Analysing multiple time series and extending significance testing in wavelet analysis

Tristan Rouyer[1,2,*], Jean-Marc Fromentin[1], Nils Chr. Stenseth[2,3], Bernard Cazelles[4,5]

[1]IFREMER, Centre de Recherche Halieutique Méditerranéenne et Tropicale, Avenue Jean Monnet, BP 171, 34203 Sète cedex, France

[2]Centre for Ecological and Evolutionary Synthesis (CEES), Department of Biology, University of Oslo, PO Box 1066, Blindern, 0316 Oslo, Norway

[3]Institute of Marine Research, Department of Coastal Zone Studies, Flødevigen Research Station, 4817 His, Norway

[4]CNRS UMR 7625, École Normale Supérieure (ENS), 46 rue d'Ulm, 75230 Paris cedex 05, France

[5]IRD, UR 079, GEODES Centre IRD Ile de France, 32 avenue Henri Varagnat, 93143 Bondy cedex, France

ABSTRACT: In nature, non-stationarity is rather typical, but the number of statistical tools allowing for non-stationarity remains rather limited. Wavelet analysis is such a tool allowing for non-stationarity but the lack of an appropriate test for statistical inference as well as the difficulty to deal with multiple time series are 2 important shortcomings that limits its use in ecology. We present 2 approaches to deal with these shortcomings. First, we used $1/f^\beta$ models to test cycles in the wavelet spectrum against a null hypothesis that takes into account the highly autocorrelated nature of ecological time series. To illustrate the approach, we investigated the fluctuations in bluefin tuna trap catches with a set of different null models. The $1/f^\beta$ models approach proved to be the most consistent to discriminate significant cycles. Second, we used the maximum covariance analysis to compare, in a quantitative way, the time–frequency patterns (i.e. the wavelet spectra) of numerous time series. This approach built cluster trees that grouped the wavelet spectra according to their time–frequency patterns. Controlled signals and time series of sea surface temperature (SST) in the Mediterranean Sea were used to test the ability and power of this approach. The results were satisfactory and clusters on the SST time series displayed a hierarchical division of the Mediterranean into a few homogeneous areas that are known to display different hydrological and oceanic patterns. We discuss the limits and potentialities of these methods to study the associations between ecological and environmental fluctuations.

KEY WORDS: Non-stationarity · Multivariate time series · Wavelet clustering · Wavelet significance testing · Surrogates · Maximum covariance analysis

INTRODUCTION

Following the work of Steele (1985), Pimm & Redfearn (1988) and Lawton (1988), the time–frequency properties of a signal have become of major interest to ecologists (e.g. Petchey et al. 1997). In nature, non-linear and non-stationary processes are the rule rather than the exception (Stenseth et al. 1998, Hsieh et al. 2005), and many classical tools for time series analysis, such as Fourier analysis, require stationarity (or more often second-order stationarity, Chatfield 2004). As many eco-logical time series do not meet such requirements and as growing evidence supports recognition of the importance of transient dynamics in ecological processes (Hastings 2001, Cazelles et al. 2008), the spectral properties are not always well suited to analyse ecological time series. Wavelet analysis (Daubechies 1992) is a time scale and/or time–frequency decomposition of the signal that overcomes these problems and provides a powerful tool for analysing non-stationary, aperiodic and noisy signals often found in ecological time series (Torrence & Compo 1998).

*Email: rouyer.tristan@bio.uio.no

Unlike the Fourier transform which decomposes the time series into a sum of sine and cosine functions that are not resolved in time, wavelet analysis uses a set of functions locally defined in both time and frequency domains. Wavelet analysis has spread into many fields of research, such as signal processing, geophysics and climatology, and is becoming more and more popular within ecology (e.g. Bradshaw & Spies 1992, Grenfell et al. 2001, Klvana et al. 2004, Keitt & Urban 2005, Keitt & Fischer 2006, Ménard et al. 2007). Nonetheless, wavelet analysis still displays 2 shortcomings with respect to its use in ecology. First, the time and frequency locations of the wavelet spectra are not uncorrelated, and the statistical inference is therefore difficult (Maraun & Kurths 2004). If standard boot-strapping methods allow testing the wavelet spectra with various re-sampling procedures, the underlying null hypotheses are most often invalid for ecological time series. For instance, in a number of cases, white noise and red noise are not supported by the observations. Second, although wavelet analysis is a powerful univariate or bivariate analysis, it remains limited to analysing a large number of (shorter) time series; this is, however, a situation commonly encountered by ecologists.

We present 2 approaches to circumvent these limitations and to extend the use of wavelets within the field of ecology. First, we propose to test the wavelet spectra using surrogates. The surrogates have been introduced by Theiler et al. (1992) to determine the consistency of experimental time series with various null hypotheses of simple systems. Each surrogate is consistent with a specific null hypothesis about the underlying system (e.g. a random variable or an autoregressive-like process) while retaining some of the statistical features of the original time series (e.g. mean and variance, power spectrum). Statistics are then computed for the original data and a large set of surrogates, thus allowing testing of the original data against an empirical distribution consistent with the null hypothesis (Royer & Fromentin 2006). We propose a class of surrogates that models the underlying statistical structure of the time series as $1/f$ noise (Halley 1996). Such a model is rather convenient as it allows for the 'more time, more variation' effect displayed by many ecological time series (Lawton 1988, Inchausti & Halley 2002, Vasseur & Yodzis 2004). This class of surrogates is thus adapted to ecological time series, as it is, in addition, designed to deal with short time series.

Second, understanding ecological phenomena using time series often requires the analysis of large datasets. Wavelet cross analyses allow investigation of the association between 2 signals by extending the wavelet transform to bivariate cases. The wavelet cross-spectrum and the wavelet coherence are respectively used to quantify the local covariance and correlation between 2 non-stationary signals (e.g. Cazelles et al. 2008). However, these methods only enable the study of associations between pairs of time series, and their use remains limited in the case of large datasets. In addition, wavelet analyses yield outputs in both time and frequency domains and comparing wavelet spectra of numerous time series quickly turns into a very complex issue. When dealing with large datasets of time series, a classical and useful approach in ecology is to use hierarchical clustering. A matrix of dissimilarities between the time series is constructed, over which a clustering algorithm is applied. The dissimilarity matrix can either be performed on the raw properties of the time series or on their power spectrum, according to whether interest is more in the temporal or frequency aspects. We present here an approach that combines both time and frequency domains, as the matrix of dissimilarities used for clustering is constructed from the comparison between pairs of wavelet spectra. The wavelet spectra are compared using a procedure based on the maximum covariance analysis (MCA), a multivariate method that was originally used to compare spatio-temporal fields (Bretherton et al. 1992). Both the surrogate and the clustering approaches are applied in this paper to several datasets of time series, either simulated or real, to illustrate their applicability. Wavelet analysis and the proposed approaches applied to ecological questions illustrate how these methods can bring fruitful insights to the study of coupling between environmental and biological fluctuations.

MATERIALS AND METHODS

Wavelet analysis. The wavelet methodology is well suited for signals whose frequencies change with time. This is because this methodology enables description of the variability of a time series in both time and frequency domains, and it can cope with aperiodic components, noise and transients (Daubechies 1992, Lau & Weng 1995, Torrence & Compo 1998). The wavelet transform is based on the convolution product between the time series and a mathematical function, the so-called 'daughter wavelet'. For a given set of parameters a (scale parameter related to frequency) and τ (translation parameter related to time position), the wavelet functions Ψ, are defined at time t as follows:

$$\psi_{a,\tau}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-\tau}{a}\right) \tag{1}$$

The wavelet transform, W, is then defined as a convolution integral of the time series with the wavelet function:

$$W_x(a,\tau) = \frac{1}{\sqrt{a}}\int x(t)\psi^*\left(\frac{t-\tau}{a}\right)\mathrm{d}t = \int x(t)\psi^*_{(a,\tau)}\mathrm{d}t \tag{2}$$

where * indicates complex conjugate. We used the Morlet wavelet, a continuous and complex wavelet adapted to wavelike signals, that allows extraction of time-dependent amplitude and whose scales are related to frequencies in a simple way (Mallat et al. 1998). The relative importance of frequencies for each time step may then be represented in the time/frequency plane that forms the wavelet power spectrum:

$$S_x(a, \tau) = |W_x(a, \tau)|^2 \qquad (3)$$

where S is the wavelet power spectrum and x is the raw time series. The wavelet power spectrum, $S_x(a, \tau)$, is plotted as a function of time and period in a 2-dimensional graph. This representation, classically referred to as a 'contour plot', displays contours as identical power spectrum values, coded in the figures of this paper by different colors. The wavelet transform acts as a local filter that directly relates the magnitude of the signal to time and thus enables one to track how the frequency components change over time. Therefore, wavelet analysis is particularly adapted to investigation of non-stationary and transient signals.

Statistical inference of the wavelet spectrum. *Pointwise testing:* While analytical tests for significance can be straightforwardly calculated in Fourier analysis, the validity of such tests is greatly questioned in the case of wavelet analysis (Maraun & Kurths 2004). The pointwise testing approach, used by Torrence & Compo (1998), relies on a parametric bootstrap to assess the significance of areas. A reasonable null model of a given form is first chosen (e.g. a white noise or a first order autoregressive process, AR[1]), and a large number of random realisations is produced. Computing the wavelet spectrum for each realisation generates an empirical distribution for each point under the null model hypothesis. The test is then performed by comparing the values obtained for the original time series with the empirical distributions. As for all Monte Carlo approaches, the choice of the null model used to test the wavelet spectrum is central. Making the assumption of a white noise process is generally not appropriate for ecological data, while an autoregressive process (e.g. an AR[1]) can be an acceptable null model in some cases, for both ecological and geophysical data (Steele 1985, Vasseur & Yodzis 2004). However ecological time series display a large variety of autocorrelation structures that neither a white noise, nor an autoregressive process could consistently describe (Arino & Pimm 1995, Cuddington & Yodzis 1999, Inchausti & Halley 2002, Halley & Stergiou 2005).

A traditional surrogate approach: The surrogates can be best suited to test wavelet spectra, as they produce synthetic time series that share given statistical properties with the original time series (or with a given null model). This method has already been used for testing purposes in ecology (e.g. nonlinearity in time series, recurrence patterns, similarity of rhythms between time series, phase analysis), as several null models can be chosen according to the different algorithms employed to create the surrogates (Schreiber & Schmitz 2000, Cazelles & Stone 2003, Cazelles 2004, Royer & Fromentin 2006). The surrogate approach has also been used to test the wavelet spectrum using a hidden Markov process as a null model (Klvana et al. 2004, Saitoh et al. 2006). Other null models are the 'Type 0' surrogates, equivalent to a white noise assumption, whereas the null model referred to as 'Type 1' surrogates (or Fourier surrogates), preserves the autocorrelation structure of the time series (or equivalently its power-spectrum). The more complex null model, 'Type 2' surrogates, preserves both the Fourier spectrum and the original distribution of the data (Schreiber & Schmitz 1996, Royer & Fromentin 2006). The Fourier Type 1 and Type 2 surrogates are very interesting in an ecological perspective, as they preserve the Fourier spectrum of the original time series, allowing the oscillations that cannot be produced by an autoregressive process to be tested. However, these classes of surrogates require the length of the time series to be much larger than the dominant frequency in order to lead to satisfying surrogates, a requirement often difficult to obtain for ecological time series. They also present the disadvantage of introducing spurious low-frequency effects due to the phase randomization, but also spurious high-frequency effects when the time series are short and non-stationary (for more details see Schreiber & Schmitz 1996).

Surrogates using the slope of the spectrum: We proposed a class of surrogates, the 'beta surrogates', that display a similar variance and autocorrelation structure as the original time series and that form a less constrained hypothesis than the Fourier and Type 2 surrogates. The beta surrogates display the same relative distribution of frequencies, i.e. the same slope of the Fourier spectrum, as the original time series; this allows the dominance of low frequencies often displayed by ecological time series to be taken into account. To do so, we estimated the exponent of a power law model $1/f^\beta$, often called 'beta' in the literature, fitted to the power spectrum of the time series (Halley 1996), using the multiple segmenting method proposed by Miramontes & Rohani (2002) for short time series. Like Cuddington & Yodzis (1999), we used the spectral synthesis to generate surrogates with the previously estimated exponent. In the spectral synthesis, the amplitudes of the Fourier spectrum are scaled according to the estimated spectral exponent of

the $1/f^\beta$ model, and the phases are drawn by a uniform variable on the interval $(0, 2\pi)$. A back transformation with the inverse Fourier Transform (Voss 1988) produces the surrogates. Using this approach, we thus obtain surrogates that mimic the shape of the original ecological time series by displaying a power spectrum with the same slope, but without exactly reproducing it.

Comparing the wavelet spectra with the MCA method. *The maximum covariance analysis:* Many ecological time series, especially from the marine environment, such as phytoplankton, zooplankton or fisheries time series, contain a substantial amount of zeros together with periods of low variability. This leads to wavelet spectra that often display areas with poor, or even without, power. As a consequence, a direct comparison of the wavelet spectra will lead to a classification that could be driven by the areas with poor power. We addressed this by focusing on the common time–frequency properties between the pairs of wavelet spectra, extracted by the MCA. The MCA, historically called 'singular value decomposition' (Bretherton et al. 1992), has been widely used in meteorology to compare spatio-temporal fields (von Storch & Zwiers 1999). The MCA is close to a combined empirical orthogonal function analysis, but with a better efficiency and robustness. It relies on a singular value decomposition (which here designs the solution of the matrix problem) performed on the covariance matrix between 2 spatio-temporal fields. The first axis of the MCA thus corresponds to the largest fraction of the covariance between the 2 fields, which in this case will be the wavelet spectra (for more details on MCA, see Newman & Sardeshmukh 1995, Cheng & Dunkerton 1995). In other words, the first axis extracts the most important common pattern between the 2 fields.

Constructing the covariance matrix implies, as an underlying assumption, that the frequencies compositions at different times are independent. This is generally not true but, like the eigenvalue decomposition, the singular value decomposition is rather robust to such bias and is therefore often used in this context (e.g. Cheng & Dunkerton 1995). We compute the covariance matrix $R_{i,j}$ between each pair of wavelet spectra W_i and W_j (Fig. 1):

$$R_{i,j} = W_i W_j^t \qquad (4)$$

with W_j^t denoting the transposition of W_j. Then, the singular value decomposition is applied on $R_{i,j}$:

$$R_{i,j} = U\boldsymbol{\Gamma}V^t \qquad (5)$$

The columns of the matrix U are orthogonal and contain the singular vectors for W_i; the rows of the matrix V^t are also orthogonal and contain the singular vectors for W_j. $\boldsymbol{\Gamma}$ is a diagonal matrix whose diag-

onal elements are the singular values; they are disposed in decreasing order of magnitude, and they are proportional to the squared covariance accounted for each axis. Each singular value is thus associated to common patterns of decreasing importance between the 2 spectra. The singular value decomposition finds an orthonormal basis for each spectrum, determined by their respective singular vectors maximizing their mutual covariance. The singular value
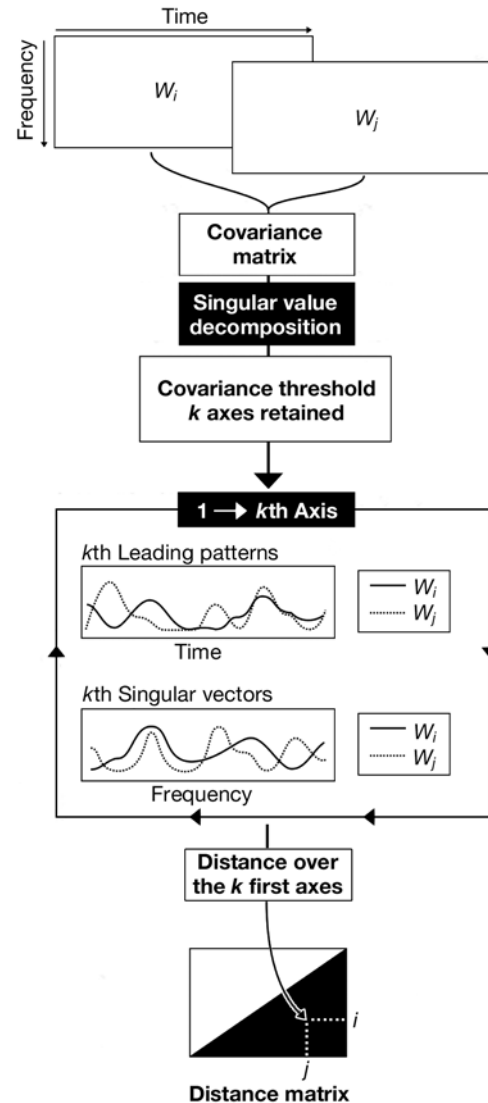

Fig. 1. Methodology employed to build the distance matrix for a set of wavelet spectra. The cross-covariance matrix is computed for a pair of wavelet spectra W_i and W_j. The singular value decomposition is applied to the covariance matrix. The k first axes are then extracted, k being chosen so that the sum of the covariance associated with the axes is below or equal to the covariance threshold. The distance index is then computed between the k first leading patterns of each wavelet spectrum and between the k first singular vectors of each wavelet spectrum

decomposition is more general than the eigenvalue decomposition used for principal component analysis, but the 2 decompositions are nevertheless related. A singular value decomposition performed on the covariance matrix between 2 similar wavelet spectra would yield the same results as an eigenvalue decomposition. The number of non-zero singular values (k) of Γ is inferior or equal to the number of frequencies analysed and each one is associated with a pair of singular vectors (frequency patterns), that are respectively associated with each spectrum.

The leading patterns show how respective frequency patterns evolve in time, and are obtained by projecting each wavelet spectrum onto its respective singular vectors. $L_i^k(t)$ and $L_j^k(t)$ are respectively the kth leading patterns for W_i and W_j, and are computed as follows:

$$L_i^k(t) = \sum_{f=1}^{f=F} U^k \times W_i(f,t)$$

$$\text{and} \quad (6)$$

$$L_j^k(t) = \sum_{f=1}^{f=F} V^k \times W_j(f,t)$$

with F the maximum frequency common to both spectra. It is then possible to reconstruct the initial wavelet spectra with a given number, N, of leading patterns by the following relationships:

$$W_i^N = \sum_{k=1}^{k=N} U^k \times L_i^k$$

$$\text{and} \quad (7)$$

$$W_j^N = \sum_{k=1}^{k=N} V^k \times L_j^k$$

These correspond to the product of a matrix formed with N singular vectors and another one formed with the N corresponding leading patterns. The reconstruction for the kth axis is therefore determined by the kth singular vector and the kth leading pattern, and the larger k is, the less important is the common covariance explained. A reconstruction corresponds to a filtered representation of the spectrum; the less important N is, the more important is the filter.

Computing the distance between the wavelet spectra: The distance between 2 wavelet spectra was measured by comparing the leading patterns and the singular vectors obtained by the MCA over a given number of axes (that correspond to a fixed percentage of the total covariance). As the relationships between the 2 singular vectors and between the 2 leading patterns were not linear, they could not be compared using a simple correlation. We thus computed the following distance (D) measure adapted from Keogh & Pazzani (1998):

$$D(L_i^k, L_j^k) = \sum_{t=1}^{n-1} \text{atan}\left[|(L_i^k(t) - L_j^k(t)) - (L_i^k(t+1) - L_j^k(t+1))|\right]$$

$$(8)$$

with n being the length of the vectors, and $L_i^k(t)$ and $L_j^k(t)$ being the kth pair of leading patterns for W_i and W_j. This metric compares 2 vectors by measuring the angle between each pair of corresponding segments, defined by the consecutive points of the 2 vectors; 2 parallel vectors will thus lead to a null distance (Fig. 2). This metric could be interpreted as a robust version of the correlation between the derivatives of the leading patterns and/or the leading vectors.

The sum of angles obtained is a comparable metric between each pair of leading patterns and singular vectors. The distance was then computed as the weighted mean of the distance for each of the k pairs of singular vectors and leading patterns retained (the weights being equal to the amount of covariance explained by each axis). For the comparison of the wavelet spectra i and j, we compute the distance $DT(i,j)$ according to the following formula:

$$DT(i,j) = \frac{\sum_{k=1}^{k=K} w_k \times (D(L_i^k, L_j^k) + D(U_i^k, V_j^k))}{\sum_{k=1}^{k=K} w_k}$$

$$(9)$$

with w_k being the weights, set equal to the amount of covariance explained by each axis. The distances, $DT(i,j)$, were then used to fill a distance matrix suitable for cluster analysis (Fig. 1). The larger the amount of covariance, the larger the number of axes retained; a large number of axes will enable to take into account more detailed common time–frequency features between the 2 spectra.

All the computations were done using R version 2.4 (R Development Core Team [2006] R: a language and environment for statistical computing). This is available online at: www.R-project.org.
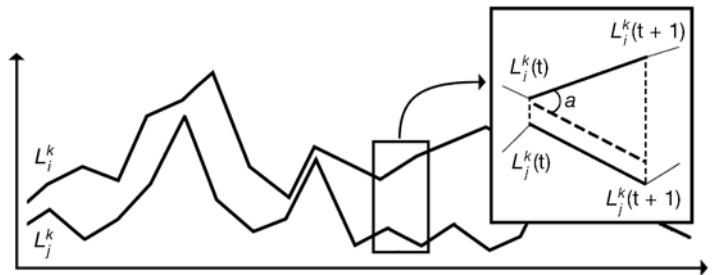

Fig. 2. Computation of the distance index used between the pair of kth leading patterns (or singular vectors), L_i^k and L_j^k (see Eq. 8). The absolute difference between the pair of vertical dashed lines for every pair of segments of the 2 series is computed and the angle a, between each pair of segments is obtained by taking the atan of the absolute differences. Summing over the segments yields the total angle between the 2 series

RESULTS

Significant cycles in bluefin tuna time series

Time series

An extensive description of the whole dataset can be found in Ravier & Fromentin (2001, 2004). These data consist of 7 time series of trap catches from various Mediterranean locations: Sicily (Favignana and Formica), Sardinia (Saline, Porto Paglia, Portoscuso, Isola Piana) and Tunisia (Sidi Daoud) (Fig. 3). These time series were selected from an extensive dataset of more than 100 time series, as they displayed a common period of 83 contiguous years (1878 to 1960) without missing values. They also displayed synchronous long-term and short-term fluctuations, and they thus constituted good examples for comparing various null models (Ravier & Fromentin 2001, Royer & Fromentin 2006). We used the beta surrogates to test whether or not the fluctuations detected in these time series were artefacts of the autocorrelation structure of the time series. The results were then compared to those obtained using Type 1 surrogates (Fourier surrogates), AR[1] and white noise, in order to assess their relevance.
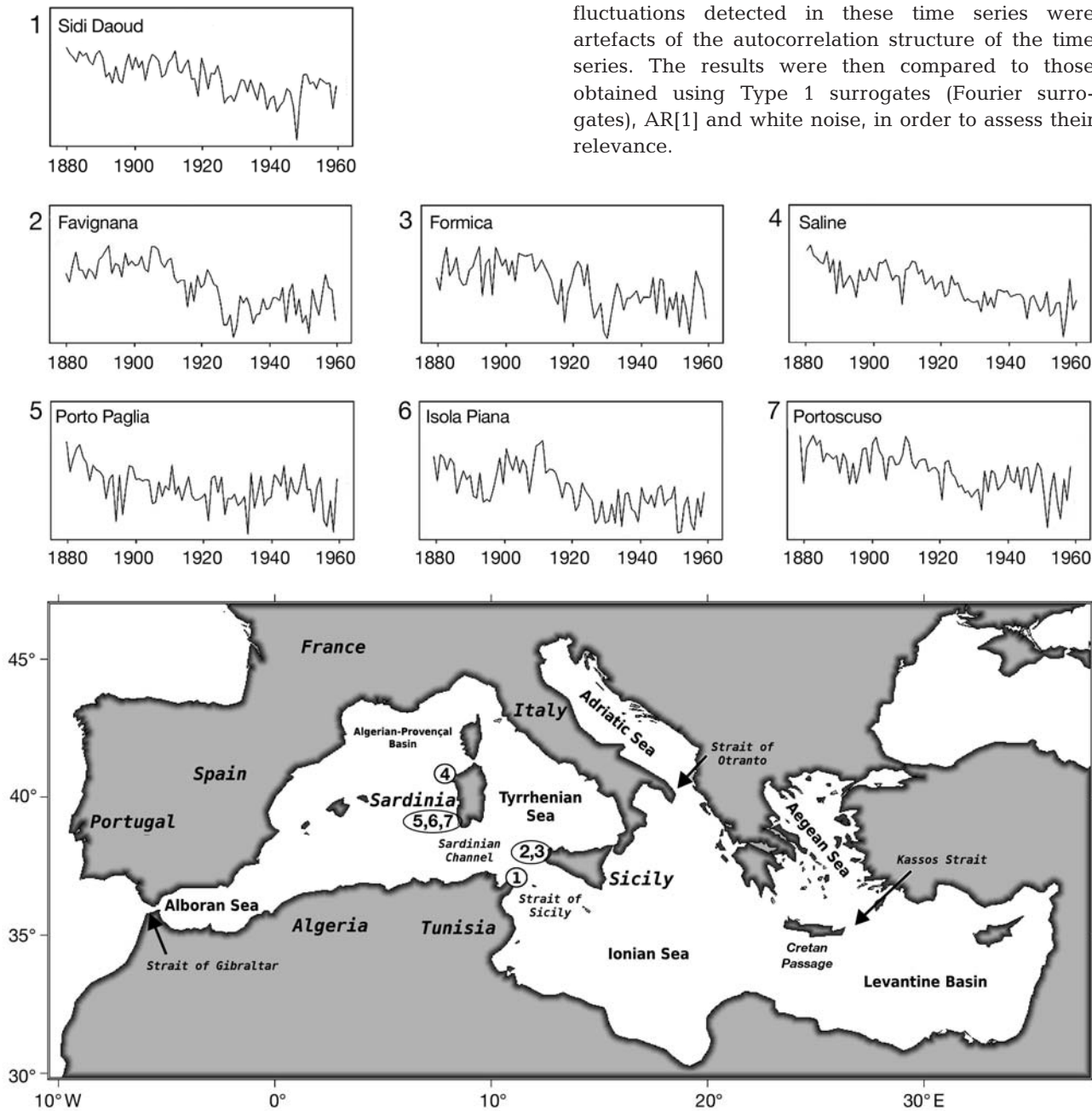

Fig. 3. Mediterranean Sea and geographic origin (1 to 7) of the Bluefin tuna trap time series. Adapted from Ravier & Fromentin (2004)
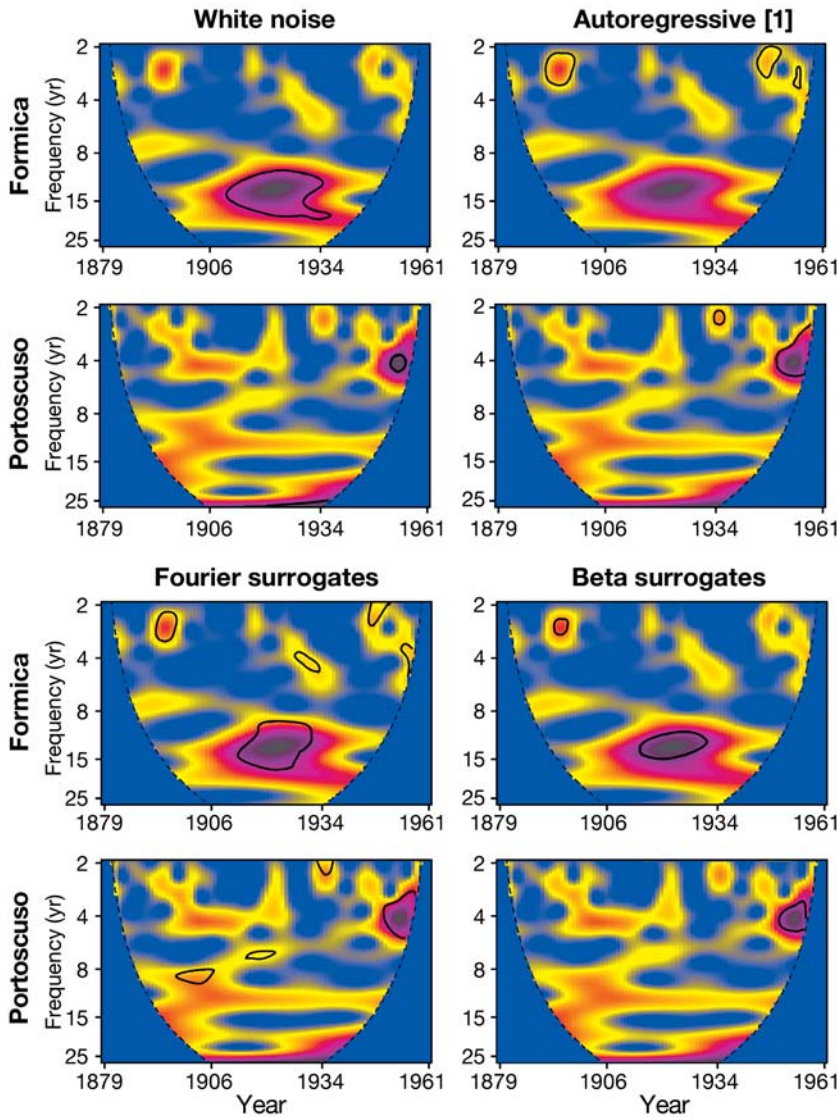
Fig. 4. Wavelet spectra of the Formica and Portoscuso time series, tested with different null hypotheses. We used a white noise process, an AR[1] process, the Fourier surrogates (Type I) and our class of surrogates (Beta surrogates). Solid black lines indicate significant areas at the 5% level. The colour gradient, from dark blue to dark red, codes for low to high power values. Curved dashed lines: limit of the cone of influence, the area where edge effects are present

Time series results

The results for the Formica and Portoscuso time series are presented to illustrate the differences between the methods (Fig. 4). The wavelet spectra displayed different significant areas (indicated by solid black lines in Fig. 4), according to the null model against which they were tested. The spectra tested against white noise displayed large significant areas at low frequencies, whereas the AR[1] process identified significant areas mostly at high frequency (e.g. Formica). These results were expected; an AR[1] is

dominated by low-frequency oscillations, whereas white noise does not produce such highly autocorrelated signals. On the contrary, the Fourier surrogates identified significant areas in the whole range of frequencies. However, their locations were sometimes not consistent (e.g. Portoscuso), and many small areas were identified. These spurious effects were likely to be created by the Fourier surrogates applied on short and non-stationary time series, as explained in 'Methods'. The beta surrogates identified significant areas in the whole range of frequencies. However, unlike the Fourier surrogates, the size of the time series and its non-stationarity did not influence the algorithm used to produce the surrogates. The locations of significant pseudo-cyclic components were well defined, and no spurious effects were detected. This approach thus allowed us to identify cycles significantly different from the expected behaviour of ecological time series displaying similarly coloured noise.

Clustering wavelet spectra

Signals with determined time–frequency properties

In order to illustrate how the classification procedure extracts common time–frequency patterns to compare the wavelet spectra, we first formed a data set with time series displaying known and controlled properties in time and frequency. Six time series displaying contrasting time–frequency properties were simulated, using sine and cosine functions (Fig. 5). The changes in frequency were either abrupt (Fig. 5, time series 2, 3, 5 and 6) or smooth (Fig. 5, time series 1 and 4). Indeed, these time series displayed different dynamics compared to real ecological time series or outputs from models (e.g. a stochastic version of the Ricker model). However, they allowed us to control both the time and frequency properties of the time series, and we could thus obtain wavelet spectra with desired time–frequency patterns.

The dataset with determined time–frequency properties was first analysed using wavelet analysis (Fig. 6). The wavelet spectra (WS) displayed clearly different patterns: (1) an increase in frequency with time; (2) a decrease in frequency with time; and (3) continuous patterns.
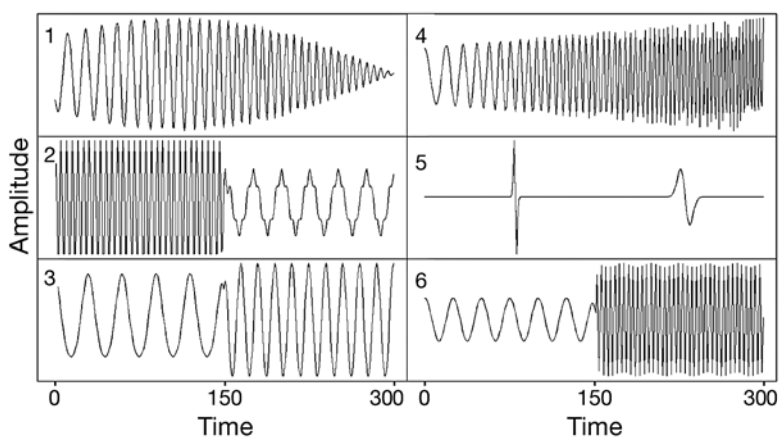
Fig. 5. Simulated dataset. The time series were designed to produce controlled and contrasting time frequency–patterns on their wavelet spectrum

Applying the classification method with a covariance threshold set at C = 99% of the total covariance, thus retaining all the details between the wavelet spectra, clustering the results produced a cluster tree revealing 2 main groups (Fig. 6). The group comprising WS 2 and 5 displayed a decrease in frequency with time, while that consisting of WS 3, 6, 1 and 4 displayed an increase in frequency with time. Examining the cluster in more detail confirmed that the procedure grouped the WS displaying similar time–frequency patterns. For instance, WS 1 and 4 displayed continuous time–frequency patterns that were the same from $t = 0$ up to $t = 150$ (see also Fig. 6); these were the closest of the dataset. On the contrary, WS 2 and 5, that displayed a discontinuous decrease in frequency with time, only shared a reduced common area that was expressed through the height of their connection in the cluster tree.

In order to illustrate how the multivariate method classifies the wavelet spectra, the first leading patterns, first singular vectors and the reconstructed WS by the first axis were plotted for the comparison of WS 1 and 4 and the comparison of WS 2 and 6 from the simulated dataset (Fig. 6). These 2 comparisons were contrasted, as the time–frequency patterns
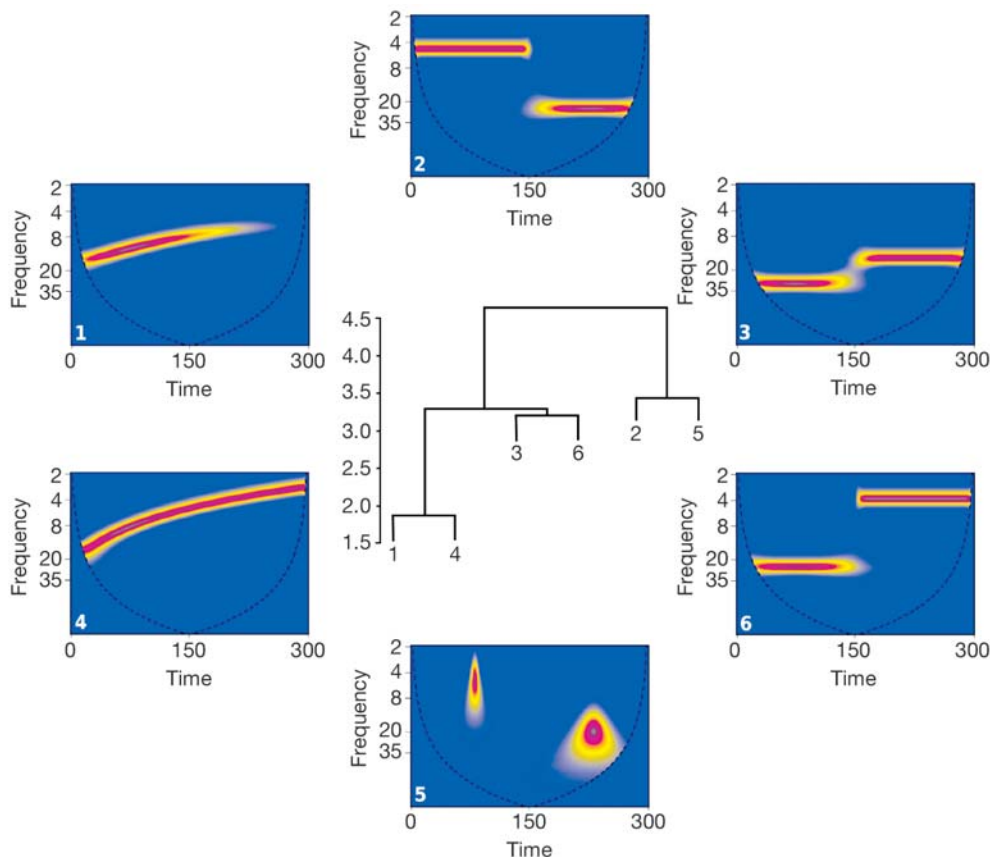

Fig. 6. Wavelet spectra (WS 1 to 6) and cluster tree for the simulated data. The colour gradient, from dark blue to dark red, codes for low to high power values. The cluster tree was obtained using the distance matrix constructed for the simulated data using the classification method. The covariance threshold was set to C = 99% of the total covariance. Curved dashed lines: limit of the cone of influence, the area where edge effects are present
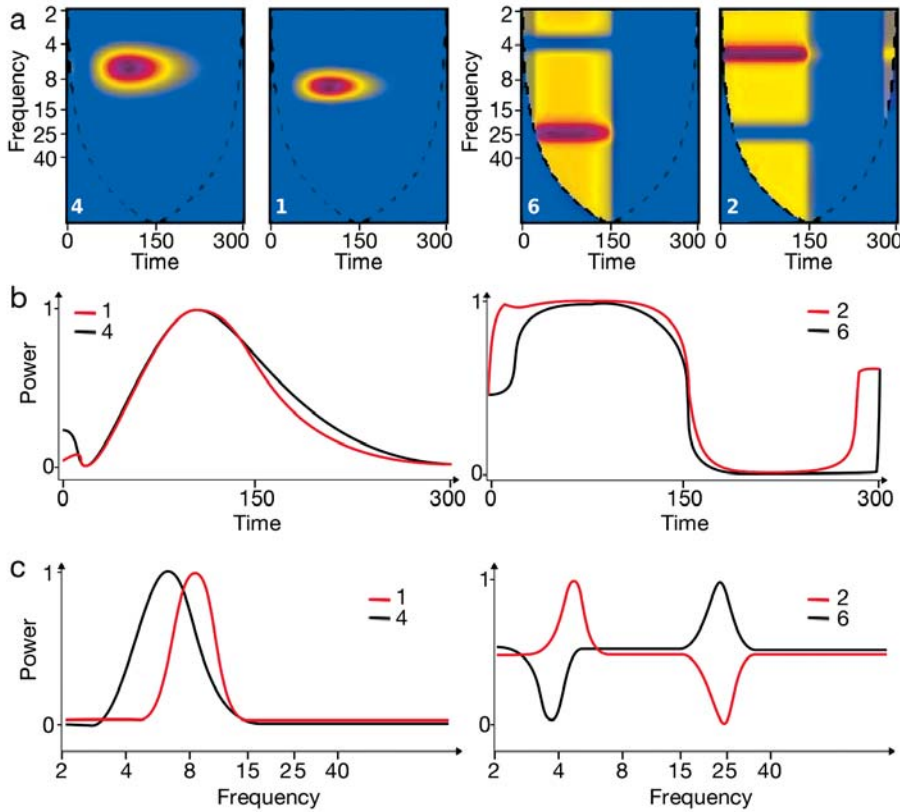
Fig. 7. Comparisons between wavelet spectra (WS) 1 and 4 (comparable patterns) and between WS 2 and 6 (dissimilar patterns) from the simulated dataset. Shown are (a) the first reconstructed WS, (b) leading patterns and (c) singular vectors for each comparison. The WS in (a) were reconstructed using the first leading pattern and the first singular vector extracted from each comparison. The 2 reconstructed WS on the left correspond to the comparison of WS 4 and 1, and the 2 on the right corresponded to the comparison of WS 6 and 2. The colour gradient, from dark blue to dark red, codes for low to high power values. Curved dashed lines: limit of the cone of influence, the area where edge effects are present

between WS 1 and 4 presented a good match, whereas WS 2 and 6 displayed opposed patterns (Fig. 7). WS 1 reconstructed by the first axis displayed a similar pattern with the reconstructed WS 4 (Fig. 7a). The leading patterns (Fig. 7b) and the singular vectors (Fig. 7c) reflected this similarity with, however, a slightly different frequency mode for the singular vectors. The difference between the patterns of the reconstructed WS 6 and 2 (Fig. 7a) was not expressed through the leading patterns (Fig. 7b), but was very well expressed through the singular vectors (Fig. 7c), as they displayed a clearly opposed fluctuation.

Sea surface temperature time series in the Mediterranean

The cluster method was applied to a large data set of Mediterranean sea surface temperature (SST). The Mediterranean is a semi-enclosed basin connected to the Atlantic Ocean by the narrow strait of Gibraltar (Fig. 3) and that consists of 2 main parts — Eastern and Western. The SST data used were extracted from the National Oceanic and Atmospheric Administration Extended Reconstructed Sea Surface Temperature, which is based on the COADS dataset (www.cdc.noaa. gov/cdc/data.noaa.ersst.html). The dataset is available

on a 2 by 2° grid from 1854 to 2005 on a monthly basis. We selected the period 1900 to 2005 in order to avoid spatial coverage problems in the historic part of the dataset. We extracted the monthly time series on each of the 80 pixels available over the Mediterranean and, in order to focus on the interannual variations, we used wavelets to filter the time series and remove the seasonal component. SST is a highly spatially-correlated variable, so that the 80 SST time series only differed by small features. Therefore, checking whether the procedure is able to discriminate spatially homogenous areas among this data set is a good test. In other words, it allows us to test whether it is powerful enough to detect differences among a large number of time series with mainly similar time–frequency properties.

The procedure was run with a covariance threshold fixed at C = 99% of total covariance and analyzed using flexible clustering. The cluster tree obtained (not presented here) was cut at 4 different heights to interpret the first levels of clustering. This led to 2, 3, 4 and 5 groups of pixels, that were then mapped (Fig. 8a–d).

The results obtained divided the Mediterranean into geographically homogenous and consistent units at each aggregation level. The first level of aggregation cut the Mediterranean into 2 clear parts near the Sicilian Strait, that separated the western Mediterranean from the eastern Mediterranean (Fig. 8a). The second
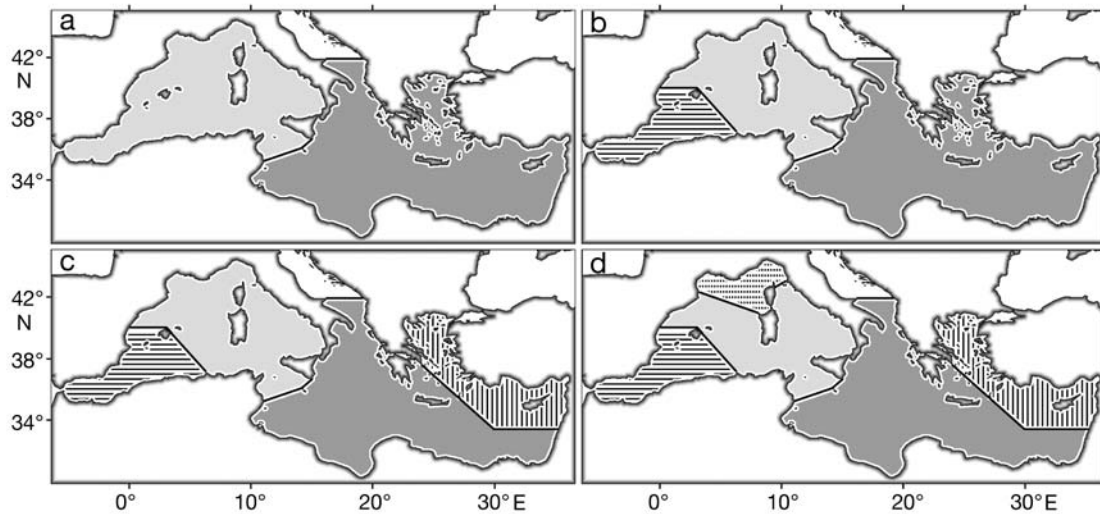
Fig. 8. Classification of the wavelet spectra for the yearly time series of sea surface temperature in the Mediterranean Sea. The classification method was applied with the covariance threshold fixed at C = 99% of the total covariance. The cluster tree was cut to identify (a) 2, (b) 3, (c) 4 and (d) 5 groups of pixels that were mapped and identified by the different shading or hatching. SST time series were extracted from the COADS dataset on a $2 \times 2°$ grid and covered the period 1900 to 2005

level of aggregation identified an area located in the western Mediterranean, that extended from the Albo-ran Sea (see Fig. 3) up to the Balearic islands (Fig. 8b). The third level of aggregation divided the eastern Mediterranean, with a group that combined the Aegean sea with the northern part of the Levantine Basin (Fig. 8c). Finally, the last level of aggregation identified pixels from the Gulf of Lions and the Lig-urian Sea, that were separated from the rest of the western Mediterranean (Fig. 8d).

DISCUSSION

Testing the wavelet spectra

The class of surrogates presented in this study offers a consistent and more powerful approach to the signif-icance testing of the wavelet spectra of ecological time series. The autocorrelation structure of time series is here described as a $1/f^\beta$ process, which takes into account the 'spectral redness' often displayed by real time series (Lawton 1988, Pimm & Redfearn 1988, Ryb-ski et al. 2006). While Fourier Type 1 and Type 2 exactly reproduce the initial spectrum, the beta surro-gate uses the fit of a $1/f^\beta$ model to generate the surro-gates. As fitting such models with the classical fast Fourier transform regression is problematic for short time series, we used the multiple segmenting method to overcome this (Miramontes & Rohani 2002). How-ever, a time series dominated by low frequency fluctu-ations must be long enough to contain sufficient long-

term fluctuations to reliably fit the model. Following Miramontes & Rohani (2002), and from an empirical perception, a size of 40 points seems to be the lesser bound to apply the beta surrogates. The generation of the $1/f^\beta$ process can then be achieved by different techniques. Even if complex techniques proved to be more optimal to generate discrete $1/f^\beta$ processes (Wor-nell 1993, Kasdin 1995), we used the spectral synthesis as it produced consistent results and displayed a good trade-off between simplicity, accuracy and computa-tional speed. The underlying null model assumed is critical for testing the wavelet spectrum. The beta sur-rogates build time series that mimic the slope of the power spectrum of the original time series; in other words, the same relative importance of frequencies in the signal. Therefore, the beta surrogates assume a large range of autocorrelation structures that are not constrained to a reduced frequency band; unlike auto-regressive processes, the beta surrogates consistently test the wavelet spectrum at both low and high fre-quencies. It can also be of further interest when there is important shift in the autocorrelation structure be-tween different time periods, as the surrogates enable an assessement of the significance of this change in the frequency content of the signal.

Time series clustering

Many approaches have been developed in the field of signal processing to compare time series by using their raw properties, the fitted parameters of auto-regressive moving average (ARMA) models or their

rhythmic properties (e.g. Henson et al. 1998, Keogh & Pazzani 1998, Xiong & Yeung 2002, Cazelles 2004). Other approaches have compared time series at various resolutions, but comparing the time series focusing on their time–frequency properties using the continuous wavelet transform as we did has not, to our knowledge, been done (Lukasik 2000, Keogh et al. 2001). Comparing time series based on their wavelet spectra is similar to comparing a set of images, and many methods from this field of research (e.g. face recognition) have used multivariate approaches coupled to wavelet transform techniques to extract features of the original image (Feng et al. 2000, Gupta & Jacobson 2006). The multivariate method used here proved to be powerful for indexing images and also for comparing spatio-temporal fields (Wu et al. 1996). It also allowed us to put the emphasis on the common time–frequency properties of the time series, as this can be of central interest to ecologists dealing with non-stationarity (Hastings 2001, Hsieh et al. 2005, Cazelles & Hales 2006). Transformation, and particularly log-transformation, is a common procedure for ecological data. This is useful for many analyses (e.g. to stabilize or rescale the variance) but it is not a requirement for wavelet analysis because it can handle non-stationary signals. Such transformations may further be applied carefully, as they affect the relative changes in amplitude between time steps and can thus distort the time–frequency patterns detected by the wavelet analysis.

Being fairly robust, the method is not expected to fail in common ecological cases. However, simulated cases showed that wavelet spectra displaying patchy patterns without a dominant frequency mode and no clear common areas could lead the procedure to produce spurious associations. One could also get counter-intuitive groupings when a wavelet spectrum shares different areas with 2 spectra, and when these 2 other spectra do not clearly share a common area (an artefact that is also known for cluster analysis computed from a distance matrix between raw time series). In fact, the main limitation of the method is that it does not specify the common patterns detected; one has to go back to the wavelet spectra to identify them. However, using the reconstruction by the first axes of the MCA to identify the common patterns between wavelet spectra may help in detecting such spurious grouping (e.g. Fig. 7a).

Classification of the Mediterranean sea surface temperature

As expected with such a spatially autocorrelated variable, the results obtained by the classification procedure on the SST time series clearly separated different areas. However, the areas obtained did not consist of random groups of contiguous locations, but rather reflected different hydrological conditions, as they were consistent with outputs from oceanic models of the Mediterranean Sea. For instance, the separation found between the eastern and the western part of the Mediterranean at the relatively shallow and narrow Sicily Strait (some 450 m depth and 140 km wide) conforms with the literature, as it separates their respective gyral circulation (e.g. Zavatarelli & Mellor 1995). The combination of the Alboran Sea and the south of the Balearic Basin (Fig. 8b) identifies an area of more dynamic circulation. The Atlantic water that enters the basin forms anticyclonic gyres in the Alboran sea (Fig. 3), and it defines a front at their eastern boundary from which mesoscale eddies arise and drift into the Balearic Basin (Millot 1985, Tintore et al. 1988). The joining of the Aegean sea with the northern part of the Levantine Basin (Fig. 8c) separates the Rhodes Gyre from the more southwestern Shikmona and Mersa-Matruh gyres (Pinardi & Masetti 2000). The last level of aggregation roughly separates the Gulf of Lions and the Ligurian Sea from the rest of the western Mediterranean (Fig. 8d). This area corresponds to a gyre but is also the source — with the northern Adriatic, that was not covered by our dataset — of deep water (Zavatarelli & Mellor 1995). If spatial groups were expected from a classification of such a spatially autocorrelated variable, the boundaries obtained well matched the different hydrological conditions over the region. This showed that the procedure was able to detect consistent groups over a large number of rather similar wavelet spectra.

CONCLUSIONS

Associations between environmental and ecological signals are often transient and difficult to identify. Wavelet analysis offers a powerful way to investigate these associations, but it has so far been restricted to univariate or bivariate analyses. Our approach in the present study allows the exterior of the use of wavelet analysis to multiple ecological and environmental time series. For instance, this procedure can be used as a basis for investigating potential statistical links between ecological and environmental time series or how an environmental variable can affect different populations. The beta surrogate addresses a second critical point, i.e. the assessment of whether or not the associations detected are likely to be an artefact caused by the intrinsic autocorrelation structure. The 2 approaches presented in this study might thus form

consistent tools for quantitative ecologists that often face non-stationary time series and their potential associations with environmental fluctuations, in a multiple time series framework.

LITERATURE CITED

Arino A, Pimm SL (1995) On the nature of population extremes. Evol Ecol 9:429–443

Bradshaw GA, Spies TA (1992) Characterizing canopy gap structure in forests using wavelet analysis. J Ecol 80:205–215

Bretherton CS, Smith C, Wallace JM (1992) An intercomparison of methods for finding coupled patterns in climate data. J Clim 5:541–560

Cazelles B (2004) Symbolic dynamics for identifying similarity between rhythms of ecological time series. Ecol Lett 7:755–763

Cazelles B, Hales S (2006) Infectious diseases, climate influences and nonstationarity. PLoS Med 3:1212–1213

Cazelles B, Stone L (2003) Detection of imperfect population synchrony in an uncertain world. J Anim Ecol 72:953–968

Cazelles B, Chavez M, Berteaux D, Ménard F, Vik JO, Jenouvrier S, Stenseth NC (2008). Wavelet analysis of ecological time series. Oecologia (in press), doi: 10.1007/s00442-008-0993-2

Chatfield C (2004). The analysis of time series: an introduction, 6th edn. Chapman & Hall/CRC, Boca Raton, FL

Cheng X, Dunkerton TJ (1995) Orthogonal rotation of spatial patterns derived from singular value decomposition. J Clim 8:2631–2643

Cuddington KM, Yodzis P (1999) Black noise and population persistence. Proc R Soc Lond B Biol Sci 266:969–973

Daubechies I (1992). Ten lectures on wavelets. Society for Industrial and Applied Mathematics, Philadelphia, PA

Feng GC, Yuen PC, Dai DQ (2000) Human face recognition using PCA on wavelet subband. J Electron Imaging 9:226–233

Grenfell BT, Bjørnstad O, Kappey J (2001) Travelling waves and spatial hierarchies in measles epidemics. Nature 414:716–723

Gupta M, Jacobson N (2006). Wavelet principal component analysis and its application to hyperspectral images. In: Proc IEEE Int Conf on Image Processing 2006, 8–11 Oct., Atlanta, GA. IEEE Signal Processing Society, Washington, DC, p 1585–1588

Halley JM (1996) Ecology, evolution and 1/f noise. Trends Ecol Evol 11:33–38

Halley JM, Stergiou KI (2005) Increasing variability of fish landings. Fish Fish 6:266–276

Hastings A (2001) Transient dynamics and persistence in ecological systems. Ecol Lett 4:215–220

Henson SM, Cushing JM, Constantino RF, Dennis B, Desharnais RA (1998) Phase switching in population cycles. Proc R Soc Lond B Biol Sci 265:2229–2234

Hsieh CH, Glaser SM, Lucas AJ, Sugihara G (2005) Distinguishing random environmental fluctuations from ecological catastrophes for the North Pacific Ocean. Nature 435:336–340

Inchausti P, Halley JM (2002) The long-term temporal variability and spectral colour of animal populations. Evol Ecol Res 4:1033–1048

Kasdin NJ (1995) Discrete simulation of colored noise and stochastic-processes and 1/f(Alpha) power-law noise generation. Proc IEEE 83:802–827

Keitt TH, Fischer J (2006) Detection of scale-specific community dynamics using wavelets. Ecology 87:2895–2904

Keitt TH, Urban DL (2005) Scale-specific inference using wavelets. Ecology 86:2497–2504

Keogh E, Pazzani M (1998). An enhanced representation of time series which allows fast and accurate classification, clustering and relevance feedback. In: Agrawal R, Stolorz P, Piatetsky-Shapiro G (eds) Proc 4th Int Conf Knowledge Discovery and Data Mining (KDD'98). ACM Press, New York, p 239–241

Keogh E, Chakrabarti K, Pazzani M, Mehrotra S (2001) Dimensionality reduction for fast similarity search in large time series databases. Knowl Inf Syst 3:263–286

Klvana I, Berteaux D, Cazelles B (2004) Porcupine feeding scars and climatic data show ecosystem effects of the solar cycle. Am Nat 164:283–297

Lau KM, Weng H (1995) Climate signal detection using wavelet transform: how to make a time series sing. Bull Am Meteorol Soc 76:2391–2402

Lawton JH (1988) More time means more variation. Nature 334:563

Lukasik E (2000). Wavelet packets based features selection for voiceless plosives classification. In: ICASSP '00: Proc IEEE Int Conf on Acoustic, Speech and Signal Processing, Istanbul, Turkey, 5–9 June, 2000, Vol 2. IEEE Computer Society, Washington, DC, p II689–II692

Mallat S, Papanicolaou G, Zhang Z (1998) Adaptive covariance estimation of locally stationary processes. Ann Stat 26:1–47

Maraun D, Kurths J (2004) Cross wavelet analysis: significance testing and pitfalls. Nonlinear Processes in Geophysics 11:505–514

Ménard F, Marsac F, Bellier E, Cazelles B (2007) Climatic oscillations and tuna catch rates in the Indian Ocean: a wavelet approach to time series analysis. Fish Oceanogr 16:95–104

Millot C (1985) Some features of the Algerian current. J Geophys Res 90:7169–7176

Miramontes O, Rohani P (2002) Estimating 1/f scaling exponents for short time-series. Physica D 166:147–154

Newman M, Sardeshmukh P (1995) A caveat concerning singular value decomposition. J Clim 8:352–360

Petchey OL, Gonzalez A, Wilson HB (1997) Effects on population persistence: the interaction between environmental noise colour, intraspecific competition and space. Proc R Soc Lond B Biol Sci 264:1841–1847

Pimm S, Redfearn A (1988) The variability of animal populations. Nature 334:613–614

Pinardi N, Masetti E (2000) Variability of the large scale general circulation of the Mediterranean Sea from observations and modelling: a review. Palaeogeogr Palaeoclimatol Palaeoecol 158:153–173

Ravier C, Fromentin JM (2001) Long-term fluctuations in the eastern Atlantic and Mediterranean bluefin tuna population. ICES J Mar Sci 58:1299–1317

Ravier C, Fromentin JM (2004) Are the long-term fluctuations in Atlantic bluefin tuna (*Thunnus thynnus*) population related to environmental changes? Fish Oceanogr 13:145–160

Royer F, Fromentin JM (2006) Recurrent and density-dependent patterns in long-term fluctuations of Atlantic bluefin

tuna trap catches. Mar Ecol Prog Ser 319:237–249

Rybski D, Bunde A, Havlin S, Von Storch H (2006) Long-term persistence in climate and the detection problem. Geophys Res Lett 33:L06718

Saitoh T, Cazelles B, Vik JO, Viljugrein H, Stenseth NC (2006) Effects of the regime shift on population dynamics of the grey-sided vole in Hokkaido, Japan. Clim Res 32:109–118

Schreiber T, Schmitz A (1996) Improved surrogate data for nonlinearity tests. Phys Rev Lett 77:635–638

Schreiber T, Schmitz A (2000) Surrogate time series. Physica D 142:346–382

Steele JH (1985) A comparison of terrestrial and marine ecological systems. Nature 313:355–358

Stenseth NC, Falck W, Chan KS, Bjørnstad ON and others (1998) From patterns to processes: phase and density dependencies in the Canadian lynx cycle. Proc Natl Acad Sci USA 95:15430–15435

Theiler J, Eubank S, Longtina A, Galdrikiana B, Farmer JD (1992) Testing for nonlinearity in time series: the method of surrogate data. Physica D 58:77–94

Tintore J, La Violette P, Blade I (1988) A study of an intense density front in the eastern Alboran Sea: the Almeria-Oran front. J Phys Oceanogr 18:1384–1397

Torrence C, Compo GP (1998) A practical guide to wavelet analysis. Bull Am Meteorol Soc 79:61–78

Vasseur DA, Yodzis P (2004) The color of environmental noise. Ecology 85:1146–1152

von Storch H, Zwiers F (1999). Statistical analysis in climate research. Cambridge University Press, Cambridge

Voss RF (1988) Fractals in nature: from characterization to simulation. In: Peitgen HO, Saupe D (eds). The science of fractal images. Springer-Verlag, New York, p 21–70

Wornell GW (1993) Wavelet-based representations for the 1/f family of fractal processes. Proc IEEE 81:1428–1450

Wu D, Agrawal D, El Abbadi A, Singh A, Smith TR (1996). Efficient retrieval for browsing large image databases. In: Proc 5th Int Conf on Knowledge Information. ACM Press, New York, p 11–18

Xiong Y, Yeung DY (2002). Mixtures of ARMA models for model-based time series clustering. In: Proc 2nd IEEE Int Conf on Data Mining (ICDM'02), Maebashi City, Japan, 9–12 Dec 2002. IEEE Computer Society, Los Alamitos, CA, p 717–720

Zavatarelli M, Mellor GL (1995) A numerical study of the Mediterranean Sea circulation. J Phys Oceanogr 25:1384–1414