

AUDREY LUSTIG

CYCLOSTATIONNARITY ANALYSIS OF ENSO MEMORY

Part of the INC (“Incendies Nouvelle Caledonie”) ANR project: improvement of fire management for biodiversity conservation in New Caledonia by developing a comprehensive understanding of the climatic, ecological and social drivers of fire.

STUDENT RESEARCH REPORT

Presented to the department of Bioinformatics and Modeling, Institut National des Sciences Appliquées de Lyon, France

And to the department of Computer Sciences in collaboration with the Complex Systems Institute IXXI, Ecole Normale Supérieure de Lyon, France

Lyon, France, September 2011

SUPERVISOR

CÉDRIC GAUCHEREL

Head of the Department of Ecology

French Institute of Pondicherry (UMIFRE 21 CNRS-MAEE)
Pondicherry 605001, India



ABSTRACT: Most climate fluctuations may be modulated by a variety of periodic or quasi-periodic deterministic forcing (e.g. diurnal, seasonal, Milankovitch cycle). This process modulation often induce cyclostationary (CS) behaviour defined as periodic correlations. We proposed here that quasi-periodic modulations may appear when the inertia (memory) of a process is modulating its pure CS property. Such process was said to exhibit an Extended AutoCorrelation (EAC) if its time-variant autocorrelation was periodic 'in average' in time and frequency domains. A new statistical tool has been developed to analyze EAC in a time series with the help of a comfortable time-frequency visualization, called Digital X Model (DXM). We examined the seasonal-to-interannual variability of El Niño/Southern Oscillation (ENSO) signals by the DXM. Results were compared to those obtained by the Monte-Carlo Singular Spectrum Analysis (SSA), the wavelet analysis and the Empirical Mode Decomposition (EMD). This comparative approach applied to SOI and SST ($\text{Niño}3$) indices confirmed the well-known interannual periodicities discussed in literature: the quasi-biennial (QB), 29 months, and the quasi-quadrennial (QQ), 50 months. The main result presented here was that these significant cycles were the only one detected by the four methods, suggesting that the QB and the QQ would be the fundamental modes of ENSO. In addition, the QB was strongly associated to the atmospheric component of ENSO and the QQ associated to the oceanic component of ENSO. This results lead to the hypothesis that the QB and QQ could be separate oscillations generated by differing physics. The interaction of these fundamental modes with the seasonal cycle could produce three additional minor periodicities by combination of tones, the 15-, 20- and 69 months oscillations, detected by the DXM and sometimes found in the three other methods.

Contents

1 Acknowledgements	3
2 Introduction and motivation	4
3 Data sets	5
3.1 Simulated data sets	5
3.2 Climatic data sets	6
4 Methods	7
4.1 Advanced spectral method for climatic time series	7
4.1.1 Singular Spectrum Analysis (SSA)	7
4.1.2 Wavelet Transform (WT)	9
4.1.3 Empirical Mode Decomposition (EMD)	11
4.2 Digital X Model (DXM)	13
4.2.1 Short and long range memory in climatology	13
4.2.2 The Digital X Model (DXM)	15
4.2.3 Test procedure and confidence level	17
4.3 Methodology	18
5 Results	19
5.1 Test results	19
5.2 Monte Carlo - SSA (MC-SSA) analysis	20
5.3 Wavelet analysis	21
5.4 Hilbert-Huang analysis	23
5.5 DXM analysis	25
6 Discussion and perspectives	26
6.1 DXM and Extended AutoCorrelation (EAC) function	26
6.2 Method comparison	27
6.2.1 Identifying modes of interannual variability of ENSO	27
6.2.2 Identifying ENSO events	29
6.3 Limits and perspectives	30
7 Conclusion	31

1 Acknowledgements

It was a pleasure for me to work at the French Institute of Pondicherry (IFP). The work environment that the staff provides is exceptional. Everyone is friendly, open, and willing to help in any way they can. The accommodation and support provided have exceeded my expectations and I am very appreciative. I would like to acknowledge the following people for their support and assistance with this internship.

I extend abundant and sincere thanks to my advisor CÉDRIC GAUCHEREL for the pleasure of working together and exchanging ideas. It was due to his valuable guidance, cheerful enthusiasm and ever-friendly nature that I was able to complete my research work in a respectable manner. His wide knowledge and his logical way of thinking have been of great value for me.

I would like to thank Professor VÉLAYOUDOM MARIMOUTOU for his contribution to this project, for his useful suggestions, for continuously challenging us with new ideas! I hope this report gives credit to the joint work and does not betray our occasionally different ideas.

I would like to express my deep gratitude to RENAUD BARBERO and Professor LAURENT PUJO-MENJOUET for taking the time to chat with me and giving me such great advices.

My warm thanks to VINCENT BONHOMME, MATHIEU CASTET, SALIMA EL MOKHTARI, JULIE ALET, DIANE BOUCHET, CAMILLE COUX, JULES MOREL, DEVAYANI KHARE and colleagues for their encouragement. They let me own a happy family in Pondicherry.

The financial support of the INC ANR project, of the Ecole Normale Supérieur (ENS) de Lyon and of the région Rhône-Alpes (Explora grant) are gratefully acknowledged.

2 Introduction and motivation

Fire is recognised as an important agent of biodiversity disturbance in New Caledonia [8]. As weather conditions, such as long-term or medium term droughts, have a prominent impact on fire occurrence and behaviour [29], the analysis of trends in the climatic time series in general and, specifically for drought indices, has become a subject of considerable interest [14, 41]. We now have evidence that rainfall variations of New Caledonia are largely affected by *El Niño Southern Oscillation* (ENSO)[32], one of the best examples of complex climatic phenomenon (see annex 1) that is subject to numerous studies [6, 30]. Basic principles of how the coupled ocean-atmosphere system works have been proposed, and appear to provide reasonable explanations for these phenomena [22] and related precipitation changes in New Caledonia [32]. However, the causes of ENSO events and their fluctuations remain unclear and understanding ENSO variability will aid providing important insights of the behaviour of the large scale precipitation system and related fires across New Caledonia.

ENSO is often thought as an irregular oscillator that swing between warm (El Niño) and cold (La Niña) sea surface temperature (SST) and low and high surface pressure over the central eastern tropical pacific [34]. Its variations are a part of the climate system's seasonal-to-interannual variability and are dominated by complex feedbacks between subsystems active on different time scales [14, 21, 39]. ENSO exhibits some degree of periodicity and is phase-locked to the annual cycle to a certain extent [35, 39]. It also possesses, however, an enigmatic irregularity that renders it a modelling and forecasting challenge [39]. Its complexity has inspired the introduction of new techniques for data analysis into climate dynamics that take into account the nonstationarity [12, 26] of the data, and in particular, their cyclostationarity (CS) [19, 23]. Such a tendency is natural since most of the climatic processes are influenced by a variety of periodic or quasi-periodic deterministic forcings (e. g. diurnal, seasonal or Milankovitch cycles) [45].

A central problem in climatic research is to reduce the numerous space-time degrees of freedom of the climate system to a minimum number of climatic modes that can explain a maximal part of its variability. Different methods have been developed to identify quasi-periodic time evolution, among them the singular spectrum analysis (SSA) and extensions have been widely applied to the reconstruction of climatic time series [14, 44, 49]. The method is a flexible data-adaptive method allowing spectral decomposition in short, noisy and chaotic time series without particular characteristics. In addition to these extremely popular approaches, the Wavelet Transform (WT) has been successfully used in many climatological studies [30, 48]. The method consists in the simultaneous decomposition of time series into the time/frequency space and consequently proves very useful in the analysis of multiscale nonstationary processes [15, 49]. More recently, the Empirical Mode Decomposition (EMD) have been introduced into climate studies [38] and consists in extracting the energy associated with various intrinsic time scales.

Despite this improvements and our knowledge about ENSO functioning, the possible causes of ENSO remain largely unknown. This motivate us in developing a new statistical tool to reveal new regularities in nonstationary time series. We propose here an unusual time-frequency transformation, called DXM (Digital X Model), which transforms the 1D vector dataset as a function of two temporal axes with different time resolution, making easier its visualisation. The aim of such a map is to transform the autocorrelation averaged in time and frequency into a two-dimensional ‘spatial’ autocorrelation. In many

ways, by taking advantage of the spatial representation, we extend what the mind and eyes do intuitively, to assess spatial patterns (map regularities) and trends. This idea emerges from discussions about the possible memory of ENSO, suggesting that the climatic process presents an inertia or a kind of intrinsic cohesion, linking time steps between each other's [12]. This observation is reinforced by the CS analysis, highlighting the fact that ENSO could no more be modulated by each month of a season, rather than several months together [23]. The hypothesis made here is that ENSO could be a phenomenon with CS behaviour 'in average' only, that is a kind of CS that extends over time and that can be quantified by computing autocorrelation averaged over time.

The DXM has been applied by [12] to the quantification of ENSO quasi-cycles and confirmed the well-known interannual periodicities (the quasi-biennial, 28 months, and the quasi-quadrennial, 45 months) and their evolutions discussed in literature [21, 28]. Those two significant quasi-cycles were the only one detected throughout the 50-year datasets, suggesting that various periodicities were detected in the past because the possible ENSO memory, coined as Extended CycloStationarity (ECS), was not taking into account. The DXM is thus a promising tool that is able to detect complementary information to existing methods dedicated to other nonstationary properties in a time series. However, a more complete analytical study of the method is still necessary [12]. Hence, the objective of this work is to support theoretically and empirically the method. The question addressed here are as follows. First, does the DXM methodology performs an extended CS analysis? Second, does it differ from any existing or already proposed methods? Third and finally, does ENSO variability presents new features in signals ? As work hypotheses we will assume that (i) the DXM performs an extended CS analysis, (ii) the DXM differs from other methods by taking into account the memory of ENSO phenomena and (iii) ENSO is a process with only two interannual quasi-periodicities.

In this report we review a few techniques of time series analysis that have been widely used in climatological studies and test their respective performances on synthetic and observed climatic signals. We do not intend to be exhaustive. Instead we present a panel of method commonly used for climatic time series analysis with different characteristics, which attempt to explore the various features of a time series. In this way our goal is to provide a general framework and strategy to extract as much information as possible from a climatic time series and to compare it to DXM.

Section 3 explains how we generated simulated time series and briefly describe two climatic time series. Section 4 is devoted to the descriptions of recent spectral techniques, the SSA, the WT and the EMD, and introduces the DXM. We show how each method highlight a particular feature of a time series. In Section 5 and 6, we analyse five time series and show the power and limitations for each technique; the climatic implication are also discussed. We conclude this report in Section 7.

3 Data sets

3.1 Simulated data sets

The general approach in interpreting the spectra of climate records is to look for frequency bands in which the power rises above the level of a random process. The level of confidence we attribute to a potential signal depends quite sensitively on our a priori assumptions

regarding the nature of the background noise and the success of signal detection will thus hinge on the proper isolation of the underlying noise [28]. In a crude approximation, background climate variability can be described by a simple random forcing called *white noise* process. White noise was constructed by producing pseudo-random data vector of length of the climate records sampled from a normal distribution (with a mean of zero and variance equal to one). Two different random data generators of MATLAB® have been tried, `wgn` and `randn`. We found that the results were not sensitive to the random number generator. Therefore we decided to report only the results from the random generator `randn`. White noise has a flat frequency spectrum, however, spectra of climatic time series frequently show a continuous decrease of spectral amplitude with increasing frequency [14, 41]. Thus, a second and more stringent case is that of *red noise* [1, 17].

The simplest statistical model for a discrete red noise series is the first-order autoregressive AR(1) process,

$$x_{t+1} = \alpha[x_t - \mu] + \epsilon_t + \mu, \quad (1)$$

where μ is the mean of the process, ϵ_t are normally distributed independent random variables with zero mean value and variance σ^2 and the constant coefficient α is related to the correlation time τ of the time series by $\tau \approx \frac{1}{1-\alpha}$; the α term causes the reddening of the time series [1]. There is theoretical justification for a red noise description of the noise background in climatic time series, as the thermal inertia of the oceans has been shown in models [17] to provide memory, effectively integrating atmospheric weather forcing. There is also considerable empirical evidence that the red noise model provides a reasonable description of the noise spectra for a variety of climatic time series, including ENSO time series [38]. Parameters for an AR(1) process were estimated from the climatic data sets and used to generate realization of AR(1) noise. We used the algorithm described by [3] which maximises the likelihood of our failing to reject the red noise null-hypothesis¹.

More elaborate models might seek to describe the noise background of the climate records such as higher order autoregressive or more detailed model of the physics governing the climate system (see [1] for a detailed discussion of this point). But many studies have directly argued that a robustly estimated AR(1) red noise background approximates quite well the noise background spectrum in a variety of climate time series [1, 28].

Finally, to test DXM performances we synthesized a simple 2D gauss circle curve of the form:

$$g(x, y) = \frac{1}{\sqrt{2}} e^{-\left(\frac{x^2+y^2}{2}\right)}. \quad (2)$$

As advocated in introduction, the DXM methodology transforms the studied 1D dataset as a function of two temporal axes with different time resolution and assess map regularities. By simulating a 2D gauss curve we directly create a signal with well-understood spatial properties that we want to quantify through DXM.

3.2 Climatic data sets

ENSO phenomenon is a coupled ocean-atmosphere system generally characterised by two indices: the Southern Oscillation Index (SOI) and the Sea Surface Temperature (SST). The SOI, commonly us to quantify the atmospheric ENSO variations, is defined as the

¹Alternative AR(1) estimators: `ar1cov`, `ar1inv`, `arbburg`, `aryule`.

normalized sea-level atmosphere pressure difference between Tahiti and Darwin, Australia. The data set used in this analysis is the monthly SOI time series extending from January 1933 though July 2011 (Figure 1a). A anomalously negative value of this index indicate a warm event (El Niño), while a highly positive value is associated with a cold event (La Niña). The oceanic component of ENSO is characterized by the monthly Niño3 SST, defined as the SST averaged over the eastern half of the tropical Pacific (5°S - 5°N , 90° - 150°W). The time interval considered goes from January 1950 to July 2011 because some values were unavailable before 1950. Both index are closely related and exhibit interannual variations [35, 29]. The data are based on the deseasonalized and normalized time series proposed by the National Oceanic and Atmospheric Administration (NOAA) Climate Prediction Center (CPC) archives². The seasonal cycle is removed by abstracting the average of the values for each calendar month over a reference interval (in this case 1981-2010, base period used by NOAA/CPC). The reason for this operation is to detect possible influences of seasons on climate phenomena rather than detecting seasons themselves. The residues from this operation, called monthly anomalies, are then normalized with respect to the standard deviation computed over the same interval to allow for a better comparison between both time series. It has been shown that working with deseasonalized and normalized data does not remove the CS properties of a process [19], the seasonal modulation being far more complex than a linear weighting operation supposed in the deseasonalized calculation.

4 Methods

4.1 Advanced spectral method for climatic time series

4.1.1 Singular Spectrum Analysis (SSA)

SSA is a fully non-parametric method especially well suited for short, noisy, and chaotic time series [44]. The SSA is similar to Principal Component Analysis (PCA) but reveals temporal variation patterns of a single time series using a predefined window length (or the maximum lag in the correlation matrix). One of SSA main advantages with respect to classical Fourier methods is its ability to detect oscillations modulated both in amplitude and phase. Thus, the original signal is no more simply decomposed into periodic sine and cosine functions, but rather into data adaptive waves possibly exhibiting non-constant amplitude and/or phase. SSA has been successfully used in many geophysical and climatological studies to reveal and predict periodic activities [30, 50]. A complete discussion of the technique is found in [44].

Strictly speaking, the objective of the SSA techniques is for the variance of the original data set to be resolved with only a few shape functions characterized by different dynamical properties. This is achieved by choosing the orthogonal SSA modes in such a way that the variance resolved by each mode decreases with mode number. Going deeper into technical details, there are three basic steps in SSA. *Step(i)*. By applying a sliding M -window to a time series $\{X(t), t = 1, 2, \dots, N\}$, we obtain a sequence of $N' = N - M + 1$ vectors $\tilde{X}_t = \{[X(t), X(t+1), \dots, X(t+M-1)], t = 1, 2, \dots, N'\}$ in the embedding space \mathcal{R}^M . The choice of the dimension M is not obvious [14]. It has been suggested that SSA is

²<http://www.cpc.ncep.noaa.gov/>

typically successful at analysing periods in the range $(\frac{M}{5}, M)$ [44]. *Step(ii).* We define the $M \times M$ lag-covariance matrix estimator \tilde{C}_X directly from the data as a Toeplitz matrix with constant diagonals; that is, its entries c_{ij} depend only on the lag $|i - j|$:

$$c_{ij} = \frac{1}{N - |i - j|} \sum_{t=1}^{N-|i-j|} \tilde{X}_t \tilde{X}_{t+|i-j|} \quad (3)$$

Step(iii). The covariance matrix, calculated from the N sample points is diagonalized and the eigenvalues are ranked in decreasing order :

$$\Lambda_X = E_X^T \tilde{C}_X E_X \quad (4)$$

where $\Lambda_X = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_M)$ is the diagonal matrix with $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_M \geq 0$ and E_X is the $M \times M$ matrix having the corresponding eigenvectors $\{\rho_k, k = 1, \dots, M\}$, named Empirical Orthogonal Functions (EOFs). Projecting the time series onto each EOF yields the corresponding principal components (PCs) A_k :

$$A_k = \sum_{j=1}^M X(t + j - 1) \rho_k(j) \quad (5)$$

To identify periodic or quasi-periodic activity in the signal, [44] use the eigenvalues spectrum, i.e. a plot of the time series eigenvalues ranked by order with the correspondent confidence bars. The near-quality of two successive eigenvalues and phase quadrature of their associated EOFs may jointly identify a fundamental oscillation [5, 44]. The nearly-equality has been characterized by [13] as follows:

$$|\lambda_{k+1} - \lambda_k| \leq \min\{\delta\lambda_k, \delta\lambda_{k+1}\} \quad (6)$$

where $\delta\lambda_k = \left(\frac{2}{N_{dof}}\right)^{\frac{1}{2}} \lambda_k$ and $N_{dof} = \frac{N}{M}$. Phase quadrature is analyzed through the lag-correlations of the corresponding PC pairs. The sum of the power spectra of the PCs is identical to the power spectrum of the time series $X(t)$; and following [44], we can study separately the spectral contribution of the various components using maximum entropy method (MEM) (see [14] for more details).

We can reconstruct that part of a time series that is associated with a single EOF or several by combining the associated PCs:

$$R_{\mathcal{K}}(t) = \frac{1}{M_t} \sum_{k \in \mathcal{K}} \sum_{j=L_t}^{U_t} A'_k(t - j + 1) \rho_k(j); \quad (7)$$

where \mathcal{K} is the set of EOFs on which the reconstruction is based and

$$(M_t, L_t, U_t) = \begin{cases} \left(\frac{1}{t}, 1, t\right), & 1 \leq t \leq M-1 \\ \left(\frac{1}{M}, 1, M\right), & M \leq t \leq N' \\ \left(\frac{1}{N-t+1}, t-N+M, M\right), & N'+1 \leq t \leq N \end{cases} \quad (8)$$

These series of length $N - M$ are called Reconstructed Components (RCs). They have the important property of preserving the phase of the time series.

SSA is designed to extract information from short and noisy time series and thus provide insight into the unknown or only partially known dynamics of the underlying system that generated the series [14]. However, standard SSA may be misleading in the case of either low signal-to-noise ratio or coloured background noise. In fact, they could make slow modes being incorrectly interpreted as nontrivial signals, and/or nontrivial signals embedded in coloured noise being neglected if their variance is smaller than that of the noise slow modes [3]. Hence, Monte Carlo (MC) approaches has been developed by [3], allowing to distinguish a given time series from any well-defined process, including red noise noise background [3, 14]. On the basis of estimated AR(1) parameters, an ensemble of simulated red noise data is generated and, for each realisation, the covariance matrix C_R and associated eigenvalues are computed. The statistical distribution of these later elements, gives confidence intervals outside which a time series can be considered to be significantly different from random realization of the red noise process.

MC-SSA in essence is an adaptive approach that provides good results for short time series [44] and allows one to extract trends from a wide class of time series. An essential disadvantage of SSA resides in the fact that PC obtained must be steady. Otherwise, it makes the explanation difficulty because the length of the PC is less than that of the series.

4.1.2 Wavelet Transform (WT)

An alternative approach is taken by WT, which does not try to find global patterns of variance but analyzes the data locally and reveals active frequencies in a time series and their changes over time. A quick review of applications to climate dynamics can be found in [30, 48], while a good introduction to WT related to climate research is given by [41].

Wavelets are functions that satisfy mathematical requirements and are used in representing data or other functions [15]. This idea is not new. Approximation using superposition of functions has existed since the early 1800's, when Joseph Fourier discovered that he could superpose sines and cosines to represent other functions. However, for many decades, scientists have wanted more appropriate functions to approximate choppy signals. By their definition, sines and cosines are non-local and stretch out to infinity. They therefore do a very poor job in approximating sharp spikes. But with WT, we can use approximating functions that are contained neatly in finite domains [15]. The core of wavelet parametric method consists in sliding window functions of particular shapes along the analyzed series, in order to obtain a time series of the projection amplitudes. The *scale* or *resolution* that we use to look at data plays a special role. In order to isolate signal discontinuities, one would like some short basis functions. At the same time, in order to obtain detailed frequency analysis, one would like some long basis functions. A way to achieve this is to have short high-frequency basis functions and long low-frequency ones. This is what you get with WT and what makes the method interesting and useful. In this sense it generates what is commonly called a phase space decomposition, where the phase is defined by two parameters (scale, time).

The wavelet procedure is to adopt a wavelet prototype function ψ , called a *mother wavelet*, with general admissibility properties, and with the more specific properties of zero mean and time and frequency localization, i.e., ψ and its Fourier transform $\hat{\psi}$ must decay rapidly outside a given interval [15]. The WT of a N -length time series $X(t)$ is a sort of a band pass filter defined as the inner product in the Hilbert space $L^2(R)$ with

respect to the mother wavelet, i.e.

$$W_n^X(s) = \sum_{n=0}^{N-1} x_{n'} \psi^* \left[\frac{(n' - n)\delta t}{s} \right], \quad (9)$$

where s represents the wavelet scale, n' the localized time index, δt the constant sampling time, and the asterisk the complex conjugate [41]. The variables s and n' are integers that scale and dilate the mother function to generate an orthogonal basis. In other words, the WT is a lossless linear transformation of data into coefficients on a basis of wavelet functions [15]. In practice it is faster to implement the convolution in Fourier space (see details in [41]).

The wavelet coefficients give a full information and a direct estimation for energies $E_n^X(s) = |W_n^X(s)|^2$ that are local in time [41]. By varying the localized time index n and the localized scale index s , one can construct a picture, the *wavelet power spectra*, showing both energy of any feature versus the scale and how this energy varies with time. Note that to make it easier to compare different wavelet power spectra, the wavelet spectrum requires a normalization (see details in [41]).

Because we are dealing with finite-length time series, errors will occur at the beginning and the end of the wavelet power spectrum. Following [41], we pad here the end of the time series with zeroes to bring the total length N up to the next higher power of two, thus limiting the edge effects. Padding with zeros introduces discontinuities at the endpoints and, as one goes to larger scales, decreases the amplitude near the edges as more zeroes enter the analysis. It is therefore useful to introduce a *Cone of Influence* (COI) in which edge effects become important. Here we take it as the area in which the wavelet power caused by discontinuity at the edge has dropped by a factor e^{-2} of the value of the edge .

The statistical significance of wavelet power is assessed relative to null hypothesis that the analysed signal is generated by a red noise process. The Fourier power spectrum of a red noise process with lag-1 autocorrelation α (estimated from the time series) is given by:

$$P_k = \frac{1 - \alpha^2}{|1 - \alpha e^{-2i\pi k}|^2}, \quad (10)$$

where k is the Fourier frequency index [41]. The WT can be thought as a consecutive series of band-pass filters applied to the time series where the wavelet scale is linearly related to the characteristic period of the filter (λ_{wt}). Hence, for a stationary process with the power spectrum P_k the variance at a given wavelet scale, by invocation of the Fourier convolution theorem, is simply the variance in the corresponding band of P_k . [41] used Monte Carlo methods to show that this approximation is very good for a red noise spectrum. As a null hypothesis, it is assumed that the studied time series has a mean power spectrum, possibly given by Eq. 10; if a peak in the wavelet power spectrum is significantly

One criticism of WT is the arbitrary choice of the mother wavelet. One should consult the related wavelet literature on this. [15, 41] studied ENSO phenomenon and used the Morlet basis and the Mexican hat basis defined as follows:

- Morlet wavelet : $\psi(\eta) = \pi^{-1/4} e^{i\omega_0 \eta} e^{-\frac{\eta^2}{2}}$
- Mexican hat wavelet (derivative of a Gaussian; m=2): $\psi(\eta) = \frac{(-1)^{m+1}}{\sqrt{\Gamma(m+\frac{1}{2})}} \frac{d^m}{d\eta^m} \left(e^{-\eta^2/2} \right)$

where ω_0 is dimensionless frequency, η is dimensionless time. Following [15, 41] we use $\omega_0 = 6$. The wavelet transform is calculated for a discrete set of 8 ($j = 0, 1, \dots, 7$) scales.

The scales are a series of fractional powers of two [41]:

$$s_j = s_0 2^{j\delta j} \quad (11)$$

where $s_0=2$ months and $\delta_j = 1/4$. This gives scales ranging from 2 months to 649 months (approximately 54 years). As mentioned before, the wavelet scale is linearly related to its Fourier period λ_{wt} . For the Morlet wavelet (with $\omega_0 = 6$) the Fourier period (λ_{wt}) is almost equal to the wavelet scale ($\lambda_{wt}^j = 1.03s_j$), for the Mexican hat case the characteristic period of the wavelet is linearly related to the scale by $\lambda_{wt}^j = 4s_j$ [41].

Since the wavelet spectrum presents a large amount of information in one image, it is often desirable to condense this information by averaging the results over a range of scales or times. One useful technique is to average the energy at every scale over the whole time series, to compare the spectral power at different scales. [41] call this the *global wavelet spectrum*. The result is a graph of averaged energies versus scale, analogous to the Fourier power spectrum, in which localization in time is lost. The Global wavelet spectrum is defined as:

$$[\bar{W}^X(s)]^2 = \frac{1}{N} \sum_{n=0}^{N-1} |W_n^X(s)|^2. \quad (12)$$

[33] shows that the global wavelet spectrum provides an unbiased and consistent estimation of the true power spectrum of a time series.

Similarly, the *scale-averaged wavelet power spectrum* examines the fluctuations in power over a range of scales [41]. This can be defined as the weighted sum of the wavelet power spectrum over scales s_{j_1} to s_{j_2} :

$$[\bar{W}_n^X]^2 = \frac{\delta j \delta t}{C_\delta} \sum_{j=j_1}^{j_2} \frac{|W_n^X(s_j)|^2}{s_j}. \quad (13)$$

where C_δ is a scaling constant for each wavelet (see [41] for more details). Following [41] we used $C_\delta = 0.776$ for the Morlet wavelet and $C_\delta = 1.966$ for the Mexican hat wavelet.

Wavelet analysis is attractive because (i) it is local, although higher frequencies are more localized; (ii) it has uniform temporal resolution for all frequency scales; and (iii) it is useful for characterizing gradual frequency changes. This is particularly suited for analyzing localized effects. However, the method is nonadaptive because the same basic wavelet is used for all data.

4.1.3 Empirical Mode Decomposition (EMD)

To overcome the shortcomings of the two precedent methods, the EMD has been recently proposed by [18] as an adaptive time-frequency data analysis method. Although it often proved remarkably effective [18, 10], the technique is faced with the difficulty of being essentially defined by an algorithm, and therefore of not admitting an analytical formulation which would allow for a theoretical analysis and performance evaluation [18].

The starting point of the EMD is to consider oscillations in signals at a very local level. We look at the evolution of a signal $X(t)$ between two consecutive extrema, then the local maxima are connected by a cubic spline as the upper envelope, and the local minima are similarly connected as the lower envelope. The mean of these two envelopes

is a function of time and designated as $m(t)$. We thus define heuristically a local low-frequency part of the signal $X(t)$, or local *trend*. For the picture to be complete, one still has to identify the corresponding local high-frequency part $d(t)$, or local *detail* so that we have $X(t) = m(t) + d(t)$. This $d(t)$ is referred to as *Intrinsic Mode Function* (IMF). Assuming that this is done in some proper way for all the oscillations composing the entire signal, the procedure can then be applied on the residual consisting of all local trends. By construction, the number of extrema is decreased when going from one residual to the next, and the whole decomposition is guaranteed to be completed with a finite number of modes. Indeed the process, referred to as *sifting*, is repeated until the detail signal $d(t)$ can be considered as zero-mean according to some stopping criteria (see [18] for more details).

Having obtained the IMF components, one will have no difficulty in applying the Hilbert transform defined as follows:

$$H_k(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{d_k(\tau)}{t - \tau} d\tau. \quad (14)$$

where d_k is the k -th IMF. $d(t)$ and $H(t)$ can be combine into a complex expression as

$$z_k(t) = A_k(t) \exp[i\theta_k(t)] \quad (15)$$

where $A_k(t) = \sqrt{H_k^2(t) + d_k^2(t)}$ is the amplitude of the nonlinear oscillatory pattern $z_k(t)$ and $\theta_k(t) = \arctan[H_k(t)/d_k(t)]$ defined the phase angle. From Eq. 15, one can obtain the instantaneous frequency through the calculation of the derivative of $\theta_k(t)$ with respect to t , i.e.:

$$\omega_k(t) = \frac{d\theta}{dt} \quad (16)$$

The instantaneous frequency represents the signal's frequency at an instance. The frequency-time distribution of the amplitude is designated as the *Hilbert amplitude spectrum*, or simply *Hilbert spectrum*. If the amplitude squared is the more preferred method to represent the energy density, then the squared values of the amplitude can be substituted to produce the Hilbert energy spectrum just as well. A description of the Hilbert transform with the emphasis on its many mathematical formalities can be found in [16]. Essentially, Eq. (14) defines the Hilbert transform as the convolution of $x(t)$ with $1/t$; therefore emphasizes the local properties of $x(t)$.

The question of the IMF's statistical significance is always an issue. In data containing noise, how can the noise be separated confidently from the information? These question were addressed by [9], [47] and [46] through a study of noise only. Before examining any results, it is necessary to list the properties of an IMF as follows: an IMF is any function having symmetric envelops defined by the local maxima and minima separately, and also having the same number of zero-crossing and extrema [46]. Based on this definition, we can determine the mean period of the function by counting the number of the peaks (local maxima) of the function. The mean period τ_m of any IMF is calculated as:

$$\tau_m = \frac{1}{L-1} \sum_{l=1}^L \rho'_l \quad (17)$$

where L is the total number of maxima and ρ'_l is the sample length between l -th and $(l+1)$ -th maxima. [9] and [10] found that when plotted the root-mean-squared³ (RMS) values of

³The RMS of the n -th IMF corresponds to its energy density, i.e. $E_n = \frac{1}{N} \sum_{j=1}^N [d_n(j)]^2$

the IMFs as a function of the mean period derived from the fractional Gaussian noise on a log-log scale, the results formed a straight line. Instead of fractional Gaussian noise, [46] studied the Gaussian white noise only. They also found the same relationship between the RMS values of the IMFs as a function of the mean period. Additionally, they also studied the statistical properties of the scattering of the data and found the bounds for the noise data distribution analytically. From the scattering, they deduced a 95% bound for the white noise. Therefore, they concluded that when a dataset is analyzed by using EMD, if the RMS-mean period values exceed the noise bound, then those IMFs must represent statistically significant information. Following these works, we proposed here to study the distribution of the the RMS-mean period values of a red noise process to deduce a 95% bound for the coloured process.

The combination of the EMD and the Hilbert spectral analysis is also known as the *Hilbert-Huang transform* (HHT). Selection of modes corresponds to an automatic and adaptive (signal-dependant) time-variant filtering [37]. Such a representation is thus well-suited to the analysis of nonlinear and nonstationary data and has been proved quite versatile in a broad range of applications for extracting signals from data generated in noisy nonlinear and nonstationary processes [20]. But one of the major drawbacks of the EMD is the frequent appearance of IMF consisting of signals of widely disparate scales or a signal of a similar scale residing in different IMF components often due to signal intermittency [20]. The intermittence could not only cause serious aliasing in time-frequency distribution, but also make the physical meaning of individual IMF unclear [18]. Several methods [18, 20] have been proposed to overcome the scale problem but are not presented in this report.

4.2 Digital X Model (DXM)

4.2.1 Short and long range memory in climatology

Many processes in nature are of such a high complexity that a description by deterministic models is difficult or not desirable. However, their characteristic behaviour is often captured by stochastic process [17]. It has been shown that climate variability in principle can be modelled by rather simple and linear AR-process (see section 3). Typical properties of AR-processes are an exponential decay of the autocorrelation function and thus a finite memory. Processes with these attributes are called short-range or short-memory process. More recently, long-range dependence have been reported in geophysical records [4]. The long-memory is an asymptotic property of a stochastic process, characterising the decay of the correlation between observations separated by increasingly large time lags.

It is now commonly believed that ENSO is the result of a complex dynamical coupling between the atmosphere and the ocean which presents 'memory' that extends over many months [42]. We implicitly manipulate this concept when talking about an El Niño events which are no more characterized by an isolated negative SOI value (positive Niño3 SST value) rather than described by integrated negative (positive) index values of a persisting process [34]. There has been works suggesting that the internal memory of the coupled system resides in the ocean thermocline state [25]. The thermal inertia of oceans is at the origin of long term memory and interannual variability, both of which lead to extreme events at different time scales. Long-range dependence reported in ENSO may not necessarily hinder time series analysis and forecasting [7], nonetheless this feature of ENSO

memory is intriguing. Hence, in this study we assume that ENSO can be described by a system which presents an internal memory resulting from the combined inertia of ocean (long term memory) and atmosphere (short term memory) and thus, that ENSO is no more modulated by each month of a season rather than several months together.

Most geophysical systems, however, are characterized by periodic cycles [45], such as the diurnal cycle, the annual cycle, the Milankovitch or the seasonal cycle. It is thus appropriate to include this external periodicity in the statistical analysis [45]; this may results in *cyclostationarity* (CS) statistics [24, 42]. CS statistics differ from stationary statistics in that the various quantities, such as the autocorrelation, are periodic functions of the phase with respect to the external periodic cycle. Let consider a time series $X(t)$ with mean $\bar{X}(t)$. Its autocorrelation is defined as follows:

$$C(t, t + \tau) = \langle (X(t) - \bar{X}(t)) (X(t + \tau) - \bar{X}(t)) \rangle, \text{ also noted } C(t, \tau), \quad (18)$$

where the angular bracket $\langle \cdot \rangle$ denotes ensemble average and τ a time lag. A process is said to be cyclostationary at order 2 if its mean and autocorrelation are periodic in t with period d for each τ [24]. That is:

$$\bar{X}(t + d) = \bar{X}(t) \text{ and } C(t + d, \tau) = C(t, \tau) \quad (19)$$

Considering that ENSO may be characterized by periodic cycle (external forcing) and that long range dependence has been detected in ENSO memory, we propose here, that ENSO could be a *CS process that extends over time*. A CS process has a memory in the sense when its autocorrelation is periodic 'in average' in time and frequency [12]. This can be expressed by:

$$\left\langle \langle C((t + u) + v \cdot d, \tau) \rangle_{u \in \Phi_{u_{max}}} \right\rangle_{v \in \Phi_{v_{max}}} = C(t, \tau), \quad (20)$$

with $\Phi_{u_{max}} = [-u_{max}, u_{max}]$ and $\Phi_{v_{max}} = [-v_{max}, v_{max}]$. u and v are integer factors ranging in their respective intervals $\Phi_{u_{max}}$ and $\Phi_{v_{max}}$. The coefficient u (resp. v) intends to take into account the possible links or cohesion between neighbouring time steps (resp. frequencies) of the process. Note that the coefficient u is summed, while the coefficient v is multiplied: the reason is that using an additive coefficient v would have blurred the frequency determination (enlargement of the peak frequency) [12]. Due to this averaging over the v values, the d parameters will further be referenced as a 'quasi-period' to not be confused with the period of a pure CS process.

The autocorrelation of a process, locally averaged in time and frequency, has been called the Extended Autocorrelation (EAC) function. Following the terminology introduced in Eq. (20), the EAC is defined as follows:

$$EAC_d(t, u_{max}, v_{max}, \tau) = \left\langle \langle C((t + u) + v \cdot d, \tau) \rangle_{u \in \Phi_{u_{max}}} \right\rangle_{v \in \Phi_{v_{max}}} \quad (21)$$

In discrete time, the EAC is expressed by:

$$EAC_d(u_{max}, v_{max}, \tau) = \sum_t EAC_d(t, u_{max}, v_{max}, \tau) \quad (22)$$

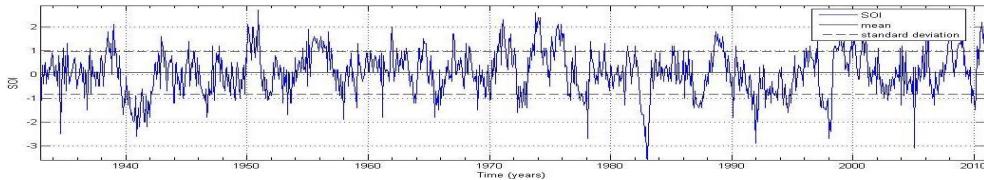
4.2.2 The Digital X Model (DXM)

An unusual time-frequency method, called the DXM, has been proposed to quantify the EAC function [12]. Such as the EMD, the technique is faced with the difficulty of being essentially defined by an algorithm. An analytical analysis remains necessary to theoretically establish the link between the DXM and the EAC. This will be the subject of future research.

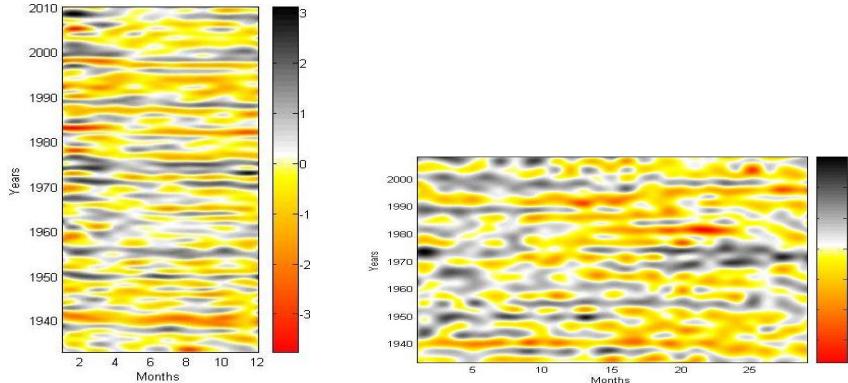
The starting point of the DXM is to transform the studied time series as a function of two temporal axes with different time resolution [11]: the sampling time along the x -axis and a chosen quasi-period d along the y -axis. This is equivalent to a projection of the time series $\{X(t) : t = 1, \dots, N\}$ on a map \mathcal{M} defined by:

$$\mathcal{M}(k - u, v) = X(v + u \cdot d), \text{ with } N = k \cdot d \quad (23)$$

where k is the number of lines v , d the quasi-period as well the number of columns u . For ease of viewing (but not for calculation), the index values are interpolated with an exact 2D polynomial functions and visualized with a colour scale (Figure 1). Such a map \mathcal{M} is called a *DXM map* and a signal represented in 2D is referred as a *folded signal*. Similarly, we say that the signal is *unfolded* to assess the transition from the 2D to the 1D representation.



(a) Southern Oscillation Index (SOI) time series from NOAA/CPC over (Jan1933 - July2011)



(b) DXM map with $d = 12$. (c) DXM map with $d = 29$.

Figure 1: SOI time series (a) and associated DXM maps with two distinct quasi-period, $d = 12$ (b) et $d = 29$ (c). Index monthly values are plotted as a function of time along the x -axis and y -axis. X-axis represents a short time scale while y shows a long one. A colour scale associated with an exact polynomial interpolation helps to visualize the index fluctuations. El Niño and La Niña events are respectively represented in orange-red and black.

The trends of successive time steps along the x - or y -axes represent the process modulation at different time resolutions. Note that choosing a different period d would lead to a different DXM map, showing different ‘spatial’ patterns (Figure 1b-1c). Here emerges

the idea of taking advantage of the spatial representation to extend what the mind and eyes do intuitively to assess map regularities and implicitly extended time regularities. In other words, the aim of such maps is to transform the supposed autocorrelation averaged in time and frequency (EAC) into a two-dimensional ‘spatial’ autocorrelation.

Let $\mathcal{M}_d(x, y)$ denotes the pixel values of a DXM map \mathcal{M} with a quasi-period d , where x and y are the coordinates in vertical and horizontal directions, respectively. For calculation the index values of the DXM map are not interpolated. Also, let k and d be the DXM map height and width, respectively, so that $1 \leq x \leq k$ and $1 \leq y \leq d$. The basic principle of the DXM method is, first, to compare a reference sub-map of size $(2u_{max} + 1) \times (2v_{max} + 1)$, also termed pixel subset or moving window, to a target sub-map of the same size, both sampled in the DXM map \mathcal{M} as illustrated in Figure 2. The acknowledged correlation criterion in between two sub-maps takes the form:

$$R_{i,j}(p, l) = \left\{ \sum_{x=i-u_{max}}^{i+u_{max}} \sum_{y=j-v_{max}}^{j+v_{max}} [\mathcal{M}(x, y) - \bar{m}_{i,j}] [\mathcal{M}(x+p, y+l) - \bar{m}_{i,p;j,l}] \right\} \times \\ \left\{ \sqrt{\sum_{x=i-u_{max}}^{i+u_{max}} \sum_{y=j-v_{max}}^{j+v_{max}} [\mathcal{M}(x, y) - \bar{m}_{i,j}]^2} \sqrt{\sum_{x=i-u_{max}}^{i+u_{max}} \sum_{y=j-v_{max}}^{j+v_{max}} [\mathcal{M}(x+p, y+l) - \bar{m}_{i,p;j,l}]^2} \right\}^{-1} \quad (24)$$

where (i, j) denotes the center location of the subset of size $(2u_{max} + 1) \times (2v_{max} + 1)$ pixels sampled on the reference map; $\mathcal{M}(x, y)$ and $\mathcal{M}(x+p, y+l)$ are the values of the studied time series at the positions (x, y) and $(x+p, y+l)$, respectively; while p and l are the relative displacement components between the two subsets; $\bar{m}_{i,j}$ and $\bar{m}_{i,p;j,l}$ are the mean values of the corresponding subsets which can be expressed as:

$$\bar{m}_{i,j} = \frac{1}{(2u_{max} + 1)(2v_{max} + 1)} \sum_{x=i-u_{max}}^{i+u_{max}} \sum_{y=j-v_{max}}^{j+v_{max}} \mathcal{M}(x, y) \quad (25)$$

$$\bar{m}_{i,p;j,l} = \frac{1}{(2u_{max} + 1)(2v_{max} + 1)} \sum_{x=i-u_{max}}^{i+u_{max}} \sum_{y=j-v_{max}}^{j+v_{max}} \mathcal{M}(x+p, y+l) \quad (26)$$

For a specified position (i, j) selected in the DXM map, as shown in figure 2, one can obtain the distribution of the correlation criterion $R_{i,j}(p, l)$ over a pre-appointed searching region of size $(4u_{max} + 1) \times (4v_{max} + 1)$ around (i, j) , by varying the local location coordinate (p, l) . In this way, the average of the correlation criterion $\bar{R}(i, j)$ is calculated over the searching region as follows:

$$\bar{R}(i, j) = \begin{cases} 0 & , \text{ si } p=0 \text{ et } l=0 \\ \frac{\sum_{p=-u_{max}}^{u_{max}} \sum_{l=-v_{max}}^{v_{max}} [\rho_{i,j}(p, l)]}{(2u_{max} + 1)(2v_{max} + 1) - 1} & , \text{ sinon.} \end{cases} \quad (27)$$

The central correlation of the searching region is systematically removed, as it is equal to 1 (autocorrelation with a null time lag) and thus can hide information on process regularities. Finally, the whole in-plane displacement field can be acquired by altering the position (i, j) sampled in the DXM map. This provides an estimation of what we called the previous section the Extended Autocorrelation (EAC):

$$\widehat{EAC}_d(u_{max}, v_{max}) = \frac{\sum_{i=2u_{max}+1}^{k-2u_{max}} \sum_{j=2v_{max}+1}^{d-2v_{max}} [\bar{R}(i, j)]}{(k - 4u_{max})(d - 4v_{max})}, \text{ with } k = N/d \quad (28)$$

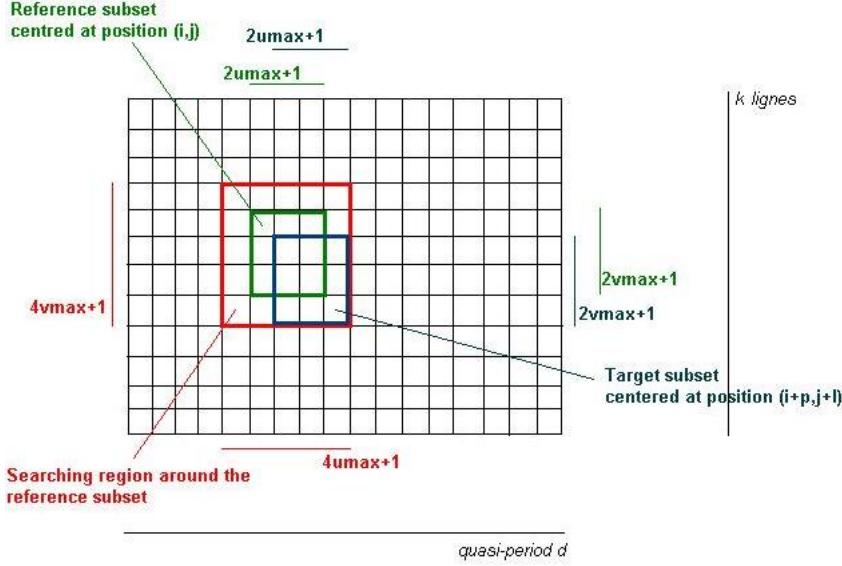


Figure 2: Schematic illustration of the DXM algorithm

Note that the borders of the map \mathcal{M} have to be avoided at a distance of u_{max} along the x -axis and v_{max} along the y -axis. Hence, maps with too short periodicities or number of periodicities are prohibited.

Consider now the same time series folded into a DXM map with a new quasi-period d' . This will indeed change the relative position of each pixel in the DXM map and therefore change the map regularities as well as the final $\widehat{EAC}_d(u_{max}, v_{max})$ value. It is then possible to define an *EAC profile* which represents the $\widehat{EAC}_d(u_{max}, v_{max})$ as a function of the quasi-period d . An extrema find at quasi-period d in the EAC profile implies that the corresponding map does present regular patterns (correlation or anti-correlation) and consequently, a relatively high or small EAC value.

4.2.3 Test procedure and confidence level

An other algorithm, based on the properties of the variogram⁴, has been tried to describe the degree of spatial dependence of a DXM map. The results were not conclusive and therefore were not reported in this report.

The confidence levels for the EAC profile were estimated using Monte-Carlo simulations. We first simulated 500 random times series. For each realisation, the EAC profile was computed for a discrete set of 40 ($d = 15, \dots, 55$) quasi-periods. We used a reference window of size (3×3) pixels, i.e., $u_{max} = 1$ and $v_{max} = 1$. The normal distribution of the EAC values allowed us to give the 95% confidence intervals outside which a time series can be considered to be significantly different from a random realization. To check whether or not statistical artefacts were introduced, we experimented with two other window size: (5×5) and (7×7) pixels. Finally, an ensemble of 500 simulated red noise data, estimated

⁴The variogram describes the degree of spatial dependence of a spatial random field or stochastic process. It is defined as the variance of the difference between field values at two locations across realizations of the field.

from both SOI and Niño3 SST series, were generated and the associated confidence levels were computed for a reference window of size (3×3) pixels.

The DXM method was experimented using a 2D discrete gauss curve described by Eq. (2). We generated a square matrix of 900 data points which is approximately the size of ENSO indices. The signal was then unfolded to create a 1D vector data set. Its EAC profile was numerically computed for a discrete set of 90 ($d = 10, \dots, 100$) quasi-periods. We used a reference window of size (3×3) pixels. The EAC profile was compared to the classical autocorrelation function (ACF). To check whether or not statistical artefacts were introduced, we experimented with two other window size: (5×5) and (7×7) pixels. This simple model was only used as a mathematical example to empirically assess characteristic properties of the DXM methodology.

4.3 Methodology

First, the four methods were tested using Monte-Carlo simulations: 500 white noise and red noise processes were simulated to define their respective confidence levels (see Annex 2). Then, we considered an application of the four methods to the SOI and Niño3 SST. As the quality of a prior unknown trend is hard to evaluate, we do not search for the best method but rather seek to evaluate some of their features.

The PCs of the studied time series and their characteristic dominant period were displayed in MC-SSA analysis. Following [14], we used a window length $M = 60$ which allowed us to capture periodicities as long as 5 years. The importance of the PC was judged by identifying which components contribute significantly more variance relative to the red noise background within a SSA eigenvalue spectrum. In addition, we performed a partial reconstruction of the studied time series to identify ENSO events. The latter were compared with the events defined by the NOAA/CPC. We used the routines implemented in Matlab by E. Breitenberger, available at: <http://pangea.stanford.edu/research/Oceans/GES290/Breitenberger-SSAMatlab/>.

The Morlet and the Mexican hat basis were selected for the wavelet analysis. We displayed the Wavelet power spectrum of the SOI and Niño3 SST series. For comparison with other methods, the wavelet scale was expressed in terms of its Fourier period. Trends were extracted for periods from 2 to 256 months for the Morlet basis and from 8 to 512 for the Mexican hat basis. We conducted tests of any oscillatory component of ENSO indexes against red noise null-hypothesis using the Monte Carlo method. To condense the information, the global wavelet spectrum and the scale-averaged wavelet spectrum were calculated for each power spectrum. We exploited the routines implemented in Matlab provided by C. Torrence and G. Compo, available at: <http://paos.colorado.edu/research/wavelets/>.

Using the EMD methodology, the SOI and Niño3 SST time series were decomposed into IMFs. The averaged period of each IMF and its density were plotted as a log-log scatter plot and compared to the reference red noise samples to determine whether a specific IMF contains any statistically significant information. To allow a spectral analysis, the Hilbert spectrum of the significant IMF was plotted. We performed a partial reconstruction of the studied climatic index to identify ENSO events. We exploited the routines implemented in Matlab by G. Rilling, available at: <http://web.mit.edu/anaghad/MEng/code/EMD/>

Finally, the EAC profiles of SOI and Niño3 SST index were computed for quasi-periods $d \in \{10, \dots, 100\}$. We used a reference window of size (3×3) pixels). We compared the

two profiles to the reference red noise samples to determine whether a specific quasi-period d contains any statistically significant information.

5 Results

5.1 Test results

The 95% confidence intervals of the EAC profile, calculated from 500 simulations of white noise, appeared quite stable (see Annex 2). Table 1 gives the confidence levels for three different window sizes.

Reference window size	u_{max}	v_{max}	Mean EAC	95% CL
3×3 pixels	1	1	-0.067	[-0.054, -0.080]
5×5 pixels	2	2	-0.024	[-0.031, -0.015]
7×7 pixels	3	3	-0.012	[-0.017, -0.005]

Table 1: **Confident levels (CL) of the EAC profile.** The first column indicates the size of the reference moving window, corresponding to the parameters presented in column 2 and 3. The fourth column is the mean of the EAC profile over a discrete set of 40 ($d = 15, \dots, 55$) quasi-periods. The fifth column is the mean of the 95% confidence interval over the same set of quasi-periods.

The negative values for the confidence levels and the influence of the window size on the mean of the EAC were intriguing. Indeed, the DXM tool has been developed to quantify a spatial autocorrelation. Under the null hypothesis of a variable randomly distributed in the DXM map, we expected the EAC values to fluctuate around zero. Besides we expected the reference window size to only influence the 95% confidence interval rather than the mean itself. Thus, the results above raise several issues: (i) Do we really generate a 2D random distributed process? (ii) Does a 2D random process present a spatial structure naturally anticorrelated? (iii) Does the DXM algorithm systematically introduce a negative bias in the estimation of the spatial autocorrelation? The influence of the window size on the mean tend to weight for the hypothesis (ii). However, a theoretical analysis of the method and extensive simulation experiments remain necessary to answer such questions.

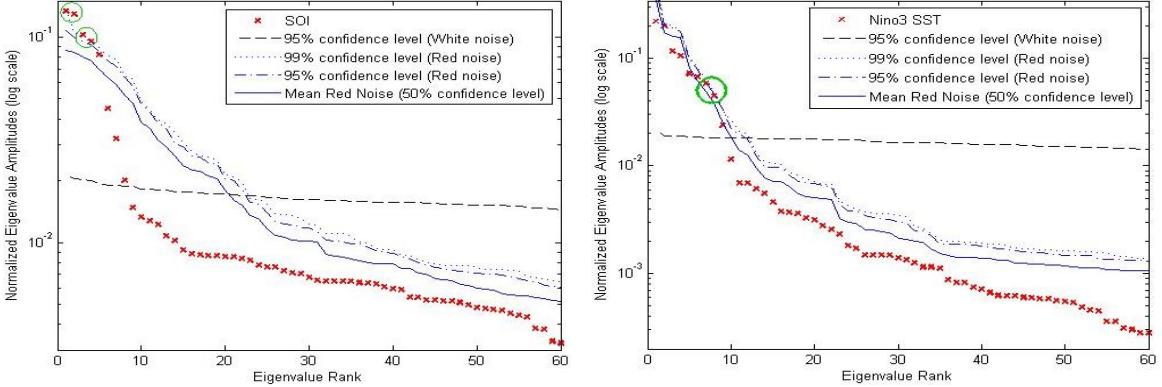
Parameters for an AR(1) process (see Eq. 1) were estimated from SOI and Niño3 SST data sets and used to generate 500 realizations of red noise. These parameters are given in Table 2. The normal distribution of the 500 EAC profiles allowed us to give the 95% and 99% confidence level.

AR(1) parameters	μ	α	σ
Estimation from SOI	0.0045	0.7452	0.6197
Estimation from Niño3 SST	0.0325	0.9214	0.3511

Table 2: **Parameters for an AR(1) process**

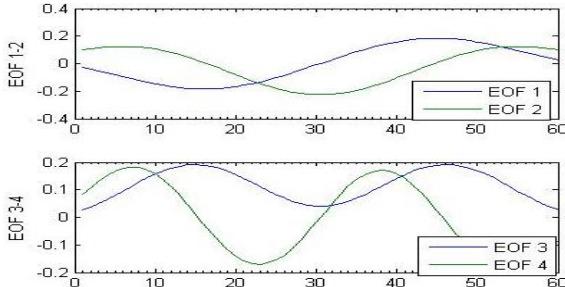
Finally, simulations experiments confirmed that the DXM was able to detect DXM map regularities and thus was able to detect extended time regularities or quasi-cycles in a time series (See annex 3). Nevertheless, the cost of this method was that quasi-cycle multiples were also detected. Hence, they may confuse or hide a longer periodicity present in the signal.

5.2 Monte Carlo - SSA (MC-SSA) analysis

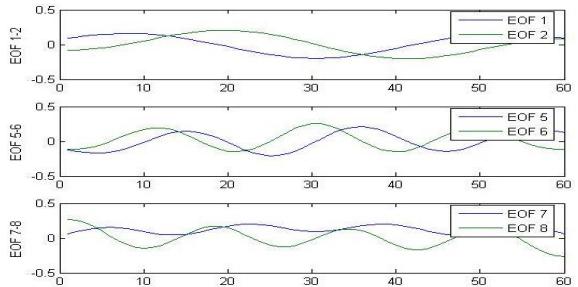


(a) **Singular Spectrum of SOI.** In order to enhance the readability of the graphs, the error bars are not represented. There is a clear grouping of the first five eigenvalues, followed by a very steep slope of three additional eigenvalues. Only the four first pass the pairing criteria and are significant (green circles) for the 99% red noise confidence level.

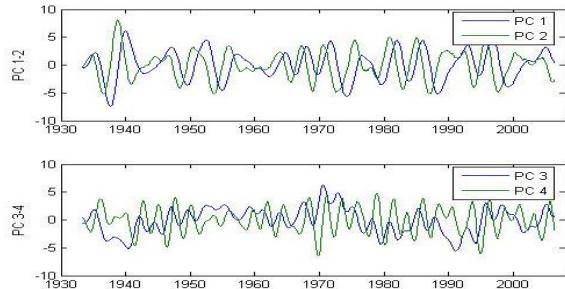
(b) **Singular Spectrum of Niño3 SST.** In order to enhance the readability of the graphs, the error bars are not represented. The first eight eigenvalues pass the pairing criteria. Only the pair 7-8 (green circle) emerges above the red-noise-like background. The second eigenvalue stands slightly above the 99% and the sixth stand slightly above the 95% red noise confidence level.



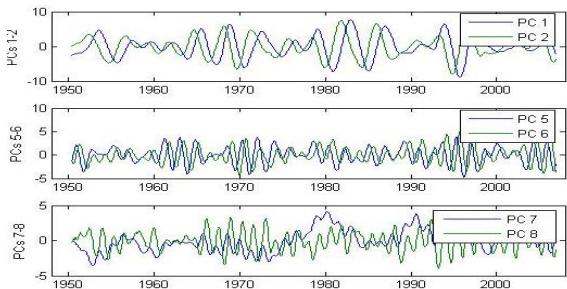
(c) **Significant EOFs of SOI.** They are grouped into two leading pairs, (EOFs 1-2) and (EOFs 3-4).



(d) **Significant EOFs of Niño3 SST.** They are grouped into three leading pairs, (EOFs 1-2), (EOFs 5-6) and (EOFs 7-8).



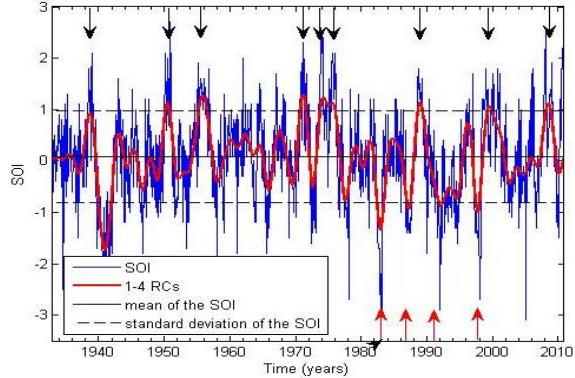
(e) **Significant PCs of SOI.** Note phase quadrature for each leading pair of PCs which corresponds in the singular spectrum analysis (figure a) to a pair of successive eigenvalues that are nearly equal. The maximum cross correlation between (PC1-2) equals 0.79 and between (PC3-4) equals 0.49. According to [5], we may identify a fundamental oscillation.



(f) **Significant PCs of Niño3 SST.** Note phase quadrature for each leading pair of PCs which corresponds in the singular spectrum analysis (figure b) to a pair of successive eigenvalues that are nearly equal. The maximum cross correlation between (PC1-2) equals 0.86, between (PC4-5) equals 0.83 and between (PC4-5) equals 0.37. According to [5], we may identify a fundamental oscillation.

Pairs	Periods (months)	Variances (%)
1-2	58.7 - 55.4 **	24.9
3-4	31.2 **	19.6
(5)	27.7	8.2
(6-7)	21.7-19.57	8
(8)	16.7	2.3

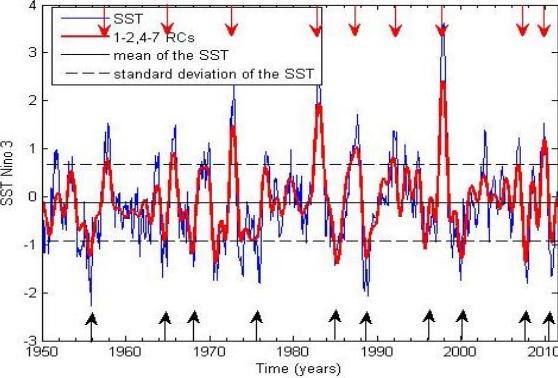
(g) **Characteristics of the oscillatory pairs of SOI.** Each row corresponds to a single or a pair of eigenvalues. Eigenvalues in parentheses does not pass the pairing criteria or does not emerge above the red-noise-like background, but are included for completeness. The third column contains the dominant period of the oscillation. The double star (**) indicates dominant periods that are highly significant. The fourth column gives the total percentage of variance associated with the oscillatory pair, calculated from the eigenvalue spectrum.



(i) **Partial reconstitution of SOI based on EOFs 1-4 (red curve).** The raw SOI is shown as the blue curve. **El Niño** events are shown as upward pointing arrows on the lower abscissa, while **La Niña** events are shown as downward pointing arrows on the upper abscissa.

Pairs	Periods (months)	Variances (%)
1-2	47.5*	42
(3-4)	29.35-27.7	22.3
5-6	20.8-19.2 *	13.4
7-8	16.4-16.1 **	9.8

(h) **Characteristics of the oscillatory pairs of Niño3 SST.** Each row corresponds to a single or a pair of eigenvalues. Eigenvalues in parentheses does not emerge above the red-noise-like background, but are included for completeness. The third column contains the dominant period of the oscillation. The double star (**) indicates dominant periods that are significant at 99%, while the single star (*) is used for those significant at 95%. The fourth column gives the total percentage of variance associated with the oscillatory pair, calculated from the eigenvalue spectrum.



(j) **Partial reconstitution of Niño3 SST based on EOFs 1-2,4-8 (red curve).** The raw Niño3 SST is shown as the blue curve. **La Niña** events are shown as upward pointing arrows on the lower abscissa, while **El Niño** events are shown as downward pointing arrows on the upper abscissa.

Figure 3: MC-SSA analysis of the SOI and SST for a red noise null-hypothesis

5.3 Wavelet analysis

We observed the variations in the frequency and the amplitude of ENSO events over time. The most noticeable difference between the two wavelet basis was the fine-scale structure using the Mexican hat. The Mexican hat was narrower in time-space yet broader in spectral-space than the Morlet. Thus, the Mexican hat wavelet emphasised the localisation and detection of fine scale features while the Morlet wavelet emphasised intermediate scale feature. Overall, the same features appeared in both plots, approximately at the same locations, and with the same power.

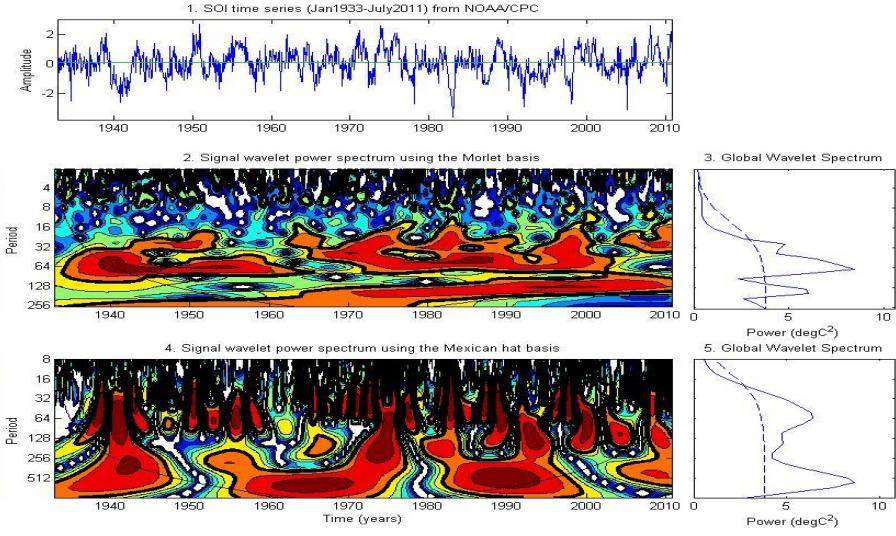


Figure 4: Wavelet analysis of the SOI time series. (1) The SOI time series from 1933 to 2011 (blue curve). The green line represents the mean of the data set. (2) The power spectrum of SOI using the Morlet basis. The thick contour encloses regions of greater than the 95% confidence level for a red-noise process. The thin line indicates the 'COI' where edge effects become important. (3) The associated global wavelet power spectra which shows dominant maxima at 27.79, 62.11 and 157.2 months. The dashed line is the significant level assuming the same background spectrum as in (2). (4) same as (2) using the Mexican hat basis. (5) same as (4) but dominant maxima occur at 63.6 and 719.3 months.

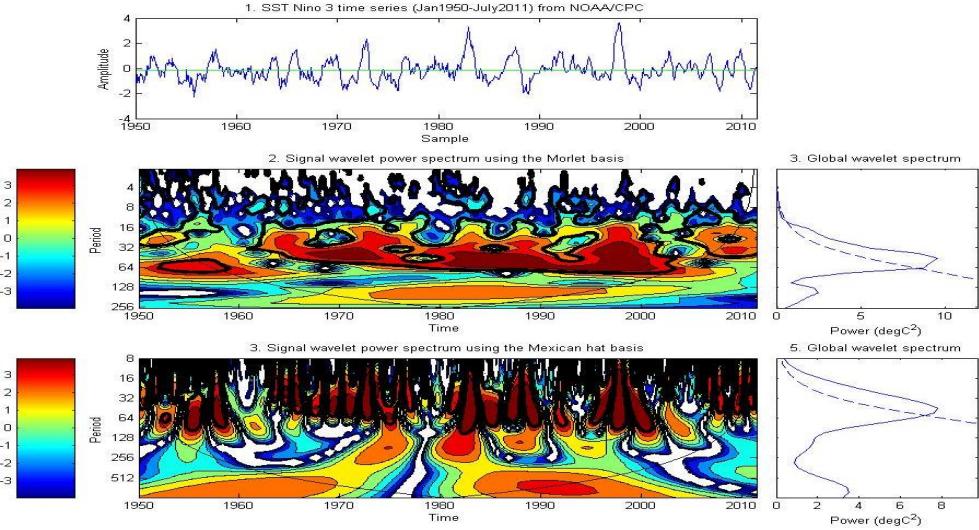


Figure 5: Wavelet analysis of the Niño3 SST time series. (1) The Niño3 SST time series from 1950 to 2011 (blue curve). The green line represents the mean of the data set. (2) The power spectrum of Niño3 SST using the Morlet basis. The thick contour encloses regions of greater than the 95% confidence level for a red-noise process. The thin line indicates the 'COI' where edge effects become important. (3) The associated global wavelet power spectra which shows a dominant maxima at 45.0 months. The dashed line is the significant level assuming the same background spectrum as in (2). (4) same as (2) for the Mexican hat basis. (5) same as (4) but dominant period occur at 46.7 months.

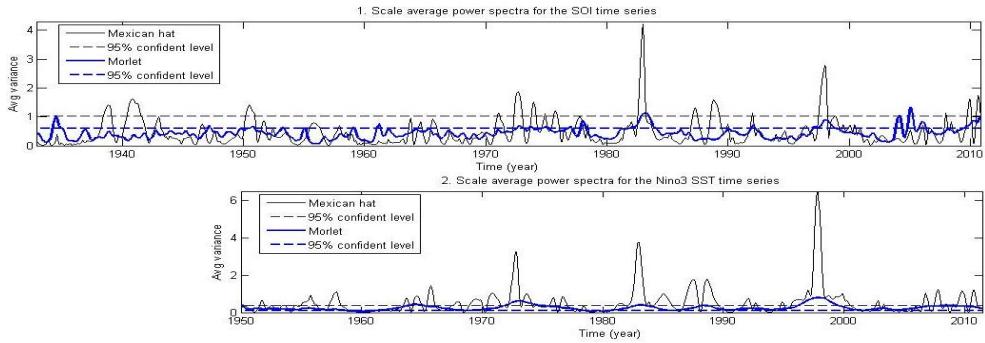
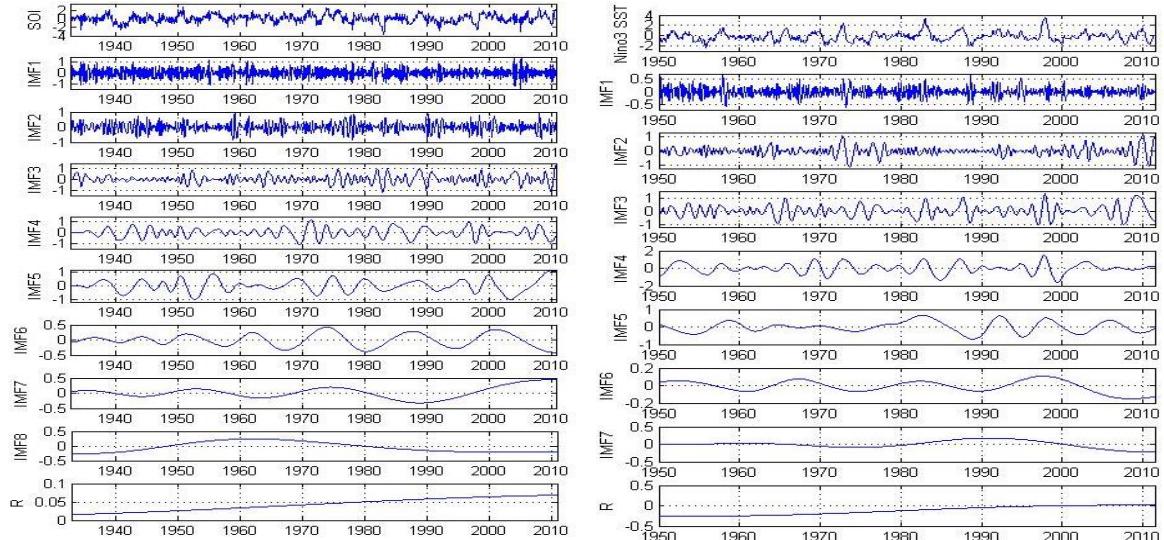


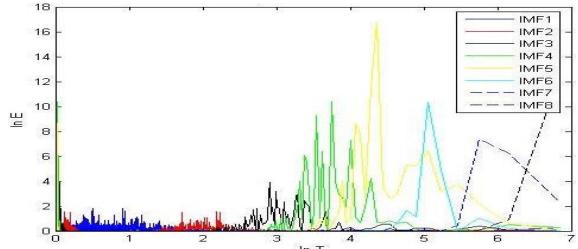
Figure 6: Scale-averaged wavelet power for (1) the SOI and (2) the Niño3 SST time series over the 8-256 months scale-band. The blue curve results from the Morlet analysis and the thick black curve from the Mexican hat analysis. The dashed line is the 95% confidence level for the Morlet analysis (blue) and the Mexican hat analysis (black).

5.4 Hilbert-Huang analysis

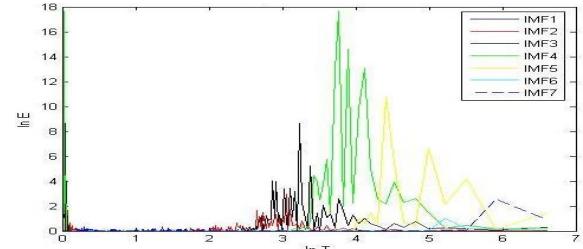


(a) **EMD decomposition of SOI.** The SOI time series (top panel), the corresponding intrinsic mode functions (IMF1-8), and the monotonic trend (R).

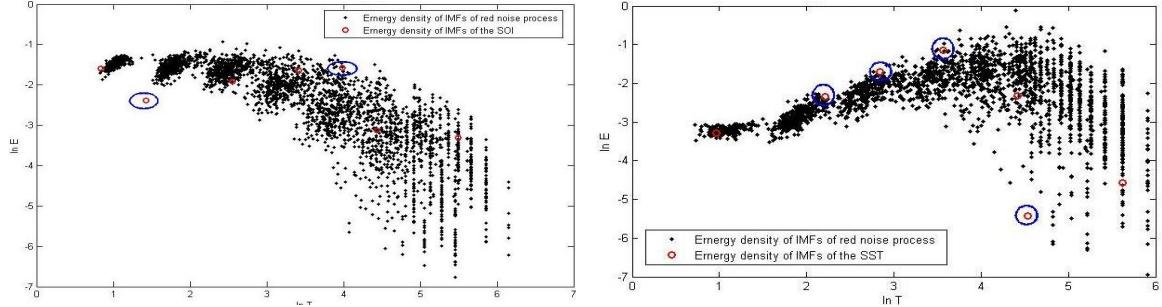
(b) **EMD decomposition of Niño3 SST.** The Niño3 SST time series (top panel), the corresponding intrinsic mode functions (IMF1-7), and the monotonic trend (R).



(c) The Fourier spectra of the IMFs as a function of the logarithm of period for the SOI time series.

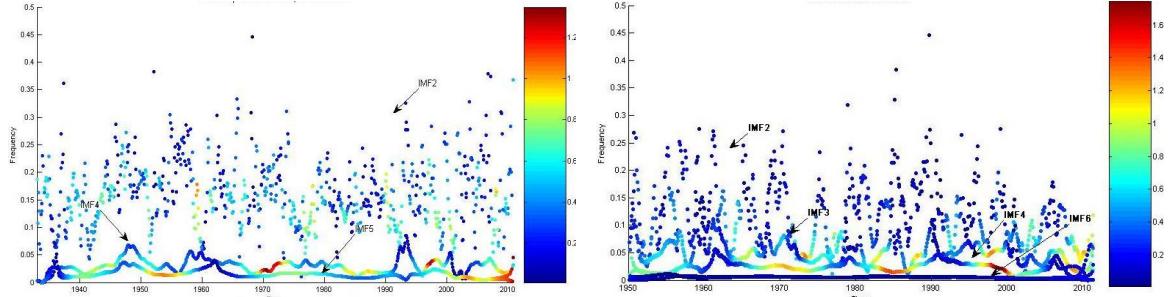


(d) The Fourier spectra of the IMFs as a function of the logarithm of period for the Niño3 SST time series.



(e) **Monte Carlo tests and energy of the IMFs as a function of the averaged period of the SOI time series.** The red-circle are the energy density of an IMF of the SOI as a function of its corresponding averaged period. The group of black dots are the energy density as function of the averaged period for IMFs 1-7 for all 500 samples of red-noise. The confidence levels have not been yet carefully estimated and necessitate further experiments. Nevertheless, IMFs 2 and 5 stand out from the scatter plot and were assumed significant.

(f) **Monte Carlo tests for a red-noise process and energy of the IMFs as a function of the averaged period of the Niño3 SST time series.** The red-circle are the energy density of an IMF of the Niño3 SST as a function of its corresponding averaged period. The group of black dots are the energy density as function of the averaged period for IMFs 1-7 for all 500 samples of red-noise. The confidence levels have not been yet carefully estimated and necessitate further experiments. Nevertheless, IMFs 2, 3, 4 and 6 stand out from the scatter plot and were assumed significant.



(g) **Hilbert spectrum of SOI.** Time-frequency distribution of the amplitude of the IMFs 2, 4 and 5 of the SOI time series. The instantaneous amplitude is visualized with a color scale. Only the IMFs 2 and 5 were detected significant. However, the IMF 4 has been often reported in literature, and was included here for completeness.

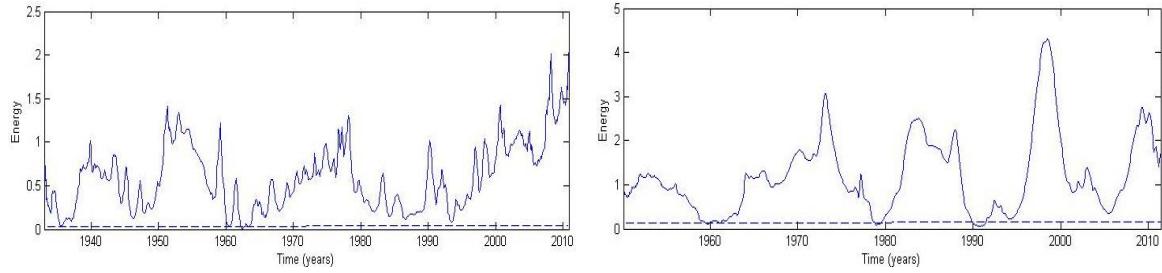
(h) **Hilbert spectrum of Niño3 SST.** Time-frequency distribution of the amplitude of the IMFs 3, 4, 5 and 6 of the Niño3 SST time series. The instantaneous amplitude are visualized with a color scale.

IMF	Dominant period (months)
2	8.9
(3)	18.4
(4)	58.7
5	76.7

(i) **Characteristics of the IMFs of SOI.** Each row corresponds to a single IMF. IMFs in parentheses did not stand out of the red-noise-like background, but are included for completeness. The second column contains the dominant period of the oscillation.

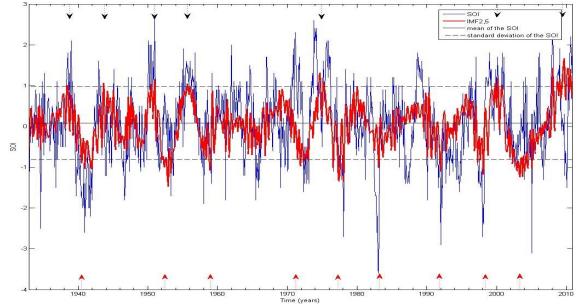
IMF	Dominant period (months)
2	21.2
3	25.6
4	43.4
(5)	83.2
6	199.6

(j) **Characteristics of the IMFs of Niño3 SST.** Each row corresponds to a single IMF. IMFs in parentheses did not stand out of the red-noise-like background, but are included for completeness. The second column contains the dominant period of the oscillation.

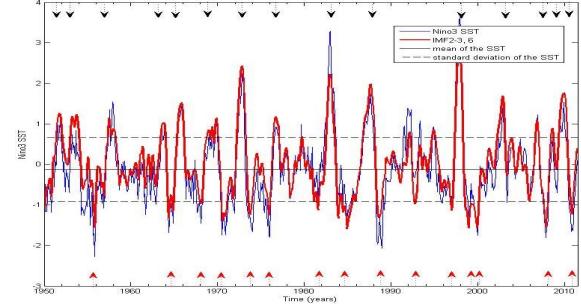


(k) Time distribution of the the energy of significant IMFs. Instantaneous energy of IMFs 2 and 5 were summed and represented has a function of time.

(l) Time distribution of the the energy of significant IMFs. Instantaneous energy of IMFs 2,3,4 and 6 were summed and represented has a function of time.



(m) Partial reconstitution of SOI based on IMFs 2 and 5 (red curve). The row SOI is shown as the blue curve. **El Niño** events are shown as upward pointing arrows on the lower abscissa, while **La Niña** events are shown as downward pointing arrows on the upper abscissa.

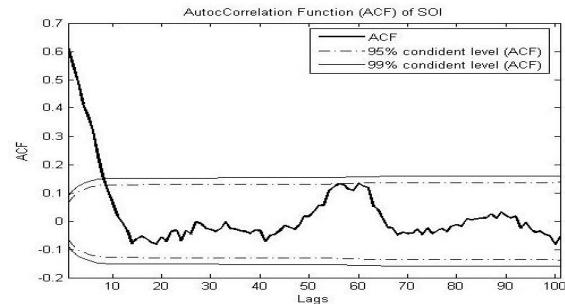


(n) Partial reconstitution of Niño3 SST based on IMFs 2,3,4 and 6 (red curve). The row SOI is shown as the blue curve. **La Niña** events are shown as upward pointing arrows on the lower abscissa, while **El Niño** events are shown as downward pointing arrows on the upper abscissa.

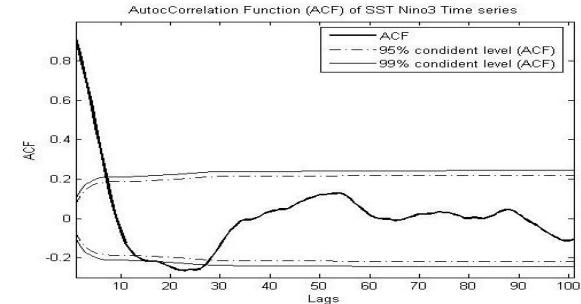
Figure 7: Hilbert-Huang analysis

5.5 DXM analysis

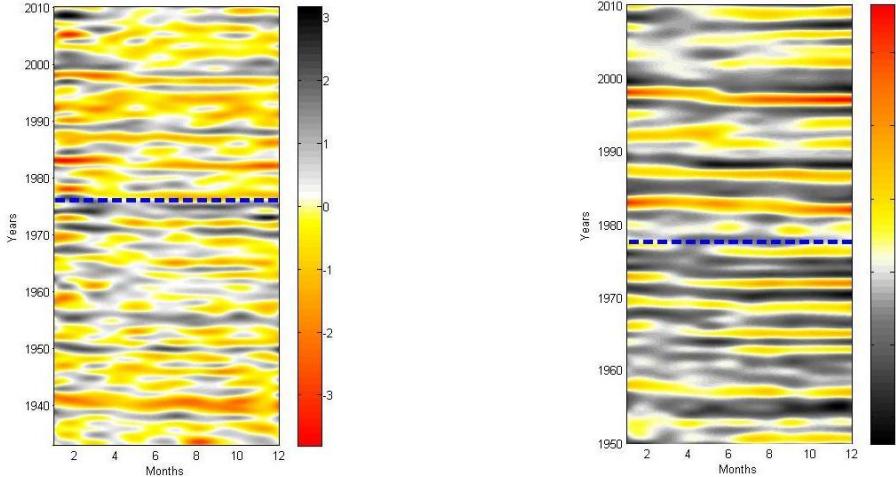
The analysis of both SOI and Niño3 SST time series is presented here. The results differed from those obtained by [12] because the algorithm to quantify the 'spatial' autocorrelation of a DXM map was carefully reviewed and tested on simulated signal.



(a) AutoCorrelation Function (ACF) of the row SOI time series.

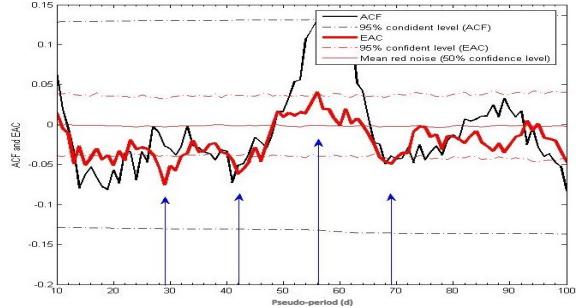


(b) AutoCorrelation Function (ACF) of the row Niño3 SST time series. Note that the series shows a significant autocorrelation (99%) for the lags 20-25.

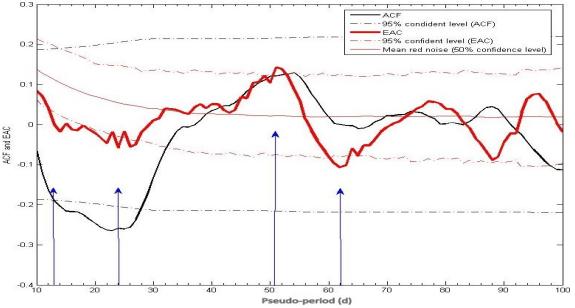


(c) **SOI DXM map with a quasi-period $d = 12$ months over the period 1933-2011.** Index monthly values are plotted as a function of time along the x -axis (short time scale) an y -axis (large time scale). A color scale associated with an exact polynomial interpolation helps to visualized index fluctuations. El Niño and La Niña events are in red (orange) and black, respectively. The blue dashed line indicates a change in spatial patterns around 1976.

(d) **Niño3 SST DXM map with a quasi-period $d = 12$ months over the period 1950-2011.** Index monthly values are plotted as a function of time along the x -axis (short time scale) an y -axis (large time scale). A color scale associated with an exact polynomial interpolation helps to visualized index fluctuations. El Niño and La Niña events are in red (orange) and black, respectively. The blue dashed line indicates a change in spatial patterns around 1977.



(e) **EAC profile of SOI.** SOI autocorrelation function (black curve) and extended autocorrelation function (red curve) as a function of, respectively, time lag and pseudo-period. The associated 95% confidence levels are shown as dashed lines. The x -axis are between 10 and 100 months. SOI show significant dominant extrema at 29, 42, 56 and 69 months. Minor peaks are observed at 13-15 and 20-22 months.



(f) **EAC profile of Niño3 SST.** Niño3 SST autocorrelation function (black curve) and extended autocorrelation function (red curve) as a function of, respectively, time lag and pseudo-period. The associated 95% confidence levels are shown as dashed lines. The x -axis are between 10 and 100 months. SOI show significant dominant extrema at 14, 26, 51 and 62 months.

Figure 8: DXM analysis

6 Discussion and perspectives

6.1 DXM and Extended AutoCorrelation (EAC) function

We proposed here to consider ENSO phenomenon as a complex coupled ocean-atmosphere system which presents memory that extends over time [12]. This property was interpreted

as a specific case of periodic statistics averaged over a short range of time steps and frequencies (Eq. (20)). Such a process was said here to exhibit an Extended AutoCorrelation (EAC) if its time-variant covariance was periodic 'in average'. A new statistical tool has been developed to analyze EAC in a time series with the help of a comfortable time-frequency visualization technique called DXM [43]. Since the method has essentially been defined by an algorithm, its performance evaluation was difficult. Indeed, the first objective of this work was to establish whether or not the DXM methodology performs an EAC analysis as defined in Eq. (21). An analytical definition of the method remains necessary to answer such question and will be the object of further researches. However, simulation experiments confirmed that the DXM method was able to detect DXM map regularities and thus was able to detect extended time regularities in the associated time series.

Hence, the DXM was thought as a promising tool that is able to detect complementary information to existing methods dedicated to other nonstationary properties in a time series. The second objective of this work was to investigate how the DXM methodology differs from the most popular techniques for data analysis into climate studies. To sharpen this, we examined the seasonal-to-interannual variability of the SOI and Niño3 SST by four different methods: the MC-SSA, the WT, the EMD and the DXM.

6.2 Method comparison

6.2.1 Identifying modes of interannual variability of ENSO

Five significant interannual oscillations were present at about 16, 20, 29, 45 and 70 months. A summary of significant oscillatory components identified by the four methods is given in Table 3.

The period around 29 months appeared to form a *quasi-biennial* (QB) oscillation following the terminology of [35]. This mode was strongly related to SOI interannual oscillations. Indeed, the quasi-period of 29 months was highly significant in the DXM analysis of SOI. As the method is sensitive to multiples, the 'double' of the QB oscillation was detected around 56 months (Figure 8e). The QB mode and its multiple were also highly significant in the MC-SSA and WT analysis of SOI (Figure 3 and 4, respectively). Indeed, the global wavelet spectrum showed significant peaks around 62 months in both Morlet and Mexican hat analysis. The QB modes is detected in the Morlet analysis which emphasized intermediate scale. In MC-SSA analysis, the QB mode (31 months and its multiple) accounted for 41,5% of the SOI variance (Figure 3i). Therefore, the QB mode may be regarded, as a fundamental ENSO mode. These results are consistent with the study of [14, 36] to some extent. Although they emphasized the role of the QB and of the oscillatory component with period of about 50 months, we considered here that the QB oscillation were the only mode detected (with its multiple) and was associated to SOI. A significant period of about 70 months may also characterized SOI interannual oscillations. It appeared as a significant oscillatory component in the EMD and the DXM analysis (Figure 4 and 8e, respectively). To our knowledge, this mode has not been reported before. While we hesitated to place too much emphasis on it, its presence was intriguing.

The period around 45 months was strongly related to the Niño3 SST index. It was detected with the four methods. Following the terminology of [36], we referred to it as the *quasi-quadrennial* (QQ) mode. It corresponds to the 'low-frequency' ENSO mode. For the MC-SSA analysis, this mode account for 42% of the variance in Niño3 SST (Figure 3h).

	SOI	Niño3 SST
SSA	58.8-55.5 months** 31.1 months**	47.6 months* 20.8-19.1 months* 16.4-16.1 months**
WT	Power spectra (local analysis) + Dominant peaks occur in the global power spectrum at ¹ : 27.8 months** 62.1-63.6 months**	Local analysis + Dominant peaks occur in the global power spectrum at ¹ : 45 - 46.7 months**
EMD	Hilbert spectrum (local analysis) + dominant frequencies of significant IMFs ¹ : 74.6 months	Hilbert spectrum (local analysis) + dominant frequencies of significant IMFs ¹ : 21.2 months 25.6 months 43.1 months
DXM	13-15 months* 20-22 months* 29 months** 42 months** 56 months* 69 months*	14 months** 24-26 months** 51 months* 62 months*

Table 3: Comparison of the significant periods. The word 'period' is used here as a general term to describe the dominant frequencies of the EOFs and IMFS, the dominant scale of the WT and the dominant quasi-period of the DXM. Periods are grouped together into five characteristic modes or ranges of frequencies : the quasi-biennial mode, 25-32 months and multiples (red), the quasi-quadrennial mode, 42-51 months (blue), 69-75 months (green), 19-22 months (fuchsia), 14-17 months (maroon). ¹The frequencies that characterize the interannual or the interdecennial variability of the ENSO phenomenon are not taking in account in this study. * Significatif for a 95% confident level ** Significatif for a 99% confident level

Hence, the QQ mode may be regarded, as another fundamental ENSO mode, in addition to the QB. This finding is consistent with the study of [21, 36].

An oscillatory component with a period of 14-16 months may also characterized Niño3 SST interannual oscillations. It appeared as a highly significant mode in the MC-SSA and the DXM analysis (Figure 3b and 8f, respectively). The latter oscillation, sometimes discussed in literature [14], accounted for 9.8% of the variance in Niño3 SST (Figure 3b). Finally, the DXM detected a 20-21 months oscillation in SOI, while the MC-SSA and the EMD found it in Niño3 SST (Figure 3b and 8f, respectively). The latter accounted for 13.4% of the variance in Niño3 SST. To our knowledge, this oscillation is new, but was at the limit of what we can reliably detect and describe.

More generally, the SOI and Niño3 SST EAC profiles appeared to be related. A peak in SOI EAC profile corresponded to a zero-crossing in Niño3 SST EAC profile, and vice-versa. This may reflect the strong coupling nature of the atmosphere and the ocean in ENSO phenomena. To our knowledge, little attention has been paid to the the 15-16 months oscillations. The 20-21 and 69-75 months oscillations were new. Although they were more controversial than the QB and the QQ modes, their identification was tentative. Indeed, these modes have a very suggestive interpretation in terms of the nonlinear interaction of

the dominant QB, QQ and annual cycle [14]. Such interaction will produce power at the difference of the:

- annual and QQ frequencies, $(12 \text{ months})^{-1}$ and $(50 \text{ months})^{-1}$, yielding **(15.8 months) $^{-1}$** ,
- annual and QB frequencies, $(12 \text{ months})^{-1}$ and $(29 \text{ months})^{-1}$, yielding **(20.5 months) $^{-1}$** ,
- QB and QQ frequencies, $(29 \text{ months})^{-1}$ and $(50 \text{ months})^{-1}$, yielding **(69 months) $^{-1}$** .

The results presented here helped us in establishing a scenario for producing multiple interannual spectral peaks. The QQ and QB could be separate oscillations, generated by differing physics, each with independent characteristic frequency [12]. Considering that the SOI and Niño3 SST indices mainly characterize the atmospheric and oceanic components of ENSO, the QB and the QQ may be interpreted has the two independent characteristic frequencies of the 'atmosphere' and of the 'Pacific ocean' themselves. The interactions of the seasonal cycle and the fundamental ENSO frequencies can produce additional peaks by combination of tones [14]. In this scenario, the QB and QQ would be the fundamental ENSO modes and their interactions with the seasonal cycle would tend to lock them to a 30 months and 50 months period, producing the more controversial 15-16, 20-21 and 69-75 months oscillations as a by-product of the nonlinear resonance.

6.2.2 Identifying ENSO events

The four methods provided highly complementary information. The MC-SSA eigenvectors carry a characteristic structure in time lag. Even more useful than examining the eigenvectors individually or in pairs was to reconstruct the space-time series associated with each pair or subset of eigenvectors (see Eq. (7)). This components represented the contribution to the original time series of a given set of eigenelements [43]. To examine the QB variability of SOI, we choose to sum over both pairs of significant eigenvalues with dominant period of about 30 months and its multiple. This yielded to a time series that appeared similar to the original in terms of accounting for major warm and cold events. Indeed, large positive peaks, those that exceeded one standard deviation, matched the strong La Niña events (Figure 3g). The only exceptions were the 1964, 1984-85, 2008-2009 cold events, which were of moderate strength. We observed the same good match between the large negative peaks and the strong El Niño events (Figure 3g). The only exceptions were the 1957-58, 1964-65 warm events and the large peak in 1977-78 which was classified by NOAA/CPC as weak events. The results were similar to those of [14]. Adding together the QQ, the 20-21 months and the 15-16 months oscillations for the reconstruction of Niño3 SST times series yielded to a time series that matched all strong events (Figure 3h). The correlation between the reconstructed Niño3 SST and the observed Niño3 SST was 0.89. This showed the very strong contribution of the combined 3 modes to ENSO.

SOI and Niño3 SST time series were also reconstructed from the IMFs of the EMD analysis (Figure 7m and 7n). A direct comparison with the MC-SSA analysis was impossible because the two methods did not detect the same significant interannual oscillatory modes. However, such as the MC-SSA (i) the EMD performed better at reconstructing the Niño3 SST time series than the SOI which showed finer temporal variations; (ii) it allowed us to detect most of the strong events and failed at recognizing weak events. In particular the 1964-1965 warm event was not detected.

The WT and the Hilbert-Huang analysis helped us at better characterizing this event because they allowed us to observe the variations in the frequency and in the amplitude of ENSO events over time. The 1964-65 event appeared sharp in the scale direction and

elongated in the time direction with small amplitude in the four wavelet power spectrum (Figure 4 and 5). The associated scale-average wavelet power showed a low amplitude when monthly SOI and SST variance were low, i.e. from 1960 to 1965. These results were consistent with previous studies [41]. They confirmed that the 1964-65 events was a long and moderate event and therefore more difficult to detect. Same features were detected with the Hilbert-Huang spectrum of both SOI and SST. Indeed, we observed a peak in frequency for all significant IMFs around 1964-65 associated with a small amplitude (Figure 7g and 7h).

Although the DXM does not allow yet to reconstruct the studied time series, this new representation facilitated the identification of regularities and irregularities in index patterns. El Niño and La Niña events, their intensity and interannual variations were clearly visible on the SOI and SST DXM maps with a quasi-period $d = 12$ (Figure 8b and 8c). SOI map showed finer temporal fluctuations than the Niño3 SST map but were in phase with the SOI. The well-studied ENSO regime shift [40] in 1976-77 was observed: both DXM map showed a change in the spatial pattern with the intensification of long and strong events after 1977. This regime shift was also detected in the four wavelet power spectrum (Figure 4 and 5) with a change in amplitude and frequencies at intermediate scales; but appeared in the early-1970s in Niño3 SST. A major change in ENSO dynamics before the mid-1970s regime shift was already reported [31] but is still controversial.

6.3 Limits and perspectives

The DXM turned out to be useful for identifying modes of interannual variability of ENSO. However, the method is still in experimental phase. As mentioned before, an analytical definition of the DXM is necessary to evaluate its performance. Moreover, several issues have been pointed out during simulation experiments:

- The DXM is a highly redundant algorithm; this should be taking into account for the definition of the confidence levels.
- The edge effects are very important because the borders of the DXM map has to be avoided during EAC calculation (See Eq. 28). Further works will aim at reducing edge effects using for example the traditional zero-padding technique or by duplicating the map along the x -axis.
- The DXM is very sensitive to the size and the shape of the reference window. Extensive simulation experiments are necessary to evaluate the influence of these parameters.
- Once the significant quasi-periods are identified though DXM, the method does not allow to reconstruct the studied time series. This is one of the main disadvantage of the methodology. Special efforts are required to extract useful information from DXM maps in order to reconstruct the space-time series.

From the climatological point of view, a scenario has been proposed to described the interannual variability of ENSO. This hypothesis has to be investigated using, for example, a coupled ocean-atmosphere climate model; and a collaboration with climatologist would be appreciate to investigate the physical process that give rise to each of the detected oscillation. In the long term, understanding ENSO variability will aid providing

important insights of the behaviour of the large scale precipitation system and related fired across New Caledonia. Indeed, it is now well establish that New Caledonia is largely affected by ENSO events. Methods, such as the multivariate SSA [14] or the cross wavelet transform [15], have been developed to extract the phase angle between two time series. An extension of the DXM methodology can be used to investigate such features [11]: from the DXM maps of two time series (with same quasi-period), one can quantify their multi-temporal similarities. The underlying hypothesis is that the impact of ENSO on the precipitation system could not only be attributed to an intrinsic chaotic behaviour or seasonal variation, but also to the inertia of the involved climatic phenomena.

7 Conclusion

This study offered a preliminary study of ENSO cyclostationarity (CS) and possible modulation of it. We proposed the definition of an extended CS, for which the CS property is averaged in a specific way to capture regularities in a signal autocorrelation with various time lags. This property, called the EAC, has been rigorously defined and offered another way to study signal memories. A new statistical tool, called DXM, was developed to study EAC in a signal and was used to examine the seasonal-to-interannual variability of ENSO time series. We performed a comparative analysis with three techniques for time series analysis, the MC-SSA, the WT and the EMD, that have been widely used in climatological studies. Two major interannual cycles of the coupled ocean-atmosphere were detected: the QB mode attributed to the atmospheric component of ENSO, and the QQ mode attributed to the ocean component of ENSO. The interactions of these two fundamental modes with the seasonal cycle lead to the apparition of three sub-harmonic periodicities, the 16-, 20- and 69-months oscillations. If confirmed, the results presented here may further help predicting and understanding the causes of ENSO events.

References

- [1] Allen, M.R., Interactions between the atmosphere and oceans on time-scales of weeks to years, Ph.D. thesis, 202pp, University of Oxford, Oxford, England, (1992)
- [2] Allen, M.R. and Smith, L.A., Investigating the origins and significance of low-frequency modes of climate variability, *Geophys. Res. Lett.*, 21:883-886, (1994)
- [3] Allen, M.R. and Smith, L.A., Monte Carlo SSA: detecting irregular oscillations in the presence of colored noise, *Journal of Climatology*, 9:3373-3404, (1996)
- [4] Bloomfield, P., Trends in global temperature, *Climatic Change*, 1:16, (1992)
- [5] Broomhead, D.S. and King, G. Extracting qualitative dynamics from experimental data, *Physica D*, 20:217-236, (1986)
- [6] Cane, M.A., The evolution of El Niño, past and future. *Earth and Planetary Science Letters* 230(3-4):227-240, (2005)
- [7] Xhen, D., Zebiak, S.E., Busalacchi, A.J., and Cane,M. A, An improved procedure for El Niño forecasting: implications for predictability, *Science*, 269:1966-1702,
- [8] Darwin 2009, Achievements Report. CSIRO Sustainable Ecosystems, Darwin, Northern Territory, www.csiro.au, (2009)
- [9] Flandrin, P., Rilling, G. and Gonçalves, P., Empirical mode decomposition as a filter bank. *IEEE Signal Process. Lett.*, 11:112-114, (2004)
- [10] Flandrin, P. and Gonçalves, P., Empirical mode decomposition as data-driven wavelet-like expansions. *Int J. Wavelets Multiresolut. inform. Process.*, 2:77-496 (2004)
- [11] Gaucherel, C. A study of the possible extended influence of the ENSO phenomenon. *Comptes Rendus Geoscience*, 33(6):175-185, (2004)
- [12] Gaucherel, C. Analysis of Enso interannual osccliations using non-stationary quasi-periodic statistics: a study of enso memory. *International journal of Climatology*, (2009)
- [13] Ghil, M. and Mo, K. Interseasonal oscillations in the global atmosphere. partI: northern hemisphere and topics. *Journal of the Atmospheric Sciences*, 48(5):752-79, (1991)
- [14] Ghil, M., Allen, M. R., Dettinger, M. D., Ide K., Kondrashov, D., Mann, M. E., Robertson, A. W., Saunders, A., Tian, Y., Varadi, F. and Yiou, P., Advanced spectral methods for climatic time series. *Reviews of Geophysics*, 40,1:1-1-41, (2002)
- [15] Grinsted, A., Moore, J.C. and Jevrejeva, S., Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear Porcesses in Geophysics*, 11:561-566, (2004)
- [16] Hann, S., Hilbert transforms of time-frequency analysis. Birkhauser, 359pp.
- [17] Hasselmann, K., Stochastic climate models. Part I: Theory, *Tellus*, 28(6):473-485, (1974)
- [18] Huang, N.E., Shen, Z., Lon, S.R., Wu, M.C., Shih, H.-H., Zheng, Q., Yen, N.-C. Tung, C,-C. and Liu, H.-H. The empirical mode decomposition and the Hilbert spectrum for nonlinear and nonstationary time series analysis, Proc. Roy. Soc. London Ser A,454:903-995, (1998)
- [19] Huang, J.P. and North, G.R. Cyclic spectral analysis of fluctuations in a GCM simulation. *Journal of the Atmospheric Sciences* 53(3):370-379, (1996)
- [20] Huang, N.E. and Wu, Z., Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Advances in Adaptive Data Analysis*, 1,1:1-41, (2009)
- [21] Jiang, N., Neelin, J.D. and Ghil, M. Quasi-quadrennial and quasi-biennial variability in the equatorial Pacific. *Climate Dynamics*, 12(2):101-112. (1995)

-
- [22] Jin, F., Neelin, D. J. and Ghil, M. El Niño/Southern Oscillation and the annual cycle: subharmonic frequency-locking and aperiodicity. *Physica D*, 98:442-465, (1996)
- [23] Kim, K.Y. and Wu, Q.G., A comparison study of EOF techniques: Analysis of nonstationary data with periodic statistics. *Journal of Climate* 12(1):185-199,(1999)
- [24] Lacoume, J.L, Amblard, P.O. and Comon, P., Statistique d'ordre superieure pour le traitement du signal, Masson: Pairs, pp290, (1995)
- [25] Latif, M., Anderson, D., Barnett, T., Cane, M., Kleeman, R., Leetmaa, A., O'Brien, J., Rosati, A., and Schneider E. A review of the predictability and prediction of ENSO. *Journal of geophysical research*, 103(C7):14,375-14,393, (1998)
- [26] Lund, R.B., Hurd, H.L., Bloomfield, P. and Smith R., Climatological time series with periodic correlation. *Journal of Climate* 8:2787-2809, (1995)
- [27] Mallat, S., A wavelet tour of signal processing, Academic press, San Diego, (1998)
- [28] Mann, M.E. and Lees, J.M. Robust estimation of background noise and signal detection in climatic time series, *Climate Change*, 33:409-445, (1996)
- [29] Moliere, A. and Rebert, J.P., Rainfall shortage and El Niño-Southern Oscillation in New Caledonia, Southwestern Pacific. American Meteorological Society, 114,6, (1986)
- [30] Moron, V., Vautard, G. and Ghil, M., Trends, interdecadal and interannual oscillations in global sea-surface temperatures. *Climate Dynamics*, 14:545-569, (1996)
- [31] Neelin, J.D., Latif, M. and Jin F-F. Dynamics of coupled ocean atmosphere models. The tropical problem, *Annu Rev Fluid Mech*, 26:617-659, (1994)
- [32] Nicet, J.B. and Delcroix, T. ENSO-related precipitation changes in New Caledonia, Southwestern Tropical Pacific: 1969-98. *American Meteorological Society*, (2000)
- [33] Percival, D.P., On estimation of the wavelet variance. *Biometrika*, 82:691-631, (1995)
- [34] Philander S.G.H., El Niño, La Niña and the Southern Oscillations. *Academic Press: San Diego*, pp 293, (1990)
- [35] Rasmusson, E.M. and Carpenter, T.H. Variations in tropical sea surface temperature and surface wind fields associated with the Southern Oscillation/EL Niño. *Mon Weather Rev*, 110:354-384. (1982)
- [36] Rasmusson, E.M., Wang,X., Ropolewsky, C.F. The biennial component of the ENSO variability. *Journal of Marine Sciences*, 1:71:96, (1990)
- [37] Rilling, G., Flandrin, P. and Gonçalvès, P. On empirical mode decomposition and its algorithms. *IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing NSIP-03*, Grado (I).
- [38] Salisbury, J.I. and Wimbush, M., Using modern time series analysis techniques to predict ENSO events from SOI time series. *Nonlinear processes in geophysics*, 9:341-345, (2002)
- [39] Saunders, A. and Ghil, M., A boolean delay equations model of ENSO variability. *Physica D*, 2801:1-25, (2001)
- [40] Setoh, T., Imawaki, S., Ostrovskii, A. and Umatani, S-I, Interdecadal variation of ENSO signals and annual cycles revealed by wavelet analysis. *Journal of Oceanography*, 55:385-394, (1999).
- [41] Torrence, C. and Compo, P. G., A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society*, (1997)

-
- [42] Torrence, C. and Webster, P.J. The annual cycle of persistence in the El Niño/Southern Oscillation, *Q. J. R. Meteorological Society*, 124:1985-2004. (1998)
 - [43] Vautard, R., and Ghil, Singular spectrum analysis in non-linear dynamics, with applications to paleoclimatic time series, *Physica D.*, 35:395-424: (1989)
 - [44] Vautard, R. and Ghil, M. Singular spectrum analysis: A toolkit for short noisy chaotic signals. *Physica D*, 58:95-126, (1992)
 - [45] Weiss, P.J. and Weiss B.J. Quantifying persistence in ENSO, *Journal of the atmospheric sciences*, 55(16):2737-2760, (1999)
 - [46] Wu, B.Z. and Huang, N.E., A study of the characteristics of white noise using the empirical mode decomposition, *Proc. R. Soc. Lond.*, 406A:1957-1611, (2004)
 - [47] Wu, Z., Schneider, E.K., Hu, Z.Z. and Cao, L. The impact of global warming on ENSO variability in climate records, COLA Technical Report, CTR 110, Oct. 2001.
 - [48] Yiou, P., Ghil, M., Jouzel, J., Paillard, D. and Vautard, R. Nonlinear variability of the climatic system, from singular and power spectra of late quaternary records. *Clim. Dyn.*, 9:371-389., (1994)
 - [49] Yiou, P., Sornette, D. and Ghil, M. Data-adaptive wavelets and multi-scales SSA. *Physica D*, (1998)
 - [50] Zhang, Y., Sheng, J. and Shabbar, A., Modes of interannual and interdecadal variability of Pacific SST, *Journal of Climate* 11:2556-2569, (1998)

ANNEX 1 : El Niño - Southern Oscillation (ENSO)

Now well known to scientists, the *El Niño-Southern Oscillation* (ENSO) was discovered in stages. The term *El Niño* ("the boy child" in Spanish) was likely coined in the 19th century by Peruvian fishermen who noticed the appearance of a warm current water every few years around Christmas. The cause of the current's appearance was a mystery to them. In 1899, India experienced a severe drought-related famine, prompting greater focus on understanding the India monsoon system, arguably the nation's most important source of water. In the early 1900's, the British Mathematician Sir Gilbert Walker noticed a statistical correlation between the monsoon's behaviour and semi-regular variation in atmospheric pressure over the tropical Pacific. He coined this variation the *Southern Oscillation*, defined as the periodic shift in atmospheric pressure differences between Tahiti and Darwin, Australia. It was not until 1969, however, that meteorologist Jacob Bjerknes proposed that *El Niño* phenomenon of the coast of South America and the *Southern Oscillation* were linked through a circulation system that he termed the *Walker circulation*. ENSO has since become recognized as the strongest and most ubiquitous source of inter-annual climate variability.

ENSO is thus defined as a quasi-periodic shift in the distribution of heat across the tropical Pacific. The phenomenon is known as a *oceanic-atmospheric coupled system* because the different components drive and affect each other. Under normal conditions, the water on the surface of the ocean is warmer than at the bottom because it is heated by the sun. In the tropical Pacific, atmospheric winds are driven by a surface pressure pattern of higher pressure in the eastern Pacific and lower pressure in the west. These winds tend to push the surface water toward the west. As the water moves west it heats up even more because it's exposed longer to the sun. Meanwhile in the eastern Pacific along the coast of South America an upwelling occurs. Upwelling is the term used to describe when deeper colder water from the bottom of the ocean moves up toward the surface away from the shore. This nutrient-rich water is responsible for supporting the large fish population commonly found in this area. Indeed, the Peruvian fishing grounds are one of the five richest in the world. In the east the water cools the air above it, and the air becomes too dense to rise to produce clouds and rain. However; in the western Pacific the air is heated by the water below it, increasing the buoyancy of the lower atmosphere thus increasing the likelihood of rain. This is why heavy rain storms are typical near Indonesia while Peru is relatively dry.

El Niño phase happens when weakening trade winds allow the warmer water from the western Pacific to flow toward the east. This flattens out the sea level, builds up warm surface water off the coast of South America, and increases the temperature of the water in the eastern Pacific. The deeper, warmer water in the east limits the amount of nutrient-rich deep water normally surfaced by the upwelling process. These nutrients are vital for sustaining the large fish populations normally found in the region and any reduction in the supply of nutrients means a reduction in the fish population. Convective clouds and heavy rains are fueled by increased buoyancy of the lower atmosphere resulting from heating by the warmer waters below. As the warmer water shifts eastward, so do the clouds and thunderstorms associated with it. Thus, rains which normally would fall over the tropical rain forests of Indonesia start falling over the deserts of Peru, causing forest fires and drought in the western Pacific and flooding in South America. Moreover the Earth's atmosphere responds to the heating of *El Niño* by producing patterns of high and low pressure which can have a profound impact on weather far away from the equatorial

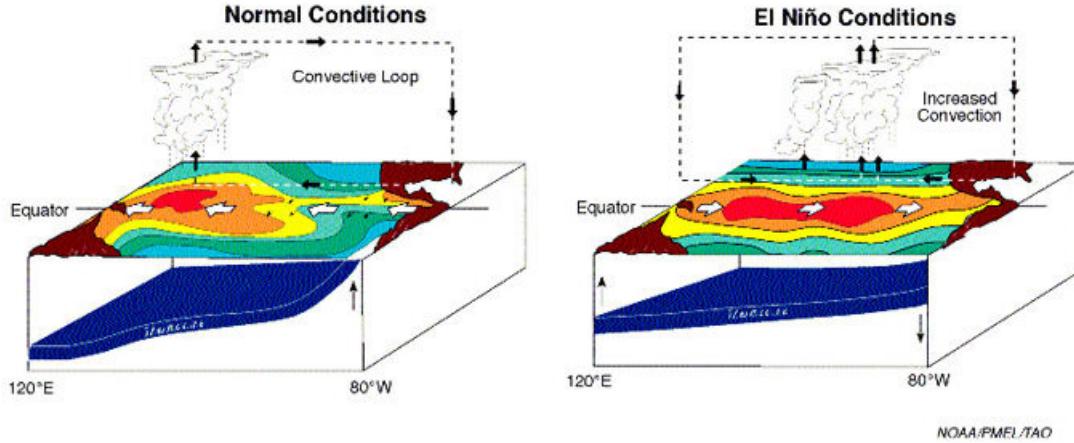


Figure 9: Normal and El Niño conditions

Pacific.

During the mid to late 1990s, these phenomena garnered much attention throughout the globe. Increased scientific investigation and dramatic meteorological impact in various parts of the world brought this complex event into media focus. ENSO offers the greatest potential for seasonal to interannual climate predictions for societal benefit. Unfortunately the causes of ENSO events and their fluctuations remain unclear although several theories have been put forward.

ANNEX 2: Definition of the confidence levels

MC-SSA confidence level

The ‘art’ of SSA is to select proper components with desired behaviour and thereby effectively filter the time series. To this purpose, one has to evaluate the statistical significance of the components found. In this study, the resulting spectral decomposition is tested against the null hypothesis that the time series is compatible with a white noise process or a red noise process. In SSA, the distinction of ‘interesting’ components (signal) from noise is based on finding a threshold (jump-down) to a ‘noise floor’ in a sequence of eigenvalues given in descending order. The SSA applied to a white noise process leads to a flat singular spectrum (Figure 10). This approach may be problematic if the signal-to-noise ratio is not sufficiently large, or the noise present in the data is not white but ‘colored’ and generally red [17]. In the classical SSA approach (comparison with a white noise) applied to red noise, the eigenvalues of the slow modes might incorrectly be interpreted as a nontrivial signal, or, on the other hand, a non trivial signal embedded in red noise might be neglected if its variance is smaller than the slow-mode eigenvalues of the background white-noise. Therefore, [47] proposed comparing SSA spectrum of the analyzed signal with the SSA spectrum of a red-noise model fitted to the studied data. In this case, the eigenvalues related to the slow modes are much larger than the eigenvalues of the modes related to higher frequencies.

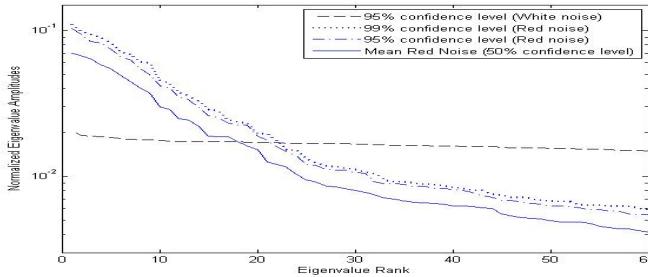


Figure 10: SSA confidence level from 500 monte carlo simulations. The blue dash line indicates the 95% and 99% confidence level for a red noise processes estimated from the SOI time series . The blue line represents the mean spectra of the 500 simulations of red noise. The black dash line is the 95% confidence level for a white noise null hypothesis.

Wavelet confidence level

To determine significance levels for wavelet spectra, one first needs to choose an appropriate background spectrum. In a crude approximation, background climate variability can be described by a simple random forcing (white noise) process. White noise present a flat frequency global spectrum (Figure 11). This provides an unbiased and consistent estimation of the true power spectrum of the time series. However, spectra of climatic time series frequently show a continuous decrease of spectral amplitude with increasing frequency [14, 41]. Peaks in the global power at small scales might incorrectly be in-

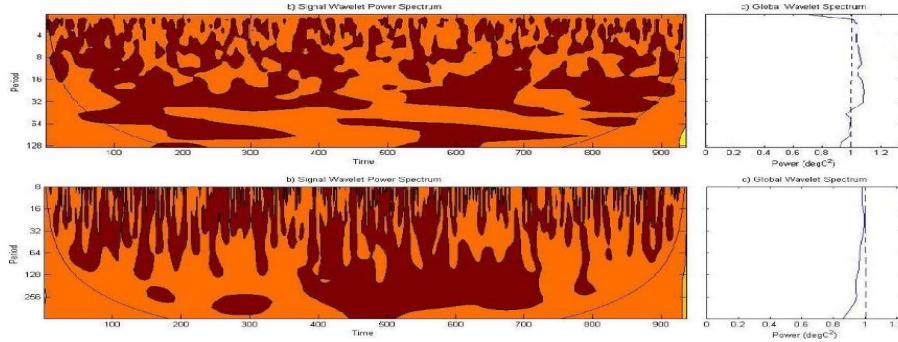


Figure 11: **Mean global power spectra of 500 realisations of white noise.** On the top the Morlet basis is used. On the bottom, the Mexican hat basis is used. The red color indicates a strong energy. Both associated global power spectra are displayed on right. The blue dashed line indicates the theoretical distribution. The blue line is the result o the Monte Carlo simulation.

terpreted as a nontrivial signal, or, on the other hand, a non trivial signal embedded in red noise might be neglected if the global power at small scale is smaller than the global power of the background with-noise. Thus, a second and more stringent case is that of red noise [1, 17]. As a null hypothesis, it is assumes that the studied time series has a mean power spectrum, possibly given by Eq. 10; if a peak in the wavelet power spectrum is significantly above this background spectrum, then it can be assumed to be a true feature with a certain confidence (Figure 12).

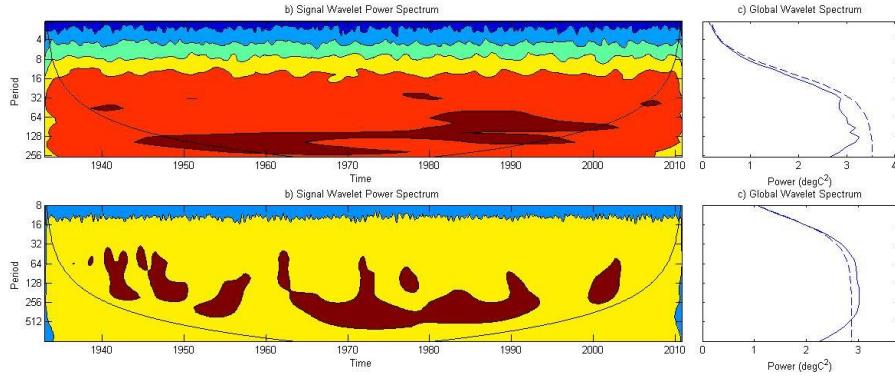


Figure 12: Mean global power spectra of 500 realisations of red noise estimated from the SOI data. On the top the Morlet basis is used. On the bottom, the Mexican hat basis is used. The red-orange color indicates strong energies while the green-blue color is used for small energies. Global power spectra are displayed on right. The blue dashed line indicates the theoretical distribution o a red noise process. The blue line is the result of the Monte Carlo simulation.

EMD confidence level

Statistical characteristics of uniformly distributed white noise, including the linear expression for the relationship between the energy density of the IMFs and the average period has been established by [46]. We present here a Monte Carlo verification of the relation between the energy density and the averaged period. The energy density of the IMFs as a

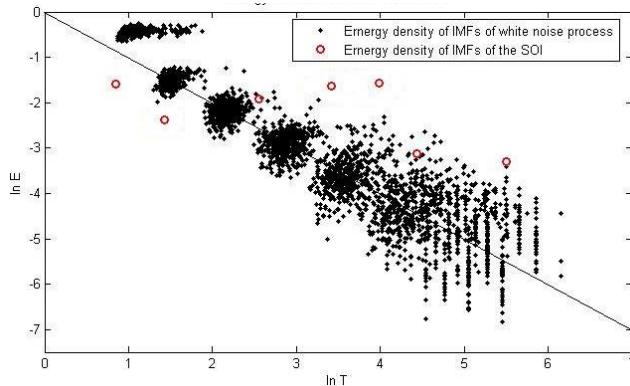


Figure 13: Monte Carlo verification of the relation between the energy density and the averaged period. The groups of the dots from upper left to the lower right are the energy density as a function of the averaged period for IMFs 1-7 for all 500 sample of white noise with identical length of 936 data points (length of SOI time series), respectively. The circle is the energy density of an IMF of the SOI as a function of its corresponding averaged period.

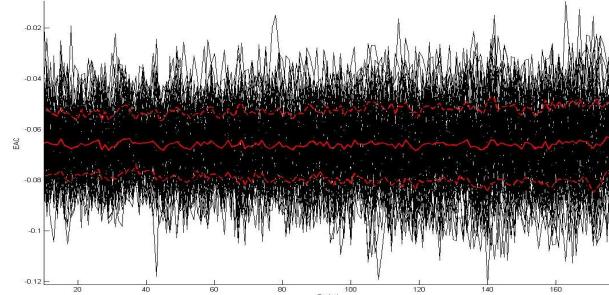
function of the mean period, derived from white noise processes, on a log-log scale formed a straight line. These results are consistent with the study of [46]. The normal distribution of the energy density allows one to deduced the 95% confidence level (not shown here). Finally by comparing, the energy density for the IMFs from SOI with the spread function, ones can identify IMFs that have its energy locates above the upper bound and below the lower bound. The latter should be considered containing information at that the selected

confidence level.

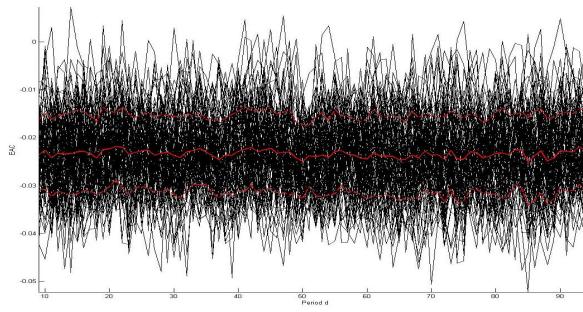
In this study, we want to determine which IMFs from SOI and Niño3 SST data set contains information, and which IMFs are red noise. As a primary study, tests have been conducted as follows: first, we decompose the target noisy data set into IMFs. Second, we construct 500 red noise records of identical length, with the target data set, as a reference. We apply the EMD method to decompose each of the 500 red noise data into IMFs. We display the energy density of the IMFs as a function of the mean period derived from white noise processes on a log-log scale. Finally, we compare the energy density for the IMFs from the data with the spread function. IMFs that stand out of the scatter plot are considered containing significant information (see Figure 7e and 7f). This experiment still necessitate a strong analysis to assess the statistical characteristics of red noise process.

DXM confidence level

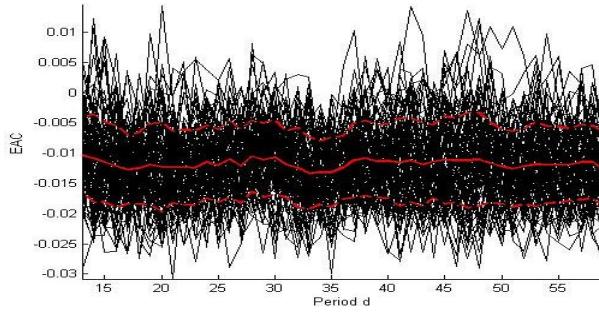
Confidence level were estimated using the Monte Carlo method. 500 random time series of the same length of the studied data were simulated. The distribution of their EAC profile (back curves) calculated with three different reference window size is displayed in Figure 7. The normal distribution allowed us to calculate the mean of the EAC profile (red curve) and the 95% confidence interval (red dash curve). They appeared quite stable. For comparison of the three simulations, the mean of the EAC profile and of the 95% confidence level was calculated for a discrete set of 40 ($d = 15, \dots, 55$) quasi-periods and are given in Table 1 (see section 5 for discussion).



(a) EAC profile of 500 white noise series calculated with a moving window of size (3×3) pixels.



(b) EAC profile of 500 white noise series calculated with a moving window of size (5×5) pixels.



(c) EAC profile of 500 white noise series calculated with a moving window of size (7×7) pixels.

Figure 14: Monte Carlo simulation and EAC profile for different reference window size.

ANNEX 3: Test procedure

The DXM method was experimented using a 2D discrete gauss curve described by Eq. (2). We generated a square matrix of 900 data points which is approximately the size of ENSO indices. This matrix represented a DXM map (not interpolated) with quasi-period $d = 30$. The signal was then unfolded to create a 1D vector data set as shown in Figure 10. In this way, we created a signal which present a strong spatial autocorrelation in its 2D form for a quasi-period $d = 30$.

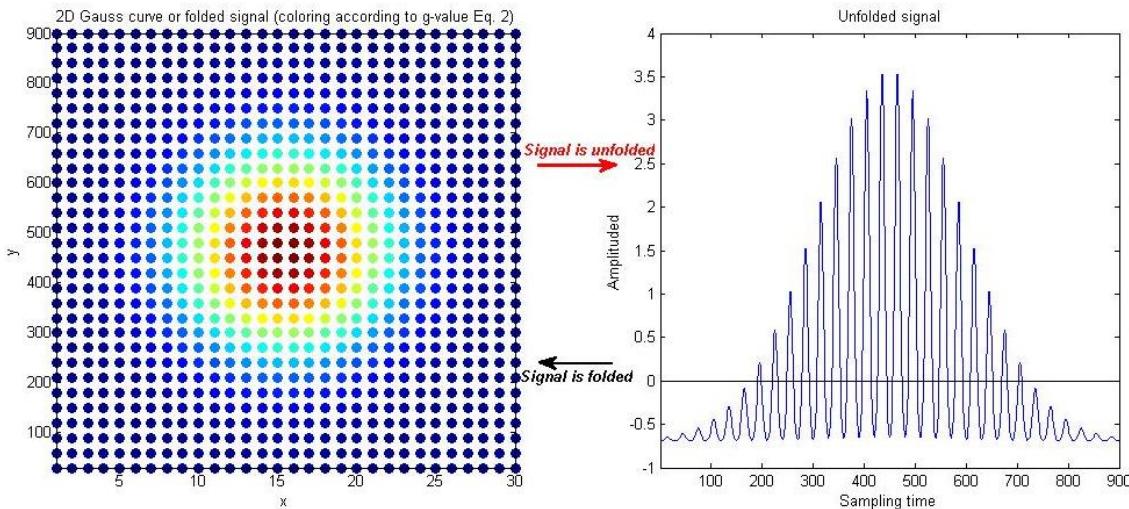


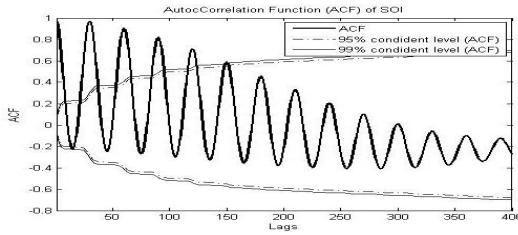
Figure 15: **Simulated signal.** DXM map of a 2D gauss on left (folded signal). A color scale is attributed to the amplitude of the signal. Unfolded signal on right.

The EAC profile of the unfolded signal was numerically computed for a set of 90 ($d = 10, \dots, 100$) quasi-periods. We used a reference window of size (3×3) pixels. The confidence level were determined using Monte Carlo simulations of 500 white noise. The EAC profile was compared to the classical autocorrelation function (ACF). The results are displayed in Figure 16.

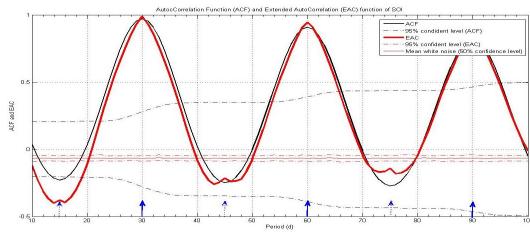
The DXM was able to detect both correlation (at quasi-period 30) and anti-correlation

(at quasi-period 14,15,16). The two peaks were highly significant. Thought the DXM was very efficient at identifying a spatial autocorrelation, the method was sensitive to the multiples. Indeed, peaks of smaller amplitude appeared at quasi-periods 44, 45, 46 and 60 as well as quasi-periods 74, 75, 76 and 90. This may confused or hide a longer periodicity in the signal.

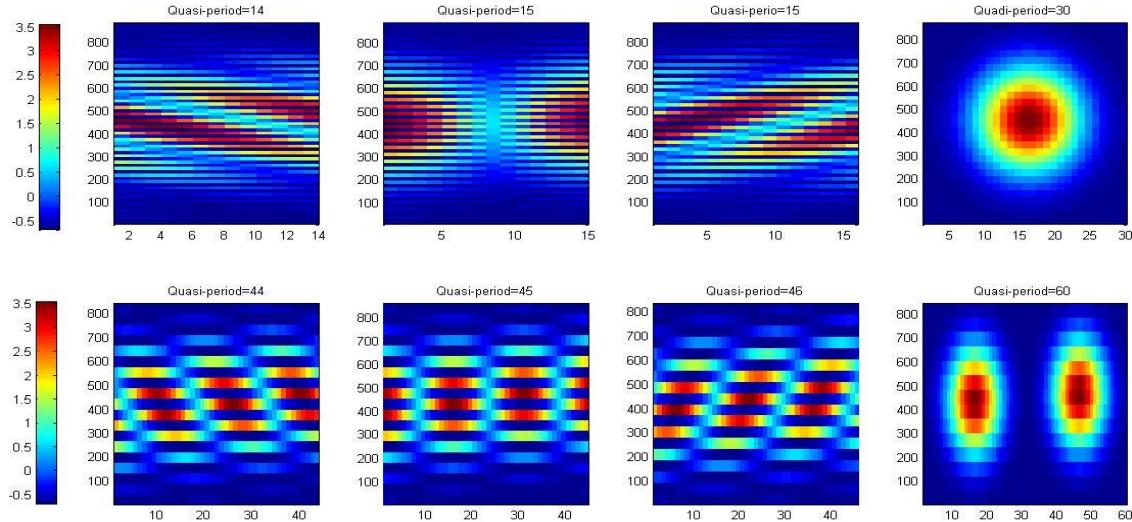
To check whether or not statistical artefacts were introduced, we experimented with a reference window of size (5×5) and (7×7) pixels (results not shown). We found that the amplitude of the peaks (extrema) in EAC profile varied somewhat with window but peaks appeared at the same quasi-periods.



(a) Autocorrelation function (ACF) of the simulated signal. The dashed and thin black lines represent the 95% and 99% confidence level, respectively.



(b) Autocorrelation (ACF) function (black) and Extended autocorrelation (EAC) function (red) of the simulated signal. The dashed lines represent the associated 95% confidence level. The thin red line is the mean EAC of the 500 white noise simulations. The x -axis corresponds is a lag-time for the ACF and a quasi-period for the EAC



(c) DXM map of the simulated signal with quasi-period d corresponding to the 8 first extrema of the EAC profile.

Figure 16: DXM analysis of the 2D gaus curve