Research Article

Independent Subspace Analysis of the Sea Surface Temperature Variability: Non-Gaussian Sources and Sensitivity to Sampling and Dimensionality

Carlos A. L. Pires¹ and Abdel Hannachi²

¹*Instituto Dom Luiz (IDL), Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisbon, Portugal*

²*Department of Meteorology, Stockholm University, Stockholm, Sweden*

Correspondence should be addressed to Carlos A. L. Pires; clpires@fc.ul.pt

Received 23 May 2017; Accepted 10 July 2017; Published 22 August 2017

Academic Editor: Davide Faranda

Copyright © 2017 Carlos A. L. Pires and Abdel Hannachi. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We propose an expansion of multivariate time-series data into maximally independent source subspaces. The search is made among rotations of prewhitened data which maximize non-Gaussianity of candidate sources. We use a tensorial invariant approximation of the multivariate negentropy in terms of a linear combination of squared coskewness and cokurtosis. By solving a high-order singular value decomposition problem, we extract the axes associated with most non-Gaussianity. Moreover, an estimate of the Gaussian subspace is provided by the trailing singular vectors. The independent subspaces are obtained through the search of “quasi-independent” components within the estimated non-Gaussian subspace, followed by the identification of groups with significant joint negentropies. Sources result essentially from the coherency of extremes of the data components. The method is then applied to the global sea surface temperature anomalies, equatorward of 65°, after being tested with non-Gaussian surrogates consistent with the data anomalies. The main emerging independent components and subspaces, supposedly generated by independent forcing, include different variability modes, namely, The East-Pacific, the Central Pacific, and the Atlantic Niños, the Atlantic Multidecadal Oscillation, along with the subtropical dipoles in the Indian, South Pacific, and South-Atlantic oceans. Benefits and usefulness of independent subspaces are then discussed.

1. Introduction

The climate system is a highly complex dynamical system with many nonlinearly interacting modes of variability. A characteristic feature of the system is its non-Gaussianity. The nonnormality of climate is an important determinant of extreme events and may provide insights into understanding the system dynamics [1–3]. For example, it is generally accepted that understanding non-Gaussian statistics of weather and climate has important consequences in atmospheric research not least because weather/climate risk assessment depends on understanding the tail of the system probability density function (PDF). Because of the complex interactions involved in the system, climate signals tend to be mixed. An important problem in climate research is to be able to disentangle main (or independent) signals from the mixture.

Empirical orthogonal functions (EOFs) and closely related methods (e.g., [4]) find signals that explain maximum variance but do not address the problem of mixing. An appropriate method that addresses the latter problem is independent component analysis (ICA). ICA is a method for separating mixed signals into their sources. It is based on the assumption that different sources, stemming from different physical and dynamical processes, can be considered to be statistically independent and can be separated by ICA.

ICA can be relevant to climate research to identify and separate possible anomaly patterns that may have different forcing [5]. ICA has been used in climate research only lately. For example, [6] applied ICA to the study of the West African vegetation on intraseasonal and interannual time scales. Reference [7] analyzed sea level pressure via ICA to identify the main contributors to the Arctic Oscillation signal.

Other authors applied ICA to the sea surface temperature (SST) and surface temperature [8–10].

Conventional ICA is essentially a one-dimensional algorithm in that the sources are unidimensional. In reality, however, and given the high nonlinear character of climate one cannot expect all source components (i.e., unidimensional) to be statistically independent. Rather, this can be relaxed in favor of finding groups of hidden (independent) sources. This has led to the development of an extension of ICA to the multidimensional ICA also known as Independent Subspace Analysis or ISA [11–14], already used in the analysis of the atmospheric variability [15]. Most common algorithms used to obtain independent sources are based on maximizing various information measures such as negentropy (NE).

The SST field is an important source of information to weather and climate variability as it constitutes the bottom boundary forcing of the overlying atmosphere and controls air-sea interaction through sensible and latent heat fluxes. SST variability is characterized by high nonlinearity and non-Gaussianity [1, 16–19]. In this manuscript, we analyze the SST using ISA to identify groups of independent sources. We use high-order cumulant tensors of coskewness and cokurtosis combined with high-order singular value decomposition (HOSVD) and minimize the generalized mutual information (MI) to obtain the independent sources. However, the performance of the methods may be hindered by sample sizes, serial correlation, or high dimensionality, well known characteristics of weather/climate. Therefore, we present statistical tests examining the biases of NE and MI and the effects of the sampling and dimensionality, which helps in finding robust non-Gaussian subspaces.

The manuscript is organized as follows. Section 2 presents the separation method and corresponding statistical tests. Section 3 presents application to the SST anomaly data and to a synthetic non-Gaussian surrogate consistent with the same data. A summary and conclusion are provided in the last section.

2. The Method

2.1. Cumulant-Based Independent Subspace Analysis. We designate by $\mathbf{X} \in \mathbb{R}^{N_p \times N_t}$ our $(N_p \times N_t)$ space-time data matrix anomaly (of rank N_r), representing N_t realizations of a random vector $\mathbf{x}(t) \in \mathbb{R}^{N_p}$ ($t = 1, \dots, N_t$). Principal Component Analysis (PCA) or Singular Value Decomposition (SVD) [20] are used to construct the PCs: $\mathbf{x}_{PC}(t) \equiv [X_{PC,1}(t), \dots, X_{PC,N_r}(t)]'$ with variances: $\lambda_1 \geq \dots \geq \lambda_{N_r}$. The method described below is applied to any whitened or isotropic random vector $\mathbf{y}$ of dimension $N \leq N_r$ obtained from $\mathbf{x}_{PC}(t)$ or to a subsequent orthogonal rotation $\mathbf{z} = \mathbf{R}'\mathbf{y}$ ($\mathbf{R}\mathbf{R}' = \mathbf{I}_N$) (note that since $\mathbf{y}$ is isotropic and standard multivariate Gaussian, then the same holds for $\mathbf{z}$). The source separation, or the factorization of the pdf of $\mathbf{y}$, $\rho_{\mathbf{y}}$ into partial pdfs is not a trivial problem when $\mathbf{y}$ is non-Gaussian. In fact, this is the case when the joint negentropy (NE), [21], which is invariant under any linear homeomorphism, is strictly positive, that is, $J(\mathbf{z}) \equiv E[\log(\rho_{\mathbf{z}}/\varphi_{\mathbf{z}})] = J(\mathbf{y}) \equiv J_{\text{rot}} > 0$ where $\varphi_{\mathbf{z}}$ is the multinormal probability density function (pdf) with the same first two moments as $\mathbf{z}$.

The pdf separation relies on the maximization of the sum of negentropies of a set of candidate sources. Precisely, let us consider the partition of the rotated vector $\mathbf{z}$ into a number N_s of subvectors or candidate sources (e.g., scalars, dyads, or tuples) as $\mathbf{z}' = [\mathbf{z}'_{(1)}, \mathbf{z}'_{(2)}, \dots, \mathbf{z}'_{(N_s)}] = (Z_1, \dots, Z_{N_s})$ with $\sum_{i=1}^{N_s} \dim(\mathbf{z}_{(i)}) = N$. Thanks to the separation theorem [13] or Negentropy Lemma [15], J_{rot} decomposes as a sum of partial N_s negentropies plus the nonnegative generalized mutual information (MI) $I(\mathbf{z}_{(1)}, \mathbf{z}_{(2)}, \dots, \mathbf{z}_{(N_s)})$ [22] as

$$\begin{aligned} J_{\text{rot}} - \sum_{i=1}^{N_s} J(\mathbf{z}_{(i)}) &= I(\mathbf{z}_{(1)}, \mathbf{z}_{(2)}, \dots, \mathbf{z}_{(N_s)}) \\ &\equiv E \left[\log \left(\frac{\rho_{\mathbf{z}}}{\prod_{i=1}^{N_s} \rho_{\mathbf{z}_{(i)}}} \right) \right] \geq 0, \\ J(\mathbf{z}_{(i)}) &= \frac{1}{2} \dim(\mathbf{z}_{(i)}) \log(2\pi e) \\ &\quad + E[\log(\rho_{\mathbf{z}_{(i)}})] \geq 0. \end{aligned} \quad (1)$$

The goal of ISA is to find a rotated vector $\mathbf{z} = \mathbf{R}'\mathbf{y}$ which is split into the largest number N_s of less statistically dependent subvectors minimizing MI. ICA is a restriction of ISA to the case of scalar sources, that is, $\dim(\mathbf{z}_{(i)}) = 1$ for all i . Gaussian distributed sources, (if they exist) are necessarily unidimensional and span the so-called Gaussian subspace $W_g(\mathbf{z})$ of dimension N_g . Its orthogonal complement (with respect to the covariance inner product), of dimension $N_{\text{ng}} = N - N_g$, is $W_{\text{ng}}(\mathbf{z}) = W_g(\mathbf{z})^\perp$, that is, $W_g(\mathbf{z}) \oplus W_{\text{ng}}(\mathbf{z}) = \mathbb{R}^N$.

The explicit MI dependence on $\mathbf{R}$ (with components $R_{ij}, \dots, i, j = 1, \dots, N$) is quite difficult to implement in general. Here, we use a cumulant-based contrast function, relying on the Edgeworth expansion approximation of the pdf of $\mathbf{y}$ [23], namely: $\rho_{\mathbf{y}} \approx \varphi(\mathbf{y})[1 + P_{\text{Ed}}(\mathbf{y})]$, where P_{Ed} is a multivariate linear combination of Hermite polynomials depending on cumulants up to a given $p = p_{\text{max}}$ maximum order, set here to $p_{\text{max}} = 4$. The third and fourth-order cumulants of $\mathbf{y}$ (components of the coskewness and cokurtosis tensors) are

$$\begin{aligned} \mathbf{S}(\mathbf{y})_{ijk} &\equiv E(Y_i Y_j Y_k); \\ \mathbf{K}(\mathbf{y})_{ijkl} &\equiv E(Y_i Y_j Y_k Y_l) - E(Y_i Y_j) E(Y_k Y_l) \\ &\quad - E(Y_i Y_k) E(Y_j Y_l) - E(Y_i Y_l) E(Y_j Y_k) \end{aligned} \quad (2)$$

for i, j, k, l in $\{1, \dots, N\}$. They are invariant under permutation of indices (hypersymmetry). After applying a tensor change of basis, the resulting cumulants are

$$\begin{aligned} \mathbf{S}(\mathbf{z})_{ijk} &\equiv \sum_{p,q,r=1}^N R_{pi} R_{qj} R_{rk} [\mathbf{S}(\mathbf{y})]_{pqr}; \\ \mathbf{K}(\mathbf{z})_{ijkl} &\equiv \sum_{p,q,r,s=1}^N R_{pi} R_{qj} R_{rk} R_{sl} [\mathbf{K}(\mathbf{y})]_{pqrs}, \end{aligned} \quad (3)$$

written in terms of $\mathbf{R}$ entries for i, j, k , and l in $\{1, \dots, N\}$. Cumulant tensors of any order $p \geq 3$ aggregate measures of

the joint non-Gaussianity of all possible subvectors $\mathbf{w}$ of $\mathbf{y}$. They vanish for all $\mathbf{w}$ when $\mathbf{y}$ is multinormal or for those $\mathbf{w}$ intersecting independent subvectors. To approximate NE, we introduce a positive defined matrix by contracting $\mathbf{S}$ and $\mathbf{K}$ as

$$\begin{aligned} \mathbf{M}(\mathbf{y})_{ij} &\equiv \frac{1}{12} \sum_{p,q=1}^N [\mathbf{S}(\mathbf{y})]_{ipq} [\mathbf{S}(\mathbf{y})]_{jq} \\ &+ \frac{1}{48} \sum_{p,q,r=1}^N [\mathbf{K}(\mathbf{y})]_{ipqr} [\mathbf{K}(\mathbf{y})]_{jqpr}, \end{aligned} \quad (4)$$

and similarly, for $\mathbf{z}$, that is, $\mathbf{M}(\mathbf{z})_{ij} = \sum_{p,q=1}^N R_{pi} R_{qj} \mathbf{M}(\mathbf{y})_{pq}$. The Edgeworth-based NE approximation is given by

$$\begin{aligned} J_{\text{Ed}}(\mathbf{y}) &\equiv \text{Tr}[\mathbf{M}(\mathbf{y})] = J_{\text{Ed}}(\mathbf{z}) = \sum_{i=1}^N \mathbf{M}_{ii}(\mathbf{z}) \\ &= \frac{1}{12} \|\mathbf{S}(\mathbf{z})\|_F^2 + \frac{1}{48} \|\mathbf{K}(\mathbf{z})\|_F^2, \end{aligned} \quad (5)$$

where, for example, $\|\mathbf{S}(\mathbf{z})\|_F^2$ stands for the Frobenius' squared norm (sum of all squared components) of the tensor $\mathbf{S}$. Note that there are examples of non-Gaussian pdfs verifying $J_{\text{Ed}}(\mathbf{y}) = 0$, including isotropic pdfs, for which nonlinear methods must be used [24, 25], such as the nonlinear-PCA (NL-PCA) [26] which has been applied to climatic data [27–31].

The approximation (5) is scaled as $J_{\text{rot}} - J_{\text{Ed}}(\mathbf{z}) = O(N_{\text{eq}}^{-1})$, if $\mathbf{z}$ behaves like an average of an equivalent number N_{eq} of independent and identically distributed (iid) random variables. $J_{\text{Ed}}(\mathbf{z})$ is appropriate to be used in ISA since it satisfies to an equivalent separation theorem (1):

$$J_{\text{Ed}}(\mathbf{z}) = \sum_{i=1}^{N_s} J_{\text{Ed}}(\mathbf{z}_{(i)}) + I_{\text{Ed}}(\mathbf{z}_{(1)}, \mathbf{z}_{(2)}, \dots, \mathbf{z}_{(N_s)}), \quad (6)$$

where I_{Ed} collects all the coskewness and cokurtosis mixing components of the different candidate sources. Therefore, ISA holds when I_{Ed} is minimum over the manifold $\text{SO}(N)$ of orthogonal rotations in $\mathbb{R}^N$. This is related to the Joint Approximated Diagonalization Eigen-matrices' (JADE) method [32].

The rotation matrix $\mathbf{R} = (\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$ composed of the Singular vectors (SVecs) of $\mathbf{M}(\mathbf{y})$ (in columns) provides the Tucker decomposition [33] or high-order singular value decomposition (HOSVD) [34]. The HOSVD applied to cumulant tensors lead to the Principal Cumulant Component Analysis (PCCA) [35] and to the Joint Moment Component Analysis (JMCA), [36]. Those techniques find an optimal reduced basis accounting for most of the variance and negentropy explained by a set of cumulants, making thus a generalization of PCA.

The singular values (SVs) of $\mathbf{M}(\mathbf{y})$: $\eta_{1,N} \geq \eta_{2,N} \geq \dots \geq \eta_{N,N}$ can be expressed as

$$\begin{aligned} \eta_{i,N} &= \frac{1}{12} \sum_{p,q=1}^N [\mathbf{S}(\mathbf{z})_{ipq}]^2 + \frac{1}{48} \sum_{p,q,r=1}^N [\mathbf{K}(\mathbf{z})_{ipqr}]^2; \\ &1 \leq i \leq N, \end{aligned} \quad (7)$$

which merge all the negentropy terms where the projection of $\mathbf{z}$ on the i th SVec: $Z_i = \mathbf{r}_i^T \mathbf{y}$ takes part. It provides the interpretation for the leading SVecs as being the main axes where non-Gaussianity is sought. Note also that there are other scalar tensor invariants characterizing the joint non-Gaussianity, relying on linear combinations of $\text{Tr}[\mathbf{M}^k(\mathbf{z})]$, $k \leq N$ (e.g., determinant of $\mathbf{M}$).

The vectors corresponding to nonnull SVs span the estimated non-Gaussian subspace $W_{\text{ng-Ed}}(\mathbf{z})$ whereas the kernel $\text{Ker}(\mathbf{M})$ spanned by the trailing SVecs associated with zero SVs spans the estimated Gaussian subspace $W_{\text{g-Ed}}(\mathbf{z})$ as far as third- and fourth-order cumulants are concerned. Note, however, that the method provides an overestimation of the Gaussian manifold $W_{\text{g-Ed}}(\mathbf{z}) \supseteq W_{\text{g}}(\mathbf{z})$ and an underestimation of the non-Gaussian manifold $W_{\text{ng-Ed}}(\mathbf{z}) \subseteq W_{\text{ng}}(\mathbf{z})$.

To summarize, the first step of source separation is to discard the Gaussian subspace and restrict the analysis to vectors in $W_{\text{ng-Ed}}(\mathbf{z}) = \text{Ker}[\mathbf{M}(\mathbf{z})]^\perp$ of dimension $N_{\text{ng-Ed}}$. Let us denote the non-Gaussian part of $\mathbf{z}$ by $\mathbf{z}_{\text{ng}} = (Z_1, \dots, Z_{N_{\text{ng-Ed}}})'$.

A given partition $\mathbf{z}'_{\text{ng}} = [\mathbf{z}'_{(1)}, \mathbf{z}'_{(2)}, \dots, \mathbf{z}'_{(N_s)}]$ that leads to separated true non-Gaussian sources, that is, $I_{\text{Ed}}(\mathbf{z}_{(1)}, \mathbf{z}_{(2)}, \dots, \mathbf{z}_{(N_s)}) = 0$, makes the invertible matrix $\mathbf{M}(\mathbf{z}_{\text{ng}})$ block-diagonal with N_s blocks. The SVD of $\mathbf{M}(\mathbf{z}_{\text{ng}})$ yields blocks of SVecs: $\mathbf{w}_{(i)}$; $i = 1, \dots, N_s$, spanning the same source's subspaces (i.e., $\text{Span}[\mathbf{w}_{(i)}] = \text{Span}[\mathbf{z}_{(i)}]$; $i = 1, \dots, N_s$) (up to rotation, permutation, or inversion of axes within each source). This, however, does not provide information on how to group the SVecs onto sources (source configuration indeterminacy), which is a complex problem.

In the presence of nonseparable multivariate sources, one will still be interested in finding, within each source, axes maximizing the single NEs. In other words, one solves the ICA problem in $W_{\text{ng-Ed}}$ by maximizing the NE sum $\sum_{i=1}^{N_{\text{ng-Ed}}} J_{\text{Ed}}(Z_i)$ in (6), that is, solving an ICA problem [37, 38] providing independent components (ICs) Z_i .

For a large class of pdfs, [13] has showed and conjectured that the ISA solution results in general from the grouping of ICs into independent subspaces.

The general problem of grouping ICs coming from HOSVD followed by ICA benefits from a special decomposition of $J_{\text{Ed}}(\mathbf{z}_{\text{ng}})$ into positive terms, hereafter called interactivities (ITIs). The decomposition runs through all possible nonnull $N_{\text{int}} \equiv 2^m - 2^{m-\min(m, p_{\text{max}})}$ subsets $\mathbf{u} \subseteq \mathbf{z}_{\text{ng}}$ with dimension $1 \leq |\mathbf{u}| \leq \min(m, p_{\text{max}} = 4)$, where $(m \equiv N_{\text{ng-Ed}})$. Therefore,

$$\begin{aligned} J_{\text{Ed}}(\mathbf{z}_{\text{ng}}) &\equiv \sum_{\mathbf{u} \subseteq \mathbf{z}} J_{\text{Ed-IT}}(\mathbf{u}), \\ J_{\text{Ed-IT}}(\mathbf{u}) &\equiv \frac{1}{12} \sum_{i,j,k: \{u_i, u_j, u_k\}=\mathbf{u}} [\mathbf{S}(\mathbf{u})_{ijk}]^2 \\ &+ \frac{1}{48} \sum_{i,j,k,l: \{u_i, u_j, u_k, u_l\}=\mathbf{u}} [\mathbf{K}(\mathbf{u})_{ijkl}]^2. \end{aligned} \quad (8)$$

As an example, if $\mathbf{u} = (U_1, U_2)'$ and keeping in mind the hypersymmetry, we get

$$\begin{aligned} J_{\text{Ed-IT}}(\mathbf{u}) &= \frac{1}{12} [3(\mathbf{S}_{112})^2 + 3(\mathbf{S}_{221})^2] \\ &\quad + \frac{1}{48} [4(\mathbf{K}_{1112})^2 + 6(\mathbf{K}_{1122})^2 + 4(\mathbf{K}_{2221})^2]. \end{aligned} \quad (9)$$

The ITIs, for 1D, that is, $\dim(\mathbf{u}) = 1$ (self-interactivities), coincide with single NEs whereas for $\dim(\mathbf{u}) \geq 2$ they coincide with joint-interactivities, which provide a measure of the non-linear statistical relationships within $\mathbf{u}$ contributing to $J_{\text{Ed}}(\mathbf{z}_{\text{ng}})$. Using cumulant properties [39], it is straight forward to show that $J_{\text{Ed-IT}}(\mathbf{u}) = 0$ when $\mathbf{u}$ has non-null intersections with different independent sources. Therefore, after computing all ITIs, the source identification consists of determining a partition $\{\mathbf{u}_{(1)}, \mathbf{u}_{(2)}, \dots, \mathbf{u}_{(N_s)}\}$ of $\mathbf{z}$ composed of subvectors of minimal dimension such that $J_{\text{Ed-IT}}(\mathbf{v}) = 0$ for all $\mathbf{v}$ intersecting at least two subvectors $\mathbf{u}_{(\alpha)}$ and $\mathbf{u}_{(\beta)}$, $\alpha \neq \beta$. We note that the NE decomposition is not unique. In fact, the concept of interaction information [40–42] provides a general frame for the decomposition of $J(\mathbf{z}_{\text{ng}})$ in terms of interaction information, defined for every $\mathbf{u} \subseteq \mathbf{z}_{\text{ng}}$, which are given by linear combinations of mutual information of subvectors $\mathbf{v} \subseteq \mathbf{u}$. See [43] for a proper application of ITIs on nonlinear fluid dynamics for the assessment of nonlinear wave resonances.

2.2. Independent Subspace Analysis from Finite Samples

2.2.1. Statistical Tests of Multivariate Non-Gaussianity. The above source separation method depends crucially on the approximation of the expectation of any function $f(\mathbf{z})$ of $\mathbf{z}$ -function: $E[f(\mathbf{z})]$ (e.g., the cumulants (2), the HOSVD matrix $\mathbf{M}(\mathbf{z})$ (4), the negentropy (5), and its interactivities (8)), through its estimation obtained by sample averages: $\bar{f}(\mathbf{z})$. The sample high-order cumulant estimators may be quite biased due to the eventual presence of single and joint outliers on short samples. This is particularly relevant to weather/climate with relatively short datasets with nonnegligible serial correlations yielding a smaller effective number of degrees of freedom: $N_{\text{dof}} \leq N_t$. The bias becomes even larger when we deal with high-order cumulants. In fact, the p th order (hypersymmetric) cumulant tensor in N dimensions has $C_p^{N+p-1} = (N+p-1)!/[(N-1)!p!]$ independent entries, which scales as a power law of N . This can lead to a high bias of the cumulants of the leading SVs of $\mathbf{M}(\mathbf{z})$ and to a possible fictitious non-Gaussianity of the pdf, resulting from the “curse of dimensionality.” Briefly speaking, the “false” non-Gaussianity may be due to either small sample size, high serial correlation, high dimension N , or high-order statistics p . To avoid non-Gaussianity overestimation, some procedures are usually applied including (i) data projection onto low-dimensional subspaces of high relevance (projection pursuit (PP) rationale [44, 45]); (ii) outlier’s removal; and (iii) use of resistant (or order) statistics considering reasonable values of the order p .

Another strategy that we will also follow here is to evaluate the distribution, moments, and quantiles of the distribution of the estimators $\bar{f}(\mathbf{z})_G$, necessary to perform ISA, under the null hypothesis (H_G) of multivariate Gaussianity of $\mathbf{z}$. All quantities under H_G will hereafter be noted with the subscript “G.” For iid realizations, all the above statistics are scaled as functions of N_t and N , for which one has obtained analytic asymptotic formulas. In the presence of serial correlation, the statistics are derived from a Monte-Carlo procedure by taking an ensemble of surrogate Gaussian time-series with the same N_{dof} , taken here as simple red noise AR(1) processes fitted to each $\mathbf{z}$ component as $g(t+1) = \hat{c}_1 g(t) + (1 - \hat{c}_1^2)^{1/2} w(t)$; $t = 0, \dots, N_t$, where $w(t)$ is a standard Gaussian white noise process and $\hat{c}_1$ is the lag-one autocorrelation. For the significance tests of non-Gaussianity, the upper quantiles (e.g., 90%–99%) statistics are taken from a large ensemble of surrogates.

The first step of source decomposition is to test the null hypothesis H_G . The finite-sample estimator of J_{Ed} , hereafter denoted as $\tilde{J}_{\text{Ed}}$ (all quantities based on sample statistics are overlined with a tilde) and others based on truncated NE approximations are conservative tests of H_G (or less powerful tests of its negation: $\sim H_G$), because they reject non-Gaussianity (Type II error) resulting from discarded cumulants of order $p > p_{\text{max}} = 4$.

Now, we discuss the behavior of the estimator $\tilde{J}_{\text{Ed};G}$ of $J_{\text{Ed};G}$ (under H_G) computed from N_t -sized samples of a vector of N independent scalars. The pdf of $\tilde{J}_{\text{Ed};G}$ is positively skewed, depending on the joint distribution of $\tilde{\mathbf{S}}_G$ and $\tilde{\mathbf{K}}_G$, which are in general not independent since they satisfy a number of inequality constraints (e.g., $\mathbf{K}_{1111} - \mathbf{S}_{111}^2 + 2 \geq 0$). The bias of $\tilde{J}_{\text{Ed};G}$ is given by

$$\begin{aligned} E(\tilde{J}_{\text{Ed};G}) &= \frac{1}{N_{\text{dof}}} P_{J_{\text{Ed}}}(N) + O\left(\frac{1}{N_{\text{dof}}^2}\right), \\ &\frac{1}{N_{\text{dof}}} P_{J_{\text{Ed}}}(N) \\ &\equiv \frac{1}{12} \left[\frac{1}{N_{\text{dof}}} (N^3 + 3N^2 + 2N) \right] \\ &\quad + \frac{1}{48} \left[\frac{1}{N_{\text{dof}}} (N^4 + 6N^3 + 11N^2 + 6N) \right], \end{aligned} \quad (10)$$

where the first and second square-bracket terms of (10) represent, respectively, scaled versions of $E[\|\tilde{\mathbf{S}}_G\|_F^2]$ and $E[\|\tilde{\mathbf{K}}_G\|_F^2]$. The bias (10) tends asymptotically to $P_{J_{\text{Ed}}}(N)/N_{\text{dof}}$. It also provides a lower-bound of the positive $\tilde{J}_{\text{Ed}}$ bias for non-Gaussian pdfs, as it occurs for any negentropy estimator [46].

The biases from the Frobenius norms were obtained by multiplying the biases of the different squared cumulant estimators by their multiplicities [36]. For instance, there are $3N(N-1)$ cumulant biases of type $E[(Y_i Y_j Y_i)^2] = 2N_{\text{dof}}^{-1} + O(N_{\text{dof}}^{-2})$; $i \neq j$ in $E[\|\tilde{\mathbf{S}}_G\|_F^2]$ based on the moments of the

FIGURE 1: (a) Contour plot of f_{REd} as a function of N_t , $\hat{c}_1$ for a fixed dimension $N = 4$. (b) Ratio $f_{\text{REd}-P_{\text{sig}}}$ for the confidence level $p_{\text{sig}} = 0.95$, as a function of the dimension N and the lag-one autocorrelation $\hat{c}_1$ for a sample size $N_t = 1000$. The ratio increases with decreasing N or increasing $\hat{c}_1$.

standard Gaussian: $E(U^{2k}) = (2k)!/(2^k k!)$. Figure 1 shows the dependence on N_t and $\hat{c}_1$ for $N = 4$, of the ratio

$$f_{\text{REd}}(N_t, \hat{c}_1, N) \equiv \frac{E(\tilde{J}_{\text{Ed:G}}) N_t}{P_{\text{Ed}}(N)} \quad (11)$$

between the actual bias and the asymptotic bias of $\tilde{J}_{\text{Ed:G}}$ (at $\hat{c}_1 = 0$), obtained from 1000 Monte-Carlo replicates. Moreover, the dependence of f_{REd} on N is quite negligible with relative deviations less than 5% (not shown).

From the analysis of Figure 1(a) and as expected, the ratio f_{REd} remains close to 1 when $\hat{c}_1$ is small and N_t is large. For $N_t \geq 500$, the ratio has reached the asymptotic value, where it becomes dependent only on $\hat{c}_1$ with $N_{\text{dof}} = N_t / f_{\text{REd}}(\infty, \hat{c}_1, N) < N_t$, where f_{REd} is interpreted as the “average interval between independent realizations.” For small sample sizes, $f_{\text{REd}} < f_{\text{REd}}(\infty, \hat{c}_1, N)$, due to a negative higher-order term in (10).

The variance of $\tilde{J}_{\text{Ed:G}}$ involves covariances between generic cumulants, for example, $\text{cm}(\mathbf{u}_1)$ and $\text{cm}(\mathbf{u}_2)$, of subvectors $\mathbf{u}_1, \mathbf{u}_2 \subseteq \mathbf{z}$. Those covariances satisfy $\text{cov}[\text{cm}(\mathbf{u}_1), \text{cm}(\mathbf{u}_2)] = kN_t^{-d}$; $k \geq 0$ with $d = 2$ if $\mathbf{u}_1 = \mathbf{u}_2$ and $d = 3$ if $\mathbf{u}_1 \supset \mathbf{u}_1 \cap \mathbf{u}_2 \neq \emptyset$ and $d = \infty$ otherwise.

As for the bias, the quantiles $p_{\text{sig}} \in [0, 1]$ of $\tilde{J}_{\text{Ed:G}}$ hereby noted $Q_{p_{\text{sig}}}(\tilde{J}_{\text{Ed:G}})$ scale asymptotically as N_t^{-1} . Figure 1(b) shows the dependence on N and $\hat{c}_1$ for $N_t = 1000$, of the ratio $f_{\text{REd}-P_{\text{sig}}} \equiv Q_{p_{\text{sig}}}(\tilde{J}_{\text{Ed:G}})[N_t / P_{\text{Ed}}(N)]$. This figure provides a kind of a rule of thumb thresholds of $\tilde{J}_{\text{Ed}}$ for rejecting H_G in the asymptotic regime.

2.2.2. Estimation of the Gaussian Subspace by HOSVD. Following a rejection of H_G , the next step is to estimate the Gaussian subspace W_g and its dimension N_g . With finite

samples, one computes the sample-based matrix $\tilde{\mathbf{M}}(\mathbf{z})$ of $\mathbf{M}(\mathbf{z})$ (4). However, W_g can no longer be estimated by $\text{Ker}(\tilde{\mathbf{M}})$ (see Section 2.1) but through the subspace spanned by SVecs associated with the trailing (not significant) SVs.

Under H_G , the (estimated) SVs $\tilde{\eta}_{i,N;G}$, (see (7)), with quantiles $Q_{p_{\text{sig}}}(\tilde{\eta}_{i,N;G})$, $i = 1, \dots, N$, are asymptotically $O(N_{\text{dof}}^{-1})$ and the SVs become undistinguishable for large N_{dof} [36]. However, the SVs are not independent and consequently their marginal quantiles cannot be estimated independently for each i . To overcome this difficulty, a sequence of statistical tests is applied, as discussed in the next subsection.

In summary, we first consider the vectors $\mathbf{z} = (Z_1, \dots, Z_N)'$ $\mathbf{R}'\mathbf{y}$ and $\mathbf{z}_s \equiv (Z_{s,1}, \dots, Z_{s,N})' = \mathbf{R}'_s \mathbf{y}_s$, where $\mathbf{y}_s$ is a generic surrogate of $\mathbf{y}$ and $\mathbf{R}$ and $\mathbf{R}_s$ are matrices filled by the ranked SVecs of $\tilde{\mathbf{M}}(\mathbf{y})$ and $\tilde{\mathbf{M}}(\mathbf{y}_s)$, respectively. For each i in $\{1, \dots, N\}$, we construct the vector $\mathbf{b}_i \equiv (Z_1, \dots, Z_{i-1}, Z_{s,i}, \dots, Z_{s,N})$, where the trailing $i, \dots, N$ components are swapped by the associated ranked surrogates. Then we compute their NE contributions $\tilde{\mathbf{M}}_{ii}(\mathbf{b}_i)$ (5) and finally the quantile $Q_{p_{\text{sig}}}[\tilde{\mathbf{M}}_{ii}(\mathbf{b}_i)]$ evaluated from a large ensemble of surrogates. The test then proceeds backward, starting from the smallest SV ($i = N$). Therefore, if $\tilde{\mathbf{M}}_{kk}(\mathbf{z}) < Q_{p_{\text{sig}}}[\tilde{\mathbf{M}}_{kk}(\mathbf{b}_k)]$, for all $k \geq i$, then the null hypothesis $H_{G,i}$ (Z_k undistinguishable from Gaussian surrogates $Z_{s,k}$ for all $k \geq i$) cannot be rejected. The smallest i for which $H_{G,i}$ is not rejected determines the estimated Gaussian subspace $W_{g-\text{Ed}}$ with dimension $N_{g-\text{Ed}}$ (spanned by the trailing $N_{g-\text{Ed}}$ SVecs of $\mathbf{R}$), and $W_{ng-\text{Ed}} \equiv W_{g-\text{Ed}}^\perp$ (spanned by the leading $N_{ng-\text{Ed}} \equiv N - N_{g-\text{Ed}}$ SVecs of $\mathbf{R}$).

2.2.3. Estimation of Multivariate Sources. The next step is to identify the non-Gaussian sources within the vector of

non-Gaussian components $\mathbf{z}_{\text{ng}} \equiv (Z_1, \dots, Z_{N_{\text{ng-Ed}}})'$ spanning $W_{\text{ng-Ed}}$. Since $\mathbf{z}_{\text{ng}}$ is a subvector of $\mathbf{z}$, we easily get $\tilde{J}_{\text{Ed}}(\mathbf{z}_{\text{ng}}) \leq \tilde{J}_{\text{Ed}}(\mathbf{z})$.

We test the separability of $W_{\text{ng-Ed}}$ into scalar and/or subvector sources by looking for a new orthogonal rotation $\mathbf{z}_{\text{ISA}} \equiv \mathbf{R}'_{\text{ISA}} \mathbf{z}_{\text{ng}} = (Z_{1,\text{ISA}}, \dots, Z_{N_{\text{ng-Ed}},\text{ISA}})' = [\mathbf{z}'_{(1),\text{ISA}}, \mathbf{z}'_{(2),\text{ISA}}, \dots, \mathbf{z}'_{(N_s),\text{ISA}}]'$ composed of a maximal number $N_s \geq 2$ of subvector candidate sources $\mathbf{z}_{(i),\text{ISA}}$, $i = 1, \dots, N_s$, such that the joint pdf satisfies

$$\rho_{\mathbf{z}_{\text{ISA}}} = (1 + \varepsilon_{\text{ISA}}) \prod_{k=1}^{N_s} \rho_{\mathbf{z}_{(k),\text{ISA}}}, \quad (12)$$

where the norm $|\varepsilon_{\text{ISA}}|$ of the pdf separation error ε_{ISA} is below the threshold of rejection of the null hypothesis of nonseparability ($N_s = 1$). There is an a priori source configuration indeterminacy. For instance, for $N_{\text{ng-Ed}} = 4$, there are five possible ways, namely: 4 scalars (ICA case); 2 scalars and a dyad; 2 dyads; a triad plus a scalar and the global quartet.

For a chosen source configuration, an appropriate contrast function to maximize is

$$\begin{aligned} F_{\text{ISA}}(\mathbf{R}) &\equiv \sum_{i=1}^{N_s} \tilde{J}_{\text{Ed}}(\mathbf{z}_{(i),\text{ISA}}) \\ &= \tilde{J}_{\text{Ed}}(\mathbf{z}_{\text{ng}}) \\ &\quad - \tilde{J}_{\text{Ed}}(\mathbf{z}_{(1),\text{ISA}}, \mathbf{z}_{(2),\text{ISA}}, \dots, \mathbf{z}_{(N_s),\text{ISA}}), \end{aligned} \quad (13)$$

where $\tilde{J}_{\text{Ed}}$ must be below some significance threshold, passing the separation statistical test. However, an undecidable situation may occur when multiple configurations pass the test. To avoid the case of multiple configuration indeterminacy, we solve instead the ICA problem

$$\mathbf{R}_{\text{ICA}} \equiv \arg \max_{\mathbf{R}_{\text{ISA}}} \left[\sum_{i=1}^{N_s} \tilde{J}_{\text{Ed}}(Z_{i,\text{ISA}}) \right], \quad (14)$$

providing the vector of best scalar sources: $\mathbf{z}_{\text{ICA}} = \mathbf{R}'_{\text{ICA}} \mathbf{z}_{\text{ng}} = (Z_{1,\text{ICA}}, \dots, Z_{N_{\text{ng-Ed}},\text{ICA}})'$. The algorithm to get $\mathbf{R}_{\text{ICA}}$ is explained in Appendix A. Finally, we build multivariate sources forming appropriate groups of $\mathbf{z}_{\text{ICA}}$ components, according to the conjecture of [13], and as below.

We start with the NE decomposition into N_{int} interactivities (8) as

$$\tilde{J}_{\text{Ed}}(\mathbf{z}_{\text{ng}}) = \tilde{J}_{\text{Ed}}(\mathbf{z}_{\text{ICA}}) = \sum_{\mathbf{u} \subseteq \mathbf{z}_{\text{ICA}}} \tilde{J}_{\text{Ed-IT}}(\mathbf{u}). \quad (15)$$

To compare and account for the statistical significance of $\mathbf{u}$ -interactivities in (15) with different dimensions $\dim(\mathbf{u})$, we introduce scaled interactivities $\tilde{J}_{\text{Ed-ITsig}}(\mathbf{u})$ through its comparison with the quantiles of the NE $\tilde{J}_{\text{Ed-IT}}(\mathbf{u}_{\text{G}})$ of the

ranked (by HOSVD) Gaussian surrogates $\mathbf{u}_{\text{G}}$ of $\mathbf{u}$. It is defined as

$$\begin{aligned} \tilde{J}_{\text{Ed-ITsig}}(\mathbf{u}) &= \frac{\tilde{J}_{\text{Ed-IT}}(\mathbf{w})}{Q_{P_{\text{sig}}}[\tilde{J}_{\text{Ed-IT}}(\mathbf{u}_{\text{G}})]}; \\ \mathbf{w} &= \begin{cases} \mathbf{u} & \text{if } \dim(\mathbf{u}) = 1 \\ \mathbf{u}_{\text{Gsi}} & \text{if } \dim(\mathbf{u}) > 1. \end{cases} \end{aligned} \quad (16)$$

When $\dim(\mathbf{u}) > 1$ in (16), $\mathbf{w}$ is the rotation of $\mathbf{u}_{\text{Gs}}$ towards isotropy (e.g., by Gram Schmidt orthogonalization) with $\mathbf{u}_{\text{Gs}}$ being composed of the standard Gaussianized components of $\mathbf{u}$. Since $\mathbf{u}_{\text{G}}$ has Gaussian marginals, this procedure leads to a fairer and unbiased test when $\mathbf{u}$ has non-Gaussian pdf marginals. This is also supported by the fact that the mutual information of $\mathbf{u}$ is $I(\mathbf{u}) = I(\mathbf{u}_{\text{Gs}}) = I(\mathbf{u}_{\text{Gsi}}) + [-(1/2) \log(\det \mathbf{C}_{\text{Gs}})] + \sum_{k=1}^N J(\mathbf{u}_{\text{Gsi},k}) \approx I(\mathbf{u}_{\text{Gsi}})$ where the second term is the Gaussian MI component [47, 48], which depends on the correlation matrix $\mathbf{C}_{\text{Gs}}$ of $\mathbf{u}_{\text{Gs}}$ and the third term is the sum of the NEs of the components of $\mathbf{u}_{\text{Gsi}}$. The above measure is therefore resistant to single outliers of the marginal pdfs.

The different N_{int} scaled interactivities (16) are then sorted by decreasing order, retaining uniquely the leading ones verifying $\tilde{J}_{\text{Ed-ITsig}}(\mathbf{u}) \geq J_{\text{sig}}$ where J_{sig} is a fixed level of acceptance. The emerging large interactivities normally yield ICs, which contribute to the multivariate sources. The NE interactivities are then grouped along the relevant interactivities until the stopping criteria are reached, for example, $\sum_{\mathbf{u}} \tilde{J}_{\text{Ed-ITsig}}(\mathbf{u}) < \tilde{J}_{\text{Ed}}(\mathbf{z}_{\text{ng}}) - E[\tilde{J}_{\text{Ed}}(\mathbf{z}_{\text{ng}})]$, which will be adopted in practice.

3. Analysis of Sea Surface Temperature Data

3.1. Data Description and Processing. We consider the monthly SST data from the “Extended Reconstruction Sea Surface Temperature” (ERSST) version v3b product (<https://www.ncdc.noaa.gov/oa/climate/research/sst/ersstv3.php>), from the International Comprehensive Ocean–Atmosphere Data Set (ICOADS) release 2.4. The SST data are on $2^\circ \times 2^\circ$ latitude-longitude resolution and span the period 1875 to present [49]. We only consider the period 1910–2011 (102 years), restricted to the region equatorward of 65° . At each grid point, the climatology and the (12-month moving average) linear trend in addition to the mean seasonal cycle are removed from the data [50], thus yielding $N_t = 1224$ monthly SST anomalies (SSTAs), on $N_p = 8714$ grid points. A EOF analysis is then applied [51] and we only use the space spanned by the leading $N_{\text{max}} = 12$ mode for the application of ICA/ISA. Its output is independent from PC variances, even when the used PCs are quasi-degenerated. The total variance is $\sigma_{\text{Total}}^2 = 1709.4 (^\circ\text{C})^2$, the local variance at point P is σ_P^2 , with an average $\sigma_{\text{Local}}^2 \equiv \sigma_{\text{Total}}^2 / [N_p c_{\text{lat}}] \sim (0.50)^2 (^\circ\text{C})^2$ (c_{lat} = average of $\cos(\text{latitude})$) and we let the fractional local variance $\hat{\sigma}_P^2 \equiv \sigma_P^2 / \sigma_{\text{Local}}^2$. Figure 2(a) and Table 2 show the % of explained variance of the 12 leading EOFs with their confidence intervals [52] and Figure 2(b) shows the fitted lag-one autocorrelation $\hat{c}_1$ based on an AR(1) model.

FIGURE 2: (a) % of explained variance confidence interval and (b) lag-one autocorrelation of the AR(1) fitting model of the leading 12 PCs.

Figure 3 shows the leading 12 EOFs which are described, with references in Table 1. Based on $\hat{c}_1$, an estimate of the “average interval between independent realizations” is $\Delta_{\text{ind}} = N_t/N_{\text{dof}} = f_{\text{REd}}(N_t, \hat{c}_1, N)$, (Figure 1(a)) which is not too different from $(-1.)/\ln(\hat{c}_1)$ (e.g., [53]).

Figure 2(b) shows that Δ_{ind} varies from about 3 months (PC10) to more than 12 months (PC2). For our simulation later, we will use the geometric average of the 12 $\hat{c}_1$ values, that is, 0.86 (see Section 3.3).

3.2. Negentropy Distribution and Non-Gaussian Subspace Estimation. Non-Gaussianity and negentropy of a field $\hat{\mathbf{x}}$ have counterparts both on a local and on a PC basis. In fact, local cumulants are obtained from cumulants of the whitened PCs composing vector $\mathbf{y}$, through linear operators. Let us consider, for example, the local-basis coskewness tensor $\mathbf{S}(\hat{\mathbf{x}})$. Its Frobenius norm is the area integral of the local squared skewness. By using the EOF orthogonality property, we may express this as a weighted sum over all grid point triads (P, Q, R) :

$$\begin{aligned} \|\mathbf{S}(\hat{\mathbf{x}})\|_F^2 &\equiv \sum_{P, Q, R=1}^{N_p} [\mathbf{S}(\hat{\mathbf{x}})_{PQR}]^2 \hat{\sigma}_P^2 \hat{\sigma}_Q^2 \hat{\sigma}_R^2 \hat{a}_P \hat{a}_Q \hat{a}_R \\ &= \sum_{i, j, k=1}^{N_r} [\mathbf{S}(\mathbf{y})_{ijk}]^2 \hat{\lambda}_i \hat{\lambda}_j \hat{\lambda}_k \\ &\simeq \frac{\|\mathbf{S}(\mathbf{y})\|_F^2}{N_r^3} \sum_{i, j, k=1}^{N_r} \hat{\lambda}_i \hat{\lambda}_j \hat{\lambda}_k, \end{aligned} \quad (17)$$

where $\hat{a}_P$ stand for the relative area at grid point P and $\hat{\sigma}_P^2$ is the fractional local variance. The contributions to (17) of a triad of whitened PCs are proportional to their

variance fractions. The Frobenius norm (17) is dominated by contributions primarily from regions of high variance and absolute skewness, namely, the positively skewed El Niño zone [19, 54], in addition to the positive skewness southwards of 50°S [17].

Now, the NE $\tilde{J}_{\text{Ed}}(\mathbf{y})$ can be decomposed using $\tilde{\mathbf{M}}(\mathbf{y})$ (5). For each $N = 1, \dots, N_{\text{max}} = 12$, we denote by $\mathbf{y}_{(N)} = (Y_1, \dots, Y_N)$ the set of N leading whitened PCs. Figure 4 shows the diagonal terms $(\tilde{\mathbf{M}}_{kk}[\mathbf{y}_{(N)}])$, Figure 4(a) and the total NE $\tilde{J}_{\text{Ed}}[\mathbf{y}_{(N)}] = \text{Tr}(\tilde{\mathbf{M}}[\mathbf{y}_{(N)}])$, Figure 4(b). Note that, for a given N , the smaller values of $\tilde{\mathbf{M}}_{kk}[\mathbf{y}_{(N)}]$ are either the least non-Gaussian PCs or the less interactive PCs with the remaining PCs ($i \neq k$), providing therefore good candidates for the Gaussian subspace of $\mathbf{y}_{(N)}$. For example, with $N = 12$, we have $k = 2, 7, 8, 9$, and 12 (Figure 4(a)). Statistical significance of NE components and totals are assessed by using quantiles computed from Gaussian surrogates (Figure 4). Notice, for instance, the confidence level <70% of $\tilde{\mathbf{M}}_{22}[\mathbf{y}_{(N)}]$ of the PC2 due to its small degrees of freedom.

Biases $E\{\tilde{J}_{\text{Ed}}[\mathbf{y}_{(N)}]\} - J_{\text{Ed}}[\mathbf{y}_{(N)}]$ may be approximated as $E(\tilde{J}_{\text{Ed};G}) = f_{\text{REd}}(N_t, \hat{c}_1, N) P_{J_{\text{Ed}}}(N)/N_t$ (see (11)) with $\hat{c}_1 \approx 0.86$, leading to $f_{\text{REd}}(N_t = 1224, \hat{c}_1 = 0.86) \sim 3.3$ (see Figure 1(a)). For instance, for $N = 12$, $\tilde{J}_{\text{Ed}} = 3.1$, the bias is 1.7 (55%) with the remaining value (1.4) being attributed to true non-Gaussianity. Therefore, one expects that biases shall be drastically reduced after discarding the Gaussian subspace.

Figure 4(a) shows that, for all N , the largest NE component is $\tilde{\mathbf{M}}_{11}[\mathbf{y}_{(N)}]$, resulting mostly from the nonlinear correlations of PC1 with other PCs and from its self-NE, due to its positive skewness (Table 2).

Using the nonnegative consecutive jumps: $\tilde{\mathbf{M}}_{kk}[\mathbf{y}_{(N)}] - \tilde{\mathbf{M}}_{kk}[\mathbf{y}_{(N-1)}]$ (Figure 4(a)), one may detect which interactivities

TABLE 1: EOF description with the main SSTA modes, along with corresponding available indices from literature and references.

EOF rank	Description with the main SSTA modes	Available index	PC/Index cor.	Refs.	Notes
1	El Niño (ENSO)	MEI	0.98	[56]	
2	Atlantic Multi-Decadal Oscillation (AMO)	AMO	0.80	[57]	
3	Pacific Decadal Oscillation (PDO): negative phase	PDO	-0.56	[58, 59]	
	Northern Pacific Gyre Oscillation (NPGO)	NPGO	0.53	[60]	
4	El Niño Modoki (EM): negative phase	EMI	-0.42	[61]	Also known as Central Pacific or noncanonical El Niño
	El Niño Modoki (EM): positive phase	EMI	0.57	[61]	
5	Subtropical Indian Ocean Dipole: negative phase	SIOD		[62, 63]	First EOF mode of the SSTA in the Indian Ocean ranging in the 15°S–45°S band
	Atlantic Niño: negative phase	ATL3	-0.49	[64, 65]	
	South Atlantic Subtropical Dipole	SASD		[66]	First EOF coupled mode of the SSTA/wind in the South Atlantic
6	Subtropical Indian Ocean Dipole: positive phase	SIOD		[62, 63]	First EOF mode of the SSTA in the Indian Ocean ranging in the 15°S–45°S band
	Zonally oriented dipole in South Pacific			[67, 68]	Second EOF of the South Pacific SSTA
7	Meridional sequence of dipoles crossing the North-South Pacific				
	South Pacific Ocean Dipole (SPOD)	SPOD		[67, 69]	First EOF of the South Pacific SSTA
8	North Atlantic tripole (NAT)	NAT		[70]	
	Tropical Atlantic SST dipole			[68]	
9	South Pacific Ocean Dipole (SPOD)	SPOD		[67, 69]	First EOF of the South Pacific SSTA
10	Zonally oriented dipole in South Atlantic			[68]	
	Zonally oriented dipole in South Pacific			[67, 68]	Second EOF of the South Pacific SSTA
	Meridional Pacific wavelike pattern				
11	Meridional Atlantic wavelike pattern				
	Subtropical Indian Ocean Dipole (positive phase)	SIOD		[62, 63]	First EOF mode of the SSTA in the Indian Ocean ranging in the 15°S–45°S band
12	Tripolar pattern in North Pacific				
	North Atlantic tripole (NAT)	NAT		[70]	

TABLE 2: % of explained variance, skewness, and kurtosis of the 12 leading PCs and the squared projections of $W_{\text{ng-Ed}}$ onto the individual whitened PCs for the analysis over $N = 11$ PCs (see the end of Section 3.2).

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12
%Var	17.6	5.3	4.6	3.8	3.3	2.7	2.5	2.2	2.2	2.0	1.9	1.8
Skew.	0.24	0.03	0.01	0.28	-0.16	0.25	-0.08	0.23	0.14	0.16	0.29	0.13
Kurt.	-0.01	-0.64	-0.33	0.19	0.02	-0.02	-0.06	0.17	-0.04	0.55	0.14	-0.16
SqPro	0.65	0.53	0.16	0.68	0.50	0.56	0.29	0.05	0.37	0.81	0.39

FIGURE 3: Maps of the 12 variance-leading, spatially normalized EOFs of the monthly SSTAs field, computed in the period 1910–2011. Ranked EOFs are read from left to right and top to bottom.

FIGURE 4: (a) NE components parcel $\tilde{M}_{kk}[\mathbf{y}_{(N)}]$ (k in abscissas) for each $N = 1, \dots, 12$ (marked in grey shaded). (b) Total NE $\tilde{J}_{\text{Ed}}[\mathbf{y}_{(N)}]$ for $N = 1, \dots, 12$. Values rejecting the null hypothesis of multi-Gaussianity at confidence levels <70%, 70%–90%, 90%–95%, and >95% are marked, respectively, by “+,” open circle, small solid circle, and large solid circle.

mostly contribute to $\tilde{J}_{\text{Ed}}[\mathbf{y}_{(N)}]$. The largest differences are for (k, N) pairs: $\{(1, 3), (1, 5), (1, 10), (1, 11), (3, 11), (4, 10), (4, 11), (5, 10), (5, 11), (6, 11)\}$ in consistency with the high confidence level of $\tilde{M}_{kk}[\mathbf{y}_{(11)}]$ for $k = 1, 3-6, 10, 11$. Those jumps are often due to particular joint cumulants in the 2D subspaces spanned by (PC_k, PC_N) , responsible for particular types of joint non-Gaussianity, for example, scatter-plots resulting for instance from the occurrence of joint extremes. For instance, $\overline{Y_k^2 Y_i} > 0$, is favored for the sign combinations $(-, +)$ and $(+, +)$ of (Y_k, Y_i) . As an illustration, Figures 5(a) and 5(b) show, respectively, scatter-plots of (Y_{1g}, Y_{3g}) and (Y_{1g}, Y_{11g}) (the marginals have been Gaussianized (subscript g) to emphasize the 2D pdf difference with respect to 2D isotropic Gaussianity). Figure 5 shows that most of the non-Gaussianity is due to the occurrence of joint coherent 2D positive extremes, for example, the 1983 and 1997 extreme El Niño events (see Figure 3 of [50]) during which a positive PCII and a negative PDO phase have occurred.

Since $\tilde{M}_{12,12}[\mathbf{y}_{(12)}]$ is quite low and not highly significant (Figure 4(a)), we have decided to apply ICA and ISA in the 11D subspace (see Section 3.4) in which the total NE reach a confidence level > 95%. Besides the H_g test to limit N [55] presents an alternative limit. The dimensionality reduction, however, is performed below via the statistical detection of the non-Gaussian subspace as detailed in Section 2.2.2.

Therefore, we now consider the state vector $\mathbf{y}_{(11)}$ and the associated vector $\mathbf{z} = (Z_1, \dots, Z_{11})'$, yielding $\tilde{J}_{\text{Ed}}[\mathbf{z}] = 2.55$, and a bias of 1.73, leaving ~ 0.82 for “true non-Gaussianity.”

FIGURE 5: Scatter-plots of PC1 versus PC3 (a) and PC1 versus PC11 (b). The marginal pdfs have been Gaussianized.

Figure 6 shows the NE components (or SVs) $\tilde{M}_{kk}(\mathbf{z})$ ($k = 1, \dots, N$), along with the threshold quantiles 90% and 95% of $Q_{P_{\text{sig}}}[\tilde{M}_{kk}(\mathbf{b}_k)]$ obtained from an ensemble of 1000 Gaussian surrogates. For comparison, we also show the SVs $\tilde{M}_{kk}(\mathbf{z}_s)$ (open squares) of surrogate realizations $\mathbf{z}_s$ of $\mathbf{z}$. The SVs

FIGURE 6: Ranked SVs $\tilde{\mathbf{M}}_{kk}(\mathbf{z})$, $k = 1, \dots, N = 11$, contributing to NE, both from real SSTA data (solid squares) and from Gaussian surrogate realizations (open squares) and 90% (red), 95% (blue) quantiles of $Q_{p_{\text{sig}}}[\tilde{\mathbf{M}}_{kk}(\mathbf{b}_k)]$ obtained from Gaussian surrogates after rank k , to be compared with $\tilde{\mathbf{M}}_{kk}(\mathbf{z})$. The trailing 6 SVs is not significant at $p_{\text{sig}} = 95\%$, leading to $N_{\text{ng-Ed}} = 5$. The NE components restricted to the non-Gaussian subspace are marked as open circles.

of $\mathbf{z}_s$ all lie below the above thresholds, supporting the consistency of the Gaussianity test. Figure 6 also suggests that the estimated Gaussian subspace $W_{\text{g-Ed}}$ of the SSTA is spanned by the last 6 SVs (with a confidence level $p_{\text{sig}} = 95\%$) and so $N_{\text{g-Ed}} = 6$ and, $N_{\text{ng-Ed}} = 5$ for the non-Gaussian subspace $W_{\text{ng-Ed}}$. The total NE restricted to $\mathbf{z}_{\text{ng}} \equiv (Z_1, \dots, Z_5)'$ is $\tilde{J}_{\text{Ed}}[\mathbf{z}_{\text{ng}}] = 1.11$ with a bias of 0.14. The “useful” true non-Gaussianity is 0.97 which is of the order of the above estimated value (0.82). The NE components $\tilde{\mathbf{M}}_{kk}(\mathbf{z}_{\text{ng}})$ (open black circles), with their low values compared to $\tilde{\mathbf{M}}_{kk}(\mathbf{z})$, $k = 1, \dots, 5$, are also shown in Figure 6.

The squared projections of $W_{\text{ng-Ed}}$ onto each of the 11 whitened PCs are given by the inner product (Appendix B) $W_{\text{ng-Ed}} \cdot \mathbf{e}_i = \sum_{k=1}^5 R_{i,k}^2$, for $i = 1, \dots, N = 11$, (Table 2), where $\mathbf{e}_i$ is the (unitary) i th vector. By definition, those values add up to $N_{\text{ng-Ed}} = 5$. Table 2 shows that some PCs mostly project onto the non-Gaussian subspace such as PC1 (El Niño), PC4 (NPGO), and PC10 while others mostly project onto the Gaussian subspace, for example, PC7, PC8, and PC3 (negative PDO). The 48% of explained variance is distributed, respectively, as $\sum_{i=1}^{11} \hat{\lambda}_i \sum_{k=1}^5 R_{i,k}^2 = 24.7\%$ and 23.3%, for the non-Gaussian and Gaussian subspaces.

3.3. Application to Synthetic Non-Gaussian Data. We generate $N_{\text{max}} = 12$ AR(1) independent series $G_k(t)$, $t = 1, \dots, N_t$, $k = 1, \dots, N_{\text{max}}$. The whitened and uncorrelated components of the non-Gaussian vector $\mathbf{y}' = (Y_1, \dots, Y_N)$ are computed following

$$Y_k = G_k; \quad k = 1, \dots, N \leq N_{\text{max}}, \quad k \neq 2, 5$$

$$Y_2 = \sigma 2^{-1/2} (G_1^2 - 1) + (1 - \sigma^2)^{1/2} G_2;$$

$$Y_5 = \sigma G_3 G_4 + (1 - \sigma^2)^{1/2} G_5$$

(18)

with $\sigma^2 \in [0, 1]$ measuring the “strength” of non-Gaussianity. So $\mathbf{y}$ is composed of $N_{\text{g}} = N - 5$ Gaussian scalar sources spanning W_{g} and $N_{\text{ng}} = 5$ non-Gaussian variables spanning W_{ng} . The latter is composed of one dyad $\mathbf{y}'_{(1)} = (Y_1, Y_2)$ and a non-Gaussian triad $\mathbf{y}'_{(2)} = (Y_3, Y_4, Y_5)$.

The next step consists of testing the null hypothesis H_G of global Gaussianity (i.e., $\sigma^2 = 0$) and evaluating $\tilde{J}_{\text{Ed}}(\mathbf{y})$ by varying σ^2 in $[0, 1]$ and N_{g} in $\{0, \dots, 7\}$.

Figure 7(a) shows the Monte-Carlo average $E[\tilde{J}_{\text{Ed}}(\mathbf{y})]$ obtained from 1000 surrogates. As expected, the mean NE increases with σ^2 (nonlinearity effect) and N_{g} (dimensionality effect), though at a much smaller rate. The bias $E[\tilde{J}_{\text{Ed}}(\mathbf{y})]$ is the value reached at $\sigma^2 = 0$ in agreement with (11).

3.3.1. Gaussian Subspace Recovery. We investigate the strength of $\tilde{J}_{\text{Ed}}(\mathbf{y})$ as a function of σ^2 and N_{g} . The null hypothesis H_G is rejected at the confidence level p_{sig} if $\tilde{J}_{\text{Ed}}(\mathbf{y}) > [P_{J_{\text{Ed}}}(N) N_t^{-1}] f_{\text{RED-}P_{\text{sig}}}$ using $f_{\text{RED-}P_{\text{sig}}}$ (Figure 1(b)). The Monte-Carlo fraction F of H_G rejection is shown in Figure 7(b) for $p_{\text{sig}} = 99\%$. The contour line at $F = 0.99$ provides the graph $\sigma_{\text{min}}^2(N_{\text{g}})$ of σ^2 for the recovery of the non-Gaussian sources. As expected, σ_{min}^2 increases with N_{g} , starting at $\sigma_{\text{min}}^2 = 20\%$ and at $N_{\text{g}} = 0$ and reaching $\sigma_{\text{min}}^2 \sim 45\%$ at $N_{\text{g}} = 6$. This means that the larger N , the stronger (through σ^2) the non-Gaussian embedded sources. Moreover, σ_{min}^2 diminishes under a less conservative test (lower p_{sig}) or when the effective number of temporal degrees of freedom increases. This puts a limit to what is recoverable from the available (often short) weather/climate data, and, therefore, less conservative tests of Gaussianity or data dimension reduction must be considered (e.g., $p_{\text{sig}} = 90\%$) [47].

3.3.2. ICA and ISA. Here, we apply ICA and ISA for $N = 11$ ($N_{\text{g}} = 6$) and compute the skill of source separation in two cases: (a) $\sigma^2 = 20\%$, high noise, and (b) $\sigma^2 = 40\%$, intermediate noise, yielding, respectively, total NE of 1.79 and 2.57.

Figures 8(a) and 8(b) are similar to Figure 6 but for the cases (a) and (b), respectively. The analysis yields the correct dimension in both cases: $N_{\text{ng-Ed}} = 5$ with $p_{\text{sig}} = 90\%$ despite the fact that the SVs in the “high noise” case (a) are smaller. The score of the non-Gaussian subspace recovery is assessed using the inner products $W_{\text{ng-Ed}} \cdot W_{\text{ng}}$ (compared to 5 in the ideal recovery). Those products are, respectively, 3.99 and 4.49 for cases (a) and (b). Lower values of σ^2 lead to a subestimation of N_{ng} and to a worse estimation of W_{ng} .

Now, by applying ICA (see Section 2.2.3 and Appendix A) we get ICs making the vector $\mathbf{z}_{\text{ICA}} = (\mathbf{R}'_{\text{ICA}} \mathbf{R}') \mathbf{y} = (Z_{1,\text{ICA}}, \dots, Z_{N_{\text{ng-Ed}},\text{ICA}})'$. Tables 3 and 4 give, respectively, for cases (a) and (b), a summary of ICA with the sorted maximized self-NEs $\tilde{J}_{\text{Ed}}(Z_{i,\text{ICA}})$, the NE components $\tilde{M}_{ii}(\mathbf{z}_{\text{ICA}})$, and the IC

TABLE 3: Contributions to the ICA contrast function (2nd column). NE components (3rd column) and IC vector loadings on the basis of original variables (18) for the high noise case. Components in which ICs are most projected are in bold font.

IC rank	$\tilde{J}_{\text{Ed}}(Z_{i,\text{ICA}})$	$\tilde{M}_{ii}(\mathbf{z}_{\text{ICA}})$	Y_1 (dyad)	Y_2 (dyad)	Y_3 (triad)	Y_4 (triad)	Y_5 (triad)
IC1: $i = 1$	0.127	0.162	0.76	-0.58	-0.02	0.01	0.16
IC2: $i = 2$	0.066	0.125	-0.07	0.07	0.47	-0.56	0.56
IC3: $i = 3$	0.042	0.102	-0.01	-0.08	-0.38	-0.68	-0.42
IC4: $i = 4$	0.014	0.053	-0.41	-0.52	0.21	0.29	0.01
IC5: $i = 5$	0.011	0.063	-0.03	0.02	-0.68	0.07	0.57
Total	0.260	0.506					

TABLE 4: Table 3 but for the case of intermediate noise (b).

IC rank	$\tilde{J}_{\text{Ed}}(Z_{i,\text{ICA}})$	$\tilde{M}_{ii}(\mathbf{z}_{\text{ICA}})$	Y_1 (dyad)	Y_2 (dyad)	Y_3 (triad)	Y_4 (triad)	Y_5 (triad)
IC1: $i = 1$	0.410	0.454	0.70	-0.70	-0.03	0.00	0.08
IC2: $i = 2$	0.256	0.372	0.00	0.05	0.45	0.65	0.58
IC3: $i = 3$	0.109	0.257	0.03	-0.03	-0.30	0.72	-0.55
IC4: $i = 4$	0.033	0.078	0.59	0.59	-0.17	-0.07	0.00
IC5: $i = 5$	0.031	0.114	-0.04	0.04	-0.76	0.11	0.54
Total	0.840	1.277

(a)

(b)

FIGURE 7: (a) Monte-Carlo average of the sample negentropy estimator $\tilde{J}_{\text{Ed}}(\mathbf{y})$ as a function of σ^2 and N_g . (b) Contour plot of the fraction F of times for which the null hypothesis H_G is rejected at $p_{\text{sig}} = 99\%$ (1% significance level) as a function of σ^2 and N_g . The contour $F = 0.99$ can mark σ_{min}^2 at the 99% confidence level.

vector loadings in $\mathbf{RR}_{\text{ICA}}$. A few conclusions can be drawn from these tables; (i) the sum of self-NEs of ICs (2nd column) cannot explain the total NE (3rd column). This corroborates the existence of NE originating from IC interactivities and (ii) Up to a certain statistical significance, a given IC projects uniquely onto a certain source as seen through the dominant projections of ICs (in bold font). From these tables, the estimated dyad $\mathbf{z}_{\text{Ed-dyad}}$ is spanned by IC1 and IC4 whereas the estimated triad $\mathbf{z}_{\text{Ed-triad}}$ is spanned by IC2, IC3, and IC5. That supports the conjecture of [13] about the source emerging by IC grouping. Recovered sources may be rotated or inverted

with respect to initial data. The distribution of ICs among sources may differ for smaller σ^2 . The squared projections onto the true sources are, respectively, 1.35 (on a maximum of 2) and 2.42 (on a maximum of 3) in case (a) and 1.68 and 2.75 in case (b). Like the non-Gaussian subspace, the recovery of individual sources is more skillful in case (b).

The final step consists of merging the ICs as outlined in Section 2.2.3. Figure 9 presents the results for both cases, where the self-interactivities (16) are the most relevant ITIs due to their maximization by ICA (except IC5 in case (a)). The cumulated NE reaches the unbiased NE value (dashed

FIGURE 8: Same as Figure 6 but for the for synthetic data with $\sigma^2 = 20\%$ (a) and $\sigma^2 = 40\%$ (b).

line) near the threshold of the ITI acceptance (scaled-ITI > 1). In case (a) (Figure 9(a)), the accepted ITIs are for ICs 1–5, and dyadic links (IC3, IC2), (IC5, IC3) yielding the emergence of the true triad (IC2, IC3, and IC5) (see Table 3). The marginally accepted ITIs holds for (IC4, IC3, and IC2) (false triad) and (IC1, IC4) (true dyad). Excluded ITIs remain at the level of the NE bias. For smaller values of σ^2 (not shown), the recovered sources may be mixed, partly restored, or even not detected.

For case (b) (Figure 9(b)) the source recovery improves. In fact, the significant accepted ITIs are for ICs 1–5, the links for (IC5, IC3), (IC5, IC2), (IC3, IC2) with the true emerging triad (IC2, IC3, and IC5), and for the true dyad (IC4, IC1). The sum of ITIs associated with ICA and ISA is, respectively, 0.84 and 1.13, showing clearly the advantage of ISA in this case.

3.4. Application of ICA and ISA to SSTA Data. Using the estimated non-Gaussian subspace (Section 3.2), we apply here ICA to the leading $N_{ICA} = 5$ SVecs of the SSTA dataset (see Figure 6). Table 5 gives a summary of the application with the characteristics of the ranked ICs $Z_{i,ICA}$, $i = 1, \dots, 5$, by decreasing self-NEs. They explain, respectively, 7.1%, 5.5%, 4.3%, 4.1%, and 3.7% of the total variance.

Most of the NE components come from the self-NEs (an effect of ICA), through their high skewness and positive kurtosis or leptokurtosis (Table 5), leading to fat long pdf tails. Extreme values of about 3–5 standard deviations are often reached as seen by the IC time-series (Figure 10).

Under those conditions, ICs are basically aggregative and cooperative combinations of single and joint PC extremes. Quasi-independence and noncorrelation between ICs means that occurrence of extremes among different ICs tends to be asynchronous (see Figure 10) and therefore extreme phenomena are independent.

Figure 11 shows the correlation maps between the ICs and SSTAs. An alternative way to interpret ICs is through the patterns displayed in the corresponding normalized loading maps (Figure 12) obtained from combinations of EOFs (Figure 3). Correlation and loading map values at the grid point P for the i th IC are weighted by the PCs standard deviation and its inverse, respectively, $\sum_{k=1}^{N=11} D_{k,P} \mathbf{e}_k(P) (\mathbf{R}\mathbf{R}_{ICA})_{k,i}$ where $D_{k,P}$ is $\hat{\lambda}_k^{1/2} / \hat{\sigma}_P$ (Figure 11) and $\hat{\lambda}_k^{-1/2}$ (Figure 12), respectively, and $\mathbf{e}_k(P)$ is the k th EOF at P (Figure 3). By definition, ICs are spatial inner products between correspondent normalized loading maps (Figure 12) and the SSTA field. Consequently, maximal positive (negative) ICs on a given month are favored by highly positive (negative) pattern correlations between their loading maps and the SSTA field. Note, however, that the patterns shown in Figures 11 and 12 are not orthogonal. On loading maps, the weighting gives rise to the impact on ICs of PCs of small explained variance. Because of this weighting, the correlation maps (Figure 11) often display much broader and large-scale patterns than loading maps (Figure 12) [15], though both share the same main features when the range of PC variances contributing to a given IC is not too wide, which is the case here.

The leading IC1 explains about 2/3 of the total NE with the remaining 4 being almost evenly distributed (Table 5). Further experiments show that IC1 is quite robust with the use of $N_{ICA} < 5$ leading SVecs with the remaining ICs being slight mixtures of those obtained with $N_{ICA} = 5$. The above mixture of PCs suggests the existence of a global nonseparable source thanks to the nonlinear interaction across scales. This strongly suggests that ICs, obtained as blind source separation, are convenient tools for identifying nonlinear variability. Dynamical constraints may be included explicitly, for example, via nonlinear dynamical modes [71, 72].

The leading IC1 reaches values of +5 standard deviations in the years 1983 and 1997 of strong El Niños, during which PC1, PC3, PC4, and PC11 are synchronised (see also Figures 5(a) and 5(b) and Table 5). The coherence between those four PCs may sometimes be partial.

The correlation pattern for IC1 (Figure 11) has an El Niño pattern shape due to the high PC1 amplitude (Table 5) and PC1 variance whereas the loading map (Figure 12) is much correlated with EOF11, showing more local features. Therefore, from Figure 12 and Table 5, IC1 is maximal under El Niño conditions, positive NPGO (colder central North Pacific), slightly negative PDO, warmer (colder) mid-latitude (subtropical) Atlantic region, a dipolar structure in the South Pacific, and cold pools southward of 30°S in the Indian and Atlantic oceans. That is consistent with the oceanic influence of ENSO on decadal timescales [73]. Moreover, that agrees with the dynamical interactions between ENSO, PDO, and NPGO which have been discussed by several authors [74–76]. The quite relevant EOF11 pattern in IC1 is mostly correlated

FIGURE 9: Sorted scaled-interactivities by decreasing order (solid circles), ITIs (solid squares) with cumulated ITIs (open crossed circles) and unbiased total NE (horizontal dashed line). The sorted IC groups are labeled (e.g., 32 means group of IC2 and IC3). The 95% confidence levels of the selected ITIs are those for which the cumulated ITIs are below the dashed line. (a) Case (a): high noise and (b) case (b): intermediate noise.

FIGURE 10: IC time-series for the period 1910–2011.

with the El Niño SST composite (asymmetric with respect to La Niña) at mid-latitudes in all oceans, during the December-February season [68, 77], which explains the above referred coherence of PC1 and PC11 during strong El Niños.

The IC2 loading map (Figure 12) mainly coincides with the symmetrical EOF6, that is, with the positive Atlantic Niño phase, while the remaining dominant EOF contributors to IC2, that is, EOF1, EOF7, and EOF9 (Table 5), practically cancel each other, especially in tropical Pacific El Niño region, despite the presence of non-null correlation in that region (Figure 11). Negative IC2 extremes (e.g., in 1912 and 1951) coincide with the strongest Atlantic Niñas in agreement with the positive skewness of PC6 (Table 2).

The IC3 loading map pattern (Figure 12) is quite similar to EOF5 (Central Pacific (CP) El Niño), combined with the NPGO (colder North Pacific for positive IC3). That is also visible in the correlation map (Figure 11). The negative skewness of IC3 (Table 5) is compatible with the primarily negative SST skewness in Central Pacific [78]. It supports the existence of strong Central Pacific La Niñas (strong negative values of IC3), as in 1943-1944 (see Figure 10), especially during the positive phase of AMO (see the negative loading factor of PC2 for IC3 in Table 5). The loading factors of PC4, PC5, and PC2 contributing to IC3 (Table 5) also reflect the tendency for CP-El Niño (CP-La Niña) during negative (positive) phases of AMO [79], which is visible through the symmetrical correlation signals of IC3 with SSTA (Figure 11), respectively, in CP and North Atlantic areas. The imprint of the characteristic South Pacific dipole shown in EOF10 is also visible in the loading map (Figure 12).

The IC4 loading map (Figure 12) is a mix of several EOFs. The inverse proportionality of IC4 to the PCs standard deviation makes EOFs 4, 5, 9, 10, and 11 the strongest contributors (Table 5 and Figure 2(a)) to the loading map of IC4 (Figure 12), showing that IC4 is highly positive when the Indian and Atlantic oceans are warm, while the eastern (western) North and South Pacific are warm (cold), including the tropical El Niño region. That is also reflected in the correlation map (Figure 11).

The last IC5 loading map pattern (Figure 12) is nearly symmetrical to EOF10 (Table 5), especially over the Austral oceans with a characteristic sequence of three dipoles, located over the Indian, Pacific, and Atlantic subtropical oceans. Negative extremes of IC5 (e.g., 1947, 1992/93) are also associated with a strong negative AMO and NPGO. The NE of IC5 result essentially from the highly positive kurtosis and long pdf tails

FIGURE 11: Correlation maps between ICs and SSTA field.

FIGURE 12: Spatially normalized loading maps of the 5 sorted ICs.

of PC10 and PC2 (Table 2). Both the loading and correlation maps are quite similar.

The results of ISA are summarized in Figure 13. They suggest 5 ICs with the emergence of one dyadic link (IC3, IC5) with $ITI = 0.015$ and a triadic link (IC1, IC4, and IC5) with $ITI = 0.024$ and finally the strongest dyad (IC1, IC5) with an $ITI = 0.045$. This interaction is even larger than the self-NEs of ICs 3, 4, and 5. This interaction is mainly due to the nonlinear correlation: $\text{cor}(\text{IC1}, \text{IC5}^2) = 0.23$, that is, extremes of IC5 occurring preferentially under $\text{PC1} > 0$ (e.g., strong

El Niños). A possible explanation is that PC4 (simulating NPGO) influence both IC1 and IC5. From the ensemble of interactions, only IC2 appears to be significantly independent of the remaining ICs. All the remaining ITIs remain at the noise level (Figure 13).

The above analysis extracts highly negentropic ICs resulting primarily from independent combinations and coherency between PC extremes. An alternative ICA and ISA analysis is obtained with Gaussianized (followed by whitening) PCs. Further experiments (not shown) indicate that the total NE

FIGURE 13: As in Figure 9 but for the ISA summary of SSTA data.

comes from cross cumulants only and ICs emerge exclusively from combinations of joint PC extremes. The two leading ICs are practically unchanged as compared to previous ICs; the third and fourth ICs are slightly mixed while the fifth IC is not significant, that is, remaining in the Gaussian manifold. In summary, a prior Gaussianization of PCs leads in general to a reduction, both of the dimension and of the strength of the non-Gaussian subspace.

3.5. Sensitivity to the PC Spanning Space. Here, we study the behavior of the non-Gaussian subspace and ICA/ISA outputs when we sequentially increase the metric space dimension $N_{\max}$ using the leading whitened PCs. In the discussion below, it is worth mentioning that, given the direct sum: $S_+ = S_- \oplus S_r$ between two orthogonal subspaces S_- , S_r , the relationships between the associated non-Gaussian subspaces and their dimensions hold:

$$\begin{aligned} W_{\text{ng}}(S_+) &\supseteq W_{\text{ng}}(S_-) \oplus W_{\text{ng}}(S_r); \\ N_{\text{ng}}(S_+) &\geq N_{\text{ng}}(S_-) + N_{\text{ng}}(S_r), \end{aligned} \quad (19)$$

which leads to the identity

$$W_{\text{ng}}(S_+) \cdot W_{\text{ng}}(S_-) = N_{\text{ng}}(S_-). \quad (20)$$

Hereafter, we focus on the case where S_- , S_r are spanned, respectively, by the leading $N_{\max} - N'$ PCs and by the remaining $N' \geq 1$ PCs up to rank $N_{\max}$. In (19), the equality (19) holds when S_- , S_r are statistically independent or equivalent when spaces S_- , S_r have separable pdfs. Moreover, under statistical independence, the ISA/ICA may be done independently. However, the pdf separability is hardly verified on multiscale nonlinear natural systems as studied here which means that $W_{\text{ng}}(S_+)$ intersects the Gaussian subspaces $W_{\text{g}}(S_-)$, $W_{\text{g}}(S_r)$ of the parcel spaces while ICA/ISA is also changing when passing from S_- to S_+ .

An equivalent relationship to (19) for $W_{\text{ng-Ed}}$, obtained with finite samples, is much harder to obtain since $W_{\text{ng-Ed}}$

may be undersized (oversized) by errors of type I (II) when considering the Gaussianity null hypothesis H_G . This means that $W_{\text{g-Ed}}(S_+) \cdot W_{\text{ng-Ed}}(S_-) \geq 0$; that is, there may be some “leakage” of the “prior” non-Gaussian subspace ($W_{\text{ng-Ed}}(S_-)$), especially through the less non-Gaussian ICs, to the “posterior” Gaussian subspace $W_{\text{g-Ed}}(S_+)$.

Next, we analyze how the ICA/ISA applied to SSTAs varies with $N_{\max}$. To this end, we have considered the sequence $N_{\max} = 3, 9, 11, 15, 17$ in which the number of ICs $N_{\text{ng-Ed}}$, obtained by the method of Section 2.2.2 using a confidence level $p_{\text{sig}} = 90\%$ (see, e.g., Figure 6), is $N_{\text{ng-Ed}} = 1, 4, 5, 6, 7$, respectively. The above $N_{\max}$ are those for which $N_{\text{ng-Ed}}$ have shown a jump (of one or more dimensions) with respect to that obtained for $N_{\max} - 1$, where more substantial changes are expected to occur in the ICA/ISA sources. For example, $N_{\text{ng-Ed}} = 5$ for $11 \leq N_{\max} \leq 14$. Let us consider the corresponding sequence of embedding spaces: S_- , S_+ , denoting by $\mathbf{r}_1^-, \dots, \mathbf{r}_{N_-}^-$ and $\mathbf{r}_1^+, \dots, \mathbf{r}_{N_+}^+$ the IC weighting vectors (sorted by decreasing self non-Gaussianity), spanning $W_{\text{ng-Ed}}(S_-)$ and $W_{\text{ng-Ed}}(S_+)$, respectively.

To assess the impact on ICA/ISA, we have computed some inner-product-based diagnostics, plotted in Figure 14. First, the measure $P_{-+} \equiv W_{\text{ng-Ed}}(S_+) \cdot W_{\text{ng-Ed}}(S_-) = \sum_{i=1}^{N_-} \sum_{j=1}^{N_+} (\mathbf{r}_i^- \cdot \mathbf{r}_j^+)$ while the difference $P_{-+} - N_- \geq 0$ yields how the non-Gaussian subspaces are preserved throughout the space sequence, that is, how closely (20) is preserved. Then, we identify which terms $0 \leq \mathbf{r}_i^- \cdot \mathbf{r}_j^+ \leq 1$ are mostly contributing to that, typically above a threshold (e.g., 0.08). For dimensions within the range of the selected $N_{\max}$ where $N_{\text{ng-Ed}}$ is kept constant, every $\mathbf{r}_i^-$ is very well projected onto a certain $\mathbf{r}_{j(i)}^+$, leading to well tracked and robust ICs independently of the space dimension. For instance, in the sequence $N_{\max} = 11, 12, 13, 14$ ($N_{\text{ng-Ed}} = 5$), we get $P_{-+} = 4.7, 4.6, 4.8$, respectively, (quite close to the ideal value of 5) with very low leakage of non-Gaussianity. We also get the tracks of the five ICs: $[\text{IC1} \rightarrow_{0.96} \text{IC1} \rightarrow_{0.91} \text{IC1} \rightarrow_{0.94} \text{IC1}]$; $[\text{IC2} \rightarrow_{0.99} \text{IC3} \rightarrow_{0.98} \text{IC3} \rightarrow_{0.95} \text{IC3}]$; $[\text{IC3} \rightarrow_{0.83} \text{IC2} \rightarrow_{0.77} \text{IC2} \rightarrow_{0.99} \text{IC2}]$; $[\text{IC4} \rightarrow_{0.97} \text{IC4} \rightarrow_{0.98} \text{IC4} \rightarrow_{0.97} \text{IC4}]$; and $[\text{IC5} \rightarrow_{0.90} \text{IC5} \rightarrow_{0.88} \text{IC5} \rightarrow_{0.91} \text{IC5}]$, each one within brackets, where the subscript values stand for $\mathbf{r}_i^- \cdot \mathbf{r}_{j(i)}^+$. We conclude that ICA and ISA are quite robust in the range $N_{\max} = 11-14$. Only the ranks of IC2 and IC3 were swapped in the step $N_{\max} = 11 \rightarrow 12$.

Now, in Figure 14, we analyze what happens when a highly non-Gaussian or highly nonlinearly interacting PC comes into the analysis space, corresponding, for instance, to PC3, PC9, PC11, PC15, and PC17. EOF15 and EOF17 are both dominated by subtropical modes in the South Pacific and Indian oceans (not shown), nonlinearly interacting with precedent modes.

In every sequence $S_- \rightarrow S_+$, the ICs may experience three situations: (1) the emergence of new ICs; (2) the “leakage” of non-Gaussianity to the forthcoming Gaussian subspace; and (3) the decoupling of sets of ICs (one or more) into sets of ICs of larger dimensions. In our case, the emergence of 3 new ICs appears when $N_{\max}$ changes from 3 to 9 (see Figure 14). The “leakage” is more relevant with weaker ICs (low negentropy). The decoupling of ICs becomes frequent,

FIGURE 14: Schematic of the ICs (rank and negentropy shown in rectangles), aligned vertically for each $N_{\max} = 3, 9, 11, 15, 17$ (circles) with values of P_{+} (bulk arrows) between the sequential PC spanning spaces. Values with the dominant $\mathbf{r}_i \cdot \mathbf{r}_j^+$ are written next to the corresponding thin arrows. The red straight vertical lines indicate the significant dyadic links with corresponding NE interactivities.

mainly at higher dimensions, being the main mechanism of IC generation. For instance IC3 at $N_{\max} = 9$ decouples into (IC4, IC5) with $N_{\max} = 11$. Multiple interactivities, like the dyadic ones (expressed by red segments in Figure 14), may emerge on incoming ICs or transfer to the tracked ICs (e.g., the dyad (IC1, IC6) with $N_{\max} = 15$ transfers to the dyad (IC1, IC7) with $N_{\max} = 17$). There are some ICs that remain robust without any decoupling like the case of IC1 and IC3 with $N_{\max} = 11$. In general, as $N_{\max}$ increases, ICs project progressively onto low-ranked PCs dominated by smaller spatial scales and complex sharper features. In this situation, a hybrid criterion for the choice of the truncation $N_{\max}$ (not explored here) may be favored by four factors: the negentropy and explained variance by ICs, the low dispersion of IC weighting loadings (either on PC or original basis), and a low $N_{\text{ng-Ed}}$. That justifies the choice $N_{\max} = 11$ as being the lowest dimension for which $N_{\text{ng-Ed}} = 4$.

4. Summary and Conclusion

We present in this paper an approximation of the multivariate negentropy (NE) motivated by the Edgeworth expansion of the joint probability distribution, using a linear combination of Frobenius norms of the joint cumulant tensors of order $p \geq 3$. This form is especially sensitive to the extent and frequency of extremes. The NE approximation is based on the trace of a positive definite matrix $\mathbf{M}$ resulting from the high-order singular value decomposition of an appropriate contracted form of the coskewness and cokurtosis joint tensors in consistency with the exact negentropy. First, it is a tensor invariant, that is, invariant under orthogonal rotations. Second, when variables are whitened, a version of the “Separation Source Theorem” holds; that is, the approximated NE may be explicitly decomposed into positive terms measuring self-NEs (due to skewness/kurtosis of marginals) and NEs

resulting from nonlinear associations between variables. This permits the extraction of statistically independent components (ICs) from observations through the maximization of the sum of marginal NEs. The decomposition is further generalized to multivariate sources under the Independent Subspace Analysis (ISA), through the formation of groups of disjoint ICs, associated with positive NE contributors (or interactivities). The jointly Gaussian ICs span the Gaussian subspace, being simply approximated by the kernel subspace of $\mathbf{M}$.

When cumulants are estimated from finite or serially correlated samples, the NE estimator bias is positive and grows with the dimension (Curse of Dimensionality). Therefore, the estimated NE of a Gaussian distributed finite multivariate sample is prone to produce “false non-Gaussianity,” hence the importance of Gaussian subspace identification. This subspace is estimated from the trailing singular vectors of $\mathbf{M}$ for which the null hypothesis of Gaussianity cannot be rejected. The search of non-Gaussian ICs or multiple sources is sought in the non-Gaussian subspace.

The method is then applied to real data consisting of 11 leading PCs obtained from 102 years of the SST anomaly (SSTA) field over the global ocean, equatorward of 65° . To test the effectiveness of the method, prior to using the SSTA PCs, we have used synthetic data with a prescribed amount non-Gaussianity and sharing the same dimensions, sample size, lagged-one autocorrelation, and the level of negentropy as the SSTA PCs. In order to assure the recovery of the “true non-Gaussianity,” the experiment suggests a minimum of variance of the true non-Gaussian sources, which increases as the sample gets smaller and the dimension of the embedding space gets larger.

Application of the method to SSTA data based on the leading 12 PCs (the last one going practically to the Gaussian subspace), explaining more than 50% of the monthly

variance, yields 5 ICs. The most prominent or most negative IC (or IC1) represents El Niño conditions combined with negative PDO, positive NPGO, and composite impacts of El Niño (asymmetric with respect to La Niña) on oceanic mid-latitudes, poleward of 30°. The second component, IC2, reveals a loading map mostly projecting onto the Atlantic Niño and the AMO pattern. The third component, IC3, reflects the tendency of Central Pacific El Niños during the negative phase of AMO. The fourth component, IC4, combines several EOFs and the last component, IC5, is dominated by EOF10, being positive when the SSTA field in the South Pacific exhibits a zonally oriented dipole. All ICs are basically independent combinations of PC extremes and suggest independent responses of the SST field to different forcings. Only IC2 appears to be significantly isolated from the other components. The remaining components have some statistical links dominated primarily by the dyad (IC1, IC5), possibly associated with the sharing of the NPGO effect, via PC4, among those ICs.

The identification of possible independent physical forcings, responsible for the independent sources may bring some benefits, concerning, namely, the forecasting and modelling of the SSTA variability and its impacts and teleconnections on relevant surface variables (e.g., precipitation, temperature, moisture, and wind). That opens the possibility of building phenomenological, dynamical, and stochastic models for the different ICs and independent subspaces.

Appendix

A. Method of Estimating the ICA Solution

The ICA looks for the maximum of a positive defined objective function $F_{ICA}(\mathbf{R}) = \sum_{i=1}^{N_{ng-Ed}} J_{Ed}(z_i); \mathbf{z} = \mathbf{R}\mathbf{y}$. In general, for a continuous and differentiable objective function $F_{ICA}(\mathbf{R})$, the absolute maximum $\mathbf{R}_{ICA} \equiv \arg \max_{\mathbf{R}} \{F_{ICA}(\mathbf{R})\}$ holds at a null gradient $\partial F_{ICA}(\mathbf{R})/\partial \mathbf{R}$, subjected to the orthogonality constraint $\mathbf{R}\mathbf{R}' = \mathbf{I}_{N_{ng}}$. If $\mathbf{\Lambda}$ is the matrix of the Lagrange multipliers, then we easily get

$$\frac{\partial F_{ICA}(\mathbf{R})}{\partial \mathbf{R}} = \frac{1}{2} \mathbf{\Lambda} \odot \frac{\partial (\mathbf{R}\mathbf{R}')}{\partial \mathbf{R}} = \mathbf{R}\mathbf{\Lambda}_s; \quad (\text{A.1})$$

$$\mathbf{\Lambda}_s = \frac{1}{2} (\mathbf{\Lambda} + \mathbf{\Lambda}'),$$

where $\odot$ is the component-wise multiplication operator. Eq. (A.1) means that the gradient matrix coincides with its matrix polar decomposition, that is, the product of an orthogonal matrix by a symmetric matrix [80]. For an invertible matrix $\mathbf{\Lambda}$ there may exist a finite set of solutions $\mathbf{R}$ of (A.1), beyond those obtained by permutations. They correspond to suboptimal ICA solutions. However, for a singular matrix $\mathbf{\Lambda}$, there is a continuous manifold of solutions spanned along $\text{Ker}(\mathbf{\Lambda})$. In the maximization ICA problem, the gradient matrix depends nonlinearly on $\mathbf{R}$, hence the linear eigentechniques do not work. A usual approach is to write $\mathbf{R}$ as a product of Givens matrices in terms of rotation Euler angles [81] and find the absolute maximum by a gradient-based Newton method [15].

A different approach is used here. The solutions of (A.1) are obtained from an iterative algorithm. We start from a random orthogonal matrix $\mathbf{R}_1$ and compute its gradient matrix and then its SVD, necessary to polar decomposition. Then we have

$$\frac{\partial F_{ICA}(\mathbf{R}_k)}{\partial \mathbf{R}_k} = \mathbf{W}_k \mathbf{\Sigma}_k \mathbf{V}_k' = \mathbf{R}_{k+1} \mathbf{\Lambda}_{s,k+1} \quad (k \geq 1), \quad (\text{A.2})$$

where $\mathbf{R}_{k+1} = \mathbf{W}_k \mathbf{V}_k'$ and $\mathbf{\Lambda}_{s,k+1} = \mathbf{V}_k \mathbf{\Sigma}_k \mathbf{V}_k'$. The iteration stops when the relative Frobenius squared norm of the difference between consecutive rotation matrices $(1/N_{ng}) \|\mathbf{R}_{k+1} - \mathbf{R}_k\|_F^2$ is less than a threshold, δ , say 10^{-5} (note that $\|\mathbf{R}\|_F^2 = N_{ng}$). The random initial matrix is the product of Givens matrices by random Euler angles. From the trials (~ 1000), only the convergent sequences are retained and finally their limit matrices. The matrix producing the largest objective function F_{ICA} is set to $\mathbf{R}_{ICA}$. A finite nonredundant number of different local minima are normally found in the ICA optimization.

The gradient $\partial F_{ICA}(\mathbf{R})/\partial \mathbf{R}$ is written through the chain rule; F_{ICA} is derived with respect firstly to cumulants, secondly to rotated variables, and thirdly to rotation matrix entries. More generally F_{ICA} belongs to the of objective function $F(\text{cm}_1, \dots, \text{cm}_i, \dots)$, depending on cumulants (here a sum of weighted squares): cm_i . Each cumulant is a linear combination of sampling averages: $\text{cm}_i = \sum_j a_{ij} [\overline{f_{ij}(\mathbf{z})}]$. Therefore:

$$\frac{\partial F}{\partial R_{pq}} = \sum_i \frac{\partial F}{\partial \text{cm}_i} \sum_j a_{ij} \left(\frac{\partial f_{ij}}{\partial z_q} z_p \right). \quad (\text{A.3})$$

B. Similarity Score between Sources

Let us consider two sources $\mathbf{f} = (\mathbf{f}_1, \dots, \mathbf{f}_{N_1})$, $\mathbf{t} = (\mathbf{t}_1, \dots, \mathbf{t}_{N_2})$, of dimensions N_1, N_2 not necessarily equal (e.g., the proposed one by ISA and the true one), where $\mathbf{f}_i, \mathbf{t}_j$ are orthogonal basis vectors spanning each source. A tensorial invariant measure of similarity between sources is the inner product between subspaces, hereafter denoted as $\mathbf{f} \cdot \mathbf{t}$. That is given by the sum of squared SVs of the Gram matrix $\mathbf{G} \equiv \mathbf{f}\mathbf{t}'$, that is,

$$\mathbf{f} \cdot \mathbf{t} = \text{Tr} [\mathbf{G}\mathbf{G}'] = \|\mathbf{G}\|_F^2 \in [0, \min \{N_1, N_2\}]. \quad (\text{B.1})$$

Symbols

- $\hat{c}_1$: Lag-one autocorrelation of the fitted AR(1) model
- C_b^a : Number of combinations without repetition of b among a elements
- $\text{cm}(\mathbf{u})$: Generic cumulant in terms of the vector $\mathbf{u}$
- $E()$: Expectation operator
- $\mathbf{e}_k(P)$: Value of the normalized k th EOF at grid point P

$F_{ICA}(\mathbf{R}), F_{ISA}(\mathbf{R})$:	Contrast function for ICA and ISA as a function of $\mathbf{R}$
f_{RED} :	Ratio between actual and asymptotic bias of $J_{Ed:G}$
$f_{RED-P_{sig}}$:	Quantile p_{sig} of $\tilde{J}_{Ed:G}$ in units of $P_{J_{Ed}}(N)N_t^{-1}$
H_G, H_{G_i} :	Global and partial Gaussianity hypothesis
$I(\mathbf{u}, \mathbf{v}, \dots)$:	Mutual information among vectors $\mathbf{u}, \mathbf{v}, \dots$
$I_{Ed}(\mathbf{u}, \mathbf{v}, \dots)$:	Edgeworth approximation of $I(\mathbf{u}, \mathbf{v}, \dots)$
$J()$:	Negentropy
$J_{Ed}()$:	Estimated negentropy by the Edgeworth cumulant-based approximation
$J_{Ed:G}$:	Value of J_{Ed} under the Gaussianity hypothesis
$J_{Ed-IT}(\mathbf{u})$:	Part (interactivity parcel) of J_{Ed} associated with the subvector $\mathbf{u}$
$J_{Ed-ITsig}(\mathbf{u})$:	Scaled value of J_{Ed-IT} by quantiles of $J_{Ed-IT:G}(\mathbf{u})$
J_{rot} :	Invariant negentropy of a subspace
J_{sig} :	Threshold for the acceptance of a scaled interactivity $J_{Ed-ITsig}(\mathbf{u})$
$\mathbf{K}(\mathbf{y})$:	Cokurtosis tensor of $\mathbf{y}$
$\mathbf{M}(\mathbf{y})$:	Matrix obtained by contraction products of $\mathbf{S}(\mathbf{y})$ and $\mathbf{K}(\mathbf{y})$
N, N_{max} :	Dimension of $\mathbf{y}$ and its maximum value under study
N_{dof} :	Number of temporal degrees of freedom
N_{eq} :	Number of equivalent temporal degrees of freedom
N_g, N_{ng} :	Dimension of the Gaussian and non-Gaussian subspaces
N_{g-Ed}, N_{ng-Ed} :	Dimension of the Gaussian and non-Gaussian estimated subspaces
N_{int} :	Number of interactivities forming J_{Ed}
N_P :	Number of domain points
N_r :	Rank of the dataset matrix
N_s :	Number of subvectors or candidate sources
N_t :	Time-series size
p, p_{max} :	Cumulant order and maximum cumulant order
$P_{Ed}(\mathbf{y})$:	Polynomial correction of the $\mathbf{y}$ pdf by the Edgeworth approximation
$P_{J_{Ed}}(N)$:	Polynomial dependence on N of the $J_{Ed:G}$ bias
p_{sig} :	Confidence level
$Q_{p_{sig}}()$:	Quantile at the p_{sig} confidence level, used to establish thresholds
$\mathbf{R}$:	Orthogonal rotation matrix filled by singular vectors of $\mathbf{M}(\mathbf{y})$ (SVecs)
$\mathbf{R}_{ICA}$:	Orthogonal matrix relating the SVecs and ICA vectors
$\mathbf{R}_{ISA}$:	Orthogonal matrix relating the SVecs and ISA sources
$\mathbf{r}_1, \mathbf{r}_2, \dots$:	Normalized vectors filling $\mathbf{R}$ (Singular Vectors SVecs)

$\mathbf{S}(\mathbf{y})$:	Coskewness tensor of $\mathbf{y}$
$\text{Span}(\mathbf{W})$:	Subspace spanned by orthonormal column vectors in $\mathbf{W}$
$G(t)$:	Single standard Gaussian white noise at time t
W_g, W_{ng} :	Gaussian and non-Gaussian subspaces
W_{g-Ed}, W_{ng-Ed} :	Estimated Gaussian and non-Gaussian subspaces by HOSVD of $\mathbf{M}$
$\mathbf{X}$:	Space-time dataset matrix
$\mathbf{x}(t)$:	Vector of local field values at time t
$\hat{\mathbf{x}}$:	Vector of local standardized field values
$\mathbf{x}_{PC}$:	Vector of PCs at time t , sorted by decreasing variance
$\mathbf{y}$:	Vector of whitened variables (e.g., whitened PCs)
$\mathbf{z}$:	Vector of rotated whitened PCs through SVD
$\mathbf{z}_{ng}$:	$\mathbf{z}$ restricted to the non-Gaussian components
$\mathbf{z}_{(i)}$:	Subvector of components spanning the i th candidate source
$\mathbf{z}_{ICA}, \mathbf{z}_{ISA}$:	Vector of rotated whitened PCs giving the ICA and ISA solutions
$\eta_{i,N}$:	i th singular value of $\mathbf{M}(\mathbf{y})$ of dimension N
λ_k :	k th PC variance sorted by decreasing order
ρ_z :	Joint pdf of the vector $\mathbf{z}$
$\varphi(\mathbf{z})$:	Standard isotropic Gaussian PDF as a function of joint pdf of $\mathbf{z}$
σ^2 :	Variance of the non-Gaussian signal part of a scalar
σ_{local}^2 :	Average local variance
$\hat{\sigma}_P$:	Ratio between local variance at point P and total variance
σ_{Total}^2 :	Total variance
$\overline{()}$:	Quantity estimated from sample averages
$\langle \rangle$:	Sample or time average
$\dim$:	Dimension or cardinality of a vector
$\ \cdot\ _F^2$:	Frobenius squared norm (sum of the entries squares)
$\langle \rangle_G$:	Quantity evaluated under the Gaussianity hypothesis
$\langle \rangle'$:	Transpose
$\langle \rangle \odot \langle \rangle$:	Component-wise product.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was developed at IDL with the support of the Portuguese Foundation for Science and Technology through the FCT, project UID/GEO/50019/2013, Instituto Dom Luiz (IDL). The authors acknowledge the International Meteorological Institute (IMI) at Stockholm University for hosting Carlos Pires and the omnipresent support of the family.

References

- [1] P. Sura and A. Hannachi, "Perspectives of non-gaussianity in atmospheric synoptic and low-frequency variability," *Journal of Climate*, vol. 28, no. 13, pp. 5091–5114, 2015.
- [2] C. Proistosescu, A. Rhines, and P. Huybers, "Identification and interpretation of nonnormality in atmospheric time series," *Geophysical Research Letters*, vol. 43, no. 10, pp. 5425–5434, 2016.
- [3] A. Hannachi, D. M. Straus, C. L. Franzke, S. Corti, and T. Woollings, "Low-frequency nonlinearity and regime behavior in the Northern Hemisphere extra-tropical atmosphere," *Reviews of Geophysics*, vol. 55, no. 1, pp. 199–234, 2017.
- [4] A. Hannachi, I. T. Jolliffe, and D. B. Stephenson, "Empirical orthogonal functions and related techniques in atmospheric science: a review," *International Journal of Climatology*, vol. 27, no. 9, pp. 1119–1152, 2007.
- [5] A. Hannachi, S. Unkel, N. T. Trendafilov, and I. T. Jolliffe, "Independent component analysis of climate data: a new look at EOF rotation," *Journal of Climate*, vol. 22, no. 11, pp. 2797–2812, 2009.
- [6] N. Philippon, L. Jarlan, N. Martiny, P. Camberlin, and E. Mougin, "Characterization of the interannual and intraseasonal variability of West African vegetation between 1982 and 2002 by means of NOAA AVHRR NDVI data," *Journal of Climate*, vol. 20, no. 7, pp. 1202–1218, 2007.
- [7] A. Mori, N. Kawasaki, K. Yamazaki, M. Honda, and H. Nakamura, "A reexamination of the northern hemisphere sea level pressure variability by the independent component analysis," *Scientific Online Letters on the Atmosphere*, vol. 2, pp. 5–8, 2006.
- [8] F. Aires, A. Chédin, and J.-P. Nadal, "Independent component analysis of multivariate time series: application to the tropical SST variability," *Journal of Geophysical Research Atmospheres*, vol. 105, no. 13, Article ID 2000JD900152, pp. 17437–17455, 2000.
- [9] I. Fodor and I. C. Kamath, "On the use of independent component analysis to separate meaningful sources in global temperature series," Tech. Rep., Lawrence Livermore National Laboratory, 2003, https://computation.llnl.gov/casc/sapphire/sep_climate/index.html.
- [10] S. Westra, C. Brown, U. Lall, I. Koch, and A. Sharma, "Interpreting variability in global SST data using independent component analysis and principal component analysis," *International Journal of Climatology*, vol. 30, no. 3, pp. 333–346, 2010.
- [11] J. Cardoso, "Multidimensional independent component analysis," in *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, pp. 1941–1944, Seattle, WA, USA.
- [12] A. Sharma and K. K. Paliwal, "Subspace independent component analysis using vector kurtosis," *Pattern Recognition*, vol. 39, no. 11, pp. 2227–2232, 2006.
- [13] Z. Szabó, B. Póczos, and A. Lrincz, "Separation theorem for independent subspace analysis and its consequences," *Pattern Recognition*, vol. 45, no. 4, pp. 1782–1791, 2012.
- [14] F. J. Theis, "Towards a general independent subspace analysis," in *Proceedings of Neural Information Processing Systems (NIPS '06)*, 2006.
- [15] C. A. L. Pires and A. F. S. Ribeiro, "Separation of the atmospheric variability into non-Gaussian multidimensional sources by projection pursuit techniques," *Climate Dynamics*, vol. 48, no. 3–4, pp. 821–850, 2017.
- [16] P. Sura, M. Newman, and M. A. Alexander, "Daily to decadal sea surface temperature variability driven by state-dependent stochastic heat fluxes," *Journal of Physical Oceanography*, vol. 36, no. 10, pp. 1940–1958, 2006.
- [17] P. D. Sardeshmukh and P. Sura, "Reconciling non-Gaussian climate statistics with linear dynamics," *Journal of Climate*, vol. 22, no. 5, pp. 1193–1207, 2009.
- [18] M. Perron and P. Sura, "Climatology of non-gaussian atmospheric statistics," *Journal of Climate*, vol. 26, no. 3, pp. 1063–1083, 2013.
- [19] A. Hannachi, D. B. Stephenson, and K. R. Sperber, "Probability-based methods for quantifying nonlinearity in the ENSO," *Climate Dynamics*, vol. 20, no. 2–3, pp. 241–256, 2003.
- [20] G. H. Golub and C. F. van Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, MD, USA, 3rd edition, 1996.
- [21] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, New York, NY, USA, 2nd edition, 2012.
- [22] E. Schneidman, S. Still, M. J. Berry, and W. Bialek, "Network information and connected correlations," *Physical Review Letters*, vol. 91, no. 23, pp. 238701–1–238701–4, 2003.
- [23] P. Comon, "Independent component analysis, a new concept?" *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994.
- [24] A. Hyvärinen and P. Pajunen, "Nonlinear independent component analysis: existence and uniqueness results," *Neural Networks*, vol. 12, no. 3, pp. 429–439, 1999.
- [25] L. Almeida, "MISEP—linear and nonlinear ICA based on mutual information," *Journal of Machine Learning Research*, vol. 4, no. 7–8, pp. 1297–1318, 2004.
- [26] M. Scholz, "Validation on nonlinear PCA," *Neural Processing Letters*, vol. 36, no. 1, pp. 21–30, 2012.
- [27] A. H. Monahan, "Nonlinear principal component analysis: tropical Indo-Pacific sea surface temperature and sea level pressure," *Journal of Climate*, vol. 14, no. 2, pp. 219–233, 2001.
- [28] W. W. Hsieh, "Nonlinear canonical correlation analysis of the tropical pacific climate," *Journal of Climate*, vol. 14, no. 12, pp. 2528–2539, 2001.
- [29] W. W. Hsieh and A. A. Wu, "Nonlinear multichannel singular spectrum analysis of the tropical Pacific climate variability using a neural network approach," *Journal of Geophysical Research*, vol. 107, no. C7, pp. 13–1–13–15, 2002.
- [30] A. Wu, W. W. Hsieh, A. Shabbar, G. J. Boer, and F. W. Zwiers, "The nonlinear association between the Arctic Oscillation and North American winter climate," *Climate Dynamics*, vol. 26, no. 7–8, pp. 865–879, 2006.
- [31] Q. Teng, J. C. Fyfe, and A. H. Monahan, "Northern Hemisphere circulation regimes: observed, simulated and predicted," *Climate Dynamics*, vol. 28, no. 7–8, pp. 867–879, 2007.
- [32] J. F. Cardoso and A. Souloumiac, "Blind beamforming for non-Gaussian signals," *IEEE Proceedings, Part F: Radar and Signal Processing*, vol. 140, no. 6, pp. 362–370, 1993.
- [33] L. R. Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika*, vol. 31, no. 3, pp. 279–311, 1966.
- [34] L. de Lathauwer, B. de Moor, and J. Vandewalle, "A multilinear singular value decomposition," *SIAM Journal on Matrix Analysis and Applications*, vol. 21, no. 4, pp. 1253–1278, 2000.
- [35] L. H. Lim and J. Morton, "Cumulant component analysis: a simultaneous generalization of PCA and ICA," *Computational Algebraic Statistics, Theories and Applications*, 2008.
- [36] E. Jondeau, E. Jurczenko, and M. Rockinger, "Moment component analysis: an illustration with international stock markets," *Business & Economic Statistics*, pp. 1–23, 2016.

- [37] T. Blaschke and L. Wiskott, "An improved cumulant based method for independent component analysis," in *Proceedings of the Artificial Neural Networks Proceedings (ICANN '02)*, vol. 2415 of *Lecture Notes in Computer Science, Conference Paper*, pp. 1087–1093, Madrid, Spain, August 2002.
- [38] X. Geng, L. Ji, and K. Sun, "Principal skewness analysis: algorithm and its application for multispectral/hyperspectral images indexing," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 10, pp. 1821–1825, 2014.
- [39] C. Ahlback, J. Usatine, and N. Pippenger, "A Combinatorial Interpretation of the Joint Cumulant," 2012, <https://arxiv.org/abs/1211.0652v1>.
- [40] W. J. McGill, "Multivariate information transmission," *Psychometrika*, vol. 19, no. 2, pp. 97–116, 1954.
- [41] T. Tsujishita, "On triple mutual information," *Advances in Applied Mathematics*, vol. 16, no. 3, pp. 269–274, 1995.
- [42] A. Jakulin and A. I. Bratko, "Quantifying and visualizing attribute interactions: an approach based on entropy," 2004, <https://arxiv.org/abs/cs/0308002>.
- [43] C. A. L. Pires and R. A. P. Perdigão, "Non-gaussian interaction information: estimation, optimization and diagnostic application of triadic wave resonance," *Nonlinear Processes in Geophysics*, vol. 22, no. 1, pp. 87–108, 2015.
- [44] J. H. Friedman and J. W. Tukey, "A projection pursuit algorithm for exploratory data analysis," *IEEE Transactions on Computers*, vol. 23, pp. 881–890, 1974.
- [45] P. J. Huber, "Projection pursuit," *The Annals of Statistics*, vol. 13, no. 2, pp. 435–525, 1985.
- [46] L. Paninski, "Estimation of entropy and mutual information," *Neural Computation*, vol. 15, no. 6, pp. 1191–1253, 2003.
- [47] C. A. Pires and R. A. P. Perdigão, "Non-Gaussianity and asymmetry of the winter monthly precipitation estimation from the NAO," *Monthly Weather Review*, vol. 135, no. 2, pp. 430–448, 2007.
- [48] C. A. L. Pires and R. A. P. Perdigão, "Minimum mutual information and non-gaussianity through the maximum entropy method: theory and properties," *Entropy*, vol. 14, no. 6, pp. 1103–1126, 2012.
- [49] T. M. Smith, R. W. Reynolds, T. C. Peterson, and J. Lawrimore, "Improvements to NOAA's historical merged land-ocean surface temperature analysis (1880–2006)," *Journal of Climate*, vol. 21, no. 10, pp. 2283–2296, 2008.
- [50] M. Messié and F. Chavez, "Global modes of sea surface temperature variability in relation to regional climate indices," *Journal of Climate*, vol. 24, no. 16, pp. 4314–4331, 2011.
- [51] M. B. Richman, "Rotation of principal components," *Journal of Climatology*, vol. 6, no. 3, pp. 293–335, 1986.
- [52] G. R. North, T. L. Bell, R. F. Cahalan, and F. J. Moeng, "Sampling errors in the estimation of empirical orthogonal functions," *Monthly Weather Review*, vol. 110, no. 7, pp. 699–706, 1982.
- [53] H. von Storch and F. W. Zwiers, *Statistical Analysis in Climate Research*, Cambridge University Press, 1999.
- [54] Y. M. Okumura and C. Deser, "Asymmetry in the duration of El Niño and la Niña," *Journal of Climate*, vol. 23, no. 21, pp. 5826–5843, 2010.
- [55] I. Koch and K. Naito, "Dimension selection for feature selection and dimension reduction with principal and independent component analysis," *Neural Computation*, vol. 19, no. 2, pp. 513–545, 2007.
- [56] K. Wolter and M. S. Timlin, "Monitoring ENSO in COADS with a seasonally adjusted principal component index," in *Proceedings of the 17th Climate Diagnostics Workshop, Norman, OK, NOAA/NMC/CAC and Cosponsors*, vol. 5257, pp. 52–57, 1993.
- [57] R. A. Kerr, "A North Atlantic climate pacemaker for the centuries," *Science*, vol. 288, no. 5473, pp. 1984–1986, 2000.
- [58] N. J. Mantua, S. R. Hare, Y. Zhang, J. M. Wallace, and R. C. Francis, "A Pacific interdecadal climate oscillation with impacts on salmon production," *Bulletin of the American Meteorological Society*, vol. 78, no. 6, pp. 1069–1079, 1997.
- [59] N. J. Mantua and S. R. Hare, "The pacific decadal oscillation," *Journal of Oceanography*, vol. 58, no. 1, pp. 35–44, 2002.
- [60] E. Di Lorenzo, N. Schneider, K. M. Cobb et al., "North Pacific Gyre Oscillation links ocean climate and ecosystem change," *Geophysical Research Letters*, vol. 35, no. 8, Article ID L08607, 2008.
- [61] K. Ashok, S. K. Behera, S. A. Rao, H. Weng, and T. Yamagata, "El Niño Modoki and its possible teleconnection," *Journal of Geophysical Research C: Oceans*, vol. 112, no. 11, Article ID C11007, 2007.
- [62] Y. Morioka, T. Tozuka, S. Masson, P. Terray, J.-J. Luo, and T. Yamagata, "Subtropical dipole modes simulated in a coupled general circulation model," *Journal of Climate*, vol. 25, no. 12, pp. 4029–4047, 2012.
- [63] R. R. Rodrigues, E. J. D. Campos, and R. Haarsma, "The impact of ENSO on the south Atlantic subtropical dipole mode," *Journal of Climate*, vol. 28, no. 7, pp. 2691–2705, 2015.
- [64] S. E. Zebiak, "Air-sea interaction in the equatorial Atlantic region," *Journal of Climate*, vol. 6, no. 8, pp. 1567–1586, 1993.
- [65] P. Chang, T. Yamagata, P. Schopf et al., "Climate fluctuations of tropical coupled systems: the role of ocean dynamics," *Journal of Climate*, vol. 19, no. 20, pp. 5122–5174, 2006.
- [66] S. A. Venegas, L. A. Mysak, and D. N. Straub, "Atmosphere-ocean coupled variability in the South Atlantic," *Journal of Climate*, vol. 10, no. 11, pp. 2904–2920, 1997.
- [67] R. I. Saurral, F. J. Doblas-Reyes, and J. García-Serrano, "Observed modes of sea surface temperature variability in the South Pacific region," *Climate Dynamics*, pp. 1–15, 2017.
- [68] C. Deser, M. A. Alexander, S.-P. Xie, and A. S. Phillips, "Sea surface temperature variability: patterns and mechanisms," *Annual Review of Marine Science*, vol. 2, no. 1, pp. 115–143, 2010.
- [69] Y. Guan, J. Zhu, B. Huang, Z.-Z. Hu, and J. L. Kinter, "South Pacific Ocean dipole: a predictable mode on multiseasonal time scales," *Journal of Climate*, vol. 27, no. 4, pp. 1648–1658, 2014.
- [70] C. Deser and M. S. Timlin, "Atmosphere-ocean interaction on weekly timescales in the North Atlantic and Pacific," *Journal of Climate*, vol. 10, no. 3, pp. 393–408, 1997.
- [71] D. Mukhin, A. Gavrilov, A. Feigin, E. Loskutov, and J. Kurths, "Principal nonlinear dynamical modes of climate variability," *Scientific Reports*, vol. 5, Article ID 15510, 2015.
- [72] A. Gavrilov, D. Mukhin, E. Loskutov, E. Volodin, A. Feigin, and J. Kurths, "Method for reconstructing nonlinear modes with adaptive structure from multidimensional data," *Chaos*, vol. 26, no. 12, p. 123101, 2016.
- [73] S.-I. An, J.-S. Kug, A. Timmermann, I.-S. Kang, and O. Timm, "The influence of ENSO on the generation of decadal variability in the North Pacific," *Journal of Climate*, vol. 20, no. 4, pp. 667–680, 2007.
- [74] N. Schneider and B. D. Cornuelle, "The forcing of the Pacific Decadal Oscillation," *Journal of Climate*, vol. 18, no. 21, pp. 4355–4373, 2005.

- [75] M. Newman, M. A. Alexander, T. R. Ault et al., “The pacific decadal oscillation, revisited,” *Journal of Climate*, vol. 29, no. 12, pp. 4399–4427, 2016.
- [76] E. Di Lorenzo, G. Liguori, N. Schneider, J. C. Furtado, B. T. Anderson, and M. A. Alexander, “ENSO and meridional modes: A null hypothesis for Pacific climate variability,” *Geophysical Research Letters*, vol. 42, no. 21, pp. 9440–9448, 2015.
- [77] B. Rodríguez-Fonseca, R. Suárez-Moreno, B. Ayarzagüena et al., “A review of ENSO influence on the North Atlantic. A non-stationary signal,” *Atmosphere*, vol. 7, no. 7, article 87, 2016.
- [78] J. Su, R. Zhang, T. Li, and X. Rong, “Skewness of subsurface ocean temperature in the equatorial Pacific based on assimilated data,” *Chinese Journal of Oceanology and Limnology*, vol. 27, no. 3, pp. 600–606, 2009.
- [79] J.-Y. Yu, P.-K. Kao, H. Paek et al., “Linking emergence of the central Pacific El Niño to the Atlantic multidecadal oscillation,” *Journal of Climate*, vol. 28, no. 2, pp. 651–662, 2015.
- [80] R. I. Jennrich, “A simple general procedure for orthogonal rotation,” *Psychometrika*, vol. 66, no. 2, pp. 289–306, 2001.
- [81] R. C. Raffenetti and K. Ruedenberg, “Parametrization of an orthogonal matrix in terms of generalized eulerian angles,” *International Journal of Quantum Chemistry*, vol. 4, Proceedings of the International Symposium on Quantum Biology and Quantum Pharmacology, no. S3B, pp. 625–634, 1969.

Hindawi

Submit your manuscripts at
<https://www.hindawi.com>