisie. Des résultats encourageants ont d'ores-et-déjà été obtenus (Bell et Sejnowski, 1995) avec cette hypothèse.

2.4. L'implémentation algorithmique de l'ACI dans le cas linéaire

Soient \mathbf{x} et \mathbf{y} , respectivement, l'entrée et la sortie du perceptron. Pour les fonctions de transfert, on prend la fonction logistique et la fonction de coût pour l'estimation des paramètres \mathbf{J} est l'entropie de sortie $-H(\mathbf{y})$. Nous avons vu (équation (9)) que la minimisation de $-H(\mathbf{y})$ est équivalente à la minimisation de $-\langle \ln(|df_i/dh_i|) \rangle - \ln|\det \mathbf{J}|$. La minimisation de cette fonction de coût devient, avec des fonctions logistiques : $\Delta \mathbf{J} \propto [\mathbf{J}^t]^{-1} + \bar{\mathbf{y}} \cdot \mathbf{x}^t$, où :

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

On utilise le gradient naturel (Amari, 1998), qui est plus stable numériquement qu'une descente de gradient usuelle ; la phase d'apprentissage devient alors :

$$\Delta \mathbf{J} \propto \mathbf{J} + \bar{\mathbf{y}} \cdot \mathbf{h}^t \cdot \mathbf{J}$$

Une étape de pré-traitement des données, qui vise à extraire du problème les statistiques d'ordre 2 et à réduire la dimension des données traitées, est utilisée : $\tilde{\mathbf{x}} \leftarrow \mathbf{J}_0 \cdot (\mathbf{x} - \langle \mathbf{x} \rangle)$, avec \mathbf{J}_0 la matrice $Q \times N$ de la solution DVS. L'algorithme ACI estime par la suite une matrice \mathbf{J}_{ica} de dimension $Q \times Q$. La solution globale $Q \times N$ tient compte du pré-traitement : $\mathbf{J} = \mathbf{J}_{ica} \cdot \mathbf{J}_0$. Les Q fonctions de base sont les colonnes de la matrice inverse \mathbf{J}^{-1} .

3. Analyse de la température de surface de l'océan tropical

3.1. Données utilisées

L'étude de la variabilité de la température de surface de l'océan (TSO) est particulièrement intéressante, car elle est la résultante de phénomènes importants, complexes et souvent non périodiques. La base de données GISST2.2 (*Global Sea-Ice and Sea Surface Temperature*) de l'Ukmo (*United Kingdom Meteorological Office*) contient une climatologie mensuelle globale de la TSO sur une grille régulière. Cette base de données a été créée en associant des observations satellitaires et des mesures in situ. Dans cette base de données, nous avons sélectionné la période la plus cohérente : de 1961 à 1994, la dimension temporelle des données est donc de $N = 408$ (34 ans \times 12 mois). La résolution spatiale est $1^\circ \times 1^\circ$ et l'on s'intéresse à la région tropicale (20°S – 20°N en latitude ; 0° – 360° en longitude) sur mer, ce qui correspond à

$M = 11\,275$ pixels. De plus, on travaille sur les anomalies (effets saisonniers supprimés), plutôt que sur les moyennes mensuelles de la température elle-même. Ainsi, la base de données \mathcal{D} s'écrit : $\mathcal{D} = \{x(j) \in \mathbb{R}^{408} ; j = 1, \dots, 11\,275\}$. La phase de pré-traitement est effectuée, comme indiqué précédemment, en projetant les données observées sur les $Q = 10$ premières composantes DVS, qui totalisent 55,9 % de la variance totale.

La projection de chaque série temporelle d'un pixel sur la i^{e} fonction de base donne un « champ de scores » $\{h_i(j) ; j = 1, \dots, M\}$, indiquant la localisation géographique de la fonction de base. Les signes et les amplitudes intrinsèques des fonctions de base (et des champs de scores) n'ont pas de signification physique : l'information pertinente est donnée par les amplitudes relatives.

3.2. Résultats de l'ACI

3.2.1. Composante ACI 1

Les périodicités temporelles de cette composante sont de 5,3, 2,3 et 1 an. Le lien avec la variabilité du nord de l'Atlantique tropical est bien illustré, dans le champ de scores (*figure 2a*), par un pôle important sur cette région. De plus, nous avons calculé une corrélation de $-0,59$ avec l'indice de NAO (*North Atlantic Oscillation*), dont les périodicités sont de 2 et 1 an (Rasmusson et al., 1990). Un pôle opposé existe aussi dans le Sud-Est Pacifique tropical, qui est une région privilégiée pour le début d'un phénomène Enso. Cette téléconnection est confirmée par le travail de Delecluse et al. (1994) qui suggère que l'Enso Pacifique pourrait être une source externe dans le forçage de la variabilité interannuelle de la TSO sur l'Atlantique équatorial est.

3.2.2. Composante ACI 2

Les périodicités de cette composante sont de 4,2 et 1,8 ans. La périodicité de 4,2 ans résulte du cycle quasi-quadrannal de l'Enso. La périodicité de 1,8 ans illustre la tendance de l'Enso à un cycle biennal (Rasmusson et al., 1990). Le champ de scores (*figure 2b*) montre clairement la structure classique de l'Enso sur le Pacifique équatorial (Wallace et al., 1998). De plus, la corrélation avec un des indices climatologiques liés à l'Enso est de 0,50.

3.2.3. Composante ACI 3

Les périodicités de cette composante sont de 4, 2 et 1 an. Le champ de scores (*figure 2c*) indique clairement un lien avec l'Enso. La périodicité de 1 an correspond à la tendance de phasage de l'Enso avec le cycle annuel. Le lien avec la variabilité de l'Atlantique équatorial est certain, et la corrélation avec l'indice de NAO est de $-0,44$. Cette composante, comme la composante ACI 1, indique la corrélation entre l'Atlantique et l'Enso (Delecluse et al., 1994).

3.2.4. Composante ACI 4

Les périodicités de cette composante sont de 2,1 et 1 an. On observe aussi, sur la fonction de base, une périodicité supérieure à 10 ans (cette périodicité ne peut être obtenue

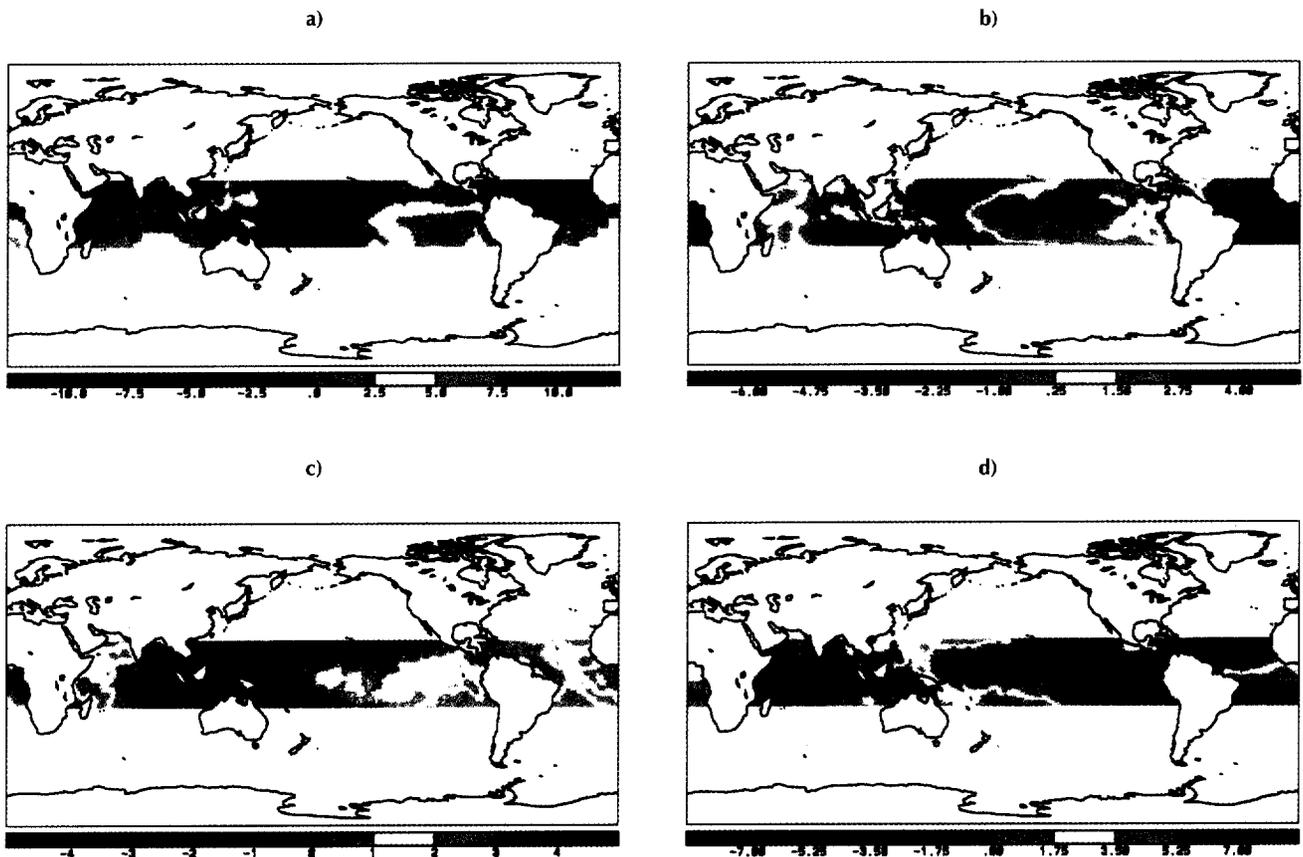


Figure 2. Champs de scores des composantes ACI : a-d.

Score maps of ACI components: a-d.

par son analyse spectrale, car les séries temporelles sont trop courtes). De plus, le champ de scores (figure 2d) indique une structure dipolaire inter-hémisphérique sur l'Atlantique tropical. Ce gradient inter-hémisphérique de TSO est bien connu comme étant la source principale de la variabilité de l'océan Atlantique. Les périodicités qui lui sont associées sont un cycle quasi-biennal et un cycle de 13 ans (Chu, 1984). Il a été démontré que ce « dipôle » possède un lien avec l'Atlantique extra-tropical et particulièrement avec la NAO (Namias, 1972). Effectivement, la corrélation avec l'indice de NAO est de $-0,60$ et la corrélation avec l'indice de SAO (*South Atlantic Oscillation*) est de $0,41$. Un autre lien a été démontré avec l'Enso (Delecluse et al., 1994 ; Enfield et Mayer, 1997). Le champ de scores montre bien une structure sur le Pacifique similaire à celle de l'Enso.

4. Conclusion

Une nouvelle technique d'extraction de composantes, basée sur la théorie de l'information, a été introduite pour l'analyse de séries temporelles géophysiques. Cette méthode a pour but d'extraire des composantes statistiquement indépendantes, une contrainte qui est plus forte que

la simple décorrélation utilisée dans les techniques classiques. De plus, il est possible de l'utiliser avec un modèle non linéaire. Cette technique permet l'analyse géographique et temporelle des variabilités d'une série temporelle géophysique.

L'utilisation de l'ACI pour l'analyse de la température de surface de l'océan (TSO), et la comparaison des résultats avec les techniques classiques (résultats de la DVS non reportés ici), permettent de conclure à la plus grande pertinence physique de la solution ACI. L'ACI montre clairement le lien entre l'Atlantique équatorial nord et sud par l'existence d'un « dipôle », entre le Pacifique et l'Atlantique via le phénomène Enso et entre l'Atlantique tropical et extra-tropical par des corrélations avec la NAO et la SAO. Le phénomène Enso a été décomposé en plusieurs composantes, pour une description géographique et temporelle plus détaillée. Notre étude montre les potentialités de l'ACI par rapport aux techniques classiques. De plus, de nombreuses perspectives sont ouvertes : le développement d'un modèle non linéaire pour l'ACI, l'étude de l'adaptation des fonctions de transfert (section 2), l'analyse simultanée de plusieurs variables physiques (en particulier l'addition de la vapeur d'eau à la TSO pour l'étude de l'Enso) et l'utilisation des résultats de l'ACI à des fins de prédiction.

Remerciements. Nous remercions N. Scott, A. Vintzileos et R. Vautard pour leurs commentaires fructueux et l'Ukmo pour la base de données GISST2.2.

5. Références

Amari S.I. 1998. Natural gradient works efficiently in learning, *Neural Computation*, 10, 251–276

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

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

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

Delecluse P., Servain J., Levy C., Arpe K. et Bengtsson L. 1994. On the connection between the 1984 Atlantic warm event and the 1982–1983 ENSO, *Tellus*, 4A, 448–464

Enfield D. et Mayer D. 1997. Tropical Atlantic Sea Surface Tem-

perature variability and its relation to El Niño–Southern Oscillation, *J. Geophys. Res.*, 102, 929–945

Nadal J.-P. et Parga N. 1994. Nonlinear neurons in the low-noise limit: a factorial code maximizes information transfer, *Network: Computation in Neural Systems*, 5, 565–581

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

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

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

Wallace J., Rasmusson E., Mitchell T., Kousky V., Sarachik E. et von Storch H. 1998. On the structure and evolution of ENSO-related climate variability in the Tropical Pacific: lessons from TOGA, *J. Geophys. Res.*, 103, C7, 14241–14259