Bayesian modeling and significant features exploration in wavelet power spectra

D. V. Divine, F. Godtliebsen

Nonlin. Processes Geophys., 14, 79–88, 2007
www.nonlin-processes-geophys.net/14/79/2007/

Nonlinear Processes in Geophysics

Bayesian modeling and significant features exploration in wavelet power spectra

D. V. Divine and F. Godtliebsen

Department of Mathematics and Statistics, Faculty of Science, University of Tromsø, 9037, Norway

Received: 29 November 2005 – Revised: 5 December 2006 – Accepted: 2 February 2007 – Published: 12 February 2007

Abstract. This study proposes and justifies a Bayesian approach to modeling wavelet coefficients and finding statistically significant features in wavelet power spectra. The approach utilizes ideas elaborated in scale-space smoothing methods and wavelet data analysis. We treat each scale of the discrete wavelet decomposition as a sequence of independent random variables and then apply Bayes' rule for constructing the posterior distribution of the smoothed wavelet coefficients. Samples drawn from the posterior are subsequently used for finding the estimate of the true wavelet spectrum at each scale. The method offers two different significance testing procedures for wavelet spectra. A traditional approach assesses the statistical significance against a red noise background. The second procedure tests for homoscedasticity of the wavelet power assessing whether the spectrum derivative significantly differs from zero at each particular point of the spectrum. Case studies with simulated data and climatic time-series prove the method to be a potentially useful tool in data analysis.

1 Introduction

A variety of different methods and tools have been developed to analyze statistical properties of data sequences. A study of time-series at different levels of time/space resolution represents a particular interest. Classical approaches, such as the Fourier transform, allow analysis of the frequency content in the signal. This implicitly presumes the harmonicity of the studied process. For most real time-series, however, this assumption is not accurate, leading to misinterpretations of the output results.

Decomposing a time-series into wavelets, in turn, allows highlighting of the variability features at different time-scales (Kaiser, 1994; Torrence and Compo, 1998; Percival and Walden, 2000), and is essentially a tool to visualize the frequency content of a signal as it varies through time. Over the last decades, wavelets have become a popular tool for data analysis. Applied fields that are now making use of wavelets include signal and image processing in physical studies, engineering, music, medicine etc.

It is known that the raw wavelet-based estimator of the time-varying power spectrum suffers from the same serious disadvantage as the periodogram in the Fourier analysis. Being an asymptotically inconsistent estimator of the true spectrum it requires some kind of smoother to be applied in a frequency domain to reduce the variance of the individual power measurements. A natural extension to wavelets would be to assume stationarity over some time interval and smooth the wavelet spectrum along the time axis.

In this study, we formulate a method in a Bayesian framework, with the smoothing procedure efficiently substituted by sampling from the posterior density. The latter is constructed basing on the prior information that can be inferred from the data themselves. In developing this approach we largely utilize ideas elaborated in a family of the so-called scale-space techniques (Chaudhuri and Marron, 1999; Park et al., 2004; Godtliebsen and Øigård, 2005). The second key issue we try to address in the paper is a search for features in the analyzed data that are "really there", or in other words, are statistically significant relative to the established hypothesis. The method implements two independent significance testing procedures for the estimated wavelet spectrum. A conventional one, introduced in Torrence and Compo (1998), hypothesizes that the background process can adequately be described by the stationary AR(1) model and tests for the presence of features inconsistent with it. For the second approach, adopted from the scale-space methods, a test for non-stationarity in a wavelet variance is developed with the decision rule based on the spectrum derivative.

Correspondence to: D. V. Divine
(dmitry.divine@npolar.no)

Published by Copernicus GmbH on behalf of the European Geosciences Union and the American Geophysical Union.

The paper is presented as follows. In Sects. 2.1 and 2.2 we present the basics of the wavelet theory and show how the application of Bayes rule can be used for modeling the wavelet coefficients. Finding the smoothing parameter β through solving the minimization problem is shown in Sect. 2.4. The procedure utilizes the estimate of the noise variance, introduced earlier in Sect. 2.3. Section 2.5 justifies the choice of the mother wavelet function. Section 2.6 briefly introduces the concept of the wavelet spectrum and provides significance tests for the smoothed wavelet power. Section 3 describes the numerical implementation of the proposed technique. In Sect. 4 we show some examples of data analysis to demonstrate the method's performance and potential, followed by conclusions in Sect. 5.

2 Method

2.1 Wavelet transform

Wavelet decompositions can be commonly divided into two principal classes following the type of the basis used for transformation. This comprises the use of an orthogonal basis in the discrete wavelet transform (DWT), a nonorthogonal basis in the maximal overlap discrete wavelet transform (MODWT), or the continuous wavelet transform (CWT). A wavelet function used for constructing the basis can be either real or complex. Thus, one can also distinguish between complex (captures better oscillatory behaviour) and real (more suitable for isolating peaks or discontinuities) wavelet transforms. We in this study restrict our analysis to real wavelets only, although the theoretical considerations are generally applicable to complex wavelets too.

Given a discrete stochastic process u_t, t=1, .., N, with a time increment δt, a continuous wavelet transform is defined as a convolution of u_t with a scaled and translated version of the "mother wavelet" ψ_0 which forms a basis of the transform. We write

$$W_t(s) = \sqrt{\frac{\delta t}{s}} \sum_{t'=1}^{N} u_{t'} \psi_0[(t'-t)\frac{\delta t}{s}]. \quad (1)$$

The wavelet transform can generally be thought of as an extension of the common discrete Fourier transform with the periodic exponential $e^{i\omega t}$ replaced with a localized wavelet function $\psi_0[(t'-t)\frac{\delta t}{s}]$. This mother wavelet function is located around time t and stretched according to the investigated scale s.

The continuous decomposition scale s of the CWT in case of the DWT is substituted by a dyadic scale 2^{j-1}, j=1, ..., J where j denotes a decomposition level. Expressing the convolution operation in terms of a linear filtering, the discrete wavelet transform writes as follows:

$$W_{jt} = \sum_{l=0}^{L_j-1} h_{j,l} u_{2^j(t+1)-1-l \mod N} \quad (2)$$

where t=1, ..., N_j, summation is over the width $L_j \equiv (2^j-1)(L-1)+1$ of the wavelet filter h_j at scale j, and L denotes the width of the wavelet filter at scale 1. Notation "$2^j(t+1)-1-l \mod N$" is defined as follows. If j is an integer such that $0 \le j \le N-1$, then $j \mod N \equiv j$; if j is any other integer, then $j \mod N \equiv j+pN$, where pN is the unique integer multiple of N such that $0 \le j+pN \le N-1$ (Percival and Walden, 2000). The number N_j of wavelet coefficients at each decomposition level j follows the law $N_j = N/2^j$ provided that the analyzed sample size $N = l2^J$ for some integers $J < J0$ and l, with $J0$ denoting the number of levels in the "full" DWT (Percival and Walden, 2000). In practice the length of the time-series may be an integer multiple of 2^J only by chance. To override this restriction the "padding" with zeroes up to a nearest integer multiple of 2^J is used, with subsequent elimination of the biased wavelet coefficients. Note that the use of the orthogonal basis ensures that the derived wavelet coefficients do not contain redundant information, i.e. they are approximately independent both along and across the scales.

2.2 Modeling the wavelet coefficients: a Bayesian approach

Suppose that the observed signal u_t, t=1, ..., N can be presented in the vector form as

$$\boldsymbol{u} = \hat{\boldsymbol{u}} + \boldsymbol{\eta} \quad (3)$$

where $\hat{\boldsymbol{u}} = [\hat{u}_1, \ldots, \hat{u}_N]^T$ is the true underlying signal. The superscript T denotes the transpose, and $\boldsymbol{\eta} = [\eta_1, \ldots, \eta_N]^T$ denotes a vector of independent Gaussian distributed errors with zero mean and a diagonal covariance matrix with elements σ^2. We assume for now that this quantity is known, although the most common situation is that it has to be estimated. Since the DWT is an orthonormal transform, the additive noise component being transformed has the same statistical properties as the untransformed noise.

In what follows below, we rest upon the property of the DWT to decorrelate efficiently the time-series even provided that the analyzed series is generated by a long-memory process (see Percival and Walden, 2000, for details). Under the reasonable approximation that W_{jt} are random samples from the Gaussian distribution, one can apply Bayes' rule for modeling these coefficients by their posterior distribution. For such an approach we adopt the ideas from the recently developed posterior smoothing technique (PS) in the scale-space framework of data representation (Godtliebsen and Øigård, 2005).

The realistic model for W_{jt} at each decomposition level j can be presented as

$$\boldsymbol{W} = \hat{\boldsymbol{W}} + \boldsymbol{\eta} \quad (4)$$

Here $\boldsymbol{W} = [W_1, \ldots, W_M]^T$ and $\hat{\boldsymbol{W}} = [\hat{W}_1, \ldots, \hat{W}_M]^T$ denote the observed and true wavelet coefficients respectively, and

$M \equiv N_j$ is the number of wavelet coefficients at the reference level of the DWT. Note that we omit further in this subsection the subscript j denoting the level of the wavelet decomposition under consideration. Also here and elsewhere the term "observed", when applying to the wavelet coefficients, highlights the fact that they are derived from observations, rather than implying that they have really been observed.

We assume that the true wavelet coefficients $\hat{\boldsymbol{W}}$ can be modeled by a Gaussian Markov Random Field, see Rue (2001), which is specified through the local characteristics

$$\mathrm{E}\left(\hat{W}_t|\hat{\boldsymbol{W}}_{-t}\right) = -\sum_{k\in\partial_t}\frac{Q_{tk}}{Q_{tt}}\hat{W}_k \quad \text{and}$$
$$\mathrm{Var}\left(\hat{W}_t|\hat{\boldsymbol{W}}_{-t}\right) = Q_{tt}^{-1}, \tag{5}$$

where $\mathrm{E}(a|b)$ and $\mathrm{Var}(a|b)$ denote conditional expectation and variance for **a** given **b**, respectively, **Q** is the inverse covariance matrix, or often referred to as the precision matrix. The **Q** matrix is nonzero if and only if $k \in \{\partial_t \cup t\}$. Here ∂_t denotes the neighbors to data point t, and $\hat{\boldsymbol{W}}_{-t}$ denotes all elements of $\hat{\boldsymbol{W}}$ apart from $\hat{W}_t$. This illustrates the Markov property, i.e.

$$p(\hat{W}_t|\hat{\boldsymbol{W}}_{-t}) = p(\hat{W}_t|\hat{\boldsymbol{W}}_{\partial_t}). \tag{6}$$

Based on these assumptions, the prior model for $\hat{\boldsymbol{W}}$ is given by

$$p(\hat{\boldsymbol{W}}) \propto \exp\left[-\beta\sum_{t\sim k}\left(\hat{W}_t - \hat{W}_k\right)^2\right], \tag{7}$$

where $t\sim k$ means that the points indexed by t and k are neighbors. In our default implementation, $\partial_t = \{t-1, t+1\}$ is used (with obvious modifications at the borders). The parameter β in Eq. (7), controls the degree of smoothness in the realizations of $\hat{\boldsymbol{W}}$ obtained from $p(\hat{\boldsymbol{W}})$. If samples are drawn from Eq. (7), large values of β will give smooth realizations of $\hat{\boldsymbol{W}}$ while small values of β will give rougher realizations.

The observed wavelet coefficients W_t now follow a Gaussian distribution with mean $\hat{W}_t$ and standard deviation σ, i.e. $W_t \sim \mathcal{N}[\hat{W}_t, \sigma^2]$. Hence, the likelihood of $\boldsymbol{W}$ given $\hat{\boldsymbol{W}}$ is

$$p(\boldsymbol{W}|\hat{\boldsymbol{W}}) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^M \exp\left[-\frac{1}{2\sigma^2}\sum_{t=1}^{M}(W_t - \hat{W}_t)^2\right]. \tag{8}$$

Using Bayes theorem (Berger, 1985), the posterior distribution of $\hat{\boldsymbol{W}}$ given $\boldsymbol{W}$ can be found from

$$\begin{aligned} p(\hat{\boldsymbol{W}}|\boldsymbol{W}) &\propto p(\boldsymbol{W}|\hat{\boldsymbol{W}})p(\hat{\boldsymbol{W}}) \\ &\propto \exp\left[-\frac{1}{2\sigma^2}\sum_{t=1}^{M}(\hat{W}_t - W_t)^2 - \beta\sum_{t\sim k}\left(\hat{W}_t - \hat{W}_k\right)^2\right]. \end{aligned} \tag{9}$$

Samples can now be drawn from the posterior distribution. An efficient exact sampling algorithm for this situation is described by Øigård (2004). The degree of smoothness in the obtained realizations for $\hat{\boldsymbol{W}}$ depends heavily on the choice of β in the same way as the degree of smoothness in the local linear kernel estimator is controlled through the bandwidth h, see Chaudhuri and Marron (1999). The choice of appropriate β can be organized in a data-driven way and is discussed further in Sect. 2.4.

2.3 Estimating the noise variance

Assessing the noise characteristics in a number of situations is not a trivial task and its detailed consideration lies beyond the scope of the present paper. No universal recipe can be proposed and each case should generally be considered individually. Besides it is yet to be decided what will be regarded as noise in the course of the analysis. In the typical climate proxy record, for example, the noise constituent is a mixture of an instrumental noise (measurement and dating errors), climatic noise, which inheres in the background process itself and some extra variability due to the postdepositional alterations of the initial profile (Fisher et al., 1985). Their separation may not be possible at all, so the question will be what part of the variability can be attributed to one or another component and subsequently filtered out.

When analyzing climatic series the problem is also often complicated by the presence of a serial correlation. If neglected, the resulting σ^2 may be substantially underestimated. We therefore propose a procedure that may be suitable when one deals with a time-series having pronounced auto-regressive characteristics. AR(1), the simplest model, is the one most commonly used. If one assumes that the analyzed time-series is generated by an AR(1) process, the noise term can be associated with residuals of the time-series and fitted AR(1) model. This readily yields the estimate of the noise variance as:

$$\sigma^2 = \mathrm{var}\left(\boldsymbol{u} - \boldsymbol{u}_{AR(1)}\right).$$

Using this approach will likely put too much conservatism in the procedure of feature detection. This, on the other hand, brings more confidence to conclusions drawn from the analysis.

2.4 Choice of β

Modeling and analysis over a broad range of the smoothing parameter simultaneously is a typical approach in the scale-space methods of data exploration. In our case, when the wavelet decomposition itself already gives the time-scale representation of a time-series, this will produce a redundant output and exert a substantial additional computational burden. An apparent way of solving this problem lies in modeling the wavelet coefficients at a single value of β rather than the range.

Given the model (Eq. 4) for the wavelet coefficients, the respective variance at each decomposition level j is the sum of the true (smoothed) coefficients variance and the noise:

$$\sigma_j^2 = \hat{\sigma_j}^2 + \sigma^2$$

where $\hat{\sigma_j}^2$ is a function of the smoothing parameter β. Having approximately Gaussian W_{jt} and $\hat{W}_{jt}$, a simple sample variance can be used as a reasonable estimator of ${\sigma_j}^2$ and $\hat{\sigma_j}^2$. The problem of finding an optimal amount of smoothing applied to the wavelet coefficients is now the problem of minimizing the relationship

$$\left|(\sigma_j^2 - \sigma^2) - \gamma\hat{\sigma_j}(\beta)^2\right| \quad (10)$$

with regard to β at each level of the wavelet decomposition. The parameter $\gamma > 0$, set to unit by default, can be used to adjust the amount of smoothing to a desired value. One should mention, however, that in some occasions the estimated noise variance may exceed the particular wavelet scale variance simply by chance, making the solution of Eq. (10) impossible. In such situations, the amount of smoothing applied to the wavelet scale is determined by a signal/noise ratio for the whole signal, namely the value of σ^2 in Eq. (10) is substituted by $\sigma_j^2 * \sigma^2/{\sigma_s}^2$, where ${\sigma_s}^2$ denotes a standard estimator of the time-series variance.

2.5 Choice of wavelet function

Since the proposed method is based on the DWT transform, a choice of a wavelet function becomes crucial. Our choice was a least asymmetric wavelet function of the width 8 (LA(8) or Sym4 in different notations). Percival and Walden (2000) argues that LA(8) often provides a good trade-off between the width of the wavelet function and its smoothness. Being relatively short, and therefore providing a narrower cone of influence in the wavelet decomposition, its shape is still a good match to the characteristic features for most of the time-series. The wavelet center frequency, 0.71, is slightly lower the optimal value of 1, suggesting its better localization in the time domain. "Least asymmetric" means that the associated wavelet filter has nearly zero phase property, i.e. the resulting features in the wavelet decomposition will be aligned in time with the features in the time-series being analyzed.

2.6 Wavelet power spectrum and significance testing

After the sampling procedure is performed, we are left with some K realizations (samples) of the true $\hat{\boldsymbol{W}}_j$ for each decomposition level j. The modeled wavelet coefficients can now be utilized for calculating the smoothed observed wavelet power spectrum (WPS), which is an estimator for the "true" WPS of the underlying process. This is defined, by analogy with Fourier analysis, as the wavelet transformation of the autocorrelation function:

$$P_{jt} = \mathrm{E}(W_{jt} W_{jt}^*), \quad (11)$$

with "*", denoting the complex conjugate, being relevant only if the complex mother wavelet is used. Commonly, when only one realization of the wavelet decomposition is given, the squared absolute values of the wavelet coefficients are used as an estimator for the true WPS. This measure is called "wavelet periodogram" and has properties similar to its counterpart in Fourier analysis (Nason et al., 2000; Maraun and Kurths, 2004). Bayesian modeling of the wavelet coefficients, in its turn, provides us with theoretically unrestricted number of independent realizations. This allows calculating the expectation value of the periodogram immediately using Eq. (11).

The wavelet power spectrum (also called the wavelet variance) decomposes the time-dependent variance of a time-series u_t on a scale-by-scale basis. Percival and Walden (2000) show that the WPS is well defined for both second order stationary time-series and non-stationary time-series with stationary backward differences as long as the mother wavelet function has the backward difference scheme embedded and its width L is large enough. Given the non-stationary time-series whose backward difference of order d is stationary, a condition $L > 2d$ is to be satisfied in order to ensure that the wavelet variance is a good approximation of the time-series variance.

Our first approach to assessing the significance of peaks in the modeled wavelet spectrum is based on testing the null-hypothesis that the analyzed signal represents samples drawn from a stationary process with a given background power spectrum $S(f)$. If a peak in the WPS is significantly above this background spectrum, then it can be claimed to be a "real" feature with a certain percent confidence. Many real time-series, in particular in geophysical studies, can be modeled using a stochastic autoregressive process of the first order, or AR(1), with a positive lag-1 autocorrelation coefficient. This model is used as default in some wavelet applications (see for example Torrence and Compo, 1998; Grinsted et al., 2004). Recall now that the wavelet coefficients at level j are nominally associated with frequencies in the interval $[f_l, f_h] = [1/2^{j+1}, 1/2^j]$ (Percival and Walden, 2000). Using the results of Torrence and Compo (1998), an α-quantile for the distribution of $\hat{W}_j^2/{\sigma_s}^2$ at the j-th level of DWT is defined as

$$q_{j,p}^{AR(1)} = \frac{Q_1(\alpha)}{\delta f} \int_{f_l}^{f_h} \frac{1-\phi^2}{1+\phi^2-2\phi\cos(2\pi f)} df, \quad (12)$$

where ${\sigma_s}^2$ is a standard estimator of the time-series variance, $\mathbf{P}[Q > Q_1(\alpha)] = \alpha$ and Q_1 is χ_1^2 distributed, $f = 0, ..., 0.5$ is the frequency and $\delta f = f_h - f_l$. We now can consider a feature to be significant if $\hat{W}_{jt}^2 > {\sigma_s}^2 q_{j,\alpha}^{AR(1)}$ with α equal to, say, 0.05.

The second approach utilizes all available realizations of the wavelet power at each particular scale of the DWT. From these realizations and a decision rule it is decided, at each (j, t) location, whether the derivative of the wavelet spectrum is significantly different from zero. We interpret the

procedure as testing for violation in homoscedasticity at the particular level of the DWT. The magnitude of the derivative is estimated from the K samples by:

$$dP_{jt,k} = \hat{W}^2_{j(t+1),k} - \hat{W}^2_{jt,k}.$$

For the chosen statistical model dP_{jt} at point (j,t) has a symmetric distribution centered at zero when the derivative at this point is zero. At each point (j,t) we therefore claim that dP_{jt} is "significantly different" from zero if the absolute value of the mean $E(dP_{jt})$, is large compared to its standard deviation $SD(dP_{jt})$:

$$\left| \frac{E(dP_{jt})}{SD(dP_{jt})} \right| > q^E_{j,\alpha}.$$

Here $q^E_{j,\alpha}$ denotes an appropriate quantile depending on the DWT scale number j, and the level of the test α. For estimating $E(dP_{jt})$ and $SD(dP_{jt})$ we use ordinary empirical estimators for the mean and standard deviation.

The quantile $q^E_{j,\alpha}$ is estimated directly from the data following the procedure proposed in Øigård (2004). For each scale number j, it is found by using the empirical distribution, obtained from the large amount of available simulated samples. We define the standardized estimates of the derivative $dP^S_{jt,k}$ by

$$dP^S_{jt,k} = \frac{dP_{jt,k} - E(dP_{jt})}{SD(dP_{jt})}.$$

Then, for each decomposition level j=1, ..., J and the estimated wavelet power in the points t=1, ..., N_j, the quantile $q_{jt,\alpha}$ is chosen to be the largest value such that 100α% of the modeled standardized realizations $\left|dP^S_{jt,k}\right|$, k=1, ..., K are greater than $q_{jt,\alpha}$. For large values of K, we have that $P\left\{\left|dP^S_{jt,k}\right| > q_{jt,\alpha}\right\} \approx \alpha$. We then proceed conservatively, and choose $q^E_{j,\alpha}$ to be the maximum of all the N_j quantiles $q_{jt,\alpha}$ for each decomposition level j, i.e. $q^E_{j,\alpha} = \max_t \left\{q_{jt,\alpha}\right\}$. This can be basically thought of as a procedure correcting for multiple testing.

3 Numerical implementation

The computational steps to analyze the signal using the proposed method are as follows:

1. Set input arguments: Specify the noise variance term in the signal model, as defined in Eq. (3), and choose a desired type of the quantile for significance testing procedure. Note that when choosing $q^{AR(1)}$ quantile, a relevant estimate of the autocorrelation parameter is to be additionally provided. By default this parameter is generated through the embedded function. If the value of the noise variance is not available, it can be estimated using the default procedure proposed in Sect. 2.3. Change γ set by default to 1 to $0<\gamma<1$ or $\gamma>1$ if the modeled spectrum needs to be under- or oversmoothed. Choose between solving the minimization problem for β (default) and specifying the value of the smoothing parameter manually.
2. Find the DWT of the original data sequence for the prescribed range of scales j. The DWT algorithm is based on the routine *wavedec* from the Wavelet toolbox for Matlab. The method implementation fitted to the use of LA(8) basis function can, in principle, be adapted to other wavelets. By default "padding" with zeroes is used to extend the analyzed time series up to a nearest integer multiple of 2^J, with subsequent elimination of the biased wavelet coefficients.
3. Solve the minimization problem (Eq. 10) with respect to β for each wavelet decomposition scale.
4. At each point (j,t) of the observed DWT draw a necessary number of realizations of the modeled wavelet coefficients from the constructed posterior distribution (Eq. 9). Our default choice is K=200. As an extra option, inverse DWT (implemented using *waverec* function) uses realizations averaged over K for reconstructing the smoothed signal from the modeled (smoothed) wavelet coefficients.
5. Calculate the smoothed wavelet periodogram P_{jt} (Eq. 11) .
6. Calculate quantiles $q^E_{j,\alpha}$ from the modeled realizations of the wavelet periodogram $P_{jt,k}$ or $q^{AR(1)}$ using the specified value of ϕ.
7. For each scale j apply the significance testing procedure to the modeled wavelet power spectrum.

A computer program that performs the above steps is available from the authors. The zip-archive includes MATLAB codes, program documentation and examples files. The program requires *Wavelet Toolbox* extension package for Matlab to be installed.

4 Results

In this section we investigate the robustness of the proposed technique. Case studies with synthetic data and two real climatic time-series demonstrate the overall performance of the method and its potential as a useful tool for data analysis. Comparison of results with outputs from other methods proves the adequacy of the proposed approach to time-series exploration.

The sample climatic time-series were preliminarily detrended using a linear fit and subsequently tested for stationarity of the expected mean. For the latter procedure we

analyzed the time-integrated wavelet spectrum following the technique proposed in Percival and Walden (2000). In order to check appropriateness of the AR(1) model, which is often used by default in geophysical applications, we fitted this model to the sample time-series. The cumulative periodogram test for randomness of the residuals (Box et al., 1994) has proved the adequacy of the proposed model for describing the time-series used in both examples.

A three-component visual display device is used for showing the results of the analysis. It comprises the raw and the smoothed time-series (panel A), estimated true power spectrum of a signal (panel B) and a feature map showing the output of the significance testing procedure (panel C). The last panel in programming implementation of the method may have two different visual representations depending on what testing procedure has been chosen. For demonstrational purposes, however, the examples shown below display the output from both of the available testing methods.

Black areas in panel Ca (testing against red noise background) highlight those parts of the power spectrum shown in panel B that are statistically significant at the prescribed significance level, according to the criterion stated above. The color map used in panel Cb is similar to that one originally introduced in Chaudhuri and Marron (1999) with areas where the wavelet power exhibits statistically significant increase and decrease flagged as red and blue, respectively. The vertical axis in panel B shows the decomposition scale number. In panel C, for convenience, these are substituted by inverse of the wavelet pseudo-frequency (i.e. pseudo-period) corresponding to the decomposition scale. This is defined as the frequency maximizing the Fourier transform of the wavelet function. This may provide a hint about the real time scale being analyzed. Semitransparent fringes of the panels outline the areas affected by the edge effects. Note that the results obtained for these parts of the decomposition and significance testing should be interpreted with caution.

4.1 Testing the method

For testing the proposed technique we ran a series of numerical simulations. As test data we used generated time-series with well-known spectral characteristics. These are purely random process, stationary autoregressive process of the first order with positive autocorrelation and a random walk process, which is a non-stationary $1/f$-Type process with stationary backward difference of the first order. We generated 500 time-series of the length 1024 for each of the categories. The assigned value of σ was equal to one in all three cases and the autocorrelation coefficient ϕ in the AR(1) process was set to 0.7. Since we initially knew all the parameters of the analyzed signals, we used them when running the program. The following testing procedures have been implemented:

(a) testing for consistency of the estimated (smoothed) spectrum with the prescribed (true) spectrum of the process

(b) testing for detection of spurious significant features in these purely random data samples.

In the first experiment we examined whether, on average, the true spectrum and its estimate are consistent. If the true spectrum at a point falls outside of the confidence interval of the modeled spectrum, then this point considered a "miss". The confidence intervals were constructed based on our prior knowledge of the process type. Using the results and notation introduced in Sect. 2.6 readily gives an approximate $100(1-\alpha)\%$ confidence interval for $\hat{W}_{jt}^2/{\sigma_s}^2$ in the form

$$\left[S_j Q_1(1-\frac{\alpha}{2}),\, S_j Q_1(\frac{\alpha}{2})\right]. \tag{13}$$

S_j here denotes the theoretical normalized discrete Fourier power spectrum of the analyzed process averaged over the proper range of frequencies and defined as

$$S_j = \begin{cases} 1, & \text{N(0,1)} \\ \dfrac{1}{\delta f}\displaystyle\int_{f_l}^{f_h} \dfrac{1-\phi^2}{1+\phi^2-2\phi\cos(2\pi f)} df, & \text{AR(1) process} \\ \dfrac{1}{\delta f}\displaystyle\int_{f_l}^{f_h} \dfrac{1}{4\sin^2(\pi f)} df, & \text{random walk} \end{cases}$$

with f and δf being the same as defined earlier in Sect. 2.6.

Based on 500 available realizations of the same process the relative number of misses were estimated for each point of the wavelet spectrum. The analysis have demonstrated that it usually does not exceed a prescribed value of α for any of the three types of processes (AR(1) with ϕ=0 and ϕ=0.7, and a random walk) considered. This indicates that the modeled wavelet spectrum of a time-series is a reasonable estimate of the theoretical one.

Both significance testing procedures were subject to verification for spurious detection of significant features in purely random data samples. Running the method in such cases should, ideally, give no significant features. In practice, the number of false identifications will depend on the level of the test, which was set to α=0.05. We used the same design as in the first series of numerical experiments. The results of the analysis were obtained in the form of the relative number of features spuriously found to be significant at each point of the wavelet spectrum. These are not shown here in order to reduce the size of the current presentation. We found that for all three types of the processes considered the average number of false identifications for testing using the empirical quantile was much below the level of the test. Such a low number of false identifications is certainly due to virtual correction for multiple testing embedded in the procedure of estimating the quantile. This approach appears thereby to be

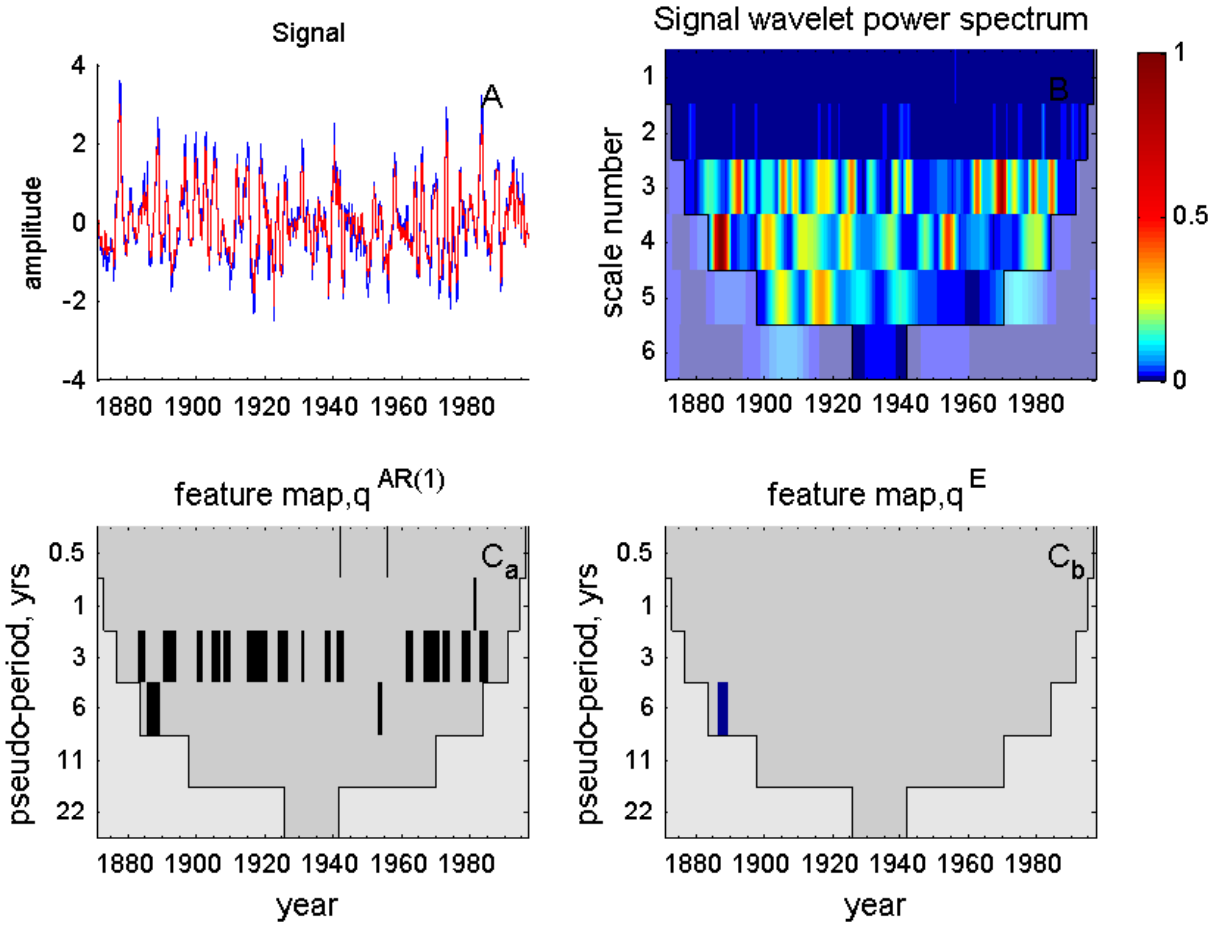

Fig. 1. (A) Raw (blue) and smoothed (red) Niño3 seasonal SST index. **(B)** normalized smoothed wavelet periodogram (estimate of the true wavelet power spectrum) of the time-series. See color bar from the right of panel (B) for spectral power gradation. **(C)** Feature maps for the power spectrum shown in B calculated using $q^{AR(1)}$ (C_a) and q^E quantiles (C_b). Grey areas in C_a highlight the features in the wavelet power spectrum of greater than 95% confidence for a red-noise process. Red and blue in panel C_b designate the areas where the wavelet power exhibits, respectively, statistically significant increase and decrease. Semitransparent fringes of panels B, C_a and C_b enclose the areas affected by the edge effects.

conservative enough to be recommended for use in situations where a suitable model for the analyzed time-series is uncertain.

The procedure of testing against the AR(1) background showed similar results (no features detected) only for purely random and AR(1) time-series, i.e. when the testing hypothesis was trivially true. Testing the random walk series treated as being AR(1), in turn, yields a persistently higher number of features marked as statistically significant (up to 20%, depending on magnitude of added random noise). The result is not unexpected keeping in mind that the theoretical spectral power of the normalized random walk time-series is generally higher than the one for the stationary AR(1) process, whatever the autocorrelation coefficient is.

4.2 Example 1: Niño3 SST index

Figure 1 shows an application of the proposed technique to the Niño3 sea surface temperature, (SST) used as a measure of the amplitude of the El Niño-Southern Oscillation (ENSO). The Niño3 SST index (panel A) is defined as the seasonal SST averaged over the central Pacific 5° S–5° N 90° W–150° W. The data for 1871–1997 is presented in the form of seasonal anomalies. A detailed analysis of this time-series using the wavelet decomposition technique is found in Torrence and Compo (1998).

The feature map calculated using $q^{AR(1)}$ quantile (ϕ=0.71) shows increased significant variability inconsistent with an AR(1) model on the time-scale of approximately 3 years before 1940 and after 1960, with somewhat fewer peaks marked as significant in between. This is in line with

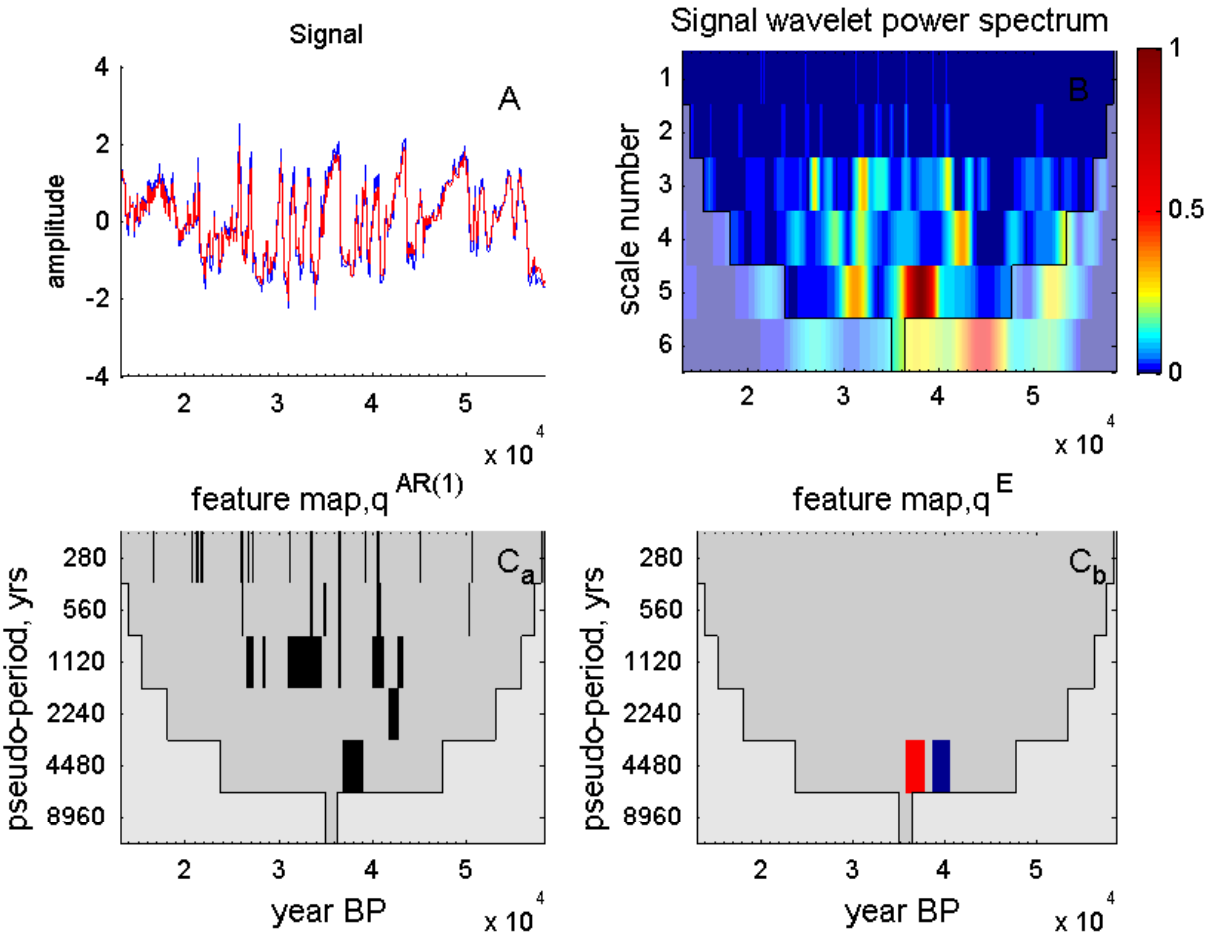

Fig. 2. Same as for Fig. 1, but for GISP2 $\delta^{18}O$ oxygen isotope record.

the conception of weaker ENSO variability during this period (Dong et al., 2006). The feature map also generally reproduces the results presented in Torrence and Compo (1998) where the real and complex CWT (Figs. 1b and c there, respectively) with testing against the red noise was used. At the same time more conservative testing using the empirical quantile with noise variance estimate of $(0.47)^2$ yielded an anticipated result, with only one peak flagged as significant.

4.3 Example 2: GISP2 δ^{18}O oxygen isotope record

As a second example, we consider the glacial part (13–59 ky BP) of the oxygen-isotope record, measured as O^{18}/O^{16} ratio, from the GISP2 ice core from Greenland (Grootes and Stuiver, 1997). This time-series reflects, to a large extent, air temperature fluctuations above Greenland during this period. Prior to applying the wavelet transform we binned initially unevenly sampled record at century resolution. Assuming the AR(1) model for the analyzed signal is generally true, the noise variance of $(0.68)^2$ is estimated from the residuals following the procedure proposed in Sect. 2.3.

Figure 2 shows the results of the analysis. Testing the null-hypothesis that the background process is AR(1) (ϕ=0.85) reveals the variability at scale 3 (1–2 kyears) inconsistent with the proposed model. Some more features appear as significant at the first two scales too. They can largely be interpreted as an extension of sharp major peaks at 1.5 ky scale (so-called Dansgaard-Oeschger oscillations) to finer scales. Two peaks are also detected as significant on the longer scales. These results are in a good agreement with spectral analyses presented in Grootes and Stuiver (1997) and Schulz and Mudelsee (2002). Testing using the more conservative empirical quantile marks as "real" only the peak close to 40 000 BP, identified as interstadial 8, according to the classification proposed in Dansgaard et al. (1993).

One needs to mention nevertheless that this inference may appear to be too conservative due to application of a simplified model for estimating σ^2. When all types of noise except the instrumental error are ruled out, its value is much reduced to a common estimate of $(0.1)^2$, weakening accordingly the conservatism of the test. The number of features detected as

significant in this case is essentially higher (not shown here); among them are Dansgaard-Oeschger oscillations at millennial scale. This example underscores the crucial role of a proper choice of σ^2 for making further inference whether the features seen in the analyzed time-series can be regarded as significant, with respect to a statistical hypothesis applied.

5 Conclusions

The motivation of this study has been to develop a wavelet-based tool for exploring structures in data sequences at different scales of resolution. We have demonstrated how the ideas elaborated in the scale-space techniques can successfully be employed in wavelet analysis. Recalling that the raw wavelet spectrum is not a consistent estimator of the local wavelet power, we put forward an idea of using the decorrelating properties of the discrete wavelet transform for Bayesian modeling of the wavelet coefficients. The smoothing to the spectrum is introduced in a natural way via a smoothing prior with parameters estimated from the data. However, as we show by an example, both the estimate of the true wavelet spectrum and analysis for significance can be quite sensitive to the estimate of the noise variance. One should therefore always consider carefully the possible impact of this choice on the final inference. It is worth mentioning that like any tool for the time-series analysis the program should not be used as a black-box without checking the properties of the data sequence prior to its analysis.

The general idea of the method – the modeling – can potentially be transferred to more conventional types of wavelet transforms, MODWT or CWT. They have a substantial advantage over the DWT, namely the results are not so sensitive to the choice of wavelet function. As shown by Percival and Walden (2000), the MODWT variance estimator is statistically more efficient and, it has much better temporal resolution and visual representation which simplifies the interpretation of derived spectra. These transforms do, however, use the non-orthogonal basis and produce a redundant output. The direct application of the method, therefore, is not possible since the basic assumption of independent errors is not satisfied any more. Modeling in this case will require, for each decomposition scale, a detailed assessment of the error covariance matrix based on the reproducing kernel of the wavelet transform (Maraun et al., 2007). The overall formalism of the method, namely Eqs. (5–9), will have to be revised too. We leave this problem for future research.

One should also notice that the algorithm involves two computer intensive procedures, namely inverting the covariance matrix and drawing samples from the posterior. The running time therefore may become relatively long when the length of the time-series exceeds some 1000 points (depending on the PC). This, together with a need to have an extra commercial software installed (wavelet package for Matlab) may potentially restrict the application of the proposed method. These problems are nevertheless planned to be solved in the next version of the program through substituting the inversion procedure in Eq. (5) by the exact solution and using the open source functions for the wavelet transform.

Acknowledgements. This study is supported by the Norwegian Research Council, project 160008/v30. The authors would like to thank two anonymous reviewers for their constructive critics given to the first version of the manuscript.

Edited by: H. A. Dijkstra
Reviewed by: two referees

References

Berger, J. (Ed.): Statistical decision theory and Bayesian analysis, Springer Verlag, 1985.

Box, E., Jenkins, G., and Reinsel, G.: Time series analysis: Forecasting and control, Prentice-Hall, Englewood Cliffs, 1994.

Chaudhuri, P. and Marron, J.: SiZer for exploration of structures in curves, J. Am. Statist. Assoc., 94, 807–823, 1999.

Dansgaard, W., Johnsen, S., Clausen, H., Dahl-Jensen, D., Gundestrup, N., Hammer, C., Hvidberg, C., Steffensen, J., Sveinbjornsdottir, A., Jouzel, J., and Bond, G.: Evidence for general instability of past climate from a 250-kyr ice-core record, Nature, 364, 218–220, doi:10.1038/364218a0, 1993.

Dong, B., Sutton, R., and Scaife, A.: Multidecadal modulation of El NiñoSouthern Oscillation (ENSO) variance by Atlantic Ocean sea surface temperatures, Geophys. Res. Lett., 33, L08 705, doi:10.1029/2006GL025766, 2006.

Fisher, D. A., Reeh, N., and Clausen, H. B.: Stratigraphic noise in the time series derived from ice cores, Ann. Glaciol., 7, 76–83, 1985.

Godtliebsen, F. and Øigård, T.: A Visual Display Device for Significant Features in Complicated Signals, Comput. Statist. Data Analysis, 48, 317–343, 2005.

Grinsted, A., Moore, J., and Jevrejeva, S.: Application of the cross wavelet transform and wavelet coherence to geophysical time series, Nonlin. Processes Geophys., 11, 561–566, 2004, http://www.nonlin-processes-geophys.net/11/561/2004/.

Grootes, P. and Stuiver, M.: Oxygen 18/16 variability in Greenland snow and ice with $10^{-3}-10^{-5}$-year time resolution, J. Geophys. Res., 102(C12), 26 455–26 470, 1997.

Kaiser, G.: A friendly guide to wavelets, Birkhauser Boston Inc., Cambridge, MA, USA, 1994.

Maraun, D. and Kurths, J.: Cross wavelet analysis: significance testing and pitfalls, Nonlin. Processes Geophys., 11, 505–514, 2004.

Maraun, D., Kurths, J., and Holschneider, M.: Nonstationary Gaussian Processes in Wavelet Domain: Synthesis, Estimation and Significance Testing, Phys. Rev. E, 75(1), 016707, 2007.

Nason, G. P., von Sachs, R., and Kroisandt, G.: Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum, J. Roy. Statist. Soc., 62(B), 271–300, 2000.

Øigård, T.: Statistical analysis and representation of non-trivial signals and random processes, Ph.D. thesis, University of Tromsø, Tromsø, Norway, 2004.

Park, C., Marron, J., and Rondonotti, V.: Dependent SiZer: Goodness-of-fit tests for tune series models, J. Appl. Statist., 8, 999–1017, 2004.

Percival, D. and Walden, A.: Wavelet methods for time-series analysis, Cambridge University Press, Cambridge, 2000.

Rue, H.: Fast Sampling of Gaussian Markov Random Fields, J. Roy. Statist. Soc., 63, 325–338, 2001.

Schulz, M. and Mudelsee, M.: REDFIT: estimating red-noise spectra directly from unevenly spaced paleoclimatic time series, Computers & Geosciences, 28, 421–426, 2002.

Torrence, C. and Compo, G.: A practical guide to wavelet analysis, Bull. Am. Meteorol. Soc., 79(1), 61–78, 1998.