



**HAL**  
open science

# Multi-scale event-based mining in geophysical time series: characterization and distribution of significant time-scales into the Sea Surface Temperature anomalies relatively to ENSO periods from 1985 to 2009

Bertrand Saulquin, Ronan Fablet, Grégoire Mercier, Hervé Demarcq, Antoine Mangin, Odile Fanton d'Andon

## ► To cite this version:

Bertrand Saulquin, Ronan Fablet, Grégoire Mercier, Hervé Demarcq, Antoine Mangin, et al.. Multi-scale event-based mining in geophysical time series: characterization and distribution of significant time-scales into the Sea Surface Temperature anomalies relatively to ENSO periods from 1985 to 2009. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2014, 7 (8), pp.1939-1404. 10.1109/JSTARS.2014.2329921 . hal-01208386

**HAL Id: hal-01208386**

**<https://hal.science/hal-01208386>**

Submitted on 29 Apr 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

# Multiscale Event-Based Mining in Geophysical Time Series: Characterization and Distribution of Significant Time-Scales in the Sea Surface Temperature Anomalies Relatively to ENSO Periods from 1985 to 2009

Bertrand Saulquin, Ronan Fablet, Grégoire Mercier, Hervé Demarcq, Antoine Mangin, and Odile Hembise Fanton d'Andon

**Abstract**—In this paper, one-dimensional (1-D) geophysical time series are regarded as series of significant time-scale events. We combine a wavelet-based analysis with a Gaussian mixture model to extract characteristic time-scales of 486 144 detected events in the Sea Surface Temperature Anomaly (SSTA) observed from satellite at global scale from 1985 to 2009. We retrieve four low-frequency characteristic time-scales of Niño Southern Oscillation (ENSO) in the 1.5- to 7-year range and show their spatial distribution. High-frequency (HF) SSTA event spatial distribution shows a dependency to the ENSO regimes, pointing out that the ENSO signal also involves specific signatures at these time-scales. These fine-scale signatures can hardly be retrieved from global EOF approaches, which tend to exhibit uppermost the low-frequency influence of ENSO onto the SSTA. In particular, we observe at global scale a major increase by 11% of the number of SSTA HF events during Niño periods, with a local maximum of 80% in Europe. The methodology is also used to highlight an ENSO-induced frequency shift during the major 1997–2000 ENSO event in the intertropical Pacific. We observe a clear shift from the high frequencies toward the 3.36-year scale with a maximum shift occurring 2 months before the ENSO maximum of energy at 3.36-year scale.

**Index Terms**—Distribution of the sea surface temperature anomalies events related to the ENSO periods, event-based mining in large geophysical datasets (big data), geophysical time series as series of significant time-scale events.

## I. INTRODUCTION

**M**ANY INFORMATION sources, including instrumental *in situ* data records and satellite observations, highlight

B. Saulquin is with the ACRI-ST, Sophia-Antipolis, Sophia-Antipolis 06904, France, and also with the Institut Mines-Telecom, Télécom Bretagne, Brest 29238, Cedex, France (e-mail: bertrand.saulquin@acri-st.fr).

R. Fablet and G. Mercier are with the Institut Mines-Telecom, Télécom Bretagne, Brest 29238, France, and also with the Université Européenne de Bretagne, Rennes 35000, France.

H. Demarcq is with the Centre de Recherche Halieutique Méditerranéenne et Tropicale, Institut de Recherche pour le Développement (IRD), IRD/Ifremer/Université Montpellier, Sète 34200, France.

A. Mangin and O. H. Fanton d'Andon are with the ACRI-ST, Sophia-Antipolis, Sophia-Antipolis 06904, France.

the great variability and the nonstationarity of the earth's climate over a wide range of time-scales from months to decades. The most widely used technique to investigate the spatiotemporal variability of climate-relevant time series, such as temperature [1] and wind [2], relies on the empirical orthogonal functions (EOFs) [3], also referred to as principle component analysis (PCA) in the literature. This method combines the extraction of the main deformation modes of the covariance (correlation) of a univariate or bivariate [singular value decomposition (SVD)] datasets and the analysis of the time correlation of these principal modes with potential causing factors. An introduction to univariate and multivariate EOF analysis may be found in [4] and a thorough review of advanced EOF-based methods along with intercomparisons is presented by Bretherton *et al.* [5]. These EOF-based analyses, however, suffer from intrinsic limitations, the main one being the assumption that the considered processes are stationary. Nevertheless, geophysical dynamics widely involve nonstationary processes (e.g., emergence of extreme events including, e.g., large Niño events, time shifts of seasonal cycles, propagation phenomena, and trends), which may hardly be characterized as stationary. Environmental data also involve strong autocorrelation level [6]. For instance, a positive anomaly in the observed wind or temperature at a given day (week) is often associated with similar conditions the following days (weeks). This natural autocorrelation is the result of short time and local events but also of large-scale signals such as, e.g., the well-known El-Niño/La-Niña oscillation [7], [8]. Such autocorrelated level clearly affects the determination of correlation significance level used as input of the EOF [9]. Both nonstationarity and autocorrelation may affect the interpretation of the extracted principal modes. Nevertheless, these aspects are often overlooked by EOF-based approaches. Besides, EOF is also known to be prone to outliers [9] and can hardly reveal fine time-scales signatures, which typically involve greater nonstationary variabilities.

Wavelet analysis is particularly appealing to address these issues. In contrast to EOF-based approaches, wavelet analysis

actually addresses the decomposition of the fluctuations exhibited by nonstationary signals. An introduction to wavelet analysis related to climate research is given by Torrence and Compo [7]. Wavelet analysis has been used to investigate global climate changes in sea surface temperature (SST) [10], [11] and interactions between physical parameters such as SST and sea surface height (SSH) [12]. While EOF-based schemes aim at extracting the main patterns of the covariance (or correlation) structure, wavelet analysis identifies and characterizes local time-scale patterns. The correlation between two processes can also be decomposed in the time-scale domain based on the wavelet coherency spectrum [7], [13].

In this paper, we further investigate wavelet analysis for geophysical time series to develop an event-based representation and analysis of a geophysical dataset. Formally, we regard a time series as a collection of significant elementary time-scale events and use an unsupervised clustering method, namely a Gaussian mixture mode (GMM) [14], to characterize the significant time-scales of the dataset.

To illustrate our approach, we study the SST anomalies (SSTAs) observed over the globe from 1985 to 2009. First, we characterize four characteristic low-frequency patterns in the SSTA time-scale distribution and study their space and time distribution regarding the ENSO modes. We also focus onto the high-frequency (HF) SSTA and show a strong spatial signature of ENSO. Usually, only the low frequency in the SSTA (from 1.5 to 8 years) is attributed to ENSO [8], [15] and previous works (e.g., Enfield [15]) use an EOF decomposition of the filtered SSTA in the 1.5- to 8-year range. We underlie here that ENSO events also depict HF signatures in the SSTA. This paper does not support a full description of the ENSO phenomenon, but it uses knowledge on its interactions with the SSTA to illustrate the added value, compared to standard methods, of both the decomposition of the time series using an events-based framework and the proposed data mining approach.

## II. EVENT-BASED ANALYSIS OF GEOPHYSICAL TIMES SERIES

### A. Wavelet-Based Extraction of Elementary Time-Scale Events

Wavelet analysis aims at characterizing nonstationary signals, i.e., signals whose statistical characteristics (e.g., mean and variance) may change over time. From the decomposition of a one-dimensional (1-D) signal in the time-scale domain, significant frequencies can be detected in any given time interval. Such decompositions typically achieve a better detection and description of the characteristic time-scale variabilities of the observed phenomenon [7] and represent a real added value to unmix nonstationary scale-dependent processes compared to classical covariance-based analysis (e.g., EOF-based schemes [3] and autoregressive models [16]). Formally, the wavelet transform of a 1-D signal consists in computing the complex wavelet coefficients  $W(s, T)$  as the projection of the signal onto scaled and translated versions of the selected mother wavelet  $\Psi$  [17]

$$W(s, T) = (1/\sqrt{s}) \int z(t) \Psi * \left( \frac{t-T}{s} \right) dt \quad (1)$$

where  $s$  is the time-scale,  $t$  and  $T$  time instants, and  $\Psi^*$  stands for the conjugate complex of the mother wavelet  $\Psi$ . Since the wavelet transform computes the similarity between the wavelets and the signal, the choice of the mother wavelet is important. SST is often modeled using harmonics [19] and Gu and Philander [17] suggest that the ENSO signal may be represented by sinusoids. This supports the choice of the Morlet wavelet, stated as a time-windowed pure harmonic component. The wavelet power spectrum of the time-scale decomposition (1) is defined as

$$P(s, T) = |W(s, T)|^2. \quad (2)$$

Here, we propose to model the studied geophysical process as a collection of significant elementary time-scale events. Formally, this amounts to viewing the spectrum as a sum of  $K$  individual events and a red noise

$$P(s, T) = \sum_{j=1}^K P_j(s, T) + P_r(s, T) \quad (3)$$

where  $P_j(s, T)$  is the detected event  $j$ ,  $P_r(s, T)$ , a theoretical red noise Fourier power spectrum [7], whose relevance, compared to the white noise model, is acknowledged for geophysical processes (a positive anomaly in the observed temperature on a given day is often associated with similar conditions the following days). The first-order red noise is characterized as

$$r(t) = \alpha r(t-1) + \epsilon \quad (4)$$

where  $\alpha$  is the lag-1 autocorrelation, i.e., the mean correlation between samples at the current and preceding time steps.  $\epsilon$  is a white noise process with zero mean and variance  $\sigma^2$ . If  $\alpha = 0$  (4), it resorts to the white noise model. For a given time series, we use a robust estimation of the noise model parameters, i.e., autocorrelation coefficient  $\alpha$  [20] and variance  $\sigma^2$  [21]. Regarding the significance level, we follow Torrence and Compo [7] who showed that if the original signal's Fourier components are normally distributed, then the wavelet power spectrum  $P$  is  $\chi^2$  distributed. The associated 95% confidence level is then obtained using

$$P(s; \alpha) > = 0.5\sigma^2 P_r(s; \alpha) \chi_2^2(95\%) \quad (5)$$

where  $\sigma^2$  is the variance of the noise model (in practice, we used a robust estimation of the variance of the time series) and  $\chi_2^2$  is the chi-square distribution with two degrees of freedom.  $P_r(s; \alpha)$  is the theoretical Fourier power spectrum of the red noise

$$P_r(s; \alpha) = (1 - \alpha^2) / \left( 1 + \alpha^2 - 2\alpha \cos\left(\frac{2\pi}{s}\right) \right). \quad (6)$$

Fig. 1 shows the distribution of the theoretical Fourier power spectrum  $P_r$  (6) for both a red ( $\alpha = 0.5$ , red curve) and a white noise ( $\alpha = 0$ , blue curve). Their corresponding dashed lines represent the 95% confidence level (5) for  $\sigma$  value set to 1. As illustrated (Fig. 1), if the autocorrelation of the noise is ignored in the analysis, it typically leads to an over-detection of the

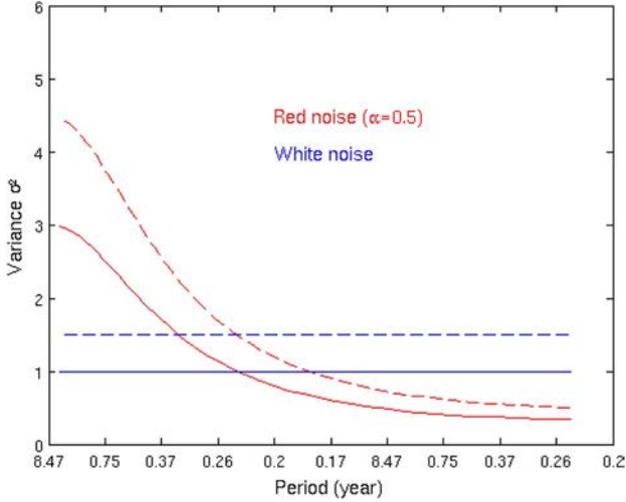


Fig. 1. Theoretical Fourier spectrum (6) as a function of the period for a white noise (blue curve) and a red noise (red curve), this latest being the representative of a geophysical time series. In dashed lines, the corresponding 95% confidence levels (5).

low-frequency component of the signal and an under-detection of its HF component.

To match the time-scale dimension of the wavelet power spectrum  $P(s, T)$  (2), the theoretical 1-D Fourier power spectrum  $P_r$  (6) is expanded for each time step  $t$

$$P_r(s, t; \alpha) = P_r(s; \alpha) \quad \forall t. \quad (7)$$

Our implementation of the wavelet analysis involves 69 time-scales  $s_j$  ranging from 0.2 to 8.5 years:

$$s_j = s_0 2^{j+1} d_j, \quad j = 0 \dots J \quad (8)$$

with  $s_0 = 1/6$  year (2 months) and the time step  $d_j = 1/12$  year (1 month) determine the smallest time-scale (2 months) and  $J$  the largest scale (8.5 years).

The determination of the elementary time-scale events from the wavelet spectrum  $P(s, T)$  [22] (see Section III-C for an example of wavelet power spectrum) involves the extraction of local regions of interest as maximal level-sets [23], [24], i.e., areas of the wavelet spectrum which depict an energy level above the significance level [6]. In our implementation, we first detect all the significant local maxima [22] in the valid part of the wavelet spectrum, i.e., out of the cone of influence [7], and then determine their associated maximal level-set. A maximal level set is the largest time-scale area in the spectrum, which contains only one maximum of energy. Finally, an ellipse is fitted onto the selected spectrum area (event).

Each ellipse is described by its time and time-scale extensions and the position of its center (local maximum of energy). Ellipse axes refer to the main axes of the time-scale covariance, which are, even if not addressed in this paper, directly linked to the frequency modulation observed during the propagation of the event [13].

### B. Event-Based Mining of the Event Database

Our event-based mining strategy involves an unsupervised analysis of the time-scale distribution of the elementary events.

This distribution is modeled as a mixture of Gaussian modes. By nature, when considering time-scale analysis, occurrences of HF events are greater than those at low-frequencies. This natural distribution of the scale of the events is referred in the literature as a fractal distribution [25]. To account for this scale-dependent sampling, the mixture model involves a scale-dependent normalization factor. Formally, the considered normalized mixture model  $f(s)$  resorts to

$$f(s) = E(s) \sum_{i=1}^I (\lambda_i N(s; \mu_i, \sigma_i)) \quad (9)$$

where  $E(s)$  is the scale-related normalization accounting for the global scale-dependent sampling of the elementary events,  $I$  is the number of modes (Gaussians),  $\lambda_i$  the prior of the mode  $i$  of the mixture,  $N$  the normal probability density function (PDF) of the time-scale events with mean  $\mu_i$ , and standard deviation  $\sigma_i$

$$N(s; \mu_i, \sigma_i) = \frac{1}{\sigma_i \sqrt{2\pi}} \exp \left[ -\frac{1}{2\sigma_i^2} (s - \mu_i)^2 \right]. \quad (10)$$

An exponential distribution for  $E(s)$  is proven meaningful (see Section IV-D)

$$E(s) = \gamma e^{-\gamma s}. \quad (11)$$

To infer the parameters of the mixture model, we first fit the normalization factor  $\gamma$  and in a second step, mixture model parameters  $\mu_i$  and  $\sigma_i$  are estimated using an EM (Expectation–Maximization) procedure [26], which aims at maximizing  $f(s)$  (9) or minimizing the log likelihood

$$L = -\log(f(s)). \quad (12)$$

For a given initialization for model parameters, the EM procedure iterates expectation steps (E-step), which compute the posterior likelihoods given current model parameters, and maximization steps (M-step) to update the model parameters given the posteriors. The algorithm iterates until numerical convergence  $|L(n+1) - L(n)| < 10^{-4}$ . The estimation of the number of modes of the mixture model proceeds as follows: given 15 initial modes in the mixture model, only the modes with  $\sigma_i > 0.05$  year are kept in the model after each EM step.

## III. APPLICATION TO THE SATELLITE-DERIVED SSTA OBSERVED FROM 1985 TO 2009

### A. Pathfinder Dataset

Satellite-derived SST data are extracted from the global AVHRR Pathfinder SST v5.2 [27] daily gridded product (<http://www.nodc.noaa.gov/SatelliteData/pathfinder4km/>). To avoid diurnal effect, we used the data acquired at nighttime. A quality control was performed by selecting pixels with a quality flag level greater than 3. This quality flag is provided in the Pathfinder v5.2 product, and its level was determined using Kilpatrick studies [28]. First, the estimation of the SSTA involves the estimation of monthly mean SST fields at 36 km

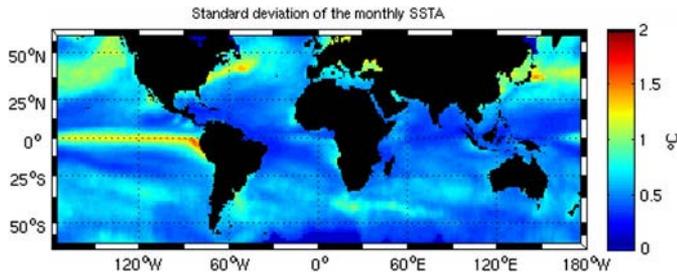


Fig. 2. The standard deviation of the monthly SSTA for the period 1985–2009 (source: Pathfinder v5.2).

resolution. A minimum of 30 observations per grid cell is used to estimate the average. The seasonal signal (climatology) must be then removed from the SST to obtain the SSTA. The harmonic-based estimation of a climatology of is more accurate than the simple average estimation [30]. Hence, for each time series, a local climatology  $S_t$  composed of four harmonics and a linear trend [29] is estimated and subtracted from the SST to remove 12, 6, 4, and 3 months periodicities

$$S_t = \sum_{i=1}^4 a_i \cdot \cos\left(\frac{2\pi it}{12}\right) + b_i \cdot \sin\left(\frac{2\pi it}{12}\right). \quad (13)$$

Finally, SSTA monthly fields were spatially averaged on a  $1^\circ \times 1^\circ$  grid. The resulting studied SSTA dataset is a  $180 \times 360 \times 300$  matrix. We removed land cells and obtained 32 047 continuous time series of 300 months (no missing data) with a view to characterizing the spatiotemporal variability of the SSTA at global scale from 1985 to 2009. Fig. 2 shows the standard deviation of the monthly SSTA for the 1985–2009 period. It highlights three types of regions of low-frequency variability: the equatorial pacific up to  $160^\circ\text{E}$  (and in a less pronounced way the equatorial Atlantic), the temperate regions, mostly of the northern hemisphere, with a maximum amplitude in the north Pacific gyre and the north-western Atlantic gyre, and the equatorial borders of the major upwelling areas.

### B. Multivariate Enso Index

The Niño/La Niña events have been thoroughly addressed in the literature and the reader may refer to <http://elNiño.noaa.gov/> as an interesting entry point to understand the ENSO and its regional impacts. The low-frequency variability in the SSTA can be associated with low-frequency atmospheric climatic variations. As a peculiar example, El-Niño-La-Niña events relate to oceanic-atmospheric oscillations of the equatorial Pacific [8]. Among the numerous ENSO-related indexes, we consider the Multivariate ENSO Index (MEI, <http://www.esrl.noaa.gov/psd/enso/mei/>), which is based on six observed variables over the tropical Pacific: sea-level pressure, zonal and meridional components of the surface wind, SST, surface air temperature, and total cloudiness fraction of the sky. As suggested by the NOAA (<http://www.esrl.noaa.gov/psd/enso/mei/rank.html>), Niño regimes were defined as the periods for which the MEI index value exceeded the percentile 30 of the positive values and conversely, La Niña regimes as significantly negative periods (below percentile 30 of the negative values).

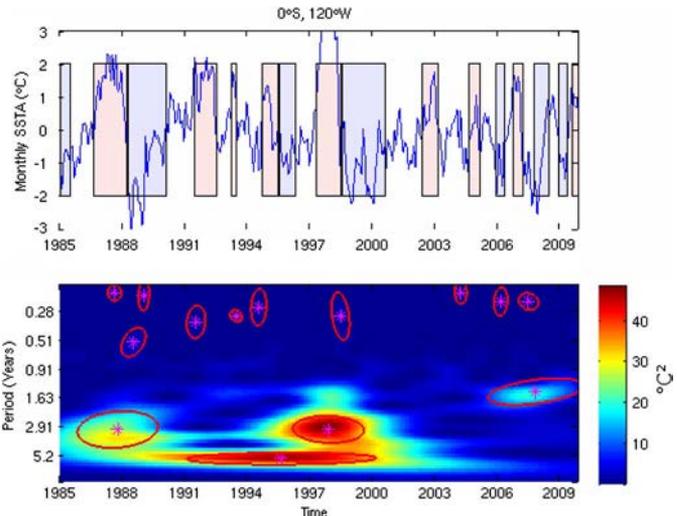


Fig. 3. Illustration of the event-based analysis of SSTA time series. Top: SSTA time series observed at  $0^\circ\text{S}$  and  $120^\circ\text{W}$ , i.e., in the eastern equatorial Pacific known to be strongly affected by ENSO processes. Bottom: the corresponding wavelet power spectrum and the detected significant elementary events delimited by ellipses with the corresponding maximum of energy indicated by a cross. See Section II-A and [23], [24] for details on the detection of the elementary events as local significant spectrum areas with respect to the theoretical energy depicted by a red noise with the same correlation and variance statistics than the considered series.

### C. Event Detection Examples in SSTA Time Series

Fig. 3 shows two SSTA time series (top) and the corresponding wavelet power spectrum (bottom) in the East Pacific. We superimposed the Niño periods (pink) and Niña periods (light blue) onto the SSTA time series (top of Fig. 3). The detected events in the power spectrum are delimited using ellipses. Events refer to wavelet spectrum areas where the energy levels are significantly greater, at 95% of confidence, than the local red noise theoretical power spectrum (5) (Fig. 1).

In the eastern Pacific (Fig. 3), at  $0^\circ\text{S}$ ,  $120^\circ\text{W}$  (center of Niño 3 region, <http://upload.wikimedia.org/wikipedia/commons/9/9d/Enso-index-map.png>), 14 significant events were detected. Two major events occur at the 3.36-year scale from 1986 to 1990 and 1998 to 2000. These two periods correspond to two well-known major ENSO events, each characterized by a succession of a strong Niño and Niña periods, the second period corresponding to the strongest Niño-Niña event recorded. As expected, the variability of the SSTA at low frequency is thus related to the ENSO signal.

### D. Characteristic Time-Scales of SSTA Elementary Events

From the 32 047 SSTA time series, we extracted 486 144 significant elementary events with estimated mean scales from 0.2 to 8 years. Fig. 4(a) shows the time-scale distribution of these events. The exponential distribution [red curve, Fig. 4(a)] is the normalization factor used to account for the scale-dependent sampling (11). Fig. 4(b) shows the normalized time-scale distribution. We divided our dataset of events in two main categories

- 1) Events with mean time-scale lower than 0.4 year showed a uniform normalized-scale distribution and were gathered in a single category, referred to the HF category.

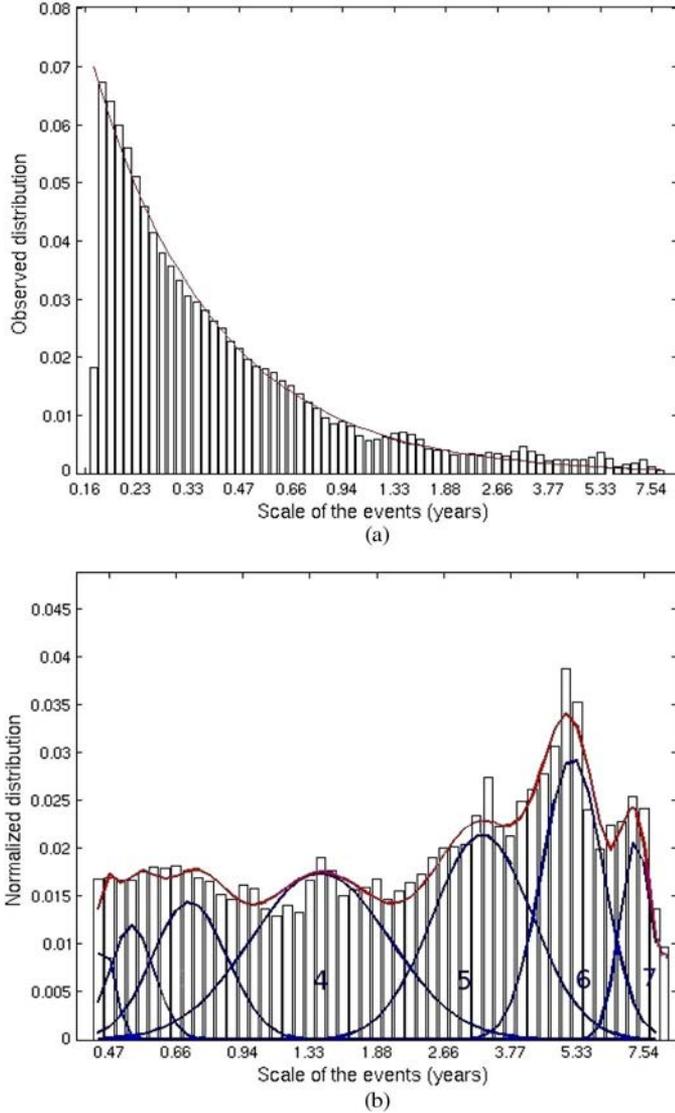


Fig. 4. Time-scale distribution and characteristic time-scales of the elementary events extracted from the SSTA dataset. (a) The initial distribution across scales of all of the extracted elementary events and the fitted exponential decay (11) corresponding to the natural fractal distribution of the event time-scales [25]. (b) The observed normalized distribution (9) with the seven Gaussian modes [(10), blue] found in the SSTA time-scale distribution. In red, the mixture model (9) fitted onto the distribution.

- 2) Events with scale greater than 0.4 year showed significant Gaussian modes in the normalized-scale-distribution. We fitted a Gaussian mixture model (9) to this dataset. The parameter estimation needed 400 EM iterations, using as convergence criterion a log likelihood threshold value of  $10^{-4}$ . The estimated model involved seven Gaussian modes [Fig. 4(b)].

Modes 3 and 4 showed mean scale of 0.70 and 1.54 years, respectively. They both refer to the interannual variability, the seasonal component (mainly the energy at a 1-year scale), being removed in the SSTA. Modes 5–7 at scales 3.36, 5.03, and 7.11 years, respectively, contain the very low-frequency changes in the SSTA caused by large space-time climatic signals such as ENSO and are therefore considered as the three characteristic low-frequency time-scales of ENSO influences onto the SSTA.

TABLE I  
MEAN AND STANDARD DEVIATION OF THE GAUSSIAN DISTRIBUTIONS FOR THE FOUR LOW-FREQUENCY REFERENCE TIME-SCALES OF THE SSTA FROM 1985 TO 2009

Mean time-scale ( $\mu$ ) in years	1.54	3.36	5.03	7.11
Sigma ( $\sigma$ ) in years	1.34	0.56	0.22	0.17

We will see later that the spatial distribution of the 1.54-year scale events also relates to ENSO region of influence. For this reason, we consider this scale as an additional characteristic ENSO time-scale. Table I summarized the characteristics (mean and standard deviation) of the Gaussian distributions for the four low-frequency reference time-scales of the SSTA. Standard deviations for reference scales greater than 1.54 are relatively low ensuring a narrow distribution and a very good confidence in these three classes.

#### E. Spatial Distribution of the SSTA Characteristic Scales

We investigate the spatial distribution of both the HF and the low-frequency characteristic time-scales. From the inferred Gaussian mixture model, we can evaluate the posterior membership probability  $\Pi_{ki} = P(Y_k = C_i)$  of the  $Y_k$ th event with a mean time-scale  $s_k$  to be assigned to the category  $C_i$

$$\Pi_{ki} = \lambda_i \cdot N(s_k | \mu_i, \sigma_i) / \sum_{j=1}^I \lambda_j \cdot N(s_k | \mu_j, \sigma_j). \quad (14)$$

The spatial distribution of the given event category  $C_i$  is estimated for each pixel (time series) using the mean number of events for  $C_i$

$$\mu(C_i) = \sum_{k=1}^K \Pi_{ki} \quad (15)$$

where  $K$  is the number of detected events in the time series. Fig. 5 shows the estimated  $\mu(C_i)$  for both the HF category [ $s < 0.4$  year, Fig. 5(a)] and the characteristic time-scales at 1.54, 3.36, and 5.03 years [Fig. 5(b)–(d)].

The HF events [Fig. 5(a)] contributed for 61% of the total number of the detected events. Overall, the mean number of detected HF events is of 10.1 over the globe for the considered 25-year period. Local maxima of 18 detections are observed on the Peruvian shores, in the South-Eastern part of the Niño3 reference area, the western Mediterranean Sea, the Arabian Sea and the Okhotsk Sea. The Arabian Sea is strongly affected by monsoon winds reversal, whose effect on SST variability from climatology is already known [32], whereas the high-latitude areas are commonly affected by winter storms that increase the SST variability, probably enhanced by the presence of the continents. No similar pattern is observed in the high latitudes of the southern hemisphere, probably because of the low interference of continental masses and the regularity of the circumpolar winds. A minimum of three events are observed in the equatorial part of the Eastern Pacific from 160 to 180°W, where the variance in the SSTA is mostly driven by the low frequency at 1.54 and 3.36-year scale [Fig. 5(b) and (c)].

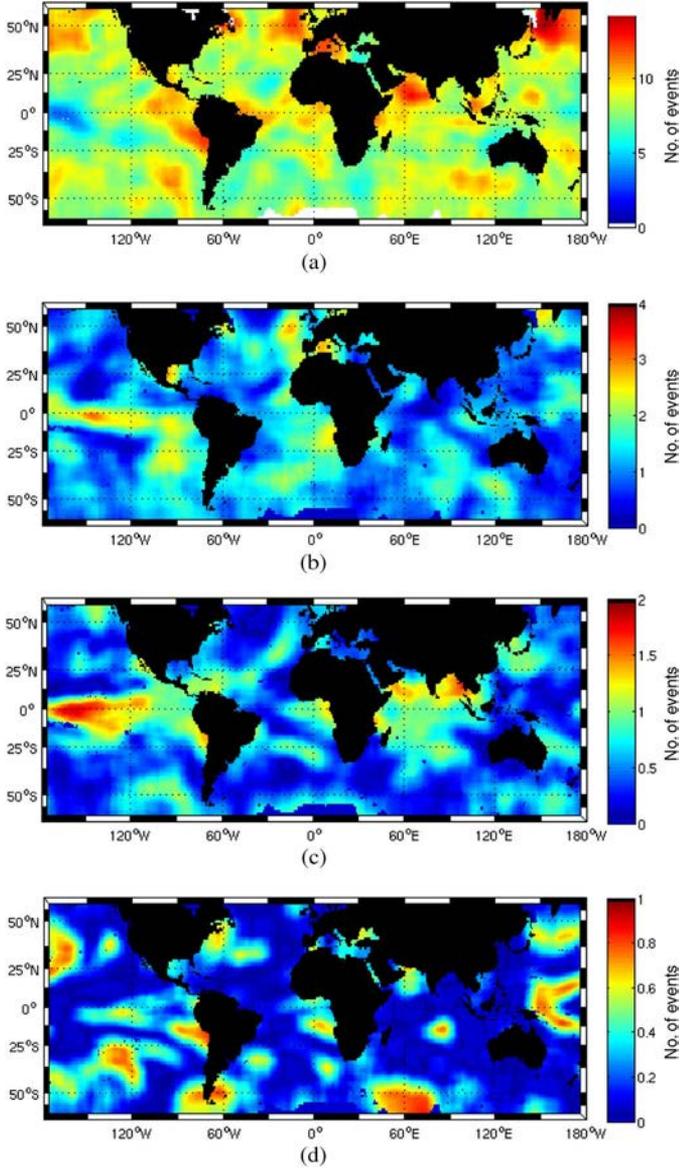


Fig. 5. Spatial distribution of the estimated SSTA characteristic time-scales. Mean number of events by time-scale categories from 1985 to 2009. (a) For the HF event category (mean time - scale < 0.4 year); (b)–(d) for characteristic time-scales of, respectively, 1.54, 3.36, and 5.03 years.

Events at 1.54-year scale [Fig. 5(b)] represent 11% of the total number of events and show a mean number of two detected events over the globe for the 1985–2009 period. Local maxima were observed in the eastern part of the Niño3 area underlying the signature of ENSO at this scale onto the SSTA. Local patches clearly appear at this scale in the Gulf of Mexico, the North Atlantic, the Namibian shores, the western Mediterranean Sea, and the Okhotsk Sea. Eastern boundary systems (Humboldt and Benguela coastal upwelling regions) specially show a higher number of events, probably caused by a high interannual variability.

The 3.36-year event category [Fig. 5(c)] accounts for 4% of all elementary events with a mean number of 0.8 event over the globe from 1985 to 2009. Its spatial distribution highlights regions known to be strongly affected by ENSO: the central equatorial Pacific, the central Humboldt system, and the northern

Indian Ocean [3], [8], [32]–[34]. In the Indian Ocean, we observe that the low-frequency signature of ENSO onto the SSTA is also observed at this time-scale. This influence of ENSO on the monsoon in this region has been largely documented [32]–[34] but often without time-scale analysis [34] or at multidecadal time-scale [33].

The 5.03-year event category [Fig. 5(d)] represents 1.5% of the elementary events with a mean number of 0.3 event over the globe. The highlighted areas are the Western part of the Pacific Ocean, the central Humboldt and southern Argentinian shores, as well as, in general, the boundaries of the regions detected at 3.36-year scales.

ENSO signal is known to propagate [15] and, e.g., ENSO signal generally occurs 4 months after it starts in the West of Peru [15] and until 9 months in the Philippines [15]. Fig. 5(b)–(d) suggests that the propagation of the ENSO signal includes time-scale shifts as already envisaged by Torrence and Webster [33]. As a peculiar example, in the Eastern intertropical Pacific [Fig. 5(c)], the detected events at 5.03-year scale [Fig. 5(d)] geographically surround the detected 3.36-year scale events suggesting the shift between frequencies during the ENSO propagation. We note that the proposed methodology (compare Section II-B) suits well to address such hypothesis compared to EOF method that does not involve such an explicit scale-related analysis [15].

#### F. Density of HF and 1.54-Year Events With Respect to ENSO Modes

To address possible HF signatures of ENSO from the analysis of the space-time distribution of HF and 1.54-years event categories, we analyze the distribution of both HF and 1.54-year event categories conditionally to the three ENSO conditions. We use the estimation of the starting and ending times of the events of category C to analyze their density  $D$  relatively to the ENSO regime  $E_r$  for normal, Niño, and Niña periods (compare Section II-B for the ENSO mode definition)

$$D(C_i|E_r) = L(C_i|E_r)/L(E_r) \quad (16)$$

where  $L(C_i|E_r)$  is the number of months spent in events of class  $C_i$  during period  $E_r$  and  $L(E_r)$  the number of months of period  $E_r$ . The density  $D(C_i|E_r)$  is a time and energy normalized representation of the energy observed for each  $C_i$  and ENSO mode. It aims at estimating the number of months where the energy is significant at this time-scale compared to the local conditions.

During normal periods, i.e., out of ENSO periods, the HF [Fig. 6(a)] and the 1.54-year frequency events [Fig. 6(b)] show, respectively, a global mean density values of 0.19 and 0.23 event month<sup>-1</sup>. Local maximum values of HF density are observed in the West of Peru, in an extended area of the Peruvian-Chile upwelling cell, i.e., for most of the Humboldt upwelling system [43], in the Arabian Sea and the Okhotsk Sea. The lowest density value (0.1 event month<sup>-1</sup>) is reached in the eastern part of the Pacific Ocean. At the 1.54-year scale [Fig. 6(b)], the West African Benguela region (0–20°W; 0–30°S) [37], characterized by a strong upwelling, displays a high density of 0.4 event month<sup>-1</sup>, whereas the East equatorial

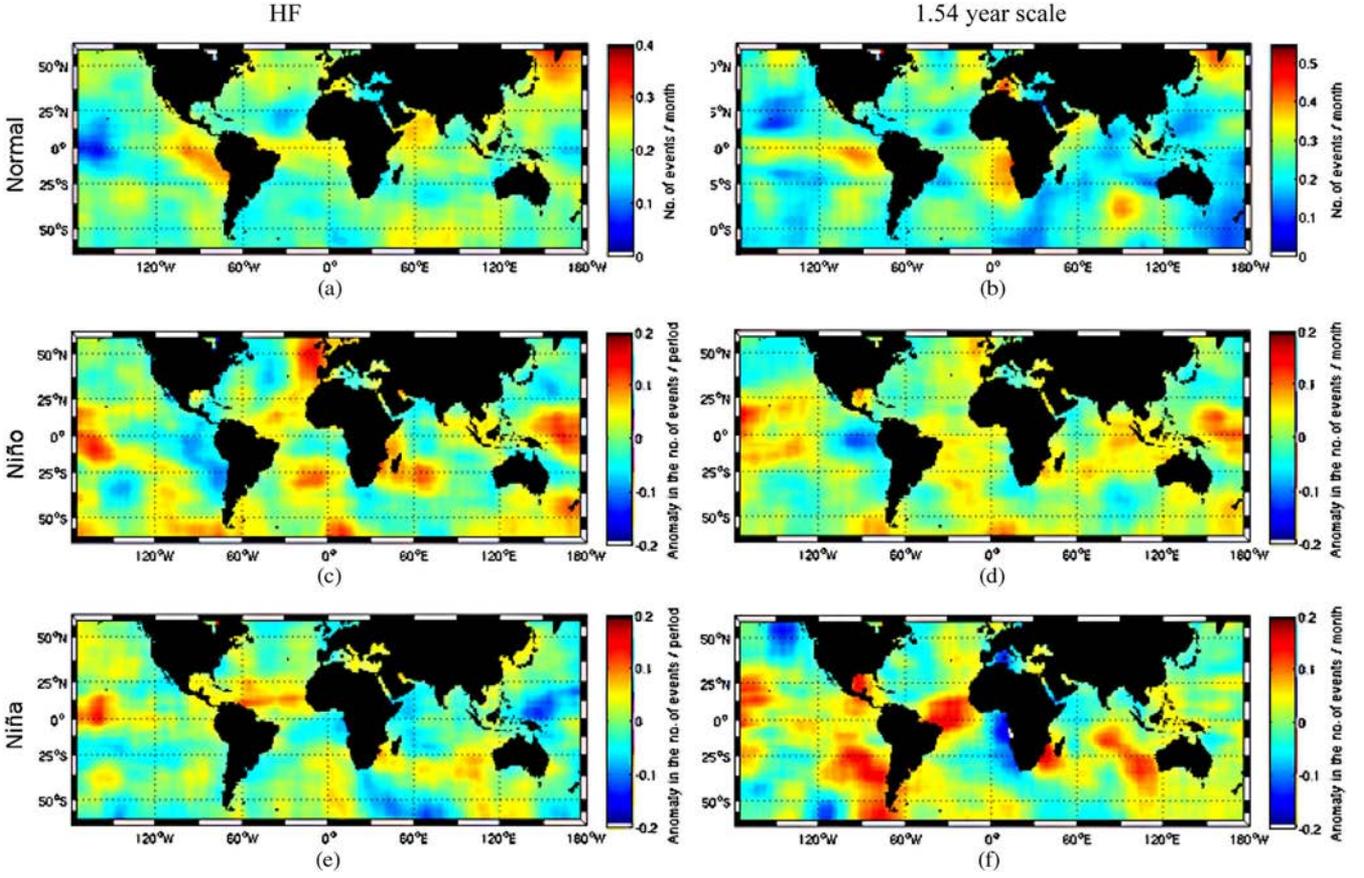


Fig. 6. Observed spatial distributions of HF and 1.54-year scale event density for normal conditions [(a) and (b)], Niño [(c) and (d)], and Niña conditions [(e) and (f)].

Pacific shows similarly high values ( $> 0.35$ ), but on a relatively moderate extent. The Okhotsk and Mediterranean Sea show a high density of event at the 1.54-year scale, whose origin is probably linked to their specific regional climate and proximity to anticyclonic areas of St. Helen and Libya [38].

To highlight the ENSO impact onto the time-scale distribution of the SSTA, event density anomaly maps are computed for both scales during Niño [Fig. 6(c) and (d)] and Niña periods [Fig. 6(e) and (f)]. Fig. 6(a) and (b) being taken as the reference state.

During Niño periods [Fig. 6(c) and (d)], the intensity of the easterlies decreases, and the warm pool of SST, usually observed in the middle of the intertropical Pacific, moves Eastward. We observe globally an increase by 11% and 6% of the HF and 1.54-year frequency events with respect to the density outside ENSO events. A large positive anomaly in the HF density of  $0.15 \text{ event month}^{-1}$  is observed over the East equatorial Pacific from  $160^\circ\text{E}$  to  $160^\circ\text{W}$  [Fig. 6(c)], in the South Atlantic, the Agulhas current [37], and in New Zealand. The last three patterns correspond to the border of the southern ocean anticyclonic regions, except for the eastern pacific, where the ENSO effects dominate. A large positive anomaly of  $0.20 \text{ event month}^{-1}$ , i.e., an increase by 80% compared to the normal HF conditions, is observed in the North East Atlantic. This is in agreement with a known influence of El Niño in the North Atlantic [35], [39]. Negative anomalies in the HF density [Fig. 6(c)] are observed in an extended area of the Chile-Peruvian upwelling system, suggesting that the decrease in the easterlies intensity, and the

resulting decrease in the upwelling intensity, tends to reduce the number of observed HF events in this area.

Niña periods [Fig. 6(e) and (f)] are characterized, in average, by a moderate positive anomaly of 8% for HF events and an increase by 6% for the 1.54-year scale, with nevertheless specific spatial patterns. In contrast to Niño phases, the easterlies strength increases during Niña periods [Fig. 6(e) and (f)] and the intertropical Pacific surface warm waters move westward. The signature of the southern Humboldt upwelling is clearly visible at the 1.54-year scale, with a positive anomaly of  $0.18 \text{ event month}^{-1}$  but off the stronger Peruvian upwelling. In the Guinea gulf and Benguela upwelling ( $5^\circ\text{N} - 30^\circ\text{S}$  and  $0^\circ - 20^\circ\text{W}$ ), Niña periods are characterized by a large negative anomaly in the SSTA events at 1.54-year scale. In this region, the SSTA is mostly dependent on the upwelling intensity, suggesting a specific stabilization of its variability during Niña periods compared to normal periods [Fig. 6(b)]. Off South Africa, we observe a clear opposite influence onto the SSTA between the West and the East shores for both Niño and Niña periods and both scales, a difference already highlighted by Rouault *et al.* [37].

It is obvious that we cannot interpret all the local differences observed in the time-scale distributions of the SSTA anomalies. ENSO phenomena are particularly complex and involve both atmospheric and oceanic processes, the SSTA being the resulting interaction between these two factors. Other large scale processes such as the Pacific Decadal Oscillation (PDO) [40], [41] and the Atlantic Multidecadal Oscillation (AMO) [42] also affect the

SSTA. It appears nevertheless that significant differences are found in the observed distributions of the HF and 1.54-year events time-scales SSTA between, on the one hand, the ENSO/normal periods, and the other hand, the Niño and Niña phases. This observation emphasizes the interest of dedicated time-scale decomposition methods to improve our understanding of processes at various spatiotemporal scales, the reference scales exhibited from the GMM [Fig. 4(b)] being used to choose the time-scales to be studied and discretize the dataset in an optimal way (compared to the standard wavelet analysis).

#### G. Investigating Frequency Shifts in the SSTA and the Intertropical Pacific During the ENSO 1997–2000 Event

Time-scale changes during Niño–Niña periods are suspected to occur [33], [36], [44]–[46] and Compo *et al.* [44] underlined that ENSO time-scale variability at decadal scales may differ substantially from one ENSO event to another. An and Wang [47] also showed a significant relationship between the observed ENSO frequency and observed SSTA structures, underlying the influence of these frequency shifts. To investigate the time-scale variability of ENSO, we focus on the 1997–2000 ENSO major event in the intertropical Pacific, the strongest ever recorded. We clearly see in Fig. 5(c), the strong signature of this ENSO event in the intertropical Pacific at 3.36-year scale. Fig. 7(a) shows the distribution of the observed mean time (center of the events) at 3.36-year scale. The maximum in the distribution is observed in August 1999, i.e., approximately in the middle of the Niño period. To investigate frequency shifts during we estimate the distribution of the variables  $\delta T_1$  and  $\delta T_2$

$$\begin{aligned}\delta T_1 &= T_{1.54} - T_{3.36} \\ \delta T_2 &= T_{HF} - T_{3.36}\end{aligned}\quad (17)$$

where  $T_{3.36}$ ,  $T_{1.54}$ , and  $T_{HF}$  are, respectively, the position (time) of the event centers (maximum of energy) at 3.36-years, 1.54-years, and HF, collocated in the same spectrum. Fig. 7(b) shows the probability density functions of both  $\delta T_1$  and  $\delta T_2$ , estimated using, respectively, 1302 and 1474 pairs of events.

In Fig. 7(b), we observe a clear increase in the number of 1.54- to 3.36-year pair of events, with a maximum observed 2 months later than the maximum at 3.36-years. Conversely, we observe a decrease in the number of HF to 3.36-year pair of events in the intertropical Pacific during the maximum of the ENSO 1997–2000 event with a minimum observed 2 months before the maximum at 3.36 years. This underlies the time-scale shifts from high to low frequencies during this ENSO. This observation motivates the determination of both the reference scales and their relative distribution to characterize ENSO compared to a classical sum of the spectrum energy between the 1.5- and 7-year scales [8]: even if the 1.5–7 year sum of energy is constant in time, the shift between scales provides a significant signature of ENSO 1997–2000.

#### IV. CONCLUSION AND FUTURE WORK

We propose in this paper an event-based mining of geophysical time series. We regard a time series as a collection of significant elementary time-scale events complemented by a red

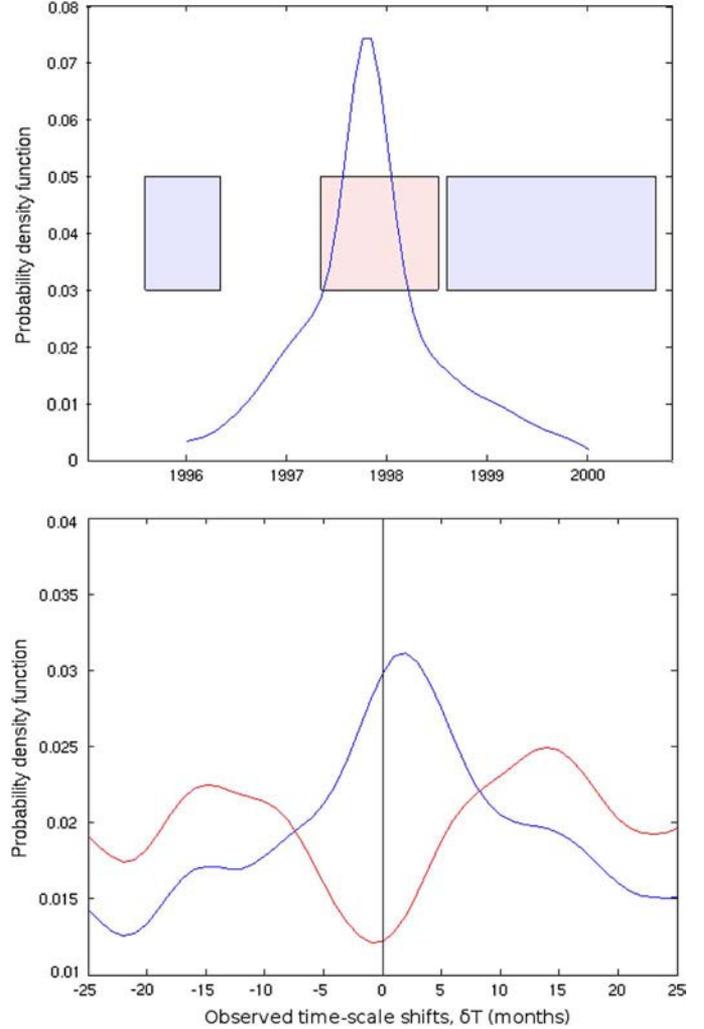


Fig. 7. (top) Temporal distribution of the maximum of energy (event centers) observed in the intertropical Pacific at 3.36-year scale [known as being a reference time-scale for ENSO, compare Fig. 5(b) and [7]]. Highlighted in pink are the Niño periods and light blue are the Niña periods. (bottom) Distribution of the observed time shifts between the maximum of energy of the events at 3.36 and 1.54 years (blue) and 3.36 and HF (red).

noise process. Our approach resorts to a normalized representation in variance of a time series through the detection of significant time-scale events. This is of key interest for SST anomalies that show a high spatial and temporal variability of the variance. The estimation of the threshold in energy to detect a significant event accounts for the autocorrelation and noise level of the local time series. This is also a key issue for geophysical time series, which depict naturally large autocorrelation noise levels (typically from 0.3 to 0.7 in the monthly SSTA dataset studied here).

The method is applied to the global SSTA observed from 1985 to 2009. We use a mixture of Gaussian to identify four reference time-scales at 1.54, 3.36, 5.03, and 7.11 years. The spatial distribution of these low-frequency reference scales highlights the intertropical Pacific, the West of Peru, the Indian Ocean, and the South of the Atlantic, regions known as being strongly influenced by ENSO. In addition, we reveal that ENSO modes are also characterized by significant space-time differences in the distribution of HF events (typically, with characteristic time-scale below 4 months). We show that the HF event density of the

SSTA increases by in mean 11% over the globe during Niño events, with a maximum of 80% in the North East of Europe and 6% during Niña periods. Even if all this HF variability may not be attributed to ENSO, this large increase is a significant signature of the Niño periods and is minimized by the EOF approach, which tends to exhibit by construction the low-frequency correlation modes. Our method also allows identifying times-scale shifts in the energy spectrum in the intertropical Pacific during the major 1997–2000 ENSO event with a maximum shift from the HF toward the reference (3.36-year scale) observed 2 months before the maximum of energy at the 3.36-year scale.

Compared to EOF-based time series analysis, the key contribution of the proposed event-based approach is to account for signal nonstationarity and noise autocorrelation in the time-scale decomposition of geophysical processes variability. While EOF-based schemes mainly reveal low-frequency patterns, our wavelet-based approach can identify both low-frequency and HF signatures and investigate their respective space-time distribution.

Compared to classical wavelet approach, our main methodological contribution lies in the characterization of significant times-scales in the SSTA taking into account the spatially varying variance and the autocorrelation of each time series. The classical wavelet approaches [7], [33] usually consider a time-scale sum of the energy between 1.5 and 7 years to depict the ENSO signatures in the SSTA [33]. Nevertheless, even if the 1.5- to 7-year sum of energy may be constant in time, the shift between scales is also a significant signature of ENSO 1997–2000 [Fig. 7(b)]. Compo [33] already pointed out that ENSO-induced changes in extratropical 500-mb height variability are time-scale dependent and An and Wang [47] showed also a relationship between ENSO frequency changes and observed structure in the SSTA, raising the crucial question of the choice of discrete frequency bands. The present methodology addresses this question providing a quantitative mean to unmix the processes and study their time-scale relationships [Fig. 7(b)]. These time-scale dependencies may evolve in time, which make them particularly difficult to deal with from an extension of the cross-wavelet spectrum [7].

Our event-based methodology opens new perspectives for the analysis of multivariate time series such as wind and SST, light, and chlorophyll-a. While we consider here the interaction between elementary events at different characteristic scales of the same geophysical variable, this methodology could be applied to two or more variables. Besides, the event-based detection could also be considered to address long-term trend estimation [29] and correlation analysis while being robust to the presence of low-frequency nonstationary signals such as ENSO.

#### ACKNOWLEDGMENT

This work has benefitted MCGS (Marine Collaborative Ground Segment; <http://www.mcgs.fr>). MCGS is a project aimed at making the most of ESA Sentinels satellites' potential for users driven services based on high-level products. MCGS addresses the need of the European Space Agency to build up data processing centers in conjunction with the Copernicus Program for the provision of services to local and national,

public and private European institutions or entities involved in marine activities.

#### REFERENCES

- [1] G. Chen, B. Shao, Y. Han, J. Ma, and B. Chapron, "Modality of semiannual to multidecadal oscillations in global sea surface temperature variability," *J. Geophys. Res. Oceans*, vol. 115, 2010.
- [2] J. Blunden, D. S. Arndt, and M. O. Baringer, "State of the climate in 2010," *Bull. Amer. Meteorol. Soc.*, vol. 92, pp. S1–S236, 2011.
- [3] R. W. Preisendorfer, *Principal Component Analysis in Meteorology and Oceanography*. Amsterdam, The Netherlands: Elsevier, 1988, p. 425.
- [4] H. Björnsson and S. A. Venegas, "A manual of EOF and SVD analysis of climatic data," McGill University, Montreal, QC, Canada, 1997.
- [5] C. S. Bretherton, C. Smith, and J. M. Wallace, "An intercomparison of methods for finding coupled patterns in climate data," *J. Clim.*, vol. 5, pp. 541–560, 2002.
- [6] C. Frankignoul and K. Hasselmann, "Stochastic climate models. Part II: Application to SST anomalies and thermocline variability," *Tellus*, vol. 29, pp. 289–305, 1977.
- [7] C. Torrence and G. P. Compo, "A practical guide to wavelet analysis," *Bull. Amer. Meteorol. Soc.*, vol. 79, pp. 61–78, 1998.
- [8] S. G. Philander, "Niño, La Niña, and the Southern Oscillation," *Int. Geophys.*, vol. 46, 1990.
- [9] D. A. Jackson and Y. Chen, "Robust principal component analysis and outlier detection with ecological data," *Environmetrics*, vol. 15, pp. 129–139, 2004.
- [10] K. Lau and H. Weng, "Climate signal detection using wavelet transform: How to make a time-series sing," *Bull. Amer. Meteorol. Soc.*, vol. 76, pp. 2391–2402, 1995.
- [11] D. M. Sonechkin and N. M. Datsenko, "Wavelet analysis of nonstationary and chaotic time-series with an application to the climate change problem," *Pure Appl. Geophys.*, vol. 157, pp. 653–677, 2000.
- [12] O. Leeuwenburgh and D. Stammer, "The effect of ocean currents on sea surface temperature anomalies," *J. Phys. Oceanogr.*, vol. 31, pp. 2340–2358, 2001.
- [13] T. Rouyer, J.-M. Fromentin, N. Stenseth, and B. Cazelles, "Analysing multiple time-series and extending significance testing in wavelet analysis," *Mar. Ecol. Prog. Ser.*, vol. 359, pp. 11–23, 2008.
- [14] B. Flannery, "Gaussian mixture models and k-means clustering," in *Numerical Recipes: The Art of Scientific Computing*, 3rd ed. Cambridge, U.K.: Cambridge Univ. Press, 2007, pp. 840–842.
- [15] D. B. Enfield and A. M. Mestas-Nunez, "Multiscale variabilities in global sea surface temperature and their relationships with tropospheric climate patterns," *J. Clim.*, vol. 12, pp. 2719–2733, 1999.
- [16] P. J. Brockwell and R. A. Davis, *Time-Series: Theory and Methods*, 2nd ed. New York, NY, USA: Springer, 2009.
- [17] D. Gu and S. G. H. Philander, "Secular changes of annual and inter-annual variability in the Tropics during the past century," *J. Clim.*, vol. 8, pp. 865–876, 1995.
- [18] S. G. Mallat, "Multifrequency channel decomposition of images and wavelet models," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 37, no. 12, pp. 2091–110, 1989.
- [19] S. Levitus, "A comparison of the annual cycle of two sea surface temperature climatologies of the World Ocean," *J. Phys. Oceanogr.*, vol. 17, no. 2, pp. 197–214, 1987.
- [20] A. J. Sarnaglia, "Estimation of periodic autoregressive processes in the presence of additive outliers," *J. Multivar. Anal. Arch.*, vol. 101, no. 9, pp. 2168–2183, 2010.
- [21] D. C. Hoaglin, F. Mosteller, and J. Tukey, *Understanding Robust and Exploratory Data Analysis*. Hoboken, NJ, USA: Wiley, 2000, pp. 404–414.
- [22] R. Fablet, "Multiscale geometric deformations along planar curves: Application to satellite tracking and ocean observation data," *IEEE Trans. Geosci. Remote Sens.*, 2012, submitted for publication.
- [23] G. L. David, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, 2004.
- [24] P. Monasse and F. Guichard, "Fast computation of a contrast invariant image representation," *IEEE Trans. Image Process.*, vol. 9, no. 5, pp. 860–872, May 2000.
- [25] B. B. Mandelbrot, *The Fractal Geometry of Nature*, 2nd ed. New York, NY, USA: W. H. Freeman, 1982, p. 460.
- [26] A. P. Dempster, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Stat. Soc.*, vol. 1, no. 39, pp. 1–38, 1977.
- [27] K. S. Casey, T. B. Brandon, P. Cornillon, and R. Evans, "The past, present and future of the AVHRR Pathfinder SST Program," in *Oceanography from Space: Revisited*, New York, NY, USA: Springer, 2010.

- [28] K. A. Kilpatrick, G. P. Podesta, and R. Evans, "Overview of the NOAA/NASA advanced very high resolution radiometer Pathfinder algorithm for sea surface temperature and associated match-up database," *J. Geophys. Res.*, vol. 106, pp. 9179–9197, 2001.
- [29] B. Saulquin *et al.*, "Detection of linear trends in multisensor time-series in the presence of autocorrelated noise: Application to the chlorophyll-a SeaWiFS and MERIS data sets and extrapolation to the incoming sentinel 3-OLCI mission," *J. Geophys. Res. Oceans*, vol. 118, 2013.
- [30] B. Narapusetty, T. DelSole, and M. K. Tippett, "Optimal estimation of the climatological mean," *J. Clim.*, vol. 22, pp. 4845–4859, 2009.
- [31] E. C. Weatherhead, *et al.*, "Factors affecting the detection of trends: Statistical considerations and applications to environmental data," *J. Geophys. Res. Atmos.*, vol. 103, no. D14, 1998.
- [32] V. Krishnamurthy and Ben P. Kirtman, "Relation between Indian monsoon variability and SST," *J. Clim.*, vol. 22, pp. 4437–4458, 2009.
- [33] C. Torrence and P. J. Webster, "Interdecadal changes in the ENSO-monsoon system," *J. Clim.*, vol. 12, no. 8, 1998.
- [34] B. Wang, J. Yang, T. Zhou, and B. Wang, "Interdecadal changes in the major modes of Asian–Australian monsoon variability: Strengthening relationship with ENSO since the late 1970 s," *J. Clim.*, vol. 21, no. 8, pp. 1771–1789, 2008.
- [35] S. Hastenrath, L. C. de Castro, and P. Aceituno, "The Southern Oscillation in the Atlantic sector," *Contrib. Atmos. Phys.*, vol. 60, no. 41987, pp. 47–463, 1987.
- [36] T. S. Kestin, D. J. Karoly, J.-I. Yano, and N. A. Rayner, "Time–frequency variability of ENSO and stochastic simulations," *J. Clim.*, vol. 11, pp. 2258–2272, 1998.
- [37] M. Rouault, B. Pohl, and P. Penven, "Coastal oceanic climate change and variability from 1982 to 2009 around South Africa," *Afr. J. Mar. Sci.*, vol. 32, no. 2, pp. 237–246, 2010.
- [38] D. E. Pedgley, D. R. Reynolds, and G. M. Tatchell, *Long-Range Insect Migration in Relation to Climate and Weather: Africa and Europe. Insect Migration: Tracking Resources Through Space and Time*. Cambridge, U.K.: Cambridge Univ. Press, 1995, pp. 3–29.
- [39] X. Wang, C. Wang, W. Zhou, D. Wang, and J. Song, "Teleconnected influence of North Atlantic sea surface temperature on the El Nino onset," *Clim. Dynam.*, vol. 37, no. 3–4, pp. 663–676, 2011.
- [40] M. A. Alexander, L. Matrosova, C. Penland, J. D. Scott, and P. Chang, "Forecasting Pacific SSTs: Linear inverse model predictions of the PDO," *J. Clim.*, vol. 21, pp. 385–402, 2008.
- [41] N. J. Mantua, "A Pacific interdecadal climate oscillation with impacts on salmon production," *Bull. Amer. Meteorol. Soc.*, vol. 78, no. 6, pp. 1069–1079, 1997.
- [42] J. R. Knight, C. K. Folland, and A. A. Scaife, "Climate impacts of the Atlantic multidecadal oscillation," *Geophys. Res. Lett.*, vol. 33, 2006, L17706.
- [43] C. Lett, P. Penven, P. Ayón, and P. Fréon, "Enrichment, concentration and retention processes in relation to anchovy (*Engraulis ringens*) eggs and larvae distributions in the northern Humboldt upwelling ecosystem," *J. Mar. Syst.*, vol. 64, no. 1–4, pp. 189–200, Jan. 2007.
- [44] G. P. Compo, P. D. Sardeshmukh, and C. Penland, "Changes of subseasonal variability associated with Niño," *J. Clim.*, vol. 14, pp. 3356–3374, 2001.
- [45] M. Latif and T. P. Barnett, "Decadal climate variability over the North Pacific and North America: Dynamics and predictability," *J. Clim.*, vol. 9, no. 10, pp. 2407–2423, 1996.
- [46] B. P. Kirtman and P. S. Schopf, "Decadal variability in ENSO predictability and prediction," *J. Clim.*, vol. 11, no. 11, pp. 2804–2822, 1998.
- [47] S. I. An and B. Wang, "Interdecadal change of the structure of the ENSO mode and its impact on the ENSO frequency," *J. Clim.*, vol. 13, no. 12, pp. 2044–2055, 2000.