

# Sea surface temperature patterns in the Tropical Atlantic: Principal component analysis and nonlinear principal component analysis

Christian Sadem Kenfack<sup>1,2,3,\*</sup>, François Kamba Mkankam<sup>2</sup>, Gaël Alory<sup>4</sup>, Yves du Penhoat<sup>5</sup>, Mahouton Norbert Hounkonnou<sup>1</sup>, Derbetini Appolinaire Vondou<sup>2</sup>, and Bawe Gerard Nfor Jr.<sup>3</sup>

<sup>1</sup>International Chair in Mathematical Physics and Applications, University of Abomey-Calavi, Cotonou, Benin

<sup>2</sup>Laboratory for Environmental Modelling and Atmospheric Physics, Department of Physics, University of Yaoundé I, Cameroon

<sup>3</sup>Mesoscopic and Multilayer Structures Laboratory (MMSL), Faculty of Science, Department of Physics, University of Dschang, Cameroon

<sup>4</sup>Laboratoire d'Etudes en Géophysique et Océanographie Spatiales (LEGOS), Toulouse, France

<sup>5</sup>IRD; LEGOS, Toulouse, France

## Article history:

Received 5 March 2016

Revised 18 August 2016

Accepted 29 August 2016

## Keywords:

PCA, NLPCA, SST, Tropical Ocean

## Citation:

Kenfack, C. S., F. K. Mkankam, G. Alory, Y. du Penhoat, M. N. Hounkonnou, D. A. Vondou, and B. G. Nfor Jr., 2017: Sea surface temperature patterns in the Tropical Atlantic: Principal component analysis and nonlinear principal component analysis. *Terr. Atmos. Ocean. Sci.*, 28, 395-410, doi: 10.3319/TAO.2016.08.29.01

## ABSTRACT

The tropical Atlantic Ocean exhibits several modes of interannual variability such as the equatorial (or Atlantic Niño) mode, and meridional (or Atlantic dipole) mode. Nonlinear principal component analysis (NLPCA) is applied on detrended monthly Sea Surface Temperature Anomaly (SSTA) data from the tropical Atlantic Ocean (30°W - 20°E, 26°S - 22°N) for the period 1950 to 2005. The objective is to compare the modes extracted through this statistical analysis to those previously extracted through simpler principal component analysis (PCA). It is shown that the first NLPCA mode explains 38% of the total SST variance compared to 36% by the first PCA while the second NLPCA mode explains 22% of the total SST variance compared to 16% by the second PCA. The first two NLPCA modes marginally explain more of the total data variance than the first two PCA modes. Our analysis confirms results from previous studies and, in addition, shows that the Atlantic El Niño structure is spatially more stable than the Atlantic dipole structure.

## 1. INTRODUCTION

Thorough knowledge of our planet and its climatic variations has increasingly become a major concern for humanity. Data collection techniques have witnessed significant progress over the past decades, especially with the use of satellites for global data collection. Given that the data density increases over time, researchers must work with more voluminous data whose analysis requires dedicated tools.

Sophisticated techniques such as principal component analysis (PCA) have become indispensable in extracting essential information from voluminous data sets (Von Storch and Zwiers 1999). The weakness of this conventional method is that only linear structures can be extracted from the

data. This limitation means that nonlinear structures are either missed or misinterpreted by these methods. Since the 1980s, Neural Network (NN) models have been developed and Kramer (1991) used them to extend PCA abilities to nonlinear relationship extraction from a dataset, which defines the nonlinear principal component analysis (NLPCA). The introduction of these two techniques has advanced environmental science.

NLPCA can be performed by a variety of methods, e.g., the NN model using multi-layer perceptrons (MLP) (Kramer 1991; Hsieh 2004, 2007) or the kernel PCA model (Schölkopf et al. 1998). NLPCA belongs to the class of nonlinear dimensionality reduction techniques, which includes principal curves (Hastie and Stuetzle 1989), Locally Linear Embedding (LLE) (Roweis and Saul 2000) and isomap (Tenenbaum et al. 2000). The latter is the more complex

\* Corresponding author  
E-mail: kevinnsadem@yahoo.fr

technique; it finds a nonlinear transformation that preserves geodesic distances between data points (Aho et al. 1983). Several extensions have since been developed to help with treating the largest datasets (De Silva and Tenenbaum 2004; Bachmann et al. 2006). Here, we focus on NLPCA based on the NN (Hecht-Nielsen 1995; Malthouse 1998). It has been successfully applied in the fields of atmospheric and oceanographic sciences (Monahan et al. 2003; Hsieh 2004).

The tropical Atlantic Ocean is characterized by a strong seasonal cycle (Li and Philander 1997; Xie and Carton 2004), particularly in the Gulf of Guinea (GoG). Here, a zone of relatively cold sea surface temperature (SST) appears from June to September along and slightly south-east of the equator: called the Atlantic Cold Tongue (ACT). This is the primary seasonal SST signal in the equatorial Atlantic Ocean (Merle et al. 1980; Bruno 1983; Xie and Carton 2004). This Atlantic signal appears almost all years in the East Equatorial Atlantic (EEA) and positions itself south of the Equator with a longitudinal extent to almost 20°W, centered and located a few degrees south of the Equator in the eastern part of the basin and slightly shifted equatorward (Caniaux et al. 2011). However, the detailed processes responsible for its variability are still the subject of considerable debate (e.g., Jouanno et al. 2011a, b).

The Atlantic Ocean has a strong impact on the West African climate, especially on precipitation. In particular, the ACT strongly affects the West African Monsoon (WAM) (Okumura and Xie 2004; Thorncroft et al. 2011). Its seasonal appearance intensifies the southerly winds in the GoG, which pushes a zonal rain band from the northern coast of the GoG to the sahelian region farther inland. It strengthens the north-south land-sea temperature contrast, enhancing the monsoonal flow that leads to a further decrease in SST. Due to the ACT effect on the WAM, suggesting potential rainfall predictability, the problem of understanding the inter-relationship between the ACT and the WAM remains at the center of many research endeavors (Okumura and Xie 2004; Hagos and Cook 2009; Druyan and Fulakeza 2015). Caniaux et al. (2011) shows that there is interannual variability in the ACT spatial pattern in GoG, which is not fully understood. Picaut et al. (1984) compared the seasonal and interannual SST variability in the tropical Atlantic. The result is that interannual changes in SST are largest in regions where the seasonal SST signal is large, including the ACT region, which suggests that the tropical Atlantic interannual variability is strongly linked to the seasonal cycle. However, while seasonal SST fluctuations are large, interannual variations in the Atlantic are modest in amplitude (Li and Philander 1997). Within the last decades, two primary modes of interannual variability have been identified in the tropical Atlantic (e.g., Servain and Merle 1993; Xie et al. 1999): the equatorial and meridional modes.

The equatorial mode is responsible mainly for SST variability in the GoG and is identified by abnormal chang-

es in the equatorial thermocline slope resulting from zonal-wind anomalies in the western tropical Atlantic. This mode coincides with the seasonal development of the equatorial cold tongue.

When the trade winds intensify in the western Atlantic, the equatorial thermocline slope increases and negative SST anomalies develop in the GoG, reinforcing the ACT. Conversely, when trade winds weaken, the equatorial thermocline slope decreases and positive SST anomalies develop in the GoG (Servain and Amault 1995). Indeed, a shallow thermocline in the GoG means that cool subsurface water is able to upwell to the surface and to create cold SST anomalies. This type of mechanism is found in Moore et al. (1978), Picaut (1983), and Katz (1997) and the relevant theory in Zebiak (1993) and Servain and Amault (1995).

Many studies emphasize the similarity between the El Niño/Southern Oscillation (ENSO) in the Pacific Ocean and this equatorial Atlantic mode, therefore also called Atlantic Niño (Horel et al. 1986; Delecluse et al. 1994; Xie and Carton 2004). However, the Atlantic Niño is much weaker than its Pacific counterpart (Zebiak 1993; Carton and Huang 1994; Latif and Grötzner 2000; Servain et al. 2003). A relationship between the Atlantic Niño and Pacific variability has also been suggested by some studies (Servain and Merle 1993; Carton and Huang 1994). The meridional mode is characterized by a north-south inter-hemispheric SST anomaly gradient that has no equivalent in the Pacific Ocean (Servain et al. 1999). It is most pronounced during the equatorial warm season from March to May (Deser et al. 2010). Xie and Tanimoto (1998) demonstrated how wind forcing induces this mode. Many observational (Enfield and Mayer 1997) and modeling (Huang et al. 2002) studies have shown that the ENSO influence on the tropical Atlantic is strongest in the North Tropical Atlantic, with Atlantic warming occurring 4 - 5 months after the mature phases of Pacific warm events (Xie and Carton 2004). In particular, this means that the meridional mode is significantly influenced by ENSO (Czaja et al. 2002). A link between the equatorial mode and the meridional mode has also been suggested by Servain et al. (1999). A recent study (Richter et al. 2012) distinguishes two types of Atlantic Niño referred as canonical Atlantic

Niño and non-canonical Atlantic Niño. While the canonical event follows the conventional ENSO mechanism, the non-canonical event involves southward temperature advection from the northern tropical Atlantic. This process could connect the meridional to the equatorial mode. Alternatively, the connection could involve a transmission of thermocline depth anomalies through equatorial waves (Foltz and McPhaden 2010). At longer time scales, an observational study (Tokinaga and Xie 2011) revealed robust changes in the equatorial Atlantic over the past six decades, most notably a locally enhanced SST warming in the GoG and a weakening of the ACT variability. Additional studies are necessary for understanding long term variations in the

tropical Atlantic, especially in ACT, equatorial and meridional modes.

The NLPCA technique has already proved useful in a climatic context. It has been applied on SST to study climate variations in the tropical Pacific (e.g., Hsieh 2001; Li et al. 2005). In particular, it has shown that the spatial variability associated with the ENSO phenomenon is non-linear (Hsieh 2004, 2007). The subsurface thermal structure of the Pacific Ocean was also studied using the NLPCA (Tang and Hsieh 2003). Moreover, this method has been used for some atmospheric studies of sea level pressure and air temperature over Canada (Monahan 2001). In the South China Sea, Chen et al. (2010) applied NLPCA on SST and showed that this statistical method accounts for more variance in the total variables in comparison with linear PCA and brings a finer interannual variability analysis. While PCA has been frequently used in the above studies of tropical Atlantic variability, NLPCA has never been applied in this region, to our knowledge. The internannual ACT modulation, which strongly affects the African climate, might be better understood using the NLPCA.

The aim of this work is to exhibit the main patterns in SST variability in the tropical Atlantic Ocean using NLPCA. This paper is an attempt, using an advanced PCA method, to evaluate and describe the main modes of tropical Atlantic Ocean. We apply PCA and NLPCA on data from the tropical Atlantic SST. The paper is organized as follows: section 2 is a brief description of the PCA, NLPCA methods and SST data we use. The third part describes the results of these methods on the Atlantic SST. The last part concludes.

## 2. METHOD AND DATA

Linear dimensionality reduction methods like PCA are useful tools for handling and interpreting high dimensional data. Several nonlinear dimensionality reduction methods such as kernel PCA (kPCA), Isomap, LLE, and NLPCA have also been developed. Many of these nonlinear methods, including most of the differential geometry based methods and some of the NN based methods, were originally developed in machine learning and machine vision communities for extracting low-dimensional information from data set applications; such as object identification and feature tracking. Both Isomap and NLPCA emanate from PCA. Isomap is a geometric/statistical method identified with (objective 1) cited in section 2.1 while NLPCA (discussed later) is an NN method identified with (objective 2) in the same paragraph. Isomap is the straightest way to use the geodesic distance for nonlinear projection. The Isomap (Tenenbaum et al. 2000) goal is to preserve the geodesics rather than the Euclidian distances. It has been broadly used in a large series of signal and image processing, pattern recognition and data analysis applications (Gámez et al. 2004). An initial dimensionality reduction step is required to apply NLPCA to reduce the di-

dimensionality of the inputs so that a NN of reasonable size can be used. Unlike some other nonlinear dimensionality reduction methods, NLPCA cannot be used directly with observed or modelled SST data sets. This contrasts with methods such as Isomap, where gridded data sets can be handled without preprocessing. The key factor that distinguishes Isomap from NLPCA is that Isomap uses a distance function that approximates geodesic distances in the data, while the latter employs Euclidean distances in the data space. In this work, we focus our attention on PCA and NLPCA.

### 2.1 PCA

PCA is a method that analyzes the variability in a single field (Rainfall, SST, etc.). It is commonly used for two objectives: (1) reducing the number of variables comprising a dataset while retaining the variability in the data and (2) identifying hidden patterns in the data, and classifying them with little loss of information (representing the variables using a small number of components). PCA extracts features from or recognizes patterns in the dataset. PCAs pattern recognition capability attracts our attention. Mathematically, the aim of PCA (Fukuoka 1951; Lorenz 1956) is to achieve decomposition of a continuous space-time field  $Y(i, t)$ , where  $i$  and  $t$  denote respectively spatial position and time, as

$$Y(i, t) = \sum_{k=1}^l \mathbf{a}_k(i) v_k(t) \quad (1)$$

where  $k$  is the index that characterizes the difference between modes and  $l$  is the number of modes contained in the field, using an optimal set of basic functions of space  $\mathbf{a}_k(i)$  and expansion function of time  $v_k(t)$ . It finds the spatial patterns in variability, their time variation, and gives a measure for the particular structure of each  $v_k$  pattern. To present this concept we suppose that we have a gridded data set composed of a space-time field  $Y(i, t)$  in the form  $\mathbf{x}$ , such as SST, at time  $t$  and spatial position  $i$ . The observed field is then represented in the form:

$$\mathbf{x}(t) = [x_1, x_2, \dots, x_l] \quad (2)$$

where  $x_i$  is a time series containing  $s$  observations. PCA is based on analysis of the covariance matrix e.g., (Hannachi et al. 2007), in which the variances are diagonal elements of the matrix and the covariance values are off diagonal terms. By dividing the covariance matrix by the variances the correlation matrix will be obtained which is the covariance matrix of the normalized variables. The eigenvalues and eigenvectors of the covariance matrix are then computed. Unlike the original data vectors, the eigenvectors are uncorrelated and orthogonal. Projection of the original data vectors onto the eigenvector space yields the principal components. In

general, fewer eigenvectors are required to sufficiently represent the data. In practice, the PCA technique aims at finding a new set of variables that capture most of the observed variance from the data through linear combinations of the original variables. All the components  $\mathbf{a}_k$  and  $v_k$  must be determined and one of the methods is to find those components successively from the first  $(v_1, \mathbf{a}_1)$  to the last  $(v_l, \mathbf{a}_l)$ . PCA then looks for  $v$ , which is a linear combination of the  $x_i$  and an associated vector, with

$$v = \mathbf{a} \cdot \mathbf{x}(t) \tag{3}$$

so that  $\langle \|\mathbf{x}(t) - \mathbf{a}v(t)\|^2 \rangle$  is minimized, and where  $\langle \dots \rangle$  denotes a time mean.

In other words, the goal is to find  $v_1$  and  $\mathbf{a}_1$  such that the last relation is minimized.  $v_1$  is called the first principal component (PC), a time series, while  $\mathbf{a}_1$  is the first eigenvector of the data covariance matrix, also called an empirical orthogonal function, (EOF), that describes a spatial pattern. From the residual,  $\mathbf{x} - \mathbf{a}v$ , subsequent modes can be calculated in the same fashion. These methods have been performed using NN and become NLPCA.

### 2.2 NLPCA

The fundamental difference between NLPCA and PCA is that NLPCA allows a nonlinear mapping from  $\mathbf{x}$  to  $v$  (denoted in this part by  $u$  to make difference between NLPCA and PCA terms) whereas PCA allows only a linear mapping. A large number of specialized NNs and learning algorithms have been proposed to perform PCA such as Isomap and NLPCA. There is no conclusive study that shows which approach is superior. The essential problem with nonlinear methods such as Isomap is that few theoretical results exist underpinning the numerical algorithms. The network model has greater flexibility than the hierarchical model for handling complex spatial relationships.

In Fig. 1, the input column vector  $\mathbf{x}$  of length  $l$  are mapped to the neurons in the hidden layer and the transfer function  $q_1$  maps from  $\mathbf{x}$  to the first hidden layer (the encoding layer), represented by  $h^{(x)}$ , a column vector of length  $m$ , with elements

$$h_k^{(x)} = q_1 [(\mathbf{W}^{(x)} \mathbf{x} + \mathbf{b}^{(x)})_k] \tag{4}$$

$(k = 1, \dots, m)$ , with the sigmoid function  $q_1 = \tanh$

where  $\mathbf{W}^{(x)}$  are weight matrices and  $\mathbf{b}^{(x)}$  is bias parameter vector. The dimensions of  $x$  and  $h^{(x)}$  are  $l$  and  $m$ , respectively, where  $\mathbf{x}$  is the input column vector of length  $l$ , and  $m$  is the number of hidden neurons in the encoding and decoding layers for  $u$ . The neurons  $u$  are calculated from a linear

combination of the hidden neurons  $h^{(x)}$ . A second transfer function  $q_2$  maps the encoding layer to the bottleneck layer containing a single neuron, which represents the Nonlinear Principal Component (NLPC)  $u$ , with

$$u = q_2 [\mathbf{W}^{(x)} \cdot h^{(x)} + \bar{\mathbf{b}}^{(x)}] \tag{5}$$

with identity function  $q_2$ . These mappings are standard feed forward NNs and are capable of representing any continuous function mappings from  $\mathbf{x}$  to  $u$ .

On the right side of Fig. 1, the top NN (a standard feed forward NN) maps  $u$  to  $\mathbf{x}'$  in two steps:

A transfer function  $q_3$  maps from  $u$  to the final hidden layer (the decoding layer)  $h^{(u)}$ ,

$$h^{(u)} = q_3 [(\mathbf{W}^{(u)} u + \mathbf{b}^{(u)})_k] \tag{6}$$

with  $q_3 = \tanh$  and  $(k = 1, \dots, m)$ ; followed by  $q_4$  mapping from  $h^{(u)}$  to  $\mathbf{x}'$ , the output column vector of length  $l$ , with

$$\mathbf{x}'_i = q_4 [(\mathbf{W}^{(u)} h^{(u)} + \bar{\mathbf{b}}^{(u)})_i] \tag{7}$$

with identity function  $q_4$ . To any given accuracy, provided large enough  $l$  and  $m$  are used to maximize by finding the optimal values of  $\mathbf{W}^{(x)}$ ,  $\mathbf{b}^{(x)}$ ,  $\mathbf{W}^{(u)}$ , and  $\bar{\mathbf{b}}^{(x)}$ .

The cost (Hsieh 2001) function

$$J_1 = \langle \|\mathbf{x}' - \mathbf{x}\|^2 \rangle \tag{8}$$

is minimized by finding the optimal values of  $\mathbf{W}^{(x)}$ ,  $\mathbf{b}^{(x)}$ ,  $\mathbf{W}^{(u)}$ , and  $\bar{\mathbf{b}}^{(u)}$ . The mean square error (MSE) between the output  $\mathbf{x}'$  and the original data  $\mathbf{x}$  is thus minimized. Without loss of generality, we impose the constraint  $\langle u \rangle = 0$ , hence

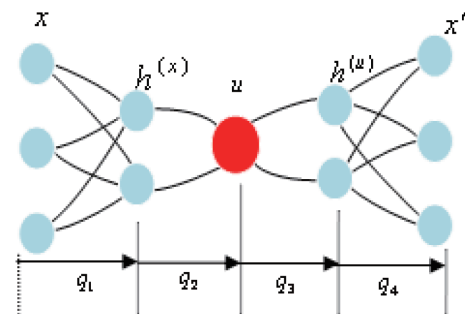


Fig. 1. The NN model for calculating nonlinear PCA. There are 3 ‘hidden’ layers of variables or ‘neurons’ (denoted by circles) sandwiched between the input layer  $\mathbf{x}$  on the left and the output layer  $\mathbf{x}'$  on the right. Next to the input layer is the encoding layer, followed by the ‘bottleneck’ layer (with a single neuron  $u$  for NLPCA), which is then followed by the decoding layer.  $q_1, q_2, q_3$ , and  $q_4$  are the transfer functions. (Color online only)



$$\bar{\mathbf{b}}^{(x)} = -\langle \mathbf{W}'^{(x)} \cdot \mathbf{h}^{(x)} \rangle \quad (9)$$

The total number of free (weight and bias) parameters used by the NLPCA is then  $2lm + 4m + l$ . Furthermore, we adopt the normalization condition that  $\langle u^2 \rangle = 1$ . This condition is approximately satisfied by modifying the cost function to

$$J = \langle \|\mathbf{x}' - \mathbf{x}\|^2 \rangle + (\langle u^2 \rangle - 1)^2 \quad (10)$$

$m$ , the number of hidden neurons in both the encoding and decoding layers, is chosen following the general parsimony principle. A larger  $m$  increases the nonlinear modeling capability of the network, but could also lead to over fitted solutions. If  $q_4$  is the identity function and  $m = 1$ , Eq. (9) then implies that all  $\mathbf{x}'$  are linearly related to a single hidden neuron, hence there can only be a linear relation between the  $\mathbf{x}'$  variables. We need to look at  $u$  for nonlinear solutions. In effect, the linear relation (3) is now generalized to  $u = f(\mathbf{x})$ , where  $f$  can be any nonlinear function represented by a feed forward NN mapping from the input layer to the bottleneck layer. Instead of  $\langle \|\mathbf{x}(t) - \mathbf{a} \cdot v(t)\|^2 \rangle$ ,  $\langle \|\mathbf{x} - g(u)\|^2 \rangle$  is minimized, where  $g$  is the general nonlinear function mapping from the bottleneck to the output layer. The residual,  $\mathbf{x} - g(u)$ , can be input into the same network to extract the second NLPCA mode, and so on for the higher modes.

That the classical PCA is indeed a linear version of this NLPCA can be readily seen by replacing all of the transfer functions with the identity function, thereby removing the NLPCA nonlinear modeling capability (Hsieh 2001). The forward map to  $u$  then involve only a linear combination of the original variables as in the PCA. A number of runs (mappings from  $u$  to  $\mathbf{x}'$ ) used to find the solution with the smallest MSE. The NLPCA here easily generalizes to more than one hidden layer mapping, as two hidden layer mappings may outperform single hidden layer mappings in modeling complicated nonlinear functions. A reconstruction using a single PCA mode is a poor approximation to the input data and the greater freedom available to the NLPCA network allows it to produce a nonlinear fit, which better represents the variability in the input data.

### 2.3 Data

We used monthly Atlantic SST data, which is the reconstruction sea surface temperature (ERSST) data set (Smith and Reynolds 2003) from the NOAA National Climatic Data Center. A climatologically annual cycle was calculated by averaging the data for each calendar month. Monthly SST anomalies (SSTAs) were defined relative to this annual cycle. This data set is a mixture of satellite and in situ data with a spatial resolution of  $2^\circ \times 2^\circ$  from 1950

- 2005. Because of the interest in El Niño, dipole and cold tongue phenomenon in the Atlantic, an empirical orthogonal analysis was made of tropical Atlantic data between  $26^\circ\text{S} - 22^\circ\text{N}$  and  $30^\circ\text{W} - 20^\circ\text{E}$ . Thus, there are  $(26 \times 25)$  650 spatial points, 215 points of missing SST data on the continent, and  $(26 \times 25 - 215)$  435 spatial and  $(56 \times 12)$  time existing points. The SSTA have long-term trends. In order to minimize the effect of these trends on the analysis, we eliminated the linear trends from datasets at each spatial location using the least squares technique. Thus, detrending the monthly anomaly, our primary dataset was formed.

### 3. TROPICAL ATLANTIC SST

Eigenvectors with the largest percentages are usually associated with physical processes. EOFs provide a series of eigenvectors, each of which contains a percentage of the temporal variability of the data. Figures 2a and b show the first two EOF modes and their corresponding principal components (PCs) (Fig. 3). The spatial patterns associated with two SST modes are shown in Fig. 2 as homogeneous correlation maps EOF1 and EOF2. The two EOF modes together account for 52% of the total monthly SST variance. Individually, they explain 36 and 16% of the total SST variance. The other components explain less than 14% of the total variance and will not be discussed further. The two modes differ in their spatial expressions. It should be noted that the EOF sign is arbitrary; however, the EOF and PC time series product recovers the correct mode polarity in any given grid box and time. For SSTA data, we choose the nonlinear PCA model  $m = 3$  and 3931 parameters which greatly exceed the number of time points. There are still far too many spatial variables (number of points in the selected space) for this dataset to be directly analyzed by the NLPCA. The SSTA data is pre-filtered by PCA to reduce the number of input variables. The NLPCA method used here is identical to that used by Hsieh (2004). Monahan (2001) used same for the study of El Niño/La Niña. The difference is that the former used 3 PCs or 3D-approximation while the latter used 10 PCs or 10D-approximation as inputs to the NLPCA. The results are similar, which means that the principal characteristics of the phenomenon (El Niño/La Niña) are contained within the first 3 PCA modes. This implies that a certain minimum number of PCs can be sufficient to capture a phenomenon. In our case, to select the number of PCs, we used the Guttman-Kaiser criterion; where only the modes with Eigen values greater than the average Eigen value were retained (Jackson 1991; Landman and Tennant 2000). This criterion gives 20 PCs whose total variance is 99%. It is true that fewer than 20 can still also give the expected results based on the rule of thumb (North et al. 1982). The first 20 PCs are then used as the input to the NLPCA network. Each input variable is normalized by subtracting its mean and then dividing by the standard deviation of the first PC (Hsieh 2001). Scatter plots

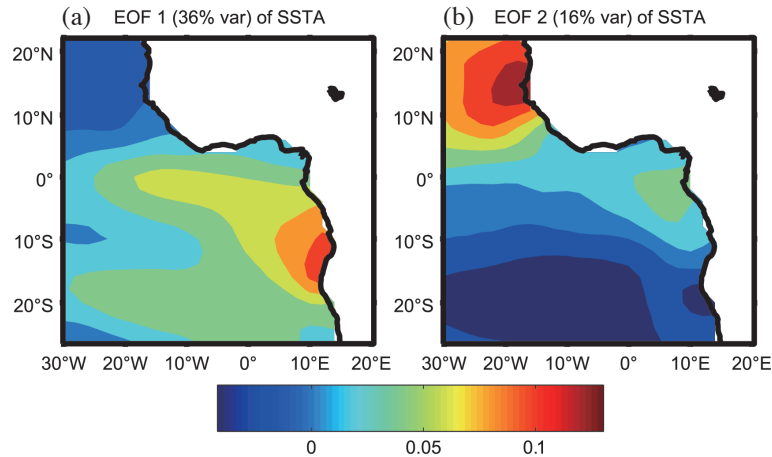


Fig. 2. Empirical orthogonal function (EOF) of detrended monthly sea surface temperature (SST) Anomalies. (a) EOF1 mode (left) and (b) EOF2 mode (right) with their explained variance in parenthesis. The contour interval is 0.01. (Color online only)

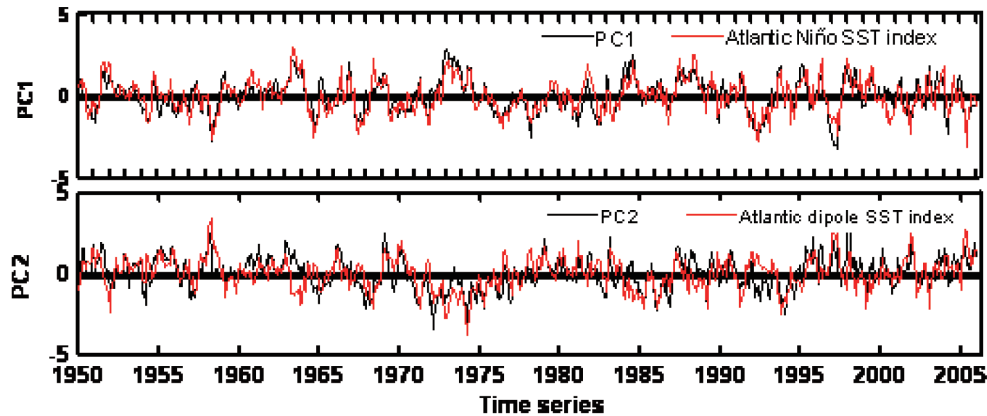


Fig. 3. The corresponding time coefficients in black, the Atlantic Niño index and Dipole index in red. Vertical lines in the time series correspond to January of the respective years, starting in January 1950 and ending in December 2005. (Color online only)

of the 3 leading principal component time series are shown by the solid blue dots in Fig. 4. All four projections are shown because it is difficult to understand the NLPCA approximation structure from a single projection. This is particularly evident in Fig. 4c, where the curve viewed edge-on, appears to be self-intersecting, whereas, in fact, the other projections demonstrate that this is not the case.

### 3.1 ACT

EOF1 exhibits (Fig. 2a) a high variance in coastal Angola, which extends from the coast to the west between 6°S and 2°N and EOF2 exhibits (Fig. 2) large variances in two places: (1) at the Equator between 5°S and 2°N and (2) along the Gabon coast. EOF1 mode (Fig. 2a) accounts for up to 36% of the total variance with the region of largest amplitude around the Angola coast. The signal decreases from East to West between 15°S and 2°N. The GoG is identified by abnormal changes in the equator and eastern part.

The SSTA signature of the first mode is mainly confined to the eastern equatorial region. This first mode presents a structure for which the pattern is more closely confined to the eastern basin and spanning a range of latitudes (Zebiak 1993) near the Angola coast. This spatial structure is characterized by a zonal pattern symmetric about the equator, with warm or cold SSTs appearing in the eastern equatorial basin to the south and north. The first mode pattern captures the zonal mode pattern in the tropical Atlantic (Zebiak 1993; Carton and Huang 1994; Huang and Shukla 1997; Handoh and Bigg 2000; Ruiz-Barradas et al. 2000). By multiplying EOF1 with its related PC, we observe some typical ACT cold years, such as 1990, 1992, 1997, 2005, and warm years 1984, 1987, and 1998 (Caniaux et al. 2011). At the end of 1964 and again in 1967, there was a general cooling over most of the Atlantic and particularly strong in the east. Likewise, in 1966 there is much less warming in these same areas. The year 1969 is revealed as being warmer, as reported in a previous study (Servain and Legler 1986). This mode

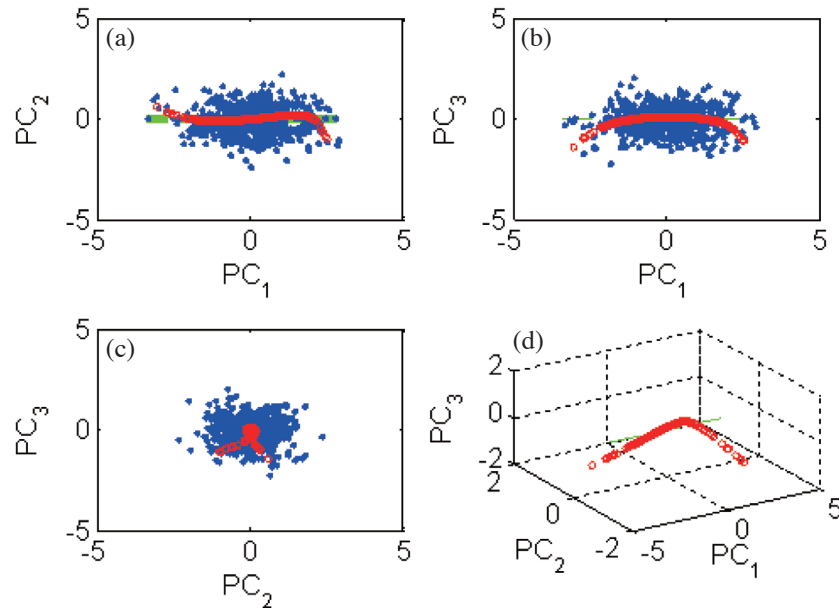


Fig. 4. Scatter plot of the SSTA data (shown as blue dots) in the principal component (PC1, PC2, and PC3) plane. The first PCA eigenvector (green color) is oriented along the horizontal line. The first mode NLPCA approximation to the data is shown by the reds circles, which traced out a Wave-shaped curve. Panel (a) displays the PC1-PC2 plane, (b) the PC1-PC3 plane, (c) the PC2-PC3 plane, and (d) gives a 3-D view. (Color online only)

mainly presents interannual variability in the tropical Atlantic. In addition, the correlation coefficient between PC1 and Atlantic Niño SST index is 0.82. The Atlantic Niño SST index is defined as the area-averaged SSTA over  $3^{\circ}\text{S} - 3^{\circ}\text{N}$ ,  $20^{\circ}\text{W} - 0^{\circ}$  (Zebiak 1993). Hence, this picture resembles the equatorial mode. The first (Fig. 4) NLPCA mode explains 38.6% of the variance of the given EOFs, i.e., 38% of the total SSTA variance compared to 36% explained by the first PCA. The NLPCA mode projection in the planes defined by pairing PC1, PC2, and PC3 are shown in red in Fig. 4. The NLPCA mode 1 is slightly coincident with the projection onto PC1 indicated by the green line in the plots. Unlike the PCA, the NLPCA shows the nonlinear form in SSTA structure. The SSTA manifests itself as a Wave-shaped curve on the non linear graph of the first nonlinear mode between the minimum and maximum values of its principal component,  $u$ . The non linearity of the sinuosity observed in Fig. 4a is modest and the maximum of PC1 is less than the absolute value of the PC1 minimum. This insinuates that we have a very cold and very warm event in the Atlantic region. It seems as if the two (ACT and Atlantic Niño) evolve together because their correlation coefficient is 0.94. Figure 5 is a plot of the normalized time series of NLPC, which bears a strong resemblance to the ACT time series. This “ACT index” is obtained by computing the inter-annual mean of SSTA in the ACT zone [in the box from  $5^{\circ}\text{S} - 2^{\circ}\text{N}$  and from the coast to  $20^{\circ}\text{W}$  during the ACT period (June - August) Caniaux et al. (2011)]. Normalization is obtained by subtracting the mean and dividing by the standard deviation. The correlation coefficient between the normalized ACT in-

dex, NLPC and linear principal component are respectively 0.85 and 0.8.

Figure 6 shows the reconstructed SST anomaly field for some values of the first nonlinear component,  $u$ . It is obtained by choosing a particular value of  $u$  associated with each of the twenty PCs and multiplied by the associated EOFs. Eight values of  $u$  are chosen equitably from minimum to maximum. To display the ACT sequences (which correspond to the value of  $u$  as Fig. 4a indicates) and only the eight equitable intervals (Fig. 6) values of  $u$  are selected: -3.14, -0.98, -0.57, -0.25, 0.06, 0.31, 0.54, and 1.08. It observed that the spatial distribution varies with each selected value of  $u$ . This variation from the minimum (strong activity of ACT) to maximum (low activity of ACT) of the eight values of  $u$  shows a contrast between the north-east and southwest of  $5^{\circ}\text{S}$ . Those patterns are repeated from (a) to (h). We note in Fig. 6 that the ACT cold year is more intense than the warm one and the ACT distribution is quasi-linear in space. This result is confirmed by the high correlation coefficient between NLPCA mode 1 and PCA1. As afore mentioned, Figs. 6a and h exhibit strong and weak ACT, respectively.

The differences in strong/weak ACT symmetry in Fig. 6 can be compared with the composite analysis results. The simplest approach is to compute the mean NLPC. Warm and cold events are defined in each time series as events whose amplitudes are greater or less than the mean NLPC. The warm and cold events have positive and negative signs, respectively. Thereafter, we compute the mean of each group; i.e., for the warm or cold years during the ACT

period (JJA). This averaged period was used for the composites because JJA displays the period in which the surface area at less than 25°C is greater than the empirical threshold surface area at  $0.40 \times 10^6 \text{ km}^2$  (Caniaux et al. 2011). The mean for the strong ACT is represented by Fig. 7a and for the weak by Fig. 7b. The largest SSTA is located in the eastern portion during average weak ACT events and centered in the same region during average strong ACT events. We may observe that the pattern of these two types of ACT is quasi symmetric.

### 3.2 Atlantic Niño

Xie and Carton (2004) showed that the Atlantic Niño is most pronounced in the boreal summer coinciding with the seasonal development of the equatorial cold tongue. Figure 6a represents the coldest year and Fig. 6h the year of the warmest ACT, which is not the warmest year but the year of weak ACT corresponding to the Atlantic Niño. The weak ACT can then be regarded as the conventional El Niño. Lübbecke and McPhaden (2013) results show that the Atlantic Niño mode is more strongly damped than the Pacific Niño. In this study, unlike in the Pacific Ocean where Hsieh (2004) showed that the Pacific El Niño is asymmetric, the spatial variability of this mode in the Atlantic Ocean is more linear than the ENSO. This is a new understanding between the Atlantic Niño and Pacific Niño.

The results presented in Figs. 6e, f, and g characterized the non canonical Atlantic Niño with negative anomalies in the northern tropical Atlantic and Fig. 6h characterized the canonical Atlantic Niño with positive anomalies in the northern tropical Atlantic. The first NLPCA mode successfully passes through the canonical El Niño states, non-canonical El Niño states and ACT states as  $u$  varies continuously from its minimum value to its maximum value. Therefore, NLPCA is capable of simultaneously capturing many more pictures in the Atlantic Ocean. In addition, it gives a more accurate description of the El Niño states (Fig. 6) than the first PCA mode (Fig. 2a), which did not fully represent the canonical and non-canonical Atlantic Niño. With a linear approach, it is generally impossible to have simultaneous results.

### 3.3 Atlantic Dipole

A north-south inter-hemispheric SST anomaly gradient is observed in the second PCA mode. The SSTAs show a remarkable meridional gradient, with significant positive anomalies spreading from the North African coast into the northwestern tropical Atlantic and significant negative anomalies in the south. There is, however, a local maximum in the tropics which has a sign, indicating that the two regions are quite strongly out of phase. PC2 and Atlantic dipole SST index are well correlated (Fig. 3) with correlation

coefficient of 0.66. EOF2 then displays an opposite phase between the North and South parts of 6°S. This picture is consistent with the Atlantic dipole (Servain et al. 1999).

We now turn to NLPCA SST mode 2 results (Figs. 8 and 9). The NLPCA network architectures used for calculating NLPCA SST mode 2 are essentially the same as those of Hsieh (2001); three hidden layer neurons with all training and fitting parameters identical to the configuration used for calculating NLPCA SST mode 1. The explained variance results show that the NLPCA mode 2 explains more variance as the second principal components. For the observational SST data, NLPCA SSTA mode 2 explains 22% of the total data variance, compared to 38% for NLPCA SST mode 1. The first two NLPCA modes between them explain marginally more of the total data variance than do the first two PCA modes.

We therefore expect that the nonlinearity will be more pronounced in Fig. 9 in the reconstruction plots for NLPCA mode 2. The reconstruction is obtained as we did in the case of NLPCA mode 1. We also display the sequences for the Atlantic dipole and only the eight equal interval values of the second NLPC mode are selected: -2.71, -1.12, -0.80, -0.38, -0.059, -0.052, 0.37, and 0.70. This variation from the minimum (strong gradient) to maximum (weak gradient) for the eight NLPC values for the second NLPCA mode shows a contrast between the north-west and south-west. There is clear nonlinearity in the distribution from NLPCA SST mode 2. The main characteristic of the Atlantic dipole is captured, with the South Pole shifted towards the coast with different signs. The main variation pattern between the large negative and positive NLPCA mode 2 values is a dipole; with each pole on either side of the equator with a different sign (Fig. 8). This meridional mode (Fig. 8) is exhibited in the second NLPCA mode. Figures 8b, c, d show the displacement of southern counterpart of the dipole, which confirms the asymmetry of the meridional mode. This shows that the variability of the second mode, which resembles the meridional mode, can change from year to year. Regardless of Fig. 8, there were interesting patterns that came out in the second NLPCA. There is some discrepancy between the Atlantic dipole structure, during the strong gradient (meridional gradient between the two lobes, Fig. 8a) events the core of the southern lobe is centered around 15°W during the strong gradient and around 20°W during some weak events (Fig. 8f). The Atlantic dipole in Fig. 8a is similar to that in Fig. 8f but the southern lobe is not located at the same place. The situation in the South lobe is unstable. We also observed that the south lobe is missing in Figs. 8b, c, d. This is not surprising because it is observed in some studies (Servain 1991; Rajagopalan et al. 1998) that the Atlantic dipole exhibits decadal time scales. This is confirmed by Xie and Carton (2004) whose analysis showed that the meridional dipole rarely occurs. Compared to Fig. 8a, the southern lobe of Fig. 8f is somewhat shifted to the east while the



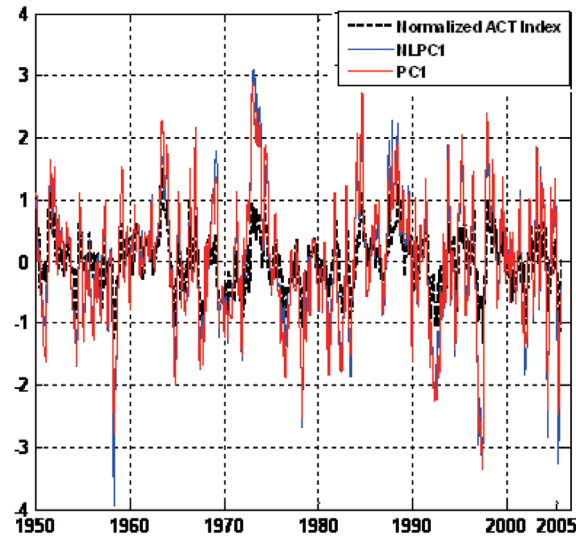


Fig. 5. Plot of NLPC1, the time series associated with SSTA NLPCA mode 1 (blue line), the normalized ACT index (black dashed line), and PC1 (red line). (Color online only)

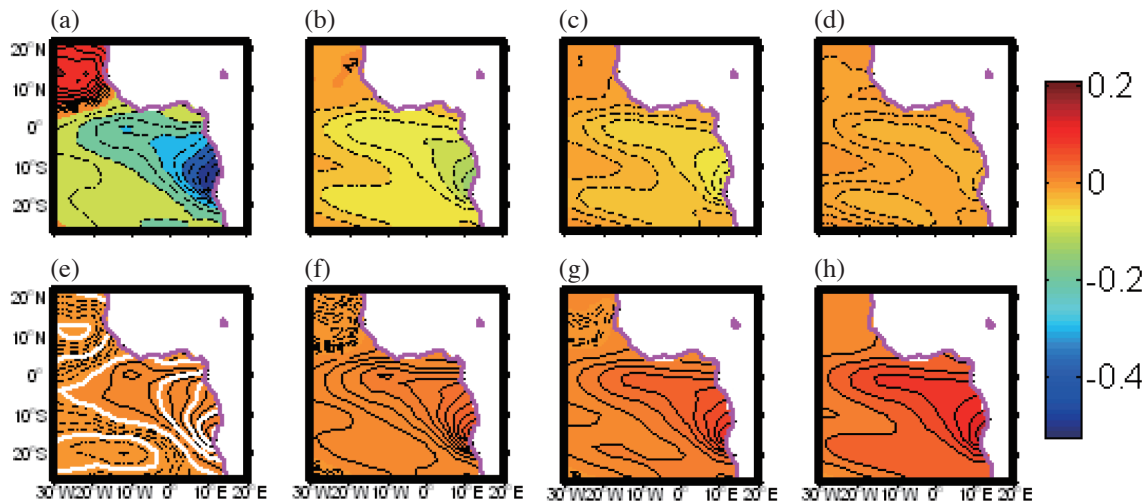


Fig. 6. The SSTA anomaly pattern (in  $^{\circ}\text{C}$ ) of the NLPCA as the minimum of NLPC u between June and August for each year of the first NLPCA mode. Considering just the eight minimum mentioned above, the anomaly pattern of the first NLPCA mode varies from (a) its minimum (strong ACT), to (b) three-quarter its minimum, to (c) half its minimum, to (d) quarter of its minimum, to (e) quarter its maximum (non-canonical Atlantic Niño), to (f) half its maximum, to (g) three-quarter its maximum, and (h) its maximum (weak ACT). Zero contours are white lines. Positive contour is bold black line and negative contours are black. The contours in pink color are the coast. (Color online only)

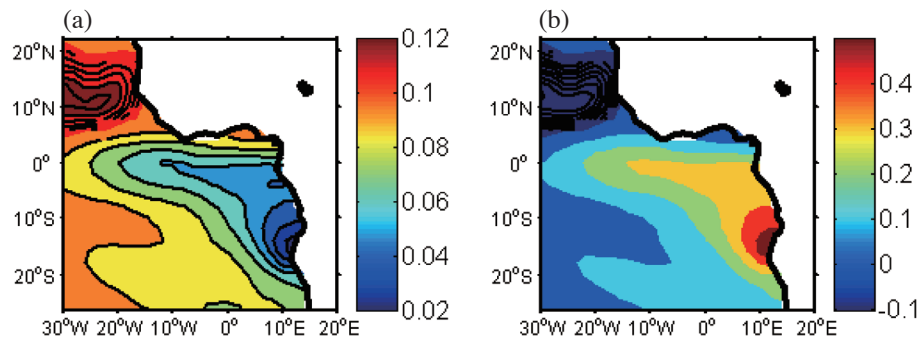


Fig. 7. Composite maps for average (a) cold ACT (left) and (b) warm ACT (right). Positive contours are black lines and negative contours are dashed black lines. (Color online only)

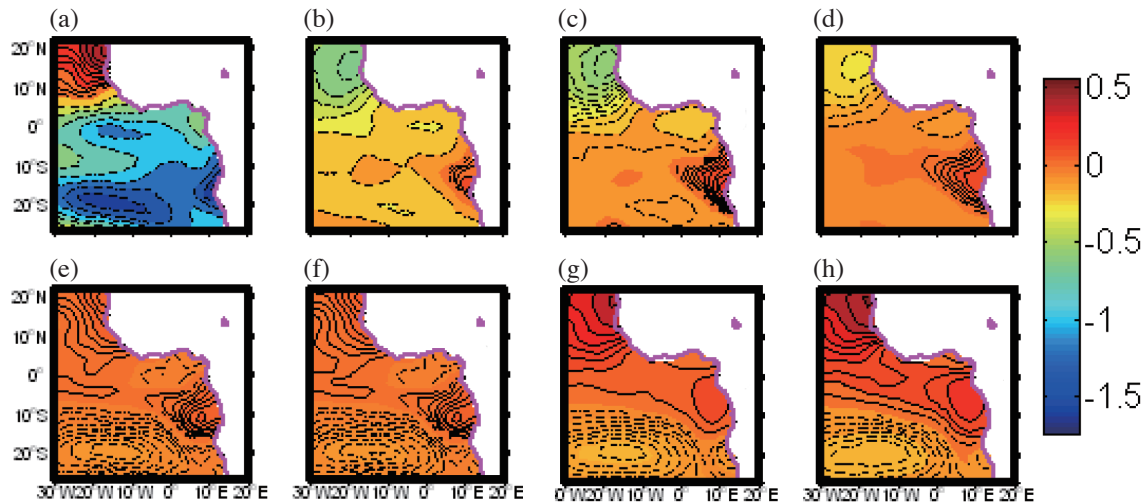


Fig. 8. The SST anomaly pattern (in  $^{\circ}\text{C}$ ) of the second NLPCA mode varies from (a) its minimum (*strong* gradient between north and South), to (b) three-quarter its minimum, to (c) half its minimum, to (d) quarter of its minimum, to (e) quarter its maximum, to (f) half its maximum, to (g) three-quarter its maximum, and (h) its maximum (*weak* gradient between north and South). Zero contours are white lines. Positives contours are black line and negative contours are dashed black line. The contours in pink color are the coast. (Color online only)

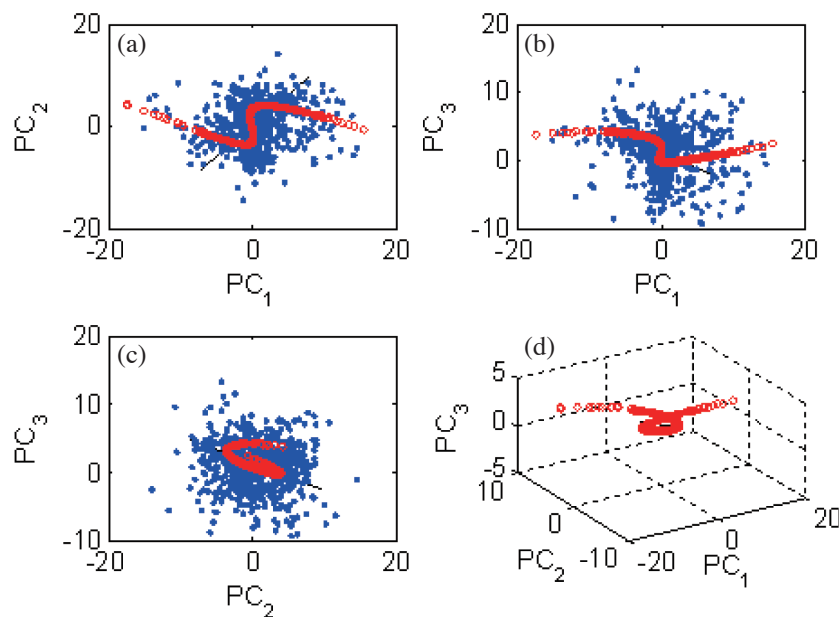


Fig. 9. Scatter plots of the SSTA data (shown as blue dots) in the principal component ( $\text{PC}_1$ ,  $\text{PC}_2$ , and  $\text{PC}_3$ ) plane. The dots show the residual data after the NLPCA mode 1 has been subtracted. This second mode NLPCA approximation to the data is shown by the reds circles, which trace out a Wave-shaped curve. (The linear solution to the dataset after NLPCA mode 1 has been removed is not the same as PCA mode 2, which is the linear solution to the dataset after PCA mode 1 has been removed.). Panel (a) displays the  $\text{PC}_1$ - $\text{PC}_2$  plane, (b) the  $\text{PC}_1$ - $\text{PC}_3$  plane, (c) the  $\text{PC}_2$ - $\text{PC}_3$  plane, and (d) gives a 3-D view. (Color online only)

northern lobes have weaker anomaly. Figure 10 shows the NLPC mode 2 and Atlantic dipole index time series with the correlation coefficient of 0.9. When observing the Atlantic dipole structure, we may realize that the sequential evolution of the phenomenon from Figs. 8a - h is interrupted, implying that the phenomena is not observed each year which is not also surprising. However, we may observe that the south counterpart of the dipole has a displacement. Thus, the spatial variation in the meridional mode changes mean-

ing that the phenomenon cannot happen in the same place every year.

#### 4. DISCUSSION

There are many possible schemes for PCA (RPCA) rotation, the varimax (Kaiser 1958) being the most popular. Both RPCA and NLPCA take the PCs from PCA as inputs. There is an important difference between PCA and rotated PCA

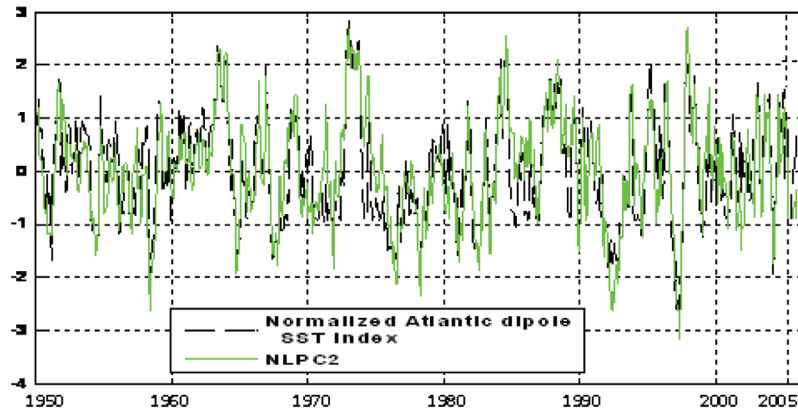


Fig. 10. Plot of NLPC2, the time series associated with SSTA NLPCA mode 2 (green line), the normalized Atlantic Dipole index (black dashed line). (Color online only)

methods; as it is generally impossible to have a simultaneous solution that explains the maximum global variance of the data and approaching pattern recognition. NLPCA can give both types of information, thus the nonlinearity in NLPCA unifies the PCA and rotated PCA approaches (Hsieh 2001). In this paper, in terms of variance, the first rotated PCA explained 31% of the total variance (not shown), versus 36% using the first PCA mode. The first PCA exhibits a more accurate description of the ACT than the first RPCA mode. The combination of two SST EOFs yields the major events in the tropical Atlantic. In this section, it is known from SSTA that PCA points out the interannual principal mode in the Atlantic Ocean. NLPCA can give more information about this variability.

The NLPCA mode 1 better represents the inter-annual variability of ACT than the linear PCA mode 1. Figure 6b corresponding to 0.75 min (u) is similar to EOF1. So, NLPCA explains variances extracted by PCA and gives more information, which is not obvious by analysis with the PCA. The linear PCA method describes the evolution in time of a standing oscillation with fixed spatial structure and time varying amplitude. NLPCA is not so constrained and therefore its power lies in characterizing more complex lower-dimensional structures (Monahan 2001). The analysis of the two curves of (Fig. 5) indicates that NLPCA can give more information about ACT. SSTA is stronger in ACT than other parts of Atlantic Ocean except in the Angola coast where it is strongest. This view appears to corroborate to the observations of Hirst and Hastenrath (1983) and of Rouault et al. (2003) that the variability in tropical southeast Atlantic SST is strongest in the Angola Benguela frontal current zone region near 15 - 20°S.

Comparison of the eight sub-figures in Fig. 6 (Figs. 6a - h, respectively), clearly show a similar pattern in the eastern part of the equatorial Atlantic. There are warm and cold SSTAs in this region, stretching further southwards along the coast of Central and South Africa. This is

consistent with ACT development. There is no big spatial difference between a fully developed strong and weak ACT. Comparing the patterns shown in Fig. 6 with that of the first EOF presented in Fig. 2a, we observe that individual PCA modes represent only a single spatial pattern of the first NLPCA mode with standing oscillations. Figure 6a is similar to Fig. 2a, hence NLPCA mode 1 includes PCA mode 1. The strong and weak ACT states (Figs. 6a and h), respectively, are confined to the eastern part of the equatorial Atlantic. One of these patterns can be captured by a conventional PCA analysis but does not capture the quasi symmetry presented by strong and weak ACTs. The spatial distribution is best described (Fig. 6) by NLPCA than PCA, and the variance may be represented by these first NLPCA modes. The asymmetry of SSTA in the NLPCA mode 1 evolution is modest. The canonical and non canonical Atlantic El Niño event is observed which cannot be pointed out by linear mode 1.

It is not necessary to study RPCA since neither the PCA nor RPCA can represent the two states of ACT simultaneously. In contrast, the first NLPCA mode successfully passes through the strong and weak states as NLPC varies continuously from its minimum to maximum value. Therefore, the rotated eigenvectors do not improve much on the nonrotated eigenvectors for the ACT and Atlantic Niño study.

The quasi-symmetry determined by a composite analysis (Fig. 7) shows just two ACT steps. These two maps correspond to the SSTA patterns of an average warm and cold ACT event, respectively. Since composite analysis is based on composite means, there will be differences in the micro-spatial patterns of strong and weak ACTs from that of the NLPCA analysis. The very weak spatial asymmetry between cold and warm ACT events observed in NLPCA mode 1 is manifest in the composite analysis. These two maps correspond to the SSTA patterns of an average weak and strong ACT, respectively. Figures 7a and b, respectively, bear a strong resemblance to Figs. 6a and h; and are strongly

similar to EOF1. This means that EOF1 exhibits strong ACT. It is more explicit in the first NLPCA mode 1, which shows both strong and weak ACTs. Note that, consistent with the maps corresponding to the 1D NLPCA approximation (Fig. 6h), the largest SSTA tend to be located on the Angola coast during the average strong and weak ACT events. The spatial correlation between the two maps is  $-0.99$ . This confirms the quasi-symmetry between the warm and cold ACT events.

Composite analysis can show only the states of the means for the weak and strong ACTs separately. But NLPCA shows different ACT states; isolating different patterns and their associated amplitudes which may be missed by the means or averaging. The NLPCA, therefore, has the nice feature of capturing a range of variability between the symmetry of different ACT states, something that is difficult to obtain using index-based SST composites. NLPCA improves on PCA by allowing low-dimensional approximations to have a structure other than that of simple standing oscillations. Both the NLPCA mode 1 and the composite analysis describe the symmetry between averages of warm and cold ACT events. However, NLPCA has the advantage of no a priori time series specification (e.g., periods) over which to composite, and provides a full 1D approximation to the data, in contrast to the absence of approximation in composite analysis. The latter can only give the mean state of ACT but PCA and NLPCA can approximate the real state. The fact that we observed that ACT is almost linear gives the impression that NLPCA is not necessary, but the implementation of only linear PCA hides this weak nonlinearity. The warm events (Atlantic Niños) are associated with reduced cold tongue development (Carton and Huang 1994).

A final comparison of the 1D NLPCA and 1D PCA approximations is given in Fig. 11, which show, respectively, maps of the pointwise correlation between the original SSTA data and the 1D PCA approximation, and of the pointwise correlation between SSTA and the 1D NLPCA approximation (Fig. 11). The two approximations are equally well correlated with the original data over the central band of the Atlantic Ocean, where the NLPCA correlations are somewhat higher than those of the PCA approximation, except

in the Northwest part of the Ocean. In the East region and in the ACT area the 1D NLPCA approximation displays a greater fidelity to the original data, as determined by the pointwise correlation, than does the 1D PCA approximation. This demonstrates the better capacity of NLPCA in representing SST data than the PCA. We observe that NLPCA represents better ACT surface on the Angolan coast than the PCA. However, the strong correlation in Fig. 11 centered in the Northwest part of the Ocean at around  $20^{\circ}\text{W}$  shows the ability of PCA to reproduce SSTA data in this region. Therefore, the NLPCA and PCA are complementary. These two statistical tools do not strongly represent globally SSTA data very close along the Guinea coast. As already mentioned above, in terms of explained variance, NLPCA mode 1 is generally greater than PCA mode 1 however the difference is not as much as that between PCA mode 2 and NLPCA mode 2, which are very large and with the NLPCA mode 2 being greater than PC mode 2.

The meridional mode (Fig. 9) is more nonlinear than that of the ACT (Fig. 6). We observe that the asymmetry of the variability of the meridional mode is more pronounced (Fig. 9a) than that of the ACT.

It was pointed out by Houghton and Tourre (1992) and confirmed by Mehta (1998) that the northern and southern hemispheres appear to act independently. Regarding the Southern lobe picture of Fig. 8 compared to other analysis we see that NLPCA gives the additional information that this lobe underwent an eastward displacement. The weak gradient between the two lobes is also observed in the sequence map. This weakening of the meridional mode is consistent with previous works (Tokinaga and Xie 2011). This means that the second NLPCA mode has fully captured the Atlantic dipole pattern, unlike PCA mode 2, which only captured part, thereby yielding a more accurate description of the Atlantic dipole (Fig. 8) than the second PCA mode (Fig. 2b), which did not fully represent the center of action displacement located in the southern part. With a linear approach, it is generally impossible to have simultaneous results. What is quite interesting is that previous statistical and diagnostic studies seem not to be interested in investigating the stability

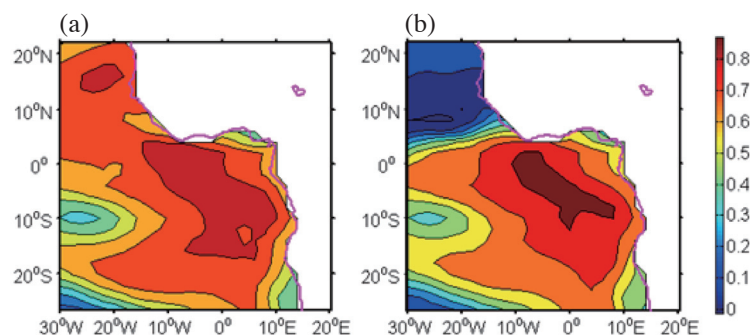


Fig. 11. Maps of point wise correlation coefficients between observed SSTA (left) and (a) 1D PCA approximation and (b) 1D NLPCA approximation (right). (Color online only)



of spatial variability in the Atlantic dipole structure.

The weakness of the gradient between the north-west and south-west means that the Atlantic dipole is insignificant during some years. This result is supported by Tokinaga and Xie (2011); they showed that CMIP3 twentieth century experiments tend to feature a weakening meridional SST gradient. Figure 9 confirms the nonlinearity observed in the characteristic of spatial variability of the meridional mode.

## 5. CONCLUSION

We investigated linear PCA and nonlinear PCA generalization application to the tropical Atlantic SST. It was observed that NLPCA mode 1 explains 38% of the total variance in the SST field, in contrast to 36% for the first PCA mode while NLPCA SST mode 2 explains 22% of the total data variance compared to 16% using the linear PCA mode 2. PCA mode 1 also characterizes average ACT variability, but only as a standing oscillation, so it is unable to evaluate the weak asymmetry in spatial pattern between average warm and cold events manifested in the 1D NLPCA. Compared to PCA, NLPCA can better point out the ACT, Atlantic dipole, Canonical and non-canonical Atlantic Niño. NLPCA can better represent all of the SST data than PCA. The Atlantic equatorial mode, which is similar to ENSO in the Pacific Ocean, has been recognized and considering the Hsieh (2004) work, it is observed that it is less linear than the latter. The non linearity in spatial variation of ACT is modest in the Atlantic Ocean in contrast to the Pacific Ocean.

NLPC mode 2 described the Atlantic dipole variability. The explained variance results show that NLPCA mode 2 explains more variance as the second principal components. The first two NLPCA modes between them marginally explain more of the total data variance than do the first two PCA modes. We observed that the Atlantic dipole is more asymmetric than Atlantic Niño. We have shown that the NLPCA serves as a useful tool to investigate various aspects of the Atlantic phenomena. For example, differences between the weak and strong ACTs, between the Atlantic canonical and non-canonical Atlantic Niño, and some of different states.

Overall, the mechanisms controlling the mean state of the tropical Atlantic are not fully understood. However, its mean state is important as an indicator for predicting future atmospheric circulation. Future analysis is needed to further explore the WAM rainfall using NLPCA and the intended SST implications for the spatio-temporal precipitation variability over West Africa.

**Acknowledgements** This work was supported by the Mwalimu Nyerere African Union Scholarship Scheme (MNAUSS). The authors would like to thank Prof. Hsieh W. William for his guidance and suggestions on an early draft of the paper.

## REFERENCES

- Aho, A. V., J. D. Ullman, and J. E. Hopcroft, 1983: Data Structures and Algorithms, Addison Wesley, Amsterdam, 427 pp.
- Bachmann, C. M., T. L. Ainsworth, and R. A. Fusina, 2006: Improved manifold coordinate representations of large-scale hyperspectral scenes. *IEEE Trans. Geosci. Rem. Sens.*, **44**, 2786-2803, doi: 10.1109/tgrs.2006.881801. [[Link](#)]
- Bruno, W., 1983: Introduction à la climatologie du golfe de guinée. *Océanographie Tropicale*, **18**, 103-138.
- Caniaux, G., H. Giordani, J. L. Redelsperger, F. Guichard, E. Key, and M. Wade, 2011: Coupling between the Atlantic cold tongue and the West African monsoon in boreal spring and summer. *J. Geophys. Res.*, **116**, C04003, doi: 10.1029/2010jc006570. [[Link](#)]
- Carton, J. A. and B. Huang, 1994: Warm events in the tropical Atlantic. *J. Phys. Oceanogr.*, **24**, 888-903, doi: 10.1175/1520-0485(1994)024<0888:WEITTA>2.0.CO;2. [[Link](#)]
- Chen, H., B. Yin, G. Fang, and Y. Wang, 2010: Comparison of nonlinear and linear PCA on surface wind, surface height, and SST in the South China Sea. *Chin. J. Oceanol. Limnol.*, **28**, 981-989, doi: 10.1007/s00343-010-9074-6. [[Link](#)]
- Czaja, A., P. V. D. Vaart, and J. Marshall, 2002: A diagnostic study of the role of remote forcing in tropical Atlantic variability. *J. Climate*, **15**, 3280-3290, doi: 10.1175/1520-0442(2002)015<3280:ADSOTR>2.0.CO;2. [[Link](#)]
- Delecluse, P., J. Servain, C. Levy, K. Arpe, and L. Bengtsson, 1994: On the connection between the 1984 Atlantic warm event and the 1982-1983 ENSO. *Tellus A*, **46**, 448-464, doi: 10.1034/j.1600-0870.1994.t01-1-00009.x. [[Link](#)]
- Deser, C., M. A. Alexander, S. P. Xie, and A. S. Phillips, 2010: Sea surface temperature variability: Patterns and mechanisms. *Annu. Rev. Mar. Sci.*, **2**, 115-143, doi: 10.1146/annurev-marine-120408-151453. [[Link](#)]
- De Silva, V. and J. B. Tenenbaum, 2004: Sparse multidimensional scaling using landmark points. Stanford Mathematics Technical Report.
- Druyan, L. M. and M. Fulakeza, 2015: The impact of the Atlantic cold tongue on West African monsoon onset in regional model simulations for 1998-2002. *Int. J. Climatol.*, **35**, 275-287, doi: 10.1002/joc.3980. [[Link](#)]
- Enfield, D. B. and D. A. Mayer, 1997: Tropical Atlantic sea surface temperature variability and its relation to El Niño-Southern oscillation. *J. Geophys. Res.*, **102**, 929-945, doi: 10.1029/96jc03296. [[Link](#)]
- Foltz, G. R. and M. J. McPhaden, 2010: Interaction between the Atlantic meridional and Niño modes. *Geophys. Res. Lett.*, **37**, L18604, doi: 10.1029/2010gl044001. [[Link](#)]

- Fukuoka, A., 1951: A study of 10-day forecast (a synthetic report). *Geophys. Mag.*, **22**, 177-208.
- Gómez, A. J., C. S. Zhou, A. Timmermann, and J. Kurths, 2004: Nonlinear dimensionality reduction in climate data. *Nonlin. Processes Geophys.*, **11**, 393-398, doi: 10.5194/npg-11-393-2004. [[Link](#)]
- Hagos, S. M. and K. H. Cook, 2009: Development of a coupled regional model and its application to the study of interactions between the west African monsoon and the eastern tropical Atlantic ocean. *J. Climate*, **22**, 2591-2604, doi: 10.1175/2008jcli2466.1. [[Link](#)]
- Handoh, I. C. and G. R. Bigg, 2000: A self-sustaining climate mode in the tropical Atlantic, 1995-1997: Observations and modelling. *Q. J. R. Meteorol. Soc.*, **126**, 807-821, doi: 10.1002/qj.49712656403. [[Link](#)]
- Hannachi, A., I. T. Jolliffe, and D. B. Stephenson, 2007: Empirical orthogonal functions and related techniques in atmospheric science: A review. *Int. J. Climatol.*, **27**, 1119-1152, doi: 10.1002/joc.1499. [[Link](#)]
- Hastie, T. and W. Stuetzle, 1989: Principal curves. *J. Am. Stat. Assoc.*, **84**, 502-516, doi: 10.1080/01621459.1989.10478797. [[Link](#)]
- Hecht-Nielsen, R., 1995: Replicator neural networks for universal optimal source coding. *Science*, **269**, 1860-1863, doi: 10.1126/science.269.5232.1860. [[Link](#)]
- Hirst, A. C. and S. Hastenrath, 1983: Atmosphere-ocean mechanisms of climate anomalies in the Angola-tropical Atlantic sector. *J. Phys. Oceanogr.*, **13**, 1146-1157, doi: 10.1175/1520-0485(1983)013<1146:AOMOCA>2.0.CO;2. [[Link](#)]
- Horel, J. D., V. E. Kousky, and M. T. Kagaro, 1986: Atmospheric conditions in the Atlantic sector during 1983 and 1984. *Nature*, **322**, 243-245, doi: 10.1038/322248a0. [[Link](#)]
- Houghton, R. W. and Y. M. Tourre, 1992: Characteristics of low frequency sea surface temperature fluctuations in the tropical Atlantic. *J. Climate*, **5**, 765-771, doi: 10.1175/1520-0442(1992)005<0765:COLFSS>2.0.CO;2. [[Link](#)]
- Hsieh, W. W., 2001: Nonlinear principal component analysis by neural networks. *Tellus A*, **53**, 599-615, doi: 10.1034/j.1600-0870.2001.00251.x. [[Link](#)]
- Hsieh, W. W., 2004: Nonlinear multivariate and time series analysis by neural network methods. *Rev. Geophys.*, **42**, RG1003, 1-25, doi: 10.1029/2002rg000112. [[Link](#)]
- Hsieh, W. W., 2007: Nonlinear principal component analysis of noisy data. *Neural Networks*, **20**, 434-443, doi: 10.1016/j.neunet.2007.04.018. [[Link](#)]
- Huang, B. and J. Shukla, 1997: Characteristics of the interannual and decadal variability in a general circulation model of the tropical Atlantic ocean. *J. Phys. Oceanogr.*, **27**, 1693-1712, doi: 10.1175/1520-0485(1997)027<1693:COTIAD>2.0.CO;2. [[Link](#)]
- Huang, B., P. S. Schopf, and Z. Pan, 2002: The ENSO effect on the tropical Atlantic variability: A regionally coupled model study. *Geophys. Res. Lett.*, **29**, 35-1-35-4, doi: 10.1029/2002GL014872. [[Link](#)]
- Jackson, J. E., 1991: A User's Guide to Principal Components, Wiley Series in Probability and Statistics, New York, 592 pp, doi: 10.1002/0471725331. [[Link](#)]
- Jouanno, J., F. Marin, Y. D. Penhoat, J. M. Molines, and J. Sheinbaum, 2011a: Seasonal modes of surface cooling in the gulf of Guinea. *J. Phys. Oceanogr.*, **41**, 1408-1416, doi: 10.1175/jpo-d-11-031.1. [[Link](#)]
- Jouanno, J., F. Marin, Y. D. Penhoat, J. Sheinbaum, and J. M. Molines, 2011b: Seasonal heat balance in the upper 100 m of the equatorial Atlantic Ocean. *J. Geophys. Res.*, **116**, C09003, doi: 10.1029/2010jc006912. [[Link](#)]
- Kaiser, H. F., 1958: The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, **23**, 187-200, doi: 10.1007/bf02289233. [[Link](#)]
- Katz, E. J., 1997: Waves along the equator in the Atlantic. *J. Phys. Oceanogr.*, **27**, 2536-2544, doi: 10.1175/1520-0485(1997)027<2536:WATEIT>2.0.CO;2. [[Link](#)]
- Kramer, M. A., 1991: Nonlinear principal component analysis using autoassociative neural networks. *AIChE J.*, **37**, 233-243, doi: 10.1002/aic.690370209. [[Link](#)]
- Landman, W. A. and W. J. Tennant, 2000: Statistical downscaling of monthly forecasts. *Int. J. Climatol.*, **20**, 1521-1532, doi: 10.1002/1097-0088(20001115)20:13<1521::AID-IJOC558>3.0.CO;2-N. [[Link](#)]
- Latif, M. and A. Grötzner, 2000: The equatorial Atlantic oscillation and its response to ENSO. *Clim. Dyn.*, **16**, 213-218, doi: 10.1007/s003820050014. [[Link](#)]
- Li, S., W. W. Hsieh, and A. Wu, 2005: Hybrid coupled modeling of the tropical Pacific using neural networks. *J. Geophys. Res.*, **110**, C09024, doi: 10.1029/2004JC002595. [[Link](#)]
- Li, T. and S. G. H. Philander, 1997: On the seasonal cycle of the equatorial Atlantic Ocean. *J. Climate*, **10**, 813-817, doi: 10.1175/1520-0442(1997)010<0813:OTSCOT>2.0.CO;2. [[Link](#)]
- Lorenz, E. N., 1956: Empirical orthogonal functions and statistical weather prediction. Scientific Report, No. 1, Dep. of Meteor., Massachusetts Institute of Technology, Cambridge, 49 pp.
- Lübbecke, J. F. and M. J. McPhaden, 2013: A comparative stability analysis of Atlantic and Pacific Niño modes. *J. Climate*, **26**, 5965-5980, doi: 10.1175/jcli-d-12-00758.1. [[Link](#)]
- Malthouse, E. C., 1998: Limitations of nonlinear PCA as performed with generic neural networks. *IEEE Trans. Neural Network.*, **9**, 165-173, doi: 10.1109/72.655038. [[Link](#)]
- Mehta, V. M., 1998: Variability of the tropical ocean surface temperatures at decadal-multidecadal timescales. Part I: The Atlantic Ocean. *J. Climate*, **11**, 2351-2375,

- doi: 10.1175/1520-0442(1998)011<2351:VOTTOS>2.0.CO;2. [[Link](#)]
- Merle, J., M. Fieux, and P. Hisard, 1980: Annual signal and interannual anomalies of sea surface temperature in the eastern equatorial Atlantic Ocean. In: Siedler, G. and J. D. Woods (Eds.), *Oceanography and Surface Layer Meteorology in the B/C Scale*, Pergamon Press, New York, 77-101, doi: 10.1016/b978-1-4832-8366-1.50023-6. [[Link](#)]
- Monahan, A. H., 2001: Nonlinear principal component analysis: Tropical Indo-Pacific Sea surface temperature and sea level pressure. *J. Climate*, **14**, 219-233, doi: 10.1175/1520-0442(2001)013<0219:NPCATI>2.0.CO;2. [[Link](#)]
- Monahan, A. H., J. C. Fyfe, and L. Pandolfo, 2003: The vertical structure of wintertime climate regimes of the northern hemisphere extratropical atmosphere. *J. Climate*, **16**, 2005-2021, doi: 10.1175/1520-0442(2003)016<2005:TVSOWC>2.0.CO;2. [[Link](#)]
- Moore, D., P. Hisard, J. McCreary, J. Merle, J. O'Brien, J. Picaut, J. M. Verstraete, and C. Wunsch, 1978: Equatorial adjustment in the eastern Atlantic. *Geophys. Res. Lett.*, **5**, 637-640, doi: 10.1029/gl005i008p00637. [[Link](#)]
- North, G. R., T. L. Bell, R. F. Cahalan, and F. J. Moeng, 1982: Sampling errors in the estimation of empirical orthogonal functions. *Mon. Weather Rev.*, **110**, 699-706, doi: 10.1175/1520-0493(1982)110<0699:SEITEO>2.0.CO;2. [[Link](#)]
- Okumura, Y. and S. P. Xie, 2004: Interaction of the Atlantic equatorial cold tongue and the African monsoon. *J. Climate*, **17**, 3589-3602, doi: 10.1175/1520-0442(2004)017<3589:IOTAEC>2.0.CO;2. [[Link](#)]
- Picaut, J., 1983: Propagation of the seasonal upwelling in the eastern equatorial Atlantic. *J. Phys. Oceanogr.*, **13**, 18-37, doi: 10.1175/1520-0485(1983)013<0018:POTSUI>2.0.CO;2. [[Link](#)]
- Picaut, J., J. Servain, A. J. Busalacchi, and M. Seva, 1984: Interannual variability versus seasonal variability in the tropical Atlantic. *Geophys. Res. Lett.*, **11**, 787-790, doi: 10.1029/gl011i008p00787. [[Link](#)]
- Rajagopalan, B., Y. Kushnir, and Y. M. Tourre, 1998: Observed decadal midlatitude and tropical Atlantic climate variability. *Geophys. Res. Lett.*, **25**, 3967-3970, doi: 10.1029/1998gl900065. [[Link](#)]
- Richter, I., S. K. Behera, Y. Masumoto, B. Taguchi, H. Sasaki, and T. Yamagata, 2012: Multiple causes of interannual sea surface temperature variability in the equatorial Atlantic Ocean. *Nat. Geosci.*, **6**, 43-47, doi: 10.1038/ngeo1660. [[Link](#)]
- Rouault, M., P. Florenchie, N. Fauchereau, and C. J. C. Reason, 2003: South East tropical Atlantic warm events and southern African rainfall. *Geophys. Res. Lett.*, **30**, doi: 10.1029/2002GL014840. [[Link](#)]
- Roweis, S. T. and L. K. Saul, 2000: Nonlinear dimensionality reduction by locally linear embedding. *Science*, **290**, 2323-2326, doi: 10.1126/science.290.5500.2323. [[Link](#)]
- Ruiz-Barradas, A., J. A. Carton, and S. Nigam, 2000: Structure of interannual-to-decadal climate variability in the tropical Atlantic sector. *J. Climate*, **13**, 3285-3297, doi: 10.1175/1520-0442(2000)013<3285:SOITDC>2.0.CO;2. [[Link](#)]
- Schölkopf, B., A. Smola, and K. R. Müller, 1998: Nonlinear component analysis as a kernel eigenvalue problem. *Neural Comput.*, **10**, 1299-1319, doi: 10.1162/089976698300017467. [[Link](#)]
- Servain, J., 1991: Simple climatic indices for the tropical Atlantic ocean and some applications. *J. Geophys. Res.*, **96**, 15137-15146, doi: 10.1029/91jc01046. [[Link](#)]
- Servain, J. and S. Amault, 1995: On forecasting abnormal climatic events in the tropical Atlantic ocean. *Ann. Geophys.*, **13**, 995-1008, doi: 10.1007/s00585-995-0995-x. [[Link](#)]
- Servain, J. and D. M. Legler, 1986: Empirical orthogonal function analyses of tropical Atlantic sea surface temperature and wind stress: 1964-1979. *J. Geophys. Res.*, **91**, 14181-14191, doi: 10.1029/jc091ic12p14181. [[Link](#)]
- Servain, J. and J. Merle, 1993: Interannual climate variations over the tropical Atlantic ocean. In: Shukla, J. (Ed.), *Prediction of Interannual Climate Variations NATO ASI Series, Vol. 6*, Springer, Berlin, Heidelberg, 153-172, doi: 10.1007/978-3-642-76960-3\_8. [[Link](#)]
- Servain, J., I. Wainer, J. P. McCreary Jr., and A. Dessier, 1999: Relationship between the equatorial and meridional modes of climatic variability in the tropical Atlantic. *Geophys. Res. Lett.*, **26**, 485-488, doi: 10.1029/1999gl900014. [[Link](#)]
- Servain, J., G. Clauzet, and I. C. Wainer, 2003: Modes of tropical Atlantic climate variability observed by PIRATA. *Geophys. Res. Lett.*, **30**, doi: 10.1029/2002GL015124. [[Link](#)]
- Smith, T. M. and R. W. Reynolds, 2003: Extended reconstruction of global sea surface temperatures based on COADS data (1854-1997). *J. Climate*, **16**, 1495-1510, doi: 10.1175/1520-0442-16.10.1495. [[Link](#)]
- Tang, Y. and W. W. Hsieh, 2003: Nonlinear modes of decadal and interannual variability of the subsurface thermal structure in the Pacific Ocean. *J. Geophys. Res.*, **108**, doi: 10.1029/2001JC001236. [[Link](#)]
- Tenenbaum, J. B., V. de Silva, and J. C. Langford, 2000: A global geometric framework for nonlinear dimensionality reduction. *Science*, **290**, 2319-2323, doi: 10.1126/science.290.5500.2319. [[Link](#)]
- Thorncroft, C. D., H. Nguyen, C. Zhang, and P. Peyrillé, 2011: Annual cycle of the West African monsoon: Regional circulations and associated water vapour

- transport. *Q. J. R. Meteorol. Soc.*, **137**, 129-147, doi: 10.1002/qj.728. [[Link](#)]
- Tokinaga, H. and S. P. Xie, 2011: Weakening of the equatorial Atlantic cold tongue over the past six decades. *Nat. Geosci.*, **4**, 222-226, doi: 10.1038/ngeo1078. [[Link](#)]
- Von Storch, H. and F. W. Zwiers, 1999: *Statistical Analysis in Climate Research*, Cambridge Univ. Press, New York, 484 pp, doi: 10.1017/cbo9780511612336. [[Link](#)]
- Xie, S. P. and J. A. Carton, 2004: Tropical Atlantic variability: Patterns, mechanisms, and impacts. In: Wang, C., S. P. Xie, and J. A. Carton (Eds.), *Earth Climate: The Ocean-Atmosphere Interaction*, Geophysical Monograph, AGU, Geophysical Monograph Series, Vol. 147, Washington, D.C., 121-142, doi: 10.1029/147GM07. [[Link](#)]
- Xie, S. P. and Y. Tanimoto, 1998: A pan-Atlantic decadal climate oscillation. *Geophys. Res. Lett.*, **25**, 2185-2188, doi: 10.1029/98gl01525. [[Link](#)]
- Xie, S. P., Y. Tanimoto, H. Noguchi, and T. Matsuno, 1999: How and why climate variability differs between the tropical Atlantic and Pacific. *Geophys. Res. Lett.*, **26**, 1609-1612, doi: 10.1029/1999gl000308. [[Link](#)]
- Zebiak, S. E., 1993: Air-sea interaction in the equatorial Atlantic region. *J. Climate*, **6**, 1567-1586, doi: 10.1175/1520-0442(1993)006<1567:AIITEA>2.0.CO;2. [[Link](#)]