



Research

Cite this article: Cazelles B, Cazelles K, Chavez M. 2014 Wavelet analysis in ecology and epidemiology: impact of statistical tests. *J. R. Soc. Interface* **11**: 20130585. <http://dx.doi.org/10.1098/rsif.2013.0585>

Received: 2 July 2013

Accepted: 5 November 2013

Subject Areas:

biometrics, environmental science

Keywords:

wavelet analysis, statistical testing, ecology, epidemiology

Author for correspondence:

Bernard Cazelles

e-mail: cazelles@biologie.ens.fr

[†]Present address: UMR 5554, CNRS-Univ. Montpellier II, Montpellier, France.

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rsif.2013.0585> or via <http://rsif.royalsocietypublishing.org>.

Wavelet analysis in ecology and epidemiology: impact of statistical tests

Bernard Cazelles^{1,2}, Kévin Cazelles^{1,†} and Mario Chavez³

¹UMR 7625, UPMC-ENS, Ecole Normale Supérieure, 46 rue d'Ulm, Paris Cedex 05 75230, France

²UMMISCO UMI 209 IRD-UPMC, Bondy 93142, France

³UMR 7225 UPMC-CNRS, Hôpital de La Pitié-Salpêtrière, 47 Bd. de l'Hôpital, Paris Cedex 13 75651, France

Wavelet analysis is now frequently used to extract information from ecological and epidemiological time series. Statistical hypothesis tests are conducted on associated wavelet quantities to assess the likelihood that they are due to a random process. Such random processes represent null models and are generally based on synthetic data that share some statistical characteristics with the original time series. This allows the comparison of null statistics with those obtained from original time series. When creating synthetic datasets, different techniques of resampling result in different characteristics shared by the synthetic time series. Therefore, it becomes crucial to consider the impact of the resampling method on the results. We have addressed this point by comparing seven different statistical testing methods applied with different real and simulated data. Our results show that statistical assessment of periodic patterns is strongly affected by the choice of the resampling method, so two different resampling techniques could lead to two different conclusions about the same time series. Moreover, our results clearly show the inadequacy of resampling series generated by white noise and red noise that are nevertheless the methods currently used in the wide majority of wavelets applications. Our results highlight that the characteristics of a time series, namely its Fourier spectrum and autocorrelation, are important to consider when choosing the resampling technique. Results suggest that data-driven resampling methods should be used such as the hidden Markov model algorithm and the 'beta-surrogate' method.

1. Introduction

Numerous studies in ecology and epidemiology consist of analysing time series to extract information and to identify scales of different patterns. However, ecological processes are typically not stationary, and there are an increasing number of papers that highlight the non-stationary features of population dynamics (see [1]), so that a global timescale decomposition may not be appropriate. Grenfell *et al.* [2] introduced wavelet analysis for characterizing non-stationary epidemiological time series. Wavelet analysis performs a local timescale decomposition of the signal, i.e. the estimation of its spectral characteristics as a function of time [3–7] or space [8]. During the past decade, wavelet analysis has been successfully applied in ecology and epidemiology, and it appears particularly attractive given the non-stationary nature of both ecological and environmental time series and the relationships between these series.

As reviewed in the electronic supplementary material, table S1, increasing numbers of works have used wavelet analysis to analyse ecological and epidemiological time series. Some of these papers were concerned with the determination of the characteristics of time series and the analysis of their possible association with environmental signals. The use of wavelets is of great interest in the case of the analysis of very long populational and environmental time series such as those that come from the rich China archives [9–12]. It is clear that the probability of finding marked non-stationarities increases with the length of the time series. This is also true for spatial analysis where the length of the analysed spatial signal can be important [13]. Wavelet analysis can also reveal an abrupt shift in the cyclic dynamics of population [14] or [15]. A major

breakthrough from wavelet analysis is linked to the existence of transient or discontinuous association between the fluctuations of population and climatic variables [16]. The recent development of wavelet clustering for multiple time series analysis represents another major innovation [17]. Other works have used the wavelet approach to compare different analysis techniques [18] or [19] and another important interest of wavelet analysis consists of comparing the frequency characteristics from observations and model simulations [20,21]. It has been demonstrated that wavelet analysis can be a powerful complementary approach when comparing modelling predictions to the ever-increasing abundance of *in situ* data [22]. Other analyses investigated the phenomenon of population synchrony where wavelets are used to extract the phase of the time-series [23–25]. Wavelet analysis was also applied to the analysis of spatial patterns in vegetation systems, but most of these works dealt with discrete wavelet (DW) transforms [8,13,26].

However, important questions concerning the wavelet techniques remain unresolved. One issue is illustrated by the following two examples, concerning the significance of wavelet quantities. The first example concerns the phase computation. Johnson *et al.* [24] computed the phase of numerous time series of larch budmoth based on wavelet decomposition for the period 1961–1998. They compared the phase differences between all time series for the full period or for different periods even after 1985 where the time series are flat without variability (see the electronic supplementary material of Johnson *et al.* [24]). This example illustrates the need for the adequate statistical test of the wavelet spectrum before extracting the phase. The second example concerns Johansson *et al.* [27], who examined the dynamic relationship between climate variables and the incidence of dengue in Thailand, Mexico and Puerto Rico using wavelet analysis. One of the main findings of Johansson *et al.* [27] is that dengue incidence oscillations are dominated by a marked seasonal component in these three countries. These results contradict previous published works where the role of the 2–3 year periodic components is highlighted [16,28]. Nevertheless, the results obtained by Johansson *et al.* [27] showed the strong dependence of the wavelet analysis on a null hypothesis using AR(p) stochastic process.

These examples clearly show that the interpretation of wavelet analysis results depends on adequate statistical testing of the wavelet quantities. Statistical hypothesis testing involves two components: a discriminant wavelet quantity, and a null hypothesis, against which observations are tested. If the null hypothesis is too simple, the explanation that we seek for analysing the data can appear inadequate. Because statistical tests cannot easily be calculated analytically for signals with time-varying spectrum, resampling techniques are frequently employed and a white noise or red noise null hypothesis is used. As for all Monte Carlo resampling techniques, the choice of the null model is central for defining adequate significance regions and then can affect the potential conclusions. Nevertheless, less work has also been done on comparing bootstrap, synthetic or surrogate data in the case of wavelet analysis (but see [7] or [8]). The validity of any test of wavelet quantities would depend heavily on the consistency of the generated synthetic data with the chosen null hypothesis and the associated stochastic process.

Here, we tested the consistency of different resampling techniques. These resampling schemes incorporate some of the key components of a time series: mean, variance and distributions of values in both time and frequency domains. By incorporating some of these components, we were able to

incorporate some degree of similarity in the synthetic time series generated by the sampling schemes and then tested different null hypothesis.

Our comparisons between the different null hypothesis and associated resampling schemes were based on the distribution of the time-series values, the classical Fourier spectrum and the wavelet power spectrum. The influence of these resampling techniques on the significance of wavelet quantities was then analysed. The performances of these resampling techniques were evaluated on the monthly reported cases of dengue haemorrhagic fever (DHF) in Thailand, and on two simulated time series.

2. Material and methods

2.1. Time series analysed

To test different null hypothesis, we have based our demonstration on the analysis of three very different non-stationary time series (figure 1). To characterize these series, we have computed the distribution of the values of these time series and their classical Fourier spectrum. These characteristics are displayed in figure 1; figure 1*a–c* shows the original time series; figure 1*d–f* displays the amplitude distributions of the time series and the Fourier spectrum are depicted in figure 1*g–i*. The three analysed series are

- (i) the monthly number of reported cases of DHF in Bangkok (figure 1*a*). Figure 1*d* displays the distribution of the time-series values. On average, the oscillations of the dengue incidence time series are dominated by the annual mode (figure 1*g*), but they also have a statistically significant common mode of oscillation around a 2–3 year period [16,28].
- (ii) an AR(2) time series with two parameter sets in a reasonable ecological range as proposed in the classical work of Royama [30]. The two sets of parameters are chosen to obtain dynamics with two periodic components. Figure 1*b* shows $n = 100$ data points, figure 1*e* shows the distribution of these data points and figure 1*h* its Fourier spectrum. One can observe the abrupt variation of the main periodicity around $t_s = 60$, and the period is *ca* eight iterations before t_s , and *ca* four iterations after.
- (iii) the time series of infectious individuals from a chaotic SEIR (susceptible–exposed–infectious–removed) model with transient dynamics [29]. Figure 1*b* shows the time series, and figure 1*e* shows the distribution of the values of this time series. Figure 1*h* displays the classical Fourier spectrum of this time series showing large peaks at a period of 1, 2 and 3–4 years.

2.2. Wavelet analysis

Wavelet analysis is based on the wavelet transform that decomposes signals over dilated and translated functions called ‘mother wavelets’, $\psi(t)$ that can be expressed as the function of two parameters, one for the time position τ , and the other for the scale of the wavelets a :

$$W_x(a, \tau) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) \psi^* \left(\frac{t - \tau}{a} \right) dt,$$

where the asterisk denotes the complex conjugate form and $x(t)$ the signal.

As in most wavelet applications of wavelets in ecology and epidemiology, we use the Morlet wavelets:

$$\psi(t) = \pi^{-1/4} \exp(-i2\omega_0 t) \exp\left(-\frac{t^2}{2}\right).$$

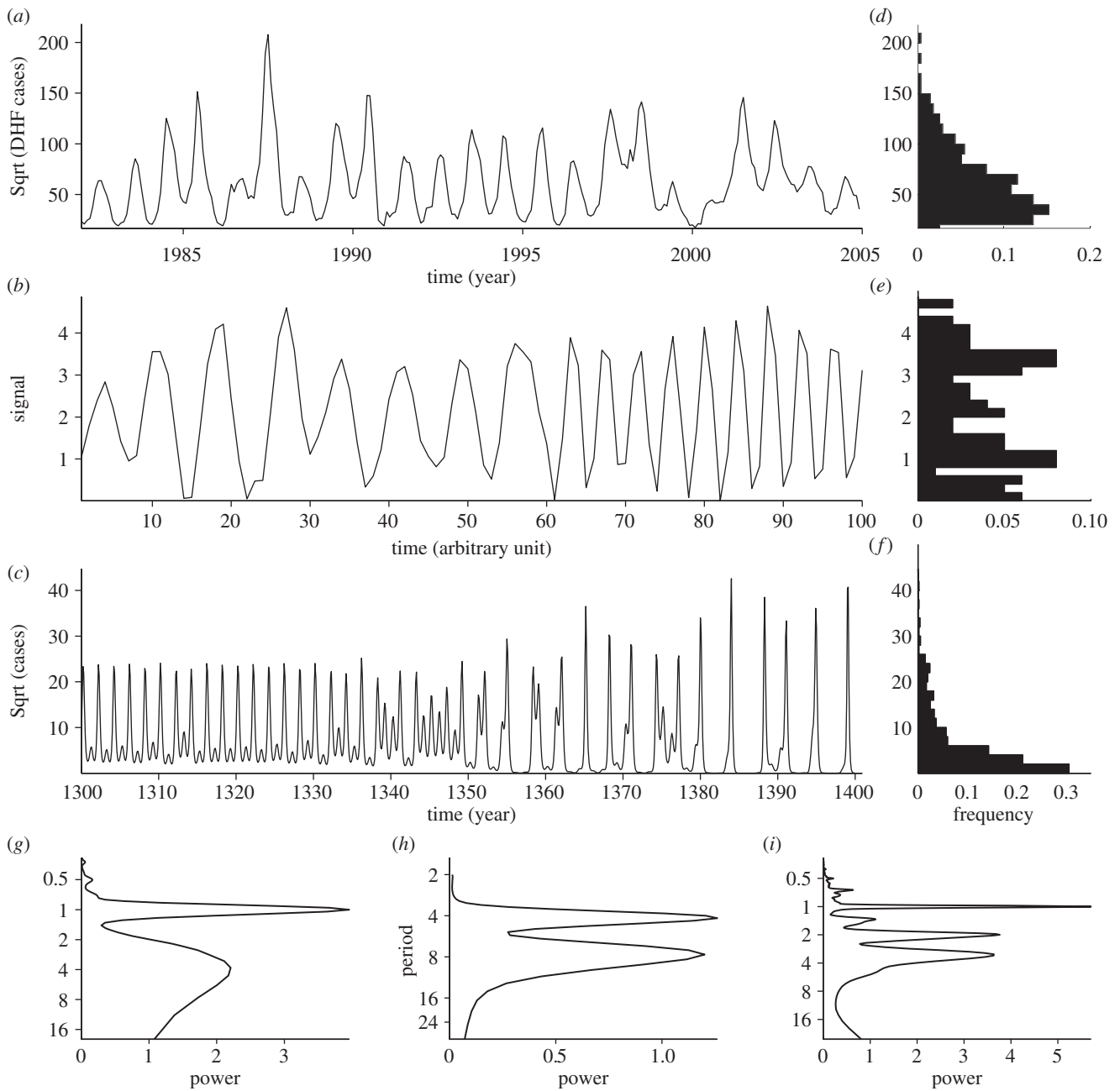


Figure 1. Analysed time series and their characteristics, the Fourier spectrum and the distribution of the values. (a) Monthly number of reported cases of dengue haemorrhagic fever in Bangkok for the time period 1987–2005. (b) AR(2) time series. The AR(2) model is: $y(t) = a_i + b_i y(t-1) + c_i y(t-2) + \varepsilon_i$. With parameters defined for the time period between $t_0 = 0$ and $t_s = 60$ as: $a_1 = 1.5$, $b_1 = 1.2$, $c_1 = -0.9$, ε_i a random Gaussian component with 0 mean and variance $\sigma_1^2 = 0.15$; and after t_s as: $a_2 = 4$, $b_2 = 0.1$, $c_2 = -1$, $\sigma_2^2 = 0.05$; $n = 100$ data points have been used. (c) Time series of the infectious population generated by a chaotic SEIR model [29]:

$$\begin{aligned} dS/dt &= \mu - \beta(t)SI - \mu S \\ dE/dt &= \beta(t)SI - (\mu + \alpha)E \\ dI/dt &= \alpha E - (\mu + \gamma)I \\ dR/dt &= \gamma I - \mu R \\ \beta(t) &= \beta_0(1 + \beta_1 \cos 2\pi t) \end{aligned}$$

with the parameters values $\alpha = 35.84$, $\mu = 0.02$, $\gamma = 100$, $\beta_0 = 1800$, $\beta_1 = 0.28$. (d–f) The distribution of the value of the three analysed time series. (g–i) The Fourier spectrum of the three analysed time series, (a–c) respectively.

Some discussion on the choice of the mother wavelet can be found in Cazelles *et al.* [6,7].

An important point with the continuous wavelet (CW) is that the relationship between the wavelet frequency and the wavelet scale that can be derived analytically (see [3]). For the Morlet

wavelet, the relation between wavelet frequencies and wavelet scales is given by

$$\frac{1}{f} = \frac{4\pi a}{\omega_0 + \sqrt{2 + \omega_0^2}}$$

When $\omega_0 \approx 2\pi$, the wavelet scale a is inversely related to the frequency, $f \approx 1/a$. This greatly simplifies the interpretation of the wavelet analysis and one can replace, on all equations, the scale a by the period $1/f$.

In some sense, the wavelet transform can be regarded as a generalization of the Fourier transform and by analogy with spectral approaches one can compute the local wavelet power spectrum:

$$S_x(f, \tau) = \| W_x(f, \tau) \|^2.$$

CWs yield a redundant decomposition in the sense that the information extracted from a given scale slightly overlaps that extracted from neighbouring scales. Nevertheless, CWs are more robust to noise when compared with other decomposition schemes. DWs have the advantage of fast implementation, but the number of scales and the time invariant property (a filter is time invariant if shifting the input in time correspondingly shifts the output) strongly depend on the data length (see [7] for a comparison of CWs and DWs in an ecological context).

2.3. Null hypothesis tested and associated synthetic series

As with other time-series methods, it is crucial to assess the statistical significance of the patterns exhibited by the wavelet analysis. Our aim here is to test whether the wavelet-based quantities (e.g. the spectra, co-spectra, coherence or phase) observed at a particular position on the timescale plane are not due to a random process. One starts with the observed time series, which are to be tested against the hypothesis of control datasets referred to as the synthetic or surrogate data, which share some statistical properties with the original series, but which are generated by different random processes.

There exists a large range of null hypotheses and associated resampling procedures, including (i) independent and identically distributed noise; (ii) linearly filtered noise; or (iii) a monotonic nonlinear transformation of linearly filtered noise. Further details of the algorithms can be found in [31]. We want to generate surrogate time series, associated with our null hypothesis, such that they mimic some key features of the raw dataset, but with some restrictions. Then, the resampling schemes used respect some of the key constituents of the time series: (i) their mean, (ii) their variance, (iii) the distribution of their values and (iv) their autocorrelation function that describes the time distribution of the variance (or the Fourier spectrum that equivalently defines the frequency distribution of the time-series variance).

Here, we have tested seven different null hypotheses with the same mean and variance as the raw time series but with different degree of similarity with the key components of the raw time series:

- (i) The first synthetic time series are simple bootstrapped series that can be assimilated to series generated by a white noise process, granting a flat power spectrum [32]. The null hypothesis tested is then: *the observed values of the wavelet quantities are identical to those that can be generated by a white noise process*. In this case, just the mean and the variance of the raw series are identical. The distributions of the raw value and the variance are, however, substantially different in many cases.
- (ii) Making the assumption of a white noise process is generally not appropriate for ecological data, while an autoregressive process with red noise characteristics, e.g. an AR(1), can be an acceptable model in numerous cases for both ecological and environmental data [33]. These synthetic time series are generated with an AR(1) stochastic process where the parameter of the model is fitted on the raw time series. The null hypothesis tested states that *the*

significant periodic characteristics are identical to those of a red noise generated by an AR(1) process. The synthetic series generated have the same mean, the same variance, a Gaussian distribution of their values and the autocorrelation function of an AR(1) process.

- (iii) As suggested by Johansson *et al.* [27], we have also used synthetic series generated by an AR(p) with $p > 1$. Fitting different models and selecting the best one based on an information criterion such as the AIC determine the p order. The null hypothesis is *identity between periodic characteristics of the observed time series and those of an AR(p) process*. The synthetic series generated have the same mean, the same variance, a Gaussian distribution of their values and the autocorrelation function of an AR(p) process.
- (iv) Ecological time series display a large variety of autocorrelation structures that can be described by neither a white noise, nor an autoregressive process [33,34]. For this reason, we used the synthetic series proposed by Rouyer *et al.* [17], called 'beta-surrogates', which display a similar autocorrelation structure to the original time series and the same relative distribution of frequencies. This allows the dominance of low frequencies often displayed by ecological time series to be taken into account. Using this approach, we obtain surrogates that mimic the shape of the original ecological time series by displaying a power spectrum with the same slope in the log scale, but without exactly reproducing it (figure 2). In this case, the synthetic series have identical mean, variance and value distribution; moreover, the time distribution of the variance is similar to those of the raw series. The null hypothesis associated with this resampling scheme states: *the observed time series can be generated by a stochastic processes, such that it keeps the same distribution of values and similar Fourier spectrum*.
- (v) To preserve the short-term temporal correlations, a resampling scheme based on a Markov process has been used. These synthetic series are computed in the following way (see [35] or [36]): the raw time series are binned to form a frequency histogram of b equal-sized bins; a transition matrix M that describes the transition from values of bin _{i} to value in bin _{j} is then estimated based on the actual relative frequencies of the data. This transition matrix is then used to generate surrogate time series of the same length as the original data. The null hypothesis underlying this synthetic series stipulates that *the observed significant values of the wavelet quantities are identical to those that can be generated by a stochastic process that has the same distribution of the values and identical short-term autocorrelation structure*. Then, the synthetic series have identical mean, variance and value distribution. For their variance distribution, they just mimic the high-frequencies components.
- (vi) The block bootstrap is the most general method to improve the accuracy of bootstrap for dependent data as time series mainly because it can preserve the original time series structure within a block [37]. We have used block bootstrap but with random block length. As there is no proper diagnostic tool to choose the optimal block length, we have defined a mean block length in accordance with the main periodicity of the raw time series. This sampling scheme generated synthetic series with identical mean, variance, value distribution, but just the dominant period is respected. This scheme is associated with a null hypothesis defined by: *the time order of quasi-identical periodic components has no consequential influence on the significance of the time distribution of the variance*.
- (vii) The last surrogate datasets correspond to the amplitude-adjusted Fourier transform (AAFT) algorithm proposed by Theiler *et al.* [38]. This algorithm generates synthetic

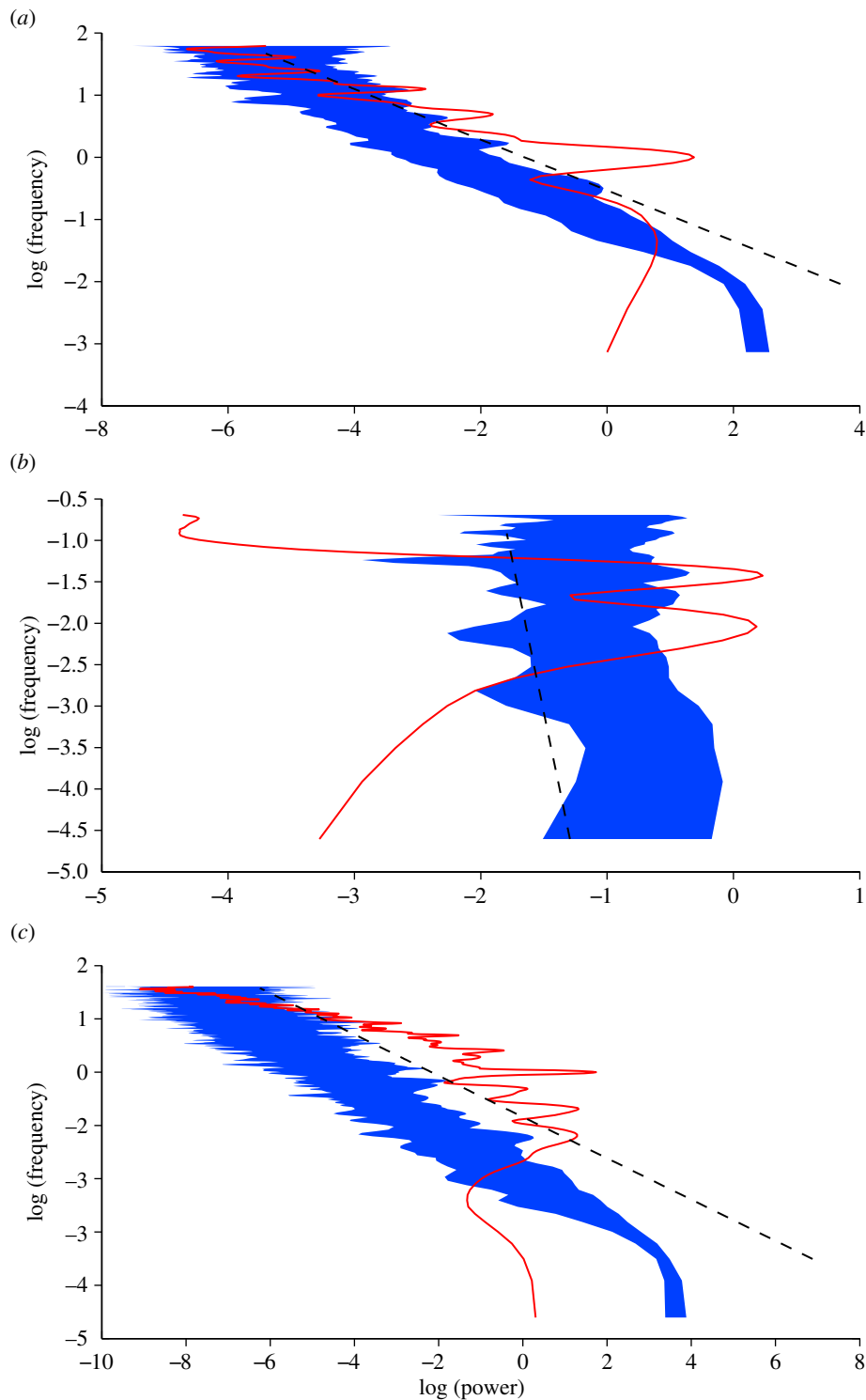


Figure 2. The ‘beta-surrogate’ synthetic time series. On each graph, the black line is the Fourier spectrum of the raw time series on a log–log scale, the black dashed line is the slope of the Fourier spectrum, and the grey area is the envelop of the Fourier spectrum of 1000 synthetic time series. (a) Monthly number of reported cases of dengue haemorrhagic fever in Bangkok (figure 1a). (b) AR(2) time series (figure 1b). (c) Time-series of the infectious population generated by a chaotic SEIR model (figure 1c). (For the online colour figure: on each graph, the red line is the Fourier spectrum of the raw time series on a log–log scale, the black dashed line is the slope of the Fourier spectrum and the blue area is the envelop of the Fourier spectrum of 1000 synthetic time series.) (Online version in colour.)

time series preserving the mean, the variance and both the value distribution and the autocorrelation function of the original time series. The AAFT algorithm approximates the sample power spectrum based on a phase randomization of the raw spectrum and a back transformation with the inverse Fourier transform producing the surrogate data [38]. The null hypothesis associated is then: *the significant periodic components are identical to those of a*

stochastic processes that has the same distribution of the values and identical Fourier spectrum.

The details of the non-classical synthetic methods (iv, v, vii) are described in the electronic supplementary material.

For each synthetic series, the wavelet transform and ‘related quantities’ are computed. As the process is repeated n_s times, the distribution of these ‘wavelet quantities’ under each null

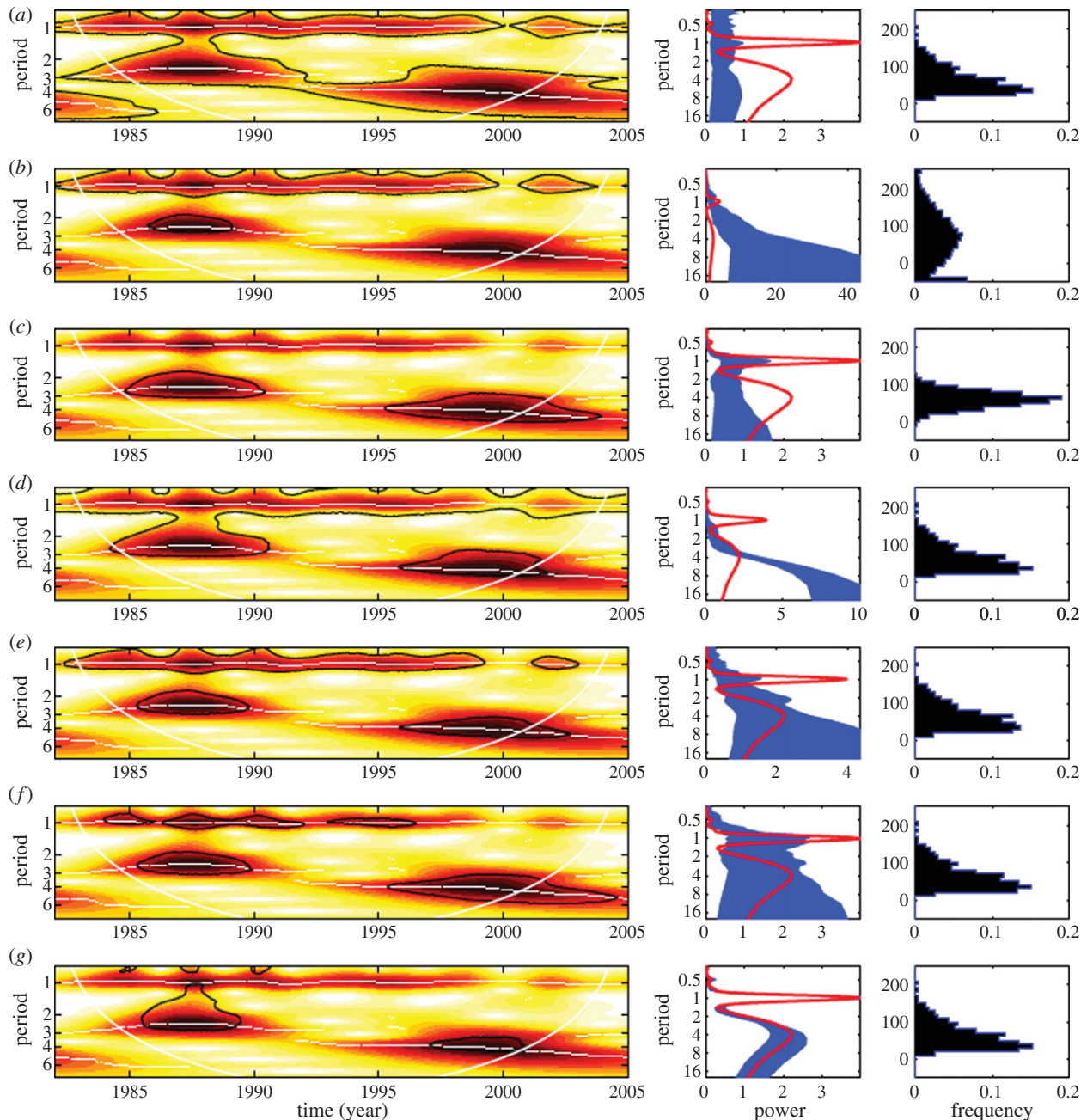


Figure 3. Wavelet analysis of the monthly number of reported cases of dengue haemorrhagic fever in Bangkok and characteristic of both the raw data and the synthetic time series used for each statistical test. (a) Test with simple bootstrapped time series. (b) Red noise test with an AR(1) process. (c) Test AR(p) process with $p > 1$, p is determined by the optimal AIC value of different models. (d) 'Beta-surrogates test'. (e) Test with HMM synthetic series. (f) Test with block bootstrap with random block length. (g) Test with AAFT synthetic series. Three panels compose each line. The first panel is the wavelet power spectrum of the raw series for each statistical test. The colours code for power values from white (low values) to dark grey (high values). The bold black lines show the 5% significant levels computed based on 1000 synthetic series. Dotted white lines represent the maxima of the undulations of the wavelet power spectrum, and the large white line indicates the cone of influence that delimits the region not influenced by edge effects. The second panel corresponds to the Fourier spectrum of the raw time series in black and the 1000 synthetic series in grey. The third panel is the distribution of the values of the 1000 synthetic time series. (For the online colour figure: in the first panel the colours code for power values from white (low values) to dark red (high values). The second panel corresponds to the Fourier spectrum of the raw time series in red and the 1000 synthetic series in blue.) (Online version in colour.)

hypothesis is constructed. We can compare the wavelet quantities of the raw series with their distribution under the null hypothesis, extracting, for instance, the 95th or the 99th percentiles of this distribution. We are then able to determine how strongly significant periodic components compared with those generated by the null hypothesis.

3. Results

Our main results are presented in figures 3 and 4 and in the electronic supplementary material, figure S1.

3.1. Consistency between synthetic and observed time series

Considering the spectral characteristic of the synthetic series, our analyses show that the simple method (simple bootstrapping and AR(1)) generates synthetic series with an inadequate Fourier spectrum (figures 3a,b and 4a,b and the electronic supplementary material, figure S1A,B). This is also verified for the AR(0) synthetic series (not shown). On the other extremity of the complexity of the resampled series, the AAFT synthetic series have a quasi-identical Fourier spectrum (figures 3g and 4g and

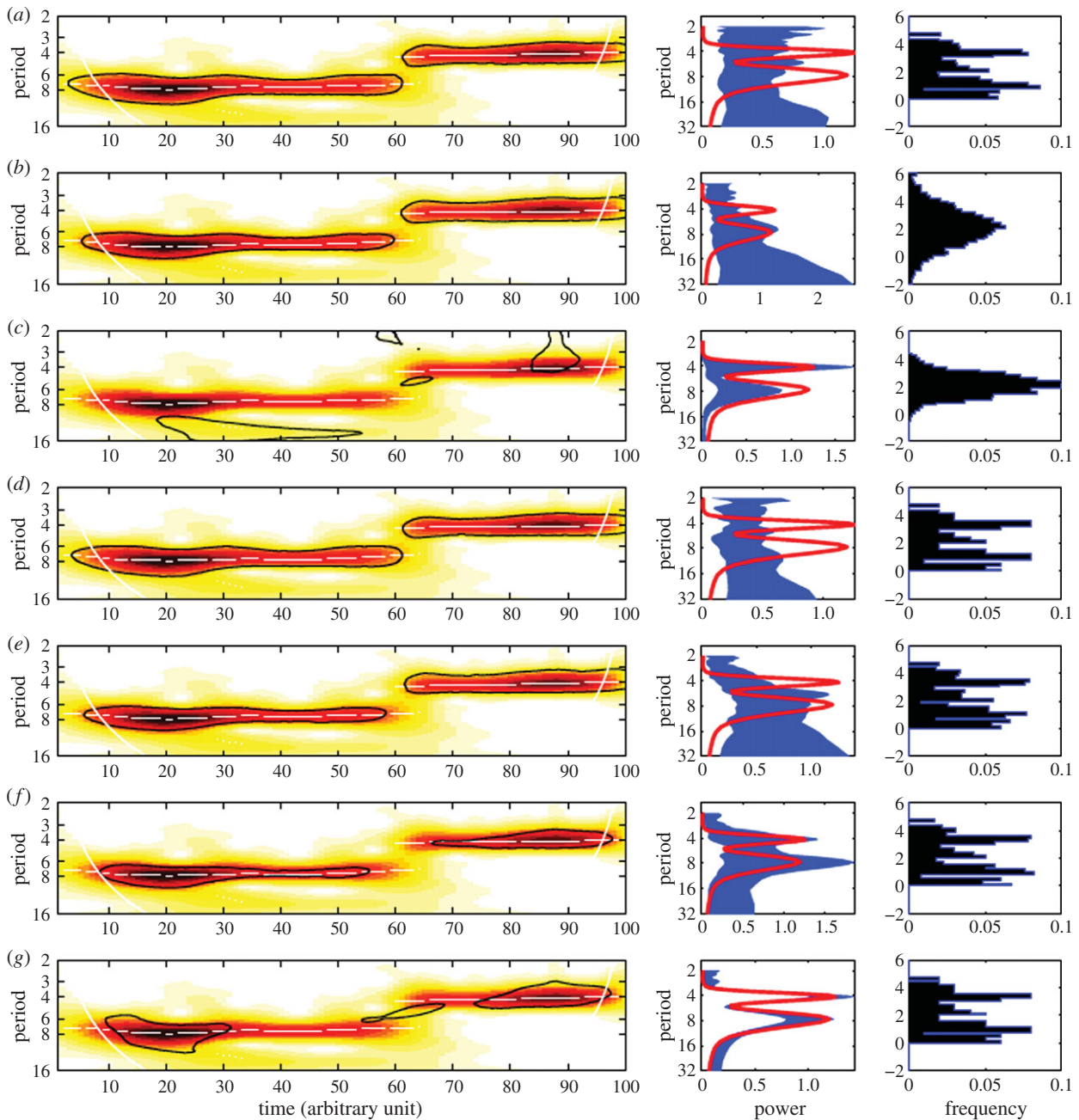


Figure 4. As figure 3 but for the AR(2) time series. (Online version in colour.)

the electronic supplementary material, figure S1G). Between these two extremes, the spectra of the series generated by the AR(p) process (except in the case of AR(2) analysed series (figure 4c)) are not strongly in accordance with the spectra of the raw series (figure 3c and the electronic supplementary material, figure S1C). Surprisingly, this is also the case for the 'beta-surrogate' (figures 3d and 4d and the electronic supplementary material, figure S1D). This last point underlines the difficulty of fitting a linear model on an observed Fourier spectrum (figure 2). Conversely, block bootstrap and HMM synthetic series have Fourier spectrum more in accordance with the observed spectrum but with an interesting variability that will generate synthetic series different from the raw series (figures 3e,f and 4e,f and the electronic supplementary material, figure S1E,F).

We observe from the third column of figures 3 and 4 and electronic supplementary material, figure S1 that the distributions of the values generated based on AR(1) or AR(p) processes are Gaussian and not consistent with those of the analysed time series. This is also the true for synthetic series

generated with a white noise process define by an AR(0) process (not shown). A visual investigation of the realizations of the different synthetic series (electronic supplementary material, figure S2) stress the inadequacy of simple bootstrap as well as the series generated by AR(1) and AR(p) processes for null hypothesis tests. Further, these simple inspections of the basic features of the different synthetic time series clearly demonstrate the inadequacy of surrogate series generated by white noise and red noise that are nevertheless currently used in the vast majority of wavelets applications.

3.2. Significance of wavelet power spectrum

Concerning the significance of the wavelet power spectrum, clearly the white noise series give more significant patterns at all the major periodic components, whereas a 'stricter' null hypothesis gave few significant patterns. This is illustrated for the AAF method where few areas are significant (figures 3a and 4a and the electronic supplementary material,

figure S1A). In this case, as the synthetic data have on average the same Fourier spectrum and the same variance repartition in the full period range, the regions that appear significant are those that have characteristics higher than the average (figures 3g and 4g and electronic supplementary material, figure S1G). Similar observation can be done for the AR(2) analysed series when the synthetic time series are generated with AR(p) processes, as the analysed process and the synthetic time series are too close, few areas are significant (figure 4c). Overfitting to real data greatly reduces the significant periodic patterns. For very simple time series, such as the AR(2) models, any resampling scheme gives very similar significant patterns mainly, because removing the order of the time series is enough to give significance. Nevertheless, for more complex data, the significant patterns depend both on the null hypothesis and on the spectral properties of the analysed series. Thus, it is difficult to generalize the different conclusions for more complex and realistic time series. Specifically, for DHF, the white noise model yields significant periodic components for both 1 and 2–3 years, whereas the red noise model give more significance on the seasonal component (figure 3a). More consistent synthetic series based on HMM or block bootstrap resampling give a high significance to the seasonal component, and also to the 2–3 years components (figure 3e,f).

4. Discussion

Wavelet analysis can help us to interpret multi-scale, non-stationary time-series data and reveal features that could not otherwise be seen [6–8]. Wavelet analysis is thus becoming an important tool for analysing time series, and has important practical applications in environmental sciences (see electronic supplementary material, table S1). Moreover, wavelets also appear as an interesting tool for characterizing and validating simulations of models of biological and environmental processes [20–22]. Nevertheless, the test of significance of ‘wavelet quantities’ is critical to enhance the contribution of wavelet analysis for producing reliable results rather than just producing colourful images.

In this work, we proposed a comparison of different null models for wavelet testing. The associated resampling schemes incorporated some of the key components of the time series and then some degree of similarity with the observed data. The performance of the different tests with their associated resampling techniques is evaluated through three criteria: the distribution of the raw value, the Fourier power spectrum and the significant region in the wavelet power spectrum. We have tested a large range of null hypothesis and resampling techniques for simple random shuffling to more elaborated surrogate series. It is clear that a simple random shuffling of data appears as a ‘weak’ null hypothesis that could be easily rejected. At the other extreme, the surrogate by Theiler *et al.* [38] represents a ‘hard’ null hypothesis that could be rejected with difficulty as these synthetic time series have the same Fourier spectrum as the raw time series.

First, one can note that for very short time series, most of the null hypotheses and associated resampling schemes reach similar conclusions. A possible explanation of this weak sensitivity to the choice of the resampling scheme for short time series may be related to the fact that the results from the statistical tests may mainly be affected by the breakdown of the time order.

Our results clearly show that the use of resampled time series under the white noise hypothesis is insufficient. Our findings also demonstrate that the red noise hypothesis (AR(1) process) generates resampled time series with characteristics that are not in agreement with those of the raw time series. These two methods can give inadequate significant areas in the wavelet power spectrum that must be interpreted carefully. Nevertheless, these two methods are currently used in the vast majority of wavelets applications. Similarly, AR(p) (with $p \in \mathbb{N}^+$) processes give, in most cases, synthetic series with inappropriate characteristics.

Contrarily, methods driven by the data generated synthetic time series that have characteristics more in accordance with those of the raw time series. AR(p) with higher order considerably reduce the discrimination power of the hypothesis testing in case of pseudo-periodic data and are not adequate in the case of more complex periodic components such as chaotic time series (see electronic supplementary material, figure S1C).

With stochastic processes, our proposed $1/f^\beta$ process or ‘beta-surrogate’ [17] gives importance to larger periodic bands in greater accordance with the apparent reality (figure 2).

The AAFT algorithm produces synthetic time series that are highly similar to raw data, so that significant regions in the wavelet power spectrum are sparse and the null hypothesis appears very hard to reject. Nevertheless, with this hypothesis, one can clearly obtain information about when the spectral characteristics of the time series differ from the mean characteristics. This is the case not only for very large values of the power spectrum but also for the regions where there is a large changing frequency behaviour (figure 4g). This can be a particular advantage when the objective is to characterize marked regime shifts or turning points.

Overall, the bootstrap algorithm and test based on simple stochastic processes led to smallest efficiency results in terms of characteristics of the generated resampled series and identified structure in the time-frequency analysis, whereas the HMM algorithm and the bootstrapping by block gave consistently good results. However, the accuracy of the block bootstrap is sensitive to the choice of the block length, and the optimal block length depends on the sample size, the data-generating process and the statistic considered. To date, there is no proper diagnostic tool to select the optimal block length and it still remains as an unsolved question for future studies. This appears particularly troublesome for time series with complex spectra.

We conclude that using data-driven resampling techniques to generate synthetic time series with characteristics in agreement with those of the empirical dataset is often preferable when wavelet approaches are applied. Considering the difficulty in bootstrapping by block, the use of the HMM algorithm ([35] or [36]) is an interesting compromise. It worth noting that these methods (HMM, bootstrap by block and beta-surrogate) are data-driven approaches and carry no *a priori* assumptions about the intrinsic processes generating these data, they just display similar statistical properties of the raw data. This conclusion can be moderated by the fact the statistical test used may also be dictated by the research question underlying the analysis of a time series. This is the case in the regime shift studies for which we have shown that the test associated with the AAFT algorithm has the principal characteristic of focusing mainly on the time position of significant ‘wavelet quantities’ (figure 4g). A regime shift can be found by demonstrating a marked modification of the

periodic components at a given time (see [14] or [15]), but the use of the AAFT algorithm would be successful for confirming the time of the shift without exploring the significance of the other part of the spectrum.

Wavelets present other limitations and weakness (see Maraun & Kurths [39]). One of these weaknesses is related to the different normalization methods used. Contrary to Fourier analysis, in wavelet analysis, there are difficulties in obtaining adequate normalization that preserves all interesting features, namely a flat spectrum for white noise and an identical spectrum for sine waves of same amplitude but different frequency. In most applications, one uses the normalization proposed by Torrence & Compo [5], which preserves a flat white noise spectrum but sine waves of equal amplitude exhibit different integrated power proportional to their oscillation scale. Nevertheless, Kaiser [40] has suggested other normalizations, which only preserve one of the mentioned features.

Considering its growing use (see electronic supplementary material, table S1), wavelet analysis is an important addition to

time-series methods with practical applications to population biology and ecology. The wavelet approach can be used alongside other time series tools to examine directly the relationship (association, dependence, synchrony) between studied time-series. We believe that wavelet analysis can contribute to future advances in our understanding of ecological or epidemiological processes in our changing world, but this cannot be done without adequate statistical testing. We suggest favouring statistical significance tests based on resampling techniques for wavelet-based quantities and hope that our present analyses will favour this development.

Acknowledgements. We thank Hedvig Nenzén and Trevor Bedford for insightful comments.

Funding statement. B.C. was partially supported by a CNRS grant 'Projets Exploratoires Pluridisciplinaires Inter-Instituts' 2011–2012 and B.C. and K.C. were partially funded by the European Commission Seventh Framework Programme (FP7/2007-2013) for the DENFREE project under grant agreement no. 282-378.

References

- Cazelles B, Hales S. 2006 Infectious diseases, climate influences, and nonstationarity. *PLoS Med.* **3**, e328. (doi:10.1371/journal.pmed.0030328)
- Grenfell BT, Bjørnstad ON, Kappey J. 2001 Travelling waves and spatial hierarchies in measles epidemics. *Nature* **414**, 716–723. (doi:10.1038/414716a)
- Meyers SD, Kelly BG, O'Brien JJ. 1993 An introduction to wavelet analysis in oceanography and meteorology: with application to the dispersion of Yanai waves. *Mon. Weather Rev.* **121**, 2858–2866. (doi:10.1175/1520-0493(1993)121<2858:AITWAI>2.0.CO;2)
- Lau K-M, Weng H. 1995 Climate signal detection using wavelet transform: how to make a time series sing. *Bull. Am. Meteorol. Soc.* **76**, 2391–2402. (doi:10.1175/1520-0477(1995)076<2391:CSDUWT>2.0.CO;2)
- Torrence C, Compo G. 1998 A practical guide to wavelet analysis. *Bull. Am. Meteorol. Soc.* **79**, 61–78. (doi:10.1175/1520-0477(1998)079<0061:APGTWA>2.0.CO;2)
- Cazelles B, Chavez M, Constantin de Magny G, Guégan J-F, Hales S. 2007 Time-dependent spectral analysis of epidemiological time-series with wavelets. *J. R. Soc. Interface* **4**, 625–636. (doi:10.1098/rsif.2007.0212)
- Cazelles B, Chavez M, Berteaux D, Ménard F, Vik JO, Jenouvrier S, Stenseth NC. 2008 Wavelet analysis of ecological time series. *Oecologia* **156**, 287–304. (doi:10.1007/s00442-008-0993-2)
- James PMA, Fleming RA, Fortin M-J. 2010 Identifying significant scale-specific spatial boundaries using wavelets and null models: spruce budworm defoliation in Ontario, Canada as a case study. *Landscape Ecol.* **25**, 873–887. (doi:10.1007/s10980-010-9465-2)
- Zhang Z, Cazelles B, Tian H, Stige LC, Bräuning A, Stenseth NC. 2009 Periodic temperature-associated drought/flood drives locust plagues in China. *Proc. R. Soc. B* **276**, 823–831. (doi:10.1098/rspb.2008.1284)
- Zhang Z, Tian H, Cazelles B, Kausrud KL, Bräuning A, Guo F, Stenseth NC. 2010 Periodic climate cooling enhanced natural disasters and wars in China during AD 10–1900. *Proc. R. Soc. B* **277**, 3745–3753. (doi:10.1098/rspb.2010.0890)
- Tian H, Stige LC, Cazelles B, Kausrud KL, Svarverud R, Stenseth NC, Zhang Z. 2011 Reconstruction of a 1,910-y-long locust series reveals consistent associations with climate fluctuations in China. *Proc. Natl Acad. Sci. USA* **108**, 14 521–14 526. (doi:10.1073/pnas.1100189108)
- Ben-Ari T *et al.* 2012 Identification of Chinese plague foci from long-term epidemiological data. *Proc. Natl Acad. Sci. USA* **109**, 8196–8201. (doi:10.1073/pnas.1110585109)
- Bradshaw GA, Spies TA. 1992 Characterizing canopy gap structure in forests using wavelet analysis. *J. Ecol.* **80**, 205–215. (doi:10.2307/2261007)
- Jenouvrier S, Weimerskirch H, Barbraud C, Park Y-H, Cazelles B. 2005 Evidence of a shift in the cyclicity of Antarctic seabird dynamics linked to climate. *Proc. R. Soc. B* **272**, 887–895. (doi:10.1098/rspb.2004.2978)
- Saitoh T, Cazelles B, Vik JO, Viljugrein H, Stenseth NC. 2006 Effects of regime shifts on the population dynamics of the grey-sided vole in Hokkaido, Japan. *Clim. Res.* **32**, 109–118. (doi:10.3354/cr032109)
- Cazelles B, Chavez M, McMichael AJ, Hales S. 2005 Nonstationary influence of El Niño on the synchronous dengue epidemics in Thailand. *PLoS Med.* **2**, e106. (doi:10.1371/journal.pmed.0020106)
- Rouyer T, Fromentin J, Stenseth N, Cazelles B. 2008 Analysing multiple time series and extending significance testing in wavelet analysis. *Mar. Ecol. Prog. Ser.* **359**, 11–23. (doi:10.3354/meps07330)
- Rodríguez-Arias MA, Rodó X. 2004 A primer on the study of transitory dynamics in ecological series using the scale-dependent correlation analysis. *Oecologia* **138**, 485–504. (doi:10.1007/s00442-003-1464-4)
- Grados D, Fablet R, Ballón M, Bez N, Castillo R, Lezama-Ochoa A, Bertrand A, Brickman D. 2012 Multiscale characterization of spatial relationships among oxycline depth, macrozooplankton, and forage fish off Peru using geostatistics, principal coordinates of neighbour matrices (PCNMs), and wavelets. *Can. J. Fish. Aquat. Sci.* **69**, 740–754. (doi:10.1139/f2012-017)
- Pascual M, Cazelles B, Bouma MJ, Chaves LF, Koelle K. 2008 Shifting patterns: malaria dynamics and rainfall variability in an African highland. *Proc. R. Soc. B* **275**, 123–132. (doi:10.1098/rspb.2007.1068)
- Laneri K, Bhadra A, Ionides EL, Bouma M, Dhiman RC, Yadav RS, Pascual M. 2010 Forcing versus feedback: epidemic malaria and monsoon rains in northwest India. *PLoS Comput. Biol.* **6**, e1000898. (doi:10.1371/journal.pcbi.1000898)
- Kara EL *et al.* 2012 Time-scale dependence in numerical simulations: assessment of physical, chemical, and biological predictions in a stratified lake at temporal scales of hours to months. *Environ. Model. Softw.* **35**, 104–121. (doi:10.1016/j.envsoft.2012.02.014)
- 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. (doi:10.1086/423431)
- Johnson DM, Bjørnstad ON, Liebhold AM. 2004 Landscape geometry and travelling waves in the larch budmoth. *Ecol. Lett.* **7**, 967–974. (doi:10.1111/j.1461-0248.2004.00659.x)
- Broutin H, Guégan J-F, Elguero E, Simondon F, Cazelles B. 2005 Large-scale comparative analysis of pertussis population dynamics: periodicity, synchrony, and impact of vaccination.

- Am. J. Epidemiol.* **161**, 1159–1167. (doi:10.1093/aje/kwi141)
26. James PMA, Sturtevant BR, Townsend P, Wolter P, Fortin M-J. 2011 Two-dimensional wavelet analysis of spruce budworm host basal area in the border lakes landscape. *Ecol. Appl.* **21**, 2197–2209. (doi:10.1890/09-1876.1)
 27. Johansson MA, Cummings DAT, Glass GE. 2009 Multiyear climate variability and dengue: El Niño southern oscillation, weather, and dengue incidence in Puerto Rico, Mexico, and Thailand: a longitudinal data analysis. *PLoS Med.* **6**, e1000168. (doi:10.1371/journal.pmed.1000168)
 28. Cummings D, Irizarry R, Huang N, Endy T, Nisalak A, Ungchusak K, Burke DS. 2004 Travelling waves in the occurrence of dengue haemorrhagic fever in Thailand. *Nature* **427**, 344–347. (doi:10.1038/nature02225)
 29. Tidd CW, Olsen LF, Schaffer WM. 1993 The case for chaos in childhood epidemics. II. Predicting historical epidemics from mathematical models. *Proc. R. Soc. Lond. B* **254**, 257–273. (doi:10.1098/rspb.1993.0155)
 30. Royama T. 1992 *Analytical population dynamics*. London, UK: Chapman & Hall.
 31. Schreiber T, Schmitz A. 2000 Surrogate time series. *Phys. D Nonlinear Phenom.* **142**, 346–382. (doi:10.1016/S0167-2789(00)00043-9)
 32. Efron B, Tibshirani RJ. 1993 *An introduction to the bootstrap*. London, UK: Chapman & Hall.
 33. Vasseur DA, Yodzis P. 2004 The color of environmental noise. *Ecology* **85**, 1146–1152. (doi:10.1890/02-3122)
 34. Cuddington KM, Yodzis P. 1999 Black noise and population persistence. *Proc. R. Soc. Lond. B* **266**, 969–973. (doi:10.1098/rspb.1999.0731)
 35. Cazelles B, Stone L. 2003 Detection of imperfect population synchrony in an uncertain world. *J. Anim. Ecol.* **72**, 953–968. (doi:10.1046/j.1365-2656.2003.00763.x)
 36. Cazelles B. 2004 Symbolic dynamics for identifying similarity between rhythms of ecological time series. *Ecol. Lett.* **7**, 755–763. (doi:10.1111/j.1461-0248.2004.00629.x)
 37. Politis DN. 2003 The impact of bootstrap methods on time series analysis. *Stat. Sci.* **18**, 219–230. (doi:10.1214/ss/1063994977)
 38. Theiler J, Eubank S, Longtin A, Galdrikian B, Doynne Farmer J. 1992 Testing for nonlinearity in time series: the method of surrogate data. *Phys. D Nonlinear Phenom.* **58**, 77–94. (doi:10.1016/0167-2789(92)90102-5)
 39. Maraun D, Kurths J. 2004 Nonlinear processes in geophysics cross wavelet analysis: significance testing and pitfalls. *Nonlinear Proc. Geophys.* **11**, 505–514. (doi:10.5194/npg-11-505-2004)
 40. Kaiser G. 1994 *A friendly guide to wavelets*. Boston, MA: Birkhäuser.