Nonstationary Gaussian processes in wavelet domain: Synthesis, estimation, and significance testing

D. Maraun* and J. Kurths
Nonlinear Dynamics Group, Institute of Physics, University of Potsdam, 14415 Potsdam, Germany

M. Holschneider
Institute of Mathematics, University of Potsdam, 14415 Potsdam, Germany

(Received 13 July 2006; published 22 January 2007)

We propose an equivalence class of nonstationary Gaussian stochastic processes defined in the wavelet domain. These processes are characterized by means of wavelet multipliers and exhibit well-defined time-dependent spectral properties. They allow one to generate realizations of any wavelet spectrum. Based on this framework, we study the estimation of continuous wavelet spectra, i.e., we calculate variance and bias of arbitrary estimated continuous wavelet spectra. Finally, we develop an areawise significance test for continuous wavelet spectra to overcome the difficulties of multiple testing; it uses basic properties of continuous wavelet transform to decide whether a pointwise significant result is a real feature of the process or indistinguishable from typical stochastic fluctuations. This test is compared to the conventional one in terms of sensitivity and specificity. A software package for continuous wavelet spectral analysis and synthesis is presented.

DOI: 10.1103/PhysRevE.75.016707 PACS number(s): 02.70.Hm, 02.50.−r, 05.40.−a, 05.45.Tp

I. INTRODUCTION

Continuous wavelet transformation (CWT) is a powerful mathematical instrument that transforms a time series to the time-scale domain. Rioul and Flandrin [1] defined the wavelet scalogram to estimate the nonstationary wavelet spectrum of an underlying process. Donoho [2] used wavelet techniques for the reconstruction of unknown functions from noisy data.

The continuous wavelet spectra of paradigmatic processes as Gaussian white noise [3] or fractional Gaussian noise [4] have been calculated analytically. Continuous wavelet spectral analysis has been applied to real-world problems in physics, climatology [5], life sciences [6], and other fields of research. Hudgins *et al.* [7] defined the wavelet cross spectrum to investigate scale- and time-dependent linear relations between different processes. This measure has been applied, e.g., in the analysis of atmospheric turbulence [7] and time-varying relations between El Niño/Southern Oscillation and the Indian monsoon [8].

Wavelet spectral analysis is an inverse problem: One aims to estimate the wavelet spectrum of an unknown underlying process. However, to characterize the quality of the estimator in terms of variance and bias, a theory for the direct problem is required: One has to formulate a framework to synthesize realizations of a known wavelet spectrum to derive to which grade the estimated wavelet spectrum reconstructs this underlying spectrum. Hitherto, such a formulation of the direct problem does not exist for continuous wavelet spectral analysis. Thus, the following questions are still unresolved: How can realizations of a specific wavelet spectrum be synthesized? How do these realizations depend on the wavelet chosen for the synthesis? What is the relation between an arbitrary stationary wavelet spectrum and the corresponding Fourier spectrum? What are the variance and the bias of an arbitrary wavelet sample spectrum? How sensitive is a significance test for the wavelet spectrum?

In their seminal paper, Torrence and Compo [9] placed wavelet spectral analysis into the framework of statistical data analysis by formulating pointwise significance tests against reasonable background spectra. However, Maraun and Kurths [10] highlighted a serious deficiency of pointwise significance testing: Given a realization of white noise, large patches of spurious significance are detected, making it—without further insight—impossible to judge which features of an estimated wavelet spectrum differ from background noise and which are just artifacts of multiple testing. This demonstrates the necessity to advance the significance testing of continuous wavelet spectra and to evaluate it in terms of sensitivity and specificity.

In this study, we suggest and elaborate a framework of nonstationary Gaussian processes defined in the wavelet domain; these processes are characterized by their time-dependent spectral properties. Based on these processes, we present the following results: First, we formulate the direct problem of continuous wavelet *synthesis*. This means that we present a concept to generate realizations of an arbitrary nonstationary wavelet spectrum and study the dependency of the realizations on the wavelets used for the synthesis. We derive the relation of an arbitrary stationary wavelet spectrum to the corresponding Fourier spectrum. An asymptotic theory for small scales is presented. Second, we study the inverse problem of continuous wavelet spectral *analysis*, i.e., estimating the wavelet spectrum and significance testing it against a background spectrum. This means that we study bias and variance of an arbitrary estimated wavelet spectrum. To overcome the problems arising from pointwise significance testing, we develop an areawise significance test, taking advantage of basic properties of the CWT. We evaluate this test in terms of sensitivity and specificity within the framework suggested.

*Electronic address: maraun@agnld.uni-potsdam.de

 ©2007 The American Physical Society

The paper is divided into three main parts. Preceded by a short review of CWT in Sec. II, we develop the new framework of nonstationary Gaussian processes in the wavelet domain in Sec. III (the direct problem). In Secs. IV and V, we study the inverse problem. Section IV deals with the estimation of wavelet spectra and presents the variance and bias of an arbitrary estimated wavelet spectrum. The results for the new areawise significance test are given separately in Sec. V.

II. CONTINUOUS WAVELET TRANSFORMATION

The CWT of a time series $s(t)$, $W_g s(t)[b,a]$, at time b and scale a (scale refers to 1/frequency) with respect to the chosen wavelet $g(t)$ is given as

$$W_g s(t)[b,a] = \int dt \frac{1}{\sqrt{a}} \bar{g}\left(\frac{t-b}{a}\right) s(t), \tag{1}$$

where the overbar denotes complex conjugate. The brackets $[\ldots]$ denote dependencies of a variable, whereas $(\ldots)$ denote dependencies of the resulting transformation. Here, we choose the L^2-normalization $1/a^{1/2}$.

For every wavelet in a strict sense $g(t)$, a reconstruction wavelet $h(t)$ can be found [3]. Using this, one can define an inverse transformation of a function $r(b,a)$ from the positive half plane $\mathbb{H}$ to the time domain,

$$M_h r(b,a)[t] = \int_{\mathbb{H}} \frac{dbda}{a^2} r(b,a) \frac{1}{\sqrt{a}} h\left(\frac{t-b}{a}\right). \tag{2}$$

The CWT from one dimension to two dimensions does not produce any new information, i.e., a continuous wavelet transform is not uncorrelated. For the wavelet transformation of Gaussian white noise $\eta(t)$, the intrinsic correlations between the wavelet coefficients at (b,a) and (b',a') are given by the reproducing kernel $K_{g,h}((b-b')/a', a/a')$ (for details and an example, see Appendix A 2) moved to the time b' and stretched to the scale a' [3,12],

$$C(b,a;b',a') \sim K_{g,h}\left(\frac{b-b'}{a'}, \frac{a}{a'}\right). \tag{3}$$

The reproducing kernel represents a time-scale uncertainty.

For a detailed discussion of CWT basics, please refer to the comprehensive literature [3,11,12]. Percival and Walden [13] give a good overview of discrete wavelet transformation (DWT) and maximum overlap discrete wavelet transformation (MODWT).

III. GAUSSIAN PROCESSES IN THE WAVELET DOMAIN

The direct problem of wavelet synthesis corresponds to a framework to generate realizations of an arbitrary nonstationary wavelet spectrum. In this section, we develop this framework and present *a priori* spectral measures.

Stationary Gaussian processes are completely defined by their Fourier spectrum $S(\omega)$. A realization of any such process can be simulated by transforming Gaussian white noise to the Fourier domain, multiplying it with a function $f(\omega)$, and transforming it back to the time domain (see, e.g., [14]); the spectrum of this process is then given by $|f(\omega)|^2$, where $f(\omega)$ is called a Fourier multiplier.

We extend this concept to synthesize nonstationary Gaussian processes using wavelet multipliers $m(b,a)$ as a function of time b and scale a. Besides the possibility to generate realizations of an arbitrary wavelet spectrum, this framework allows one to generate surrogate data of nonstationary Gaussian processes.

A complementary approach is given by the recently suggested time-frequency ARMA (TFARMA) processes [15], which are special parametric versions of quasi-nonparametric time-varying ARMA (TVARMA) processes. Syntheses based on discrete wavelet transformation are that of Nason *et al.* [16], who define stochastic processes by superimposing weighted wavelet "atoms," or that of Percival and Walden [17], which are both confined to dyadic scales.

A. Definitions

We define an equivalence class of nonstationary Gaussian processes in the wavelet domain by the wavelet multipliers $m(b,a)$. An individual process is defined by its multipliers and a synthesizing wavelet pair $(g(t),h(t))$. Realizations $s(t)$ are given as

$$s(t) = M_h m(b,a) W_g \eta(\tau), \tag{4}$$

i.e., a driving Gaussian white noise $\eta(t) \sim \mathcal{N}(0,1)$ with $\langle \eta(t_1)\eta(t_2)\rangle = \delta(t_1 - t_2)$ is transformed to the wavelet domain, multiplied with $m(b,a)$, and transformed back to the time domain. Following Appendix A 1, the realization $m(b,a)W_g\eta(\tau)$ in the wavelet domain is, in general, not a wavelet transformation and, thus, realizations $s(t)$ in the time domain depend (usually weakly [18]) on the chosen wavelet $g(t)$ and the reconstruction wavelet $h(t)$, respectively.

1. Asymptotic behavior

To ensure at least asymptotic independence of the wavelet g and the reconstruction wavelet h, one has to demand a certain asymptotic behavior of the process $m(b,a)$. As wavelet analysis is a local analysis, the behavior for long time series is not of interest. Hence, we consider the limit of small scales. We demand the following behavior:

$$|\partial_a m(b,a)| < O(a^{-1+\epsilon})$$

$$|\partial_b m(b,a)| < O(a^{-1+\epsilon}). \tag{5}$$

This means that looking with a microscope into finer and finer scales, the derivatives of $m(b,a)$ grow slower and slower (in comparison to the scale) such that the process looks more and more stationary and white. In other words, for smaller and smaller scales, more and more reproducing kernels [19] fit into local structures of the process. These relations are derived in Appendix B 1. The previous discussion shows, that the notion of a time-scale component makes only sense in the limit of small scales.

2. Relation to the Fourier spectrum

Consider a stationary Gaussian process defined by $m(b,a) \equiv m(a)$ in the wavelet domain. In this special case, the stationary Fourier spectrum $|f(\omega)|^2$ exists

$$f(\omega) \approx m\left(\frac{2\pi}{\omega}\right)C_1 + \frac{2\pi}{\omega}m'\left(\frac{2\pi}{\omega}\right)C_2, \tag{6}$$

with $a=2\pi/\omega$ and C_1 and C_2 being constants, depending on the localization of the used wavelets. As expected, the Fourier spectrum is given by the wavelet spectrum plus a correction term. The latter depends on the slope of the wavelet spectrum $m'(b,a)$ (for details, refer to Appendix B 2). For processes exhibiting the asymptotic behavior defined in Eq. (5), the correction term vanishes for high frequencies.

B. Spectral measures

Hitherto, continuous wavelet spectral measures have been defined as the expectation value of the corresponding estimator, e.g., $E(|W_g s(t)[b,a]|^2)$ for the wavelet spectrum [10]. However, these measures are not defined *a priori*, but depend on realizations $s(t)$ and the analyzing wavelet $g(t)$. Also, in general, one does not have access to the ensemble average $E(.)$ [20]. Using wavelet multipliers, one can define time-dependent spectral measures that elegantly overcome these difficulties.

1. Spectrum

Given wavelet multipliers $m(b,a)$, one can define the spectrum as

$$S(b,a) = |m(b,a)|^2. \tag{7}$$

It quantifies the variance of the process at a certain time b and scale a. White noise is given by $S(b,a)=|m(b,a)|^2 = \text{const}$.

2. Cross spectrum

Consider two linearly interacting processes $m_{1c}(b,a)$ and $m_{2c}(b,a)$, i.e., both are driven by the same noise realization: $s_1(t)=M_h m_{1c}[b,a]W_g \eta_c(t)$ and $s_{2c}(t)=M_h m_2[b,a]W_g \eta_c(t)$ respectively. Then, the cross spectrum reads

$$S_{\text{cross}}(b,a) = m_{1c}(b,a)\bar{m}_{2c}(b,a). \tag{8}$$

In general, this is a complex function that may be decomposed into amplitude and phase:

$$S_{\text{cross}}(b,a) = |S_{\text{cross}}(b,a)|\exp(i\arg(S_{\text{cross}}(b,a))). \tag{9}$$

The cross spectrum denotes the covarying power of two processes, i.e., the predictive information between each other. Possibly, a superimposed independent variance only appears in the single spectra but not in the cross spectra; this also implies that the cross spectrum vanishes for two independent processes.

3. Coherence

The coherence (sometimes coherency) is defined as the modulus of the cross spectrum, normalized to the single spectra. Exhibiting values between zero and one, it quantifies the linear relationship between two processes. In general, one rarely finds perfect linear dependence; the single processes $m_1(b,a)$ and $m_2(b,a)$ rather consist of covarying parts $m_{1c}(b,a)$ and $m_{2c}(b,a)$ as well as superimposed independent contributions m_{1i} and m_{2i}: $s_1(t)=M_h m_{1c}[b,a]W_g \eta_c(t) + M_h m_{1i}[b,a]W_g \eta_{1i}(t)$ and $s_2(t)=M_h m_{2c}[b,a]W_g \eta_c(t) + M_h m_{2i}[b,a]W_g \eta_{2i}(t)$. Then, the squared coherence reads

$$\mathcal{C}^2(b,a) = \frac{|S_{\text{cross}}(b,a)|^2}{S_1(b,a)S_2(b,a)} = \frac{|m_{1c}(b,a)\bar{m}_{2c}(b,a)|^2}{|m_1(b,a)|^2|m_2(b,a)|^2}, \tag{10}$$

with $m_1(b,a)=m_{1c}(b,a)+m_{1i}(b,a)$ and $m_2(b,a)=m_{2c}(b,a) + m_{2i}(b,a)$.

All these measures in combination with a synthesizing wavelet pair are *a priori* definitions of individual processes; the procedure to estimate them will be addressed in Sec. IV.

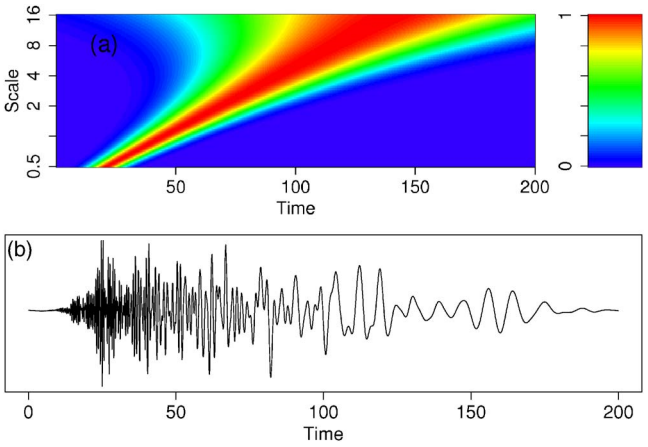

FIG. 1. (Color online) Stochastic chirp with $\epsilon=0.3$ (arbitrary units). (a) The spectrum $|m(b,a)|^2$. (b) A typical realization in the time domain, calculated with a Morlet wavelet, $\omega_0=6$.

C. Example

To illustrate this concept, we synthesize a stochastic chirp that is given as

$$m(b,a) = \exp\left(-\frac{(b-b_0(a))^2}{2\sigma^2(a)}\right) \tag{11}$$

with $b_0(a)=\beta_0+c\log(a)$ and $\sigma(a)=\sigma_0 a^{1-\epsilon}$, i.e., every voice (stripe of constant scale) is given by a Gaussian with time position and width varying with scale. The center of the Gaussian at scale a is given by b_0, the width as $\sigma(a)$, determined by the constants β_0, c, and σ_0. The power of $1-\epsilon$ ensures the process exhibiting the desired asymptotical behavior. Figures 1(a) and 1(b) show the spectrum $S(b,a) = |m(b,a)|^2$ and a typical realization generated from the spectrum by Eq. (4) in the time domain, respectively.

IV. ESTIMATING THE WAVELET SPECTRUM

Wavelet analysis is an inverse problem. One aims to estimate the wavelet spectrum of an unknown underlying pro-

cess. In this section, at first we briefly review the well-known wavelet spectral estimators and their distribution. Then, based on the framework developed in Sec. III, we derive the variance and bias of arbitrary estimated wavelet spectra.

A. Spectral estimators

Given a realization $s(t)$ of a nonstationary process, one can estimate its spectrum (i.e., calculate the wavelet sample spectrum) using a wavelet $g(t)$ by

$$\hat{S}_g(b,a) = A(|W_g s(t)|^2), \tag{12}$$

where A denotes an averaging operator defined in Sec. IV B and the caret marks the estimator. Following the terminology of Fourier analysis, the wavelet sample spectrum without averaging is either called a scalogram [1] or wavelet periodogram [e.g., [16]].

Given realizations $s_1(t)$ and $s_2(t)$ of two processes, the cross spectrum can be estimated as

$$\hat{S}_{\text{cross }g}(b,a) = A(W_g s_1(t) \bar{W}_g s_2(t)), \tag{13}$$

or decomposed into amplitude and phase,

$$\hat{S}_{\text{cross }g}(b,a) = |\hat{S}_{\text{cross }g}(b,a)| \exp(i \arg(\hat{S}_{\text{cross }g}(b,a))), \tag{14}$$

whereas the squared coherence is estimated as

$$\hat{C}_g^2(b,a) = \frac{|\hat{S}_{\text{cross }g}(b,a)|^2}{\hat{S}_{g,1}(b,a)\hat{S}_{g,2}(b,a)}. \tag{15}$$

For coherence, averaging is essential. Otherwise, one investigates power in a single point in time and scale, i.e., one attempts to infer covarying oscillations without observing the oscillations over a certain interval. Consequently, nominator and denominator become equal and one obtains a trivial value of one for any two processes.

B. Distribution, variance, and bias

Qui and Er [21] have studied variance and bias for deterministic periodic oscillations corrupted by white noise. For Gaussian processes, the wavelet scalogram $|W_g s(t)|^2$ and also the wavelet cross scalogram $W_g s_1(t) \bar{W}_g s_2(t)$ obey a χ^2 distribution with two degrees of freedom. The variance equals to two times the corresponding mean, $\text{Var}_S = 2\langle |W_g s(t)|^2 \rangle$.

To reduce the variance, averaging the wavelet scalogram in time or scale direction is required. This, in turn, produces a bias. Furthermore, the averaging destroys the simple χ^2 distribution [10]. This occurs (in contrast to the Fourier periodogram) because of the intrinsic correlations given by the reproducing kernel (Sec. A 2).

1. Variance of the wavelet sample spectrum

In practical applications, retaining a scale-independent variance appears to be a reasonable choice. This might be accomplished by averaging the same amount of independent information on every scale, i.e., by choosing the length of the averaging kernel according to the reproducing kernel. Following Eq. (3), this means [10]:

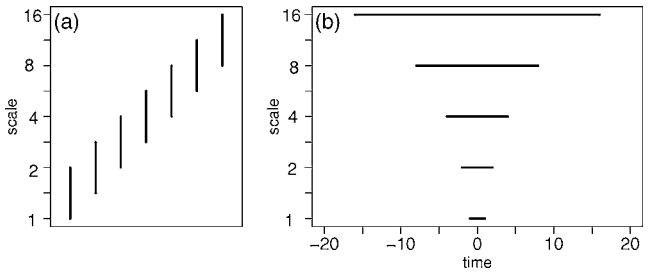

FIG. 2. Smoothing according to the reproducing kernel to provide a constant variance for all scales (arbitrary units). (a) In scale direction, the length of the smoothing window stays constant (for a logarithmic scale axis), w_a=const. (b) In time direction, the length of the smoothing window increases linearly with scale, $w_b a$.

- Averaging in scale direction should be done with a window exhibiting constant length for logarithmic scales, see Fig. 2(a). w_a denotes the half window length in the same units as N_{voice} [22].
- Averaging in time direction should be done with a window exhibiting a length proportional to scale [see Fig. 2(b)]. $w_b a$ denotes the half window length in units of time.

The (scale-independent) variance of Gaussian white noise as a function of the width of a rectangular averaging window is shown in Fig. 3. The graphs for averaging in the scale as well as in time direction resemble the shape of the reproducing kernel. An averaging window that is short compared to the effective width of the reproducing kernel includes only a minor part of the independent information and thus fails to notably reduce the variance. Thus, Fig. 3 provides guidance for choosing an appropriate length of the averaging window.

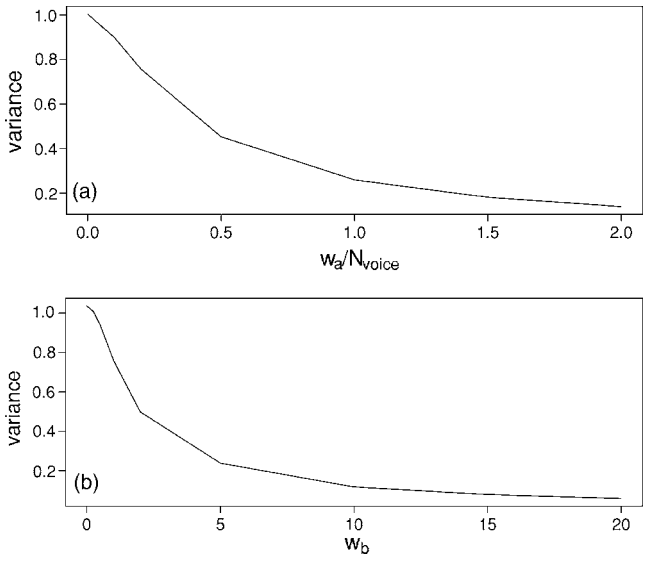

FIG. 3. Scale-independent normalized variance of the wavelet sample spectrum as a function of the lengths of a rectangular averaging window. (a) Averaging in scale direction with half window length w_a. (b) Averaging in time direction with half window length $w_b a$. The graphs resemble the shape of the reproducing kernel.

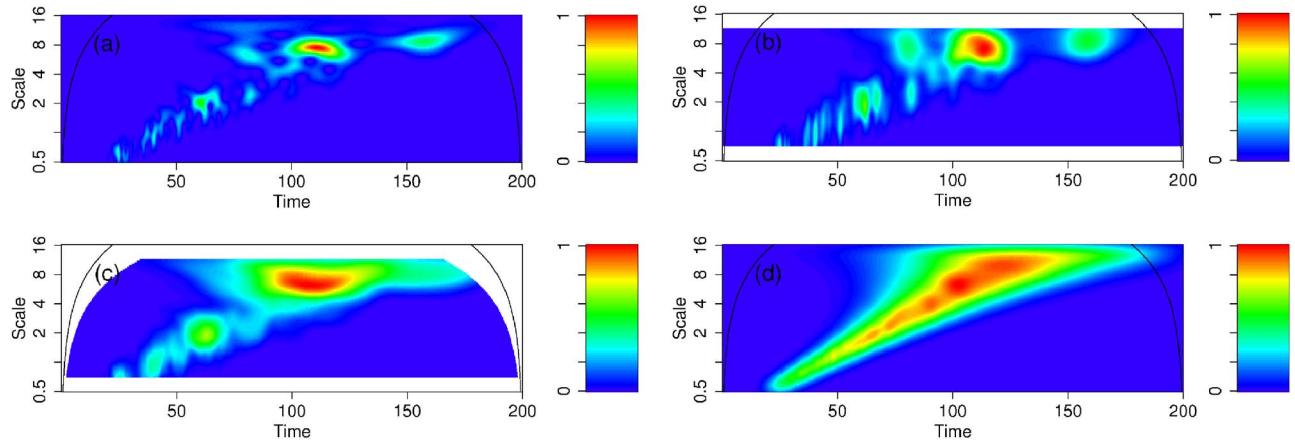

FIG. 4. (Color online) Estimation of the stochastic chirp based on the realization in Fig. 1(b) (arbitrary units). (a) The wavelet scalogram, i.e., the sample spectrum without averaging. (b) Averaged sample spectrum with $w_a/N_{\text{voice}}=0.5$. (c) Averaged sample spectrum with $w_a/N_{\text{voice}}=0.5$ and $w_b=3$. (d) The spectrum estimated as the mean of 1000 realizations without averaging.

The variance of an arbitrary wavelet sample spectrum can be estimated by constructing a bootstrap ensemble with Eq. (4). For processes following Eq. (5), the variance asymptotically vanishes for small scales without producing a bias (see Appendix B 3).

2. Bias of the wavelet sample spectrum

Given realizations of a Gaussian process defined by $m(b,a)$ and constructed with the wavelet pair $g(t)$ and $h(t)$, one can estimate the wavelet sample spectrum using a wavelet $k(t)$ and an averaging operator A. The bias at scale a and time b of the wavelet sample spectrum then reads

$$\text{Bias}(\hat{S}_g(b,a)) = \langle A(|\underbrace{W_k M_h}_{P_{h\to k}} m(b,a) W_g \eta(t)|^2)\rangle - |m(b,a)|^2, \tag{16}$$

where $P_{h\to k}$ denotes the projector defined in Appendix A 1. The bias consists of two contributions: The averaging with the operator A produces an *averaging bias* of the smoothed wavelet sample spectrum in comparison to the wavelet periodogram. Furthermore, not even the wavelet periodogram is an unbiased estimator; the projection property Appendix A 1 results in an *intrinsic bias* of the wavelet periodogram in relation to the underlying spectrum $m(b,a)$. Both the averaging bias and the intrinsic bias cause that, for finite scales, the wavelet sample spectrum is not a consistent estimator even in the limit of an infinite number of realizations. For averaging on finite scales, one has to consider the trade-off between bias and variance. For processes following Eq. (5), the bias of the estimator vanishes for small scales (see Appendix B 4).

C. Example

We recall the stochastic chirp from Sec. III C to exemplify the estimation procedure. Figure 4(a) depicts the wavelet scalogram of the realization shown in Fig. 1(b). It is easy to see that a single realization without averaging yields a rather insufficient estimation of the real spectrum. Averaging, shown in Figs. 1(b) and 1(c), reduces the variance but produces an averaging bias. The estimation based on the mean of 1000 realizations without averaging, Fig. 1(d), yields a pretty accurate result of the underlying process, which is not corrupted by the averaging bias, but only by the intrinsic bias.

V. SIGNIFICANCE TESTING

A. Pointwise testing the wavelet spectrum

To our knowledge, Torrence and Compo [9] were the first to establish significance tests for wavelet spectral measures. They assumed a red noise background spectrum for the null hypothesis and tested for every point in the time-scale plane separately (i.e., pointwise), whether the power exceeded a certain critical value corresponding to the chosen significance level. Since the critical values of an arbitrary background model are difficult to be accessed analytically [10], they need to be estimated based on a parametric bootstrap [23] as follows: Choose a significance level $1-\alpha$; choose a reasonable model (e.g., an AR[1] process in case of climate data following Hasselmann [24]) as null hypothesis H_0 and fit it to the data; estimate the $(1-\alpha)$-quantile S_{crit} (i.e., the critical value) of the corresponding background spectrum by Monte Carlo simulations. Depending on the chosen background model and the chosen normalization of the spectral estimator, the critical value in general depends on scale. Then, check for every point in the wavelet domain, whether the estimated spectrum exceeds the corresponding critical value. The set of all pointwise significant wavelet spectral coefficients is given as

$$P_{pw} = \{(b,a)|\hat{S}_g(b,a) > S_{\text{crit}}\}. \tag{17}$$

B. Areawise testing the wavelet spectrum

1. Multiple testing, intrinsic correlations, and spurious significance

The concept of pointwise significance testing always leads to the problem of multiple testing. Given a significance

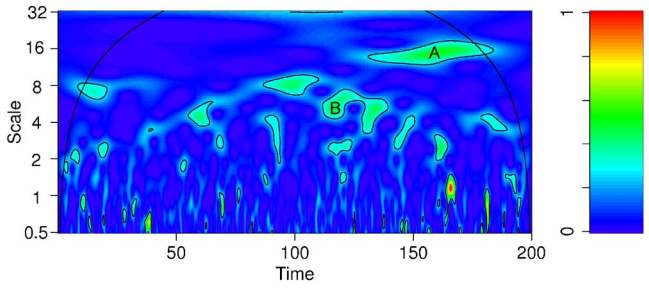

FIG. 5. (Color online) Pointwise significance test of the wavelet sample spectrum of Gaussian white noise (Morlet wavelet, $\omega_0=6$, $w_s=0$) against a white noise background spectrum of equal variance (arbitrary units). Spuriously significant patches appear.

level $1-\alpha$, a repetition of the test for N wavelet spectral coefficients leads to, on average, αN false positive results. For any time-scale-resolved analysis, a second problem comes into play. According to the reproducing kernel Eq. (3), neighboring times and scales of a wavelet transformation are correlated. Correspondingly, false positive results always occur as contiguous patches. These spurious patches reflect oscillations, which are randomly stable [25] for a short time.

For the interpretation of data from a process with an unknown spectrum, these effects mark an important problem: Which of the patches detected in a pointwise manner remain significant when considering multiple testing effects and the intrinsic correlations of the wavelet transformation?

Figure 5 illustrates that a mere visual judgment based on a sample spectrum will presumably be misleading. Even in the case of white noise, the test described in Sec. V A yields a large number of—by construction spuriously—significant patches.

2. Measuring areawise significance

We develop an areawise test that utilizes information about the size and geometry of a detected patch to decide whether it is significant or not. The main idea is as follows. If the intrinsic correlations are given by the reproducing kernel (Appendix A 2), then also the typical patch area for random fluctuations is given by the reproducing kernel. Following the dilation of the reproducing kernel Eq. (3) and as illustrated in Fig. 6, the typical patch width in time and scale direction should grow linearly with scale.

However, investigating the wavelet spectral matrices Fig. 5 reveals that many spurious patches do not have a typical form; rather, their forms are arbitrary and complex. Patches might exhibit a large extent in one direction, but be very localized in the other direction (patch A in Fig. 5). Other patches might consist of rather small patches connected by thin "bridges" (patch B in Fig. 5). These patches are spurious even though their area might be large compared to the corresponding reproducing kernel. Thus, not only the area but also the geometry has to be taken into account.

Given the set of all patches with pointwise significant values, P_{pw} [see Eq. (17)], we define *areawise significant* patches in the following way: For every (a,b), we choose a critical area $P_{\text{crit}}(b,a)$. It is given as the subset of the time-scale domain, where the reproducing kernel, dilated and

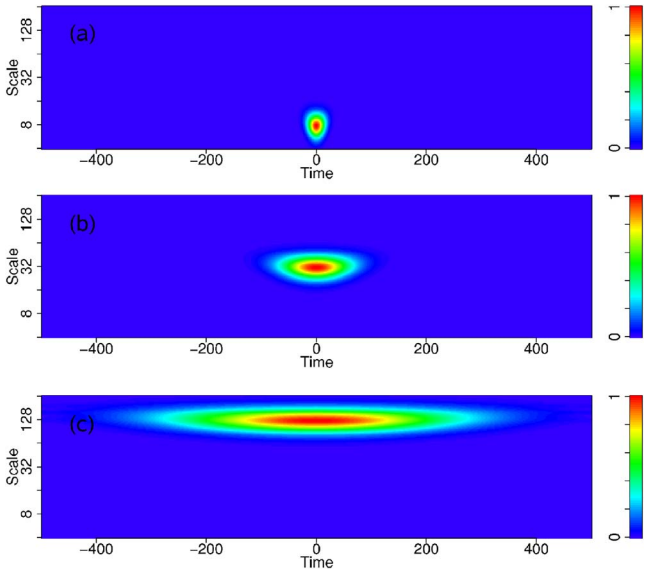

FIG. 6. (Color online) normalized reproducing kernel of the Morlet wavelet for three different scales (arbitrary units): (a) $s=8$, (b) $s=32$, and (c) $s=128$. The width in time and in scale direction increases linearly with scale (i.e., in scale direction, it appears constant on a logarithmic scale axis).

translated to (b,a), exceeds the threshold of a certain critical level K_{crit},

$$P_{\text{crit}}(b,a) = \{(b',a') | (K(b,a;b',a') > K_{\text{crit}}\}. \tag{18}$$

Then, the subset of additionally areawise significant wavelet spectral coefficients is given as the union of all critical areas that completely lay inside the patches of pointwise significant values

$$P_{aw} = \bigcup_{P_{\text{crit}}(b,a) \subset P_{pw}} P_{\text{crit}}(b,a). \tag{19}$$

In other words, given a patch of pointwise significant values, a point inside this patch is *areawise significant*, if any reproducing kernel (dilated according to the investigated scale) containing this point totally fits into the patch. Consequently, small as well as long but thin patches or bridges are sorted out as being insignificant.

3. Areawise significance level

The larger the critical area, the larger a patch needs to be to be detected by the test, i.e., the critical area is related to the significance level α_{aw} of the areawise test. We define the latter one as follows: the areas A_{pw} and A_{aw} corresponding to the pointwise and areawise patches, P_{pw} and P_{aw} result as

$$A_{pw} = \int_{P_{pw}} \frac{dbda}{a^2}$$

$$A_{aw} = \int_{P_{aw}} \frac{dbda}{a^2}. \tag{20}$$

Note that on every scale a, the area is related to the corresponding measure a^2. We now define the significance level of the areawise test as

TABLE I. Critical area P_{crit} at scale $a=1$ for different AR[1] processes with parameter a_1 for the wavelet scalogram with $\alpha_{pw}=0.05$ and $\alpha_{aw}=0.1$. The variance of the estimation is high due to the slow convergence of the stochastic rootfinding.

a_1	0.1	0.2	0.5	0.9
P_{crit}	7.01±0.06	7.21±0.36	6.94±0.31	7.00±0.14

$$1-\alpha_{aw}=1-\left\langle\frac{A_{aw}}{A_{pw}}\right\rangle, \tag{21}$$

i.e., one minus the average ratio between the areas of areawise significant patches and pointwise significant patches.

The relation between the desired areawise significance level $1-\alpha_{aw}$ and the critical area P_{crit} of the reproducing kernel is rather nontrivial. As a matter of fact, we had to estimate the corresponding critical area P_{crit} as a function of a desired significance level $1-\alpha_{aw}$ by a root-finding algorithm individually for every triplet (ω_0, w_a, w_b). The idea of this algorithm is outlined in Appendix C 1. It turns out that the critical area does not depend systematically on the chosen background model (see Table I).

4. Testing for significant areas

The actual areawise test is performed as follows: (i) Perform the pointwise test according to Sec. V A on the $1-\alpha$ level; (ii) stretch the reproducing kernel for every scale according to Eq. (3), choose a significance level $1-\alpha_{aw}$ for the areawise test and the corresponding critical area $P_{\mathrm{crit}}(b,a)$ of the reproducing kernel; (iii) slide the critical area $P_{\mathrm{crit}}(b,a)$ (for every scale the corresponding dilated version) over the wavelet matrix. A point inside a patch is defined as areawise significant, if any critical area containing this point totally lays within the patch. Figure 7 illustrates the areawise test based on the result of the pointwise test for a Gaussian white noise realization shown in Fig. 5. With $\alpha_{aw}=0.1$, the areawise test is capable of sorting out ~90% of the spuriously significant area from the pointwise test.

The areawise patch does not take into account the spectral value at a point (b,a); only information of the critical value contour line is utilized to define the patch. Thus, a strongly localized patch formed by a high peak might be sorted out. However, this problem might be handled by repeating the test for different significance levels $1-\alpha$. The higher the level, the more localized patches might be identified.

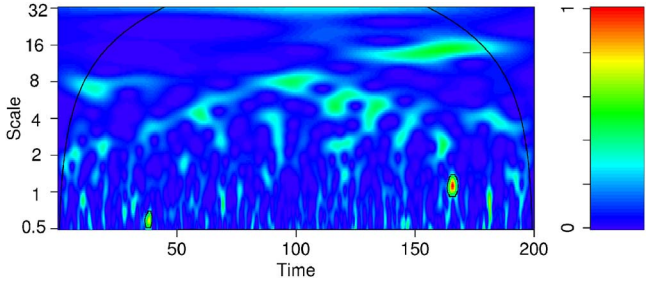

FIG. 7. (Color online) Areawise significance test performed on the example from Fig. 5 (arbitrary units). Most of the by construction spurious patches are sorted out.

C. Testing of covarying power

1. The wavelet cross spectrum

Compared to testing the single wavelet spectrum, the inference of covarying power is rather nontrivial. Such as for the stationary Fourier cross spectrum and the covariance (its time domain counterpart), no significance test for the wavelet cross spectrum exists. Assume two processes exhibiting independent power at overlapping time and scale intervals. This power does not covary, i.e., information about one of the processes is not capable of predicting the other one. Hence, the real wavelet cross spectrum is zero. By contrast, the estimated wavelet cross spectrum always differs from zero. As it is not a normalized measure, it is impossible to decide whether a cross-spectral coefficient is large because the one or the other process exhibits strong power or if actually covarying power does exist. Maraun and Kurths [10] illustrated this problem and analyzed a prominent example. To overcome this problem, one normalizes the cross spectrum and tests against zero coherence.

2. Pointwise testing of wavelet coherence

The structure of the test is similar to that developed for the wavelet spectrum. However, as the coherence is normalized to the single wavelet spectra, the critical value becomes independent of the scale as long as the smoothing is done properly, according to Sec. IV B (i.e., when the geometry of the reproducing kernel is accounted for).

In the case of Fourier analysis, the coherence critical value is independent of the processes to be compared, if they sufficiently well follow a linear description [26,27]. This independency, however, holds exactly only in the limit of a long time series. As wavelet analysis is a localized measure, this condition is not fulfilled. However, a simulation study reveals that the dependency on the process parameter a is rather marginal (see Appendix C 3).

3. Areawise testing of wavelet coherence

Also here, an areawise test can be performed to sort out false positive patches being artifacts from time and frequency resolved analysis. The procedure is exactly the same as for the wavelet spectrum, only the critical patch-size $P_{\mathrm{crit}}(b,a)$ has to be re-estimated. Areawise significant patches denote significant common oscillations of two processes. Here, common means that two processes exhibit a rather stable phase relation on a certain scale for a certain time interval.

4. Testing against random coherence

However, common oscillations do not necessarily imply coherence in a strict sense. Processes oscillating on similar frequencies trivially exhibit patches indicating an intermittently similar phase evolution. The typical lengths of these patches are determined by the decorrelation times of the

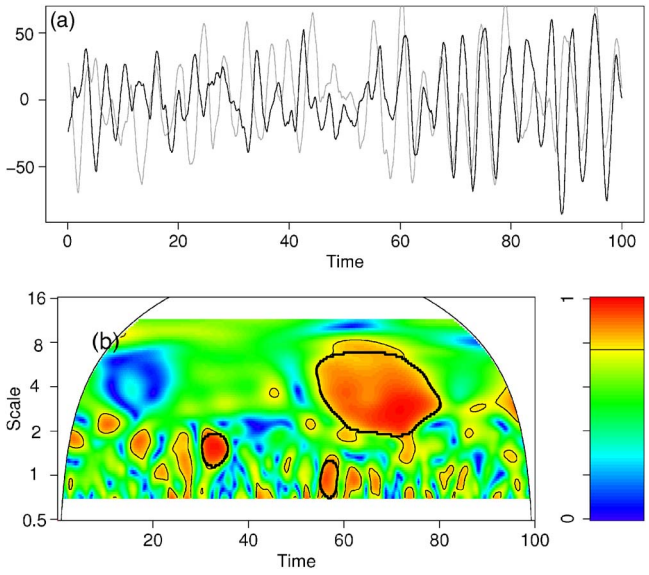

FIG. 8. (Color online) Areawise significance test for the coherence of two independent AR[2]-processes (arbitrary units). (a) Time series, (b) wavelet coherence. Thin lines: pointwise test. Thick lines: Areawise test. On the main frequency of 1/4, randomly common oscillations produce a large false positive patch.

single processes and the similarity of the concerned frequencies.

Figure 8 illustrates this discussion: We simulated realizations of two AR [2] processes with slightly different parameters: $x_i = a_1 x_{i-1} + a_2 x_{i-2} + \eta_i$, with $a_1 = 1.950\,12$, $a_2 = -0.967\,216$ for the first and $a_1 = 1.953\,03$, $a_2 = -0.967\,216$ for the second process. With a sampling time $\Delta t = 1/12$, this gives a common relaxation time of $\tau = 5$ and mean periods of $t_1 = 4$ and $t_2 = 4.4$, respectively. Even though the driving noise is independent, randomly common oscillations with a length related to the relaxation time and the difference in period occur.

If one is not only interested in deriving significant common oscillations, but also significant coherence in the sense of coupling between the processes, the areawise test has to be succeeded by another step: Using a bootstrapping approach, individually for every setting, it has to be tested if the time interval of the common oscillations is significantly long compared to typical randomly common oscillations of independent processes.

This test against random coherence has to be designed as follows: One constructs a bootstrap ensemble representing the length distribution of randomly common oscillations of the two processes under the null hypothesis (i.e., independence). On the one hand, this can be realized by a parametric bootstrap, i.e., by fitting two sufficiently complex models to the two data sets and then performing Monte Carlo simulations. Alternatively, one can apply a nonparametric bootstrap by constructing surrogate data of the two time series (e.g., using the presented wavelet synthesis). A patch with a length exceeding a certain quantile of the length distribution then signifies coherence in a strict sense. For an overview of surrogate time series (see [28]) and for bootstrapping, in general, see [29].

5. Complete test for coherence

To summarize, a complete coherence test includes the following steps: (i) Pointwise significance test as discussed in Sec. V C 2; (ii) areawise significance test as discussed in Sec. V C 3; and (iii) a bootstrap-based test against random coherence.

D. Comparison of tests

Real-world processes, in particular of geophysical or physiological nature, often exhibit power on a wide range of scales, where only a narrowband of time-localized oscillations might be interesting. The question arises as to how strong the localization in time and scale might be in relation to the background noise to be, in principle, identifiable. This question addresses the sensitivity of the test. On the other hand, it is relevant to know how many true features the test detects compared to the number of false-positive results. This question addresses the specificity of the test. (For definitions of sensitivity and specificity, see the Appendix C 2.)

To investigate these questions, we synthesized nonstationary Gaussian processes that exhibit a variance confined to a small area in the wavelet domain. We superposed white background noise to these processes with a certain signal to noise ratio $\mathcal{R}_{\text{peak}}$. The resulting processes simulate a typical situation in geophyiscs, where a signal confined in time and scale is hidden between other overlaid processes.

For different signal-to-noise ratios and signal extensions, we simulated a Monte Carlo ensemble and applied the pointwise and areawise test to every realization. From the outcomes, we estimated sensitivity and specificity and the rate of false positive and false negative results of the both tests. For details of this study, see Appendix C 4.

We summarize the following main results:

For a good signal-to-noise ratio, the specificity of both tests is very high, i.e., the sensitivity is the interesting measure; in this rather theoretical case, the pointwise test performs better. However, for a low signal-to-noise ratio, the specificity of the pointwise test is very low compared to that of the areawise test: The pointwise test produces many false-positive results, which are efficiently sorted out by the areawise test.

For data with a low signal to noise ratio, it is impossible to infer structures small compared to the reproducing kernel. They are, in principle, indistinguishable from the background noise.

Thus, for data sets exhibiting a broad spectrum (i.e., a low signal-to-noise ratio), the areawise test drastically increases the reliability of the interpretation.

VI. CONCLUSIONS

In this paper, we have presented a concept for continuous wavelet *synthesis* and *analysis*, i.e., for the direct problem of generating realizations of wavelet spectra and for the inverse problem of estimating wavelet spectra and significance testing them against a background spectrum.

(i) We have developed a framework to define nonstationary Gaussian processes in the wavelet domain; in this frame-

work, an arbitrary nonstationary wavelet spectrum is defined by wavelet multipliers in time and scale. Realizations are generated as follows: A driving Gaussian white noise is transformed to the wavelet domain, multiplied with the wavelet multipliers, and transformed back to the time domain with a suitable reconstruction wavelet. These realizations depend weakly on the wavelets used for the generation. Based on this concept, we have defined *a priori* measures for wavelet spectra and wavelet cross spectra. For the stationary case, these wavelet spectra are closely related to Fourier spectra.

Starting from the framework for the direct problem, we studied the inverse problem. (ii) We have investigated the variance and bias of continuous wavelet spectral estimators: To reduce the variance of the wavelet sample spectrum, one has to average it. Here, the extension of the averaging kernel has to be chosen corresponding to the reproducing kernel on every scale; otherwise, variance and bias will change with scale. The reproducing kernel also gives a guidance to choose an appropriate length for the averaging kernel. The wavelet sample spectrum is subject to two different types of bias: The averaging causes an averaging bias; additionally, even the wavelet periodogram exhibits an intrinsic bias.

(iii) We have proposed a new significance test. The conventional pointwise significance test produces many results that are artifacts resulting from a combination of multiple testing and intrinsic correlations given by the reproducing kernel; even white noise exhibits typical spurious patches. Thus, we have developed a new areawise significance test that subsequently assesses whether a patch exceeds a critical size given by the reproducing kernel; smaller patches are, in principle, indistinguishable from noise.

For the testing of coherence, an extra third step needs to be performed. Patches "surviving" the areawise test signify a common oscillation on a certain scale for a certain time interval. However, "common" does not mean "coherent" in the sense of coupling. Processes exhibiting oscillations on similar frequencies trivially show patches of a certain length given by the decorrelation times of the single processes. Thus, to infer coherence in a strict sense, one needs to test whether the patch is long in relation to typical randomly coherent oscillations.

We have compared the areawise significance test with the conventional pointwise significance test in terms of sensitivity and specificity. As the areawise test rejects patches small in relation to the reproducing kernel, it is slightly less sensitive but more specific. Given observations with a broad spectrum, e.g., from geophysics or physiology, the conventional test mimics a misleading structure that is successfully uncovered by the areawise test. A researcher left with a wavelet sample spectrum exhibiting many pointwise significant patches is given a measure to reject most spurious patches.

However, even though the effect of multiple testing has been dramatically reduced by the areawise test, the outcome is still merely statistical in nature. As for any statistical result, it is up to the researcher to provide a reasonable interpretation. Instead of being an end in itself, a wavelet analysis should be the starting point for a deeper physical understanding.

The presented framework is prototypically suitable for nonparametric bootstrapping in the wavelet domain. Aside from the construction of nonstationary surrogate data, this approach allows one to perform significance testing with a more complex nonstationary background spectrum. Among others, this is important for the analysis of processes with a trend in the variance. The concept might also be extended to non-Gaussian noise. These ideas, however, will be the subject of future research.

To synthesize realizations of a given wavelet spectrum, to estimate wavelet spectra and to perform areawise significance tests, we developed a free R-package based on the package Rwave by Carmona *et al.* [12]. All wavelet plots in this paper have been realized with this software [30].

ACKNOWLEDGMENTS

We are grateful for discussions with Udo Schwarz. This work has been supported by the Deutsche Forschungsgemeinschaft (DFG), projects No. SFB 555 and No. SPP 1114.

APPENDIX A: PROPERTIES OF THE TRANSFORMATION

1. Projection property

Taking an arbitrary function $f(b,a)$, the transformation $P_{g\to h}f(b,a)=W_hM_gf(b,a)$ to the time domain and back to the wavelet domain is a projector onto the space of all wavelet transformations [3]

$$P^2_{g\to h}f(b,a)=P_{g\to h}f(b,a).$$

2. Reproducing kernel

A function $r(b,a)$ is a wavelet transformation, if and only if

$$r(b,a)=\int_0^\infty \frac{da'}{a'}\int_0^\infty db'\frac{1}{a'}K_{g,h}\left(\frac{b-b'}{a'},\frac{a}{a'}\right)r(b',a')$$

and $K_{g,h}((b-b')/a',a/a')=W_gh((b-b')/a')$ is called the reproducing kernel [3]. The reproducing kernel of the Morlet wavelet is plotted in Fig. 6.

The intrinsic correlations given by the reproducing kernel constitute a fundamental difference of any time frequency (or scale) resolved analysis to time-independent Fourier analysis, where neighboring frequencies are asymptotically uncorrelated.

APPENDIX B: PROPERTIES OF GAUSSIAN PROCESSES IN WAVELET DOMAIN

1. Dependency on the wavelet

Given a realization of white noise $\eta(t)$, the difference between the realizations $s_g(t)$ and $s_h(t)$ for a certain $m(b,a)$ but different wavelets g and h reads

$$\begin{aligned}\Delta_{g,h}(t) &= M_g m W_g \eta(t) - M_h m W_h \eta(t) \\ &= \underbrace{M_h W_h} M_g m W_g \eta(t) - M_h m W_h \underbrace{M_g W_g} \eta(t) \\ &= M_h [W_h M_g m - m W_h M_g] W_g \eta(t) \\ &= M_h [P_{g\to h} * , m \cdot] W_g \eta(t),\end{aligned}$$

with $P_{g\to h} = W_h M_g$. The commutator in the previous equation is given by the integral kernel

$$[m(b',a') - m(b,a)] P_{g\to h}\left(\frac{b-b'}{a'}, \frac{a}{a'}\right).$$

Developing $m(b',a')$ into a Taylor series around (b,a) gives

$$\begin{aligned}&= [m(b,a) + (b-b')\partial_b m(b,a) + (a'-a)\partial_a m(b,a) \\ &\quad + O((a-a')^2 + (b-b')^2) - m(b,a)] P_{g\to h}\left(\frac{b-b'}{a'}, \frac{a}{a'}\right).\end{aligned}$$

For $P_{g\to h}\left(\frac{b-b'}{a'}, \frac{a}{a'}\right)$ sufficiently localized around (b,a), this reads

$$\begin{aligned}&\approx (b-b')\partial_b m(b,a) P_{g\to h}\left(\frac{b-b'}{a'}, \frac{a}{a'}\right) \\ &\quad + (a'-a)\partial_a m(b,a) P_{g\to h}\left(\frac{b-b'}{a'}, \frac{a}{a'}\right).\end{aligned}$$

With $P'_{g\to h}(b,a) = \frac{1}{a} b P_{g\to h}(b,a)$ and $P''_{g\to h}(b,a) = \left(\frac{1}{a} - 1\right) P_{g\to h}(b,a)$, we finally get

$$\begin{aligned}&= a\partial_b m(b,a) P'_{g\to h}\left(\frac{b-b'}{a'}, \frac{a}{a'}\right) \\ &\quad + a\partial_a m(b,a) P''_{g\to h}\left(\frac{b-b'}{a'}, \frac{a}{a'}\right).\end{aligned}$$

To ensure asymptotic independence of the chosen wavelet, it is necessary that $\Delta_{g,h}(t)$ vanishes for small scales. This is ensured in the following way: $P'_{g\to h}\left(\frac{b-b'}{a'}, \frac{a}{a'}\right)$ and $P''_{g\to h}\left(\frac{b-b'}{a'}, \frac{a}{a'}\right)$, given by the wavelets g and h, have to be sufficiently localized; and $a\partial_b m(b,a)$ and $a\partial_a m(b,a)$ have to vanish for small scales. This is fulfilled for processes exhibiting the asymptotic behavior given by Eq. (5).

2. Relation to Fourier spectra

For stationary processes, i.e., $m(b,a) \equiv m(a)$, Eq. (4) in the Fourier domain reads

$$\begin{aligned}\hat{s}(\omega) &= \int_0^\infty \frac{da}{a} \frac{1}{\sqrt{a}} \hat{h}(a\omega) m(a) \hat{g}(a\omega)\, \hat{\eta}(\omega) \\ &= \underbrace{\left[\int_0^\infty \frac{da}{a} \frac{1}{\sqrt{a}} G(a\omega) m(a)\right]}_{f(\omega)} \hat{\eta}(\omega).\end{aligned}$$

$$\begin{aligned}\hat{s}(\omega) &= \int_0^\infty \frac{da}{a} \frac{1}{\sqrt{a}} \hat{h}(a\omega) m(a) \hat{g}(a\omega)\, \hat{\eta}(\omega) \\ &= \left[\int_0^\infty \frac{da}{a} \frac{1}{\sqrt{a}} G(a\omega) m(a)\right]_{f(\omega)} \hat{\eta}(\omega).\end{aligned}$$

Here, we abbreviate $G(a\omega) = \hat{h}(a\omega)\hat{g}(a\omega)$. In this context, the caret refers to the Fourier transformation. The term $f(\omega)$ denotes Fourier multipliers representing the Fourier spectrum of the process $m(a)$.

Developing $m(a)$ into a Taylor series around $2\pi/\omega$, $m(a) = m(2\pi/\omega) + (a - 2\pi/\omega)m'(2\pi/\omega) + O((a-2\pi/\omega)^2)$ leads to

$$\begin{aligned}f(\omega) \approx m\left(\frac{2\pi}{\omega}\right)&\left[\int_0^\infty \frac{da}{a} \frac{1}{\sqrt{a}} G(a\omega)\right] + m'\left(\frac{2\pi}{\omega}\right) \\ &\times \left[\int_0^\infty \frac{da}{a} \frac{1}{\sqrt{a}} G(a\omega)\left(a - \frac{2\pi}{\omega}\right)\right].\end{aligned}$$

We factor out $2\pi/\omega$ in the second integral. If $G(a\omega)$ is well localized, the integrals might be considered as being constant. Finally, we obtain

$$f(\omega) \approx m\left(\frac{2\pi}{\omega}\right) C_1 + \frac{2\pi}{\omega} m'\left(\frac{2\pi}{\omega}\right) C_2.$$

As expected, the Fourier spectrum is given by the wavelet spectrum plus a correction term. The latter depends on the localization of the used wavelets and on the slope of the wavelet spectrum. For high frequencies, the difference vanishes if $m'(2\pi/\omega) < O(\omega)$ and if the process behaves as defined by Eq. (5).

3. Asymptotic variance of the wavelet sample spectrum

As discussed in Sec. III A, we constructed the class of nonstationary Gaussian processes such that they become locally stationary for small scales [see Eq. (5)]. Hence, it is possible to adapt the length w of the averaging kernel A in such a way to the process variability (given by ϵ) that it converges to zero size for small scales but at the same time includes more and more reproducing kernels. Then the variance of the spectral estimate and the bias (see Sec. IV B 2) vanish in the limit of small scales. Given the scale-dependent variance $\mathrm{Var}(a)$ of the wavelet scalogram, the following relation for the variance $\mathrm{Var}_A(a)$ of the averaged wavelet sample spectrum as a function of scale a holds

$$\mathrm{Var}_A(a) \sim \mathrm{Var}(a) a^{1-\alpha}. \tag{B1}$$

The exponent α $(1 > \alpha > 1-\epsilon)$ describes the scaling of the averaging window: $w(a) \sim a^\alpha$. The simple factor a^1 results from the width of the reproducing kernel in smoothing direction, which is proportional to scale.

Figure 9 shows the variance of the averaged wavelet sample spectrum of white noise ($\alpha = 0.75$). The solid line depicts the variance estimated from an ensemble of 1000 Gaussian chirps, the theoretically expected behavior is plotted as a dashed line.

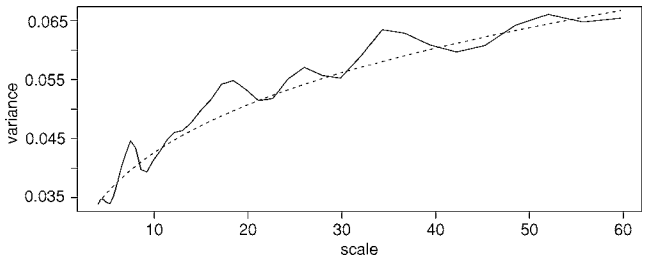

FIG. 9. Asymptotic behavior of the variance of the averaged wavelet sample spectrum of Gaussian white noise (arbitrary units). Solid line: estimation based on 1000 realizations, dashed line: theoretically expected behavior $\sim a^{1-0.75}$.

4. Asymptotic bias of the wavelet sample spectrum

The bias of the wavelet scalogram reads

$$\begin{aligned}\mathrm{Bias}(\hat{S}_g(b,a)) &= \langle |W_k M_h m(b',a') W_g \eta(t)|^2 \rangle - |m(b,a)|^2 \\ &= W_k M_h \bar{W}_k \bar{M}_h m(b_1,a_1) \bar{m}(b_2,a_2) \\ &\quad \times W_g \bar{W}_g \langle \eta(t_1)\eta(t_2)\rangle - |m(b,a)|^2.\end{aligned}$$

With $\langle \eta(t_1)\eta(t_2)\rangle = \delta(t_1-t_2)$, $\sigma_\eta = 1$ and $W_g g((t_2-b_2)/a_2) = K((b_1-b_2)/a_2, a_1/a_2)$, we get

$$= W_k M_h \bar{W}_k \bar{M}_h m(b_1,a_1)\bar{m}(b_2,a_2) K\left(\frac{b_1-b_2}{a_2}, \frac{a_1}{a_2}\right) - |m(b,a)|^2.$$

Developing $m(b_{1,2},a_{1,2})$ into a Taylor series around (a,b), i.e., $m(b_{1,2},a_{1,2}) \approx m(a,b) + (b_{1,2}-b)\partial_b m(b_{1,2},a_{1,2}) + (a_{1,2}-a)\partial_a m(b_{1,2},a_{1,2})$, and writing $(b_{1,2}-b)\partial_b m(b_{1,2},a_{1,2}) + (a_{1,2}-a)\partial_a m(b_{1,2},a_{1,2}) = f_{1,2}$, $W_k M_h \bar{W}_k \bar{M}_h K((b_1-b_2)/a_2, a_1/a_2) = C$, this leads to

$$\begin{aligned}\mathrm{Bias}(\hat{S}_g(b,a)) \approx (C-1)|m(b,a)|^2 - W_k M_h \bar{W}_k \bar{M}_h [m(b_1,a_1)\bar{f}_2 \\ + \bar{m}(b_2,a_2) f_1 + f_1 \bar{f}_2] K\left(\frac{b_1-b_2}{a_2}, \frac{a_1}{a_2}\right).\end{aligned}$$

If the wavelets are properly normalized, such that $C=1$, the bias reduces to the second term. Following the same reasoning as in Appendix B 1, the bias vanishes for $a \to 0$.

APPENDIX C: SIGNIFICANCE TESTING

1. Estimating the patch size

The significance level $1-\alpha_{aw}$ of the areawise test is a function of the critical area P_{crit}. Unfortunately, this function is not accessible analytically, such that it is impossible to choose a desired significance level $1-\alpha_{aw}$ and then straightforwardly calculate the corresponding critical area. In fact, one has to employ a root finding algorithm that solves the equation

$$f(P_{\mathrm{crit}}) - \alpha_{aw} = 0.$$

The estimation for $f(P_{\mathrm{crit}})$ results from Monte Carlo simulations and, thus, is stochastic itself—conventional root-finding algorithms fail to solve this problem. Thus, we developed an iterative procedure that is similar to stochastic approximation [31]: (i) Choose three reasonable initial guesses for P_{crit} and estimate α_{aw} based on Monte Carlo simulations. (ii) Assume a locally linear behavior of f around the root and fit a straight line to the three outcomes; (iii) As a next guess, choose the root of the straight line; (iv) Go back to step (ii), fit the straight line to all previous iterates (past iterates are given an algebraically decaying lower weight); and (v) Choose a termination criterion, e.g., a desired accuracy or a maximum number of iterations.

2. Sensitivity vs specificity

Given a population N, where a null hypothesis H_0 is right in N_R cases and wrong in $N_W = N - N_R$ cases. Applying a significance test for H_0 to measurements of every element, the numbers of true negative and false positive results are denoted as N_{TN} and N_{FP}, with $N_{TN}+N_{FP}=N_R$. The numbers of false negative and true positive results are given as N_{FN} and N_{TP}, with $N_{FN}+N_{TP}=N_W$. Then

$$\begin{aligned}\text{sensitivity} &= \frac{N_{TP}}{N_W} = \frac{N_{TP}}{N_{TP}+N_{FN}}, \\ \text{specificity} &= \frac{N_{TN}}{N_R} = \frac{N_{TN}}{N_{FP}+N_{TN}}.\end{aligned} \tag{C1}$$

The sensitivity relates the number N_{TP} of true rejections of H_0 to the total number of wrong H_0, N_W. On the other hand, the specificity measures the number N_{TN} of true acceptances of H_0 in relation to the total number of right H_0, N_R. A sensitive test rejects H_0 in preferably every case it is wrong (low β error), whereas a specific test preferably only rejects H_0 when it is definitely wrong (low α-error). For finite data, no test can be perfectly sensitive and specific, simultaneously.

3. Dependency of coherency critical values on the process

Table II shows the estimated critical values for some examples of different AR[1]-processes. The dependency on the smoothing parameters w_a and w_b can be seen comparing (a) and (b) in Table II. The dependency on the process parameter a, however, is rather marginal.

4. Comparison of tests

To investigate the sensitivity and specificity of the pointwise and the areawise significance test, we defined Gaussian bumps

$$m(b,a) = \exp\left(-\frac{(b-b_0)^2}{2\sigma_b^2}\right) \cdot \exp\left(-\frac{(c\log(a) - c\log(a_0))^2}{2\sigma_a^2}\right), \tag{C2}$$

where b_0 and a_0 denote mean time and scale, respectively, whereas σ_b and σ_b define the width in time and scale direction. The logarithm of the scale provides a Gaussian bump in the typical logarithmic scale axis wavelet matrix. Realizations were calculated according to Eq. (4) with driving noise

TABLE II. Critical values of the squared wavelet coherence $\mathcal{C}^2_{\text{crit}}$, between two AR[1]-processes with identical parameters a, $x_i = ax_{i-1} + \eta_i$, for different significance levels and different a. Estimated for a Morlet wavelet, $\omega_0=6$ with (a) $w_a=0.5$ octaves, $w_b=0$ and (b) $w_a=0.5$ octaves, $w_b=1$.

(a)	a	0	0.1	0.5	0.9
	90% level	0.861	0.862	0.868	0.880
	95% level	0.900	0.902	0.906	0.916
	99% level	0.949	0.951	0.952	0.959
(b)	a	0	0.1	0.5	0.9
	90% level	0.775	0.780	0.787	0.808
	95% level	0.827	0.832	0.838	0.856
	99% level	0.898	0.901	0.907	0.919

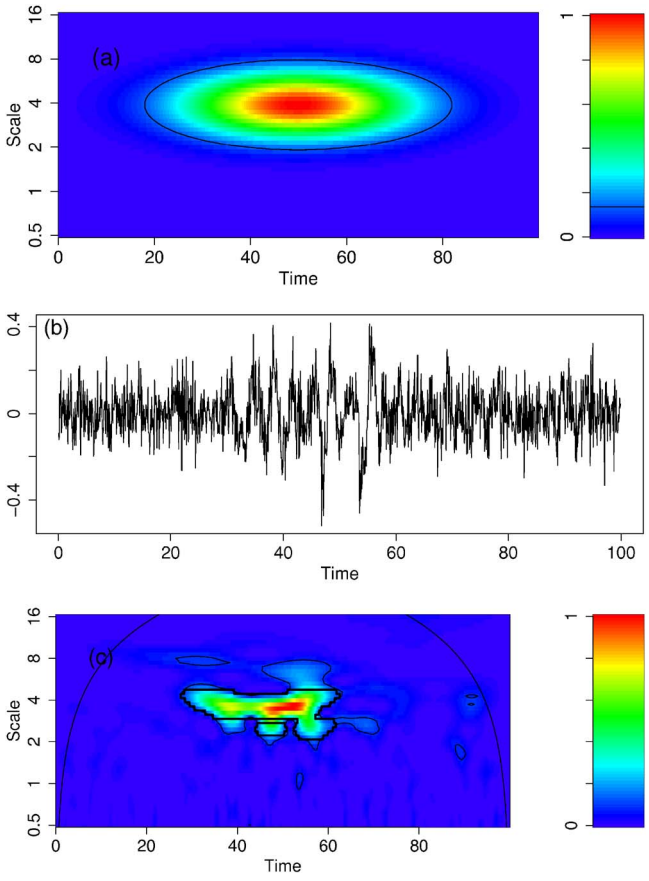

FIG. 10. (Color online) Gaussian bump with $b_0=50$, $a_0=4$, $\sigma_b=16$, and $\sigma_a=0.5$ superimposed by white noise (arbitrary units). The variance of the driving noise was $\sigma_\eta=1$, that of the background noise $\sigma_\xi=0.1$. For details see text. (a) $m(a,b)$, the contour-line marks $1/e^2$. (b) A realization in the time domain using a Morlet wavelet with $\omega_0=6$. (c) The corresponding wavelet sample spectrum calculated using the same wavelet. Thin and thick lines surround pointwise and areawise significant patches, respectively.

$\eta(t)$. The resulting time series was superimposed by independent background noise. As a simple model, we chose Gaussian white noise $\xi(t) \sim \mathcal{N}(0,\sigma)$ with zero mean and variance σ_ξ. Figure 10 displays an example. The amplitude of the driving noise was chosen as $\sigma_\eta=1$. However, the variance of

TABLE III. (a) Sensitivity of the pointwise (*pw*) and the areawise (*aw*) test, (b) ratio of false negative results from pointwise test to areawise test $A_{FN}(pw)/A_{FN}(aw)$, (c) Specificity of the *pw* and *aw* tests, (d) ratio of false-positive results from pointwise test to areawise test $A_{FP}(pw)/A_{FP}(aw)$. The signal-to-noise ratio $\mathcal{R}_{\text{peak}}$ is given as the ratio between the signal level in the peak $0.2\sigma_\eta$ and the noiselevel σ_ξ (see also Fig. 10). All values are estimated based on 1000 realizations of the corresponding bump.

(a)
$\mathcal{R}_{\text{peak}}$

		∞		20		10		2		1	
		pw	*aw*	*pw*	*aw*	*pw*	*aw*	*pw*	*aw*	*pw*	*aw*
σ_b	2	0.95	0.89	0.83	0.66	0.92	0.84	0.58	0.30	0.26	0.06
	4	0.95	0.91	0.69	0.54	0.73	0.59	0.64	0.47	0.40	0.19
	8	0.76	0.66	0.65	0.53	0.61	0.47	0.59	0.44	0.35	0.19
	12	0.79	0.71	0.56	0.47	0.53	0.41	0.52	0.38	0.43	0.27
	16	0.71	0.62	0.58	0.46	0.31	0.20	0.45	0.31	0.39	0.23

(b)
$\mathcal{R}_{\text{peak}}$

		∞	20	10	2	1
σ_b	2	0.5	0.5	0.5	0.6	0.8
	4	0.5	0.7	0.6	0.7	0.7
	8	0.7	0.7	0.7	0.7	0.8
	12	0.7	0.8	0.8	0.8	0.8
	16	0.8	0.8	0.9	0.8	0.8

(c)
$\mathcal{R}_{\text{peak}}$

		∞		20		10		2		1	
		pw	*aw*	*pw*	*aw*	*pw*	*aw*	*pw*	*aw*	*pw*	*aw*
σ_b	2	0.98	0.98	0.99	0.99	0.98	0.99	0.93	0.99	0.94	0.99
	4	0.97	0.98	1.00	1.00	1.00	1.00	0.96	0.99	0.94	1.00
	8	0.99	0.99	1.00	1.00	1.00	1.00	0.97	1.00	0.95	1.00
	12	0.98	0.99	1.00	1.00	1.00	1.00	0.99	1.00	0.94	0.99
	16	0.99	0.99	0.99	1.00	1.00	1.00	1.00	1.00	0.94	0.99

(d)
$\mathcal{R}_{\text{peak}}$

		∞	20	10	2	1
σ_b	2	1.2	1.4	1.4	8.3	10.5
	4	1.2	1.6	1.5	7.9	11.2
	8	1.6	2.0	2.5	13.9	13.2
	12	1.5	2.3	3.6	16.6	9.8
	16	1.7	2.0	11.0	21.7	9.8

the resulting bump is much lower (at the peak around $0.2\sigma_\eta$), as the bump is confined to a small spectral band. Thus, the superimposed noise with $\sigma_\xi=0.1$ represents a 50% noise level in relation to the bump itself. Therefore, we define the signal-to-noise ratio at the peak as $\mathcal{R}_{\text{peak}}=0.2\sigma_\eta/\sigma_\xi$. For $\sigma_\xi=0.2$, $\mathcal{R}_{\text{peak}}=1$.

We performed the following study. We simulated Gaussian bumps of different widths σ_b and fixed $\sigma_a=0.5$, superimposed by background noise with different variances σ_ξ. For each set of values (σ_b,σ_ξ), we simulated $N=10.000$ realizations. To every realization, we applied the pointwise $(1-\alpha_{pw}=0.95)$ and the areawise test $(1-\alpha_{aw}=0.9)$ as defined in Sec. V B 4.

Based on this experiment, we compared the sensitivity and specificity of the areawise significance test to those of the pointwise test. We define the area of the bump (i.e., the set of points where we assume H_0 as being wrong) as $P_B=\{(a,b)\,|\,m(a,b)>1/e^2\}$, the complement as $P_{NB}=\{(a,b)\,|\,m(a,b)\leq 1/e^2\}$.

The true positive patches are given as $P_{TP}=P\cap P_B$, and the true negative patches as $P_{TN}=\bar{P}\cap P_{NB}$; the false positive patches are given as $P_{FP}=P\cap P_{NB}$, and the false negative patches as $P_{FN}=\bar{P}\cap P_B$, where P stands for either P_{pw} or P_{aw} and $\bar{P}$ denotes the complement. We calculate the corresponding areas A_B, A_{NB}, A_{TP}, A_{TN}, A_{FP}, and A_{FN} as in Eq. (20). Then we can define the estimators $\frac{A_{TP}}{A_B}$ and $\frac{A_{TN}}{A_{NB}}$ for sensitivity and specificity, respectively.

On the one hand, the sensitivity of the pointwise test is higher than that of the areawise test [see (a) in Table III], as the latter one sorts out small patches in the area of the bump. The sensitivity depends strongly on the signal to noise ratio: For low background noise $\sigma_\xi \ll \sigma_\eta$, both tests perform very well [(a) in Table III], although the part of the bump area not detected by the areawise test is around twice as large than that not detected by the pointwise test (because the areawise test sorts out small patches, (b) in Table III). As the noise-level increases to the order of the bump's driving noise, $\sigma_\xi \sim \sigma_\eta$, the sensitivity decreases rapidly. For a zero signal-to-noise ratio, $\sigma_\xi \gg \sigma_\eta$ (not shown), the sensitivity of the pointwise test converges to $\alpha_{pw}=0.05$, that of the areawise test to $\alpha_{pw}\alpha_{aw}=0.005$. However, the ratio between the parts of the area not detected by the two tests ((b) in Table III) converges to $(1-\alpha_{pw})/(1-\alpha_{pw}\alpha_{aw})\approx 1-\alpha_{pw}=0.95$. In other words, for a very bad signal-to-noise ratio, the performance of the pointwise test is not really better. It just detects patches that occur spuriously because of the dominant noise.

On the other hand, the specificity of the areawise test is higher than that of the pointwise test (see (c) in Table III), as the latter one detects many more false-positive patches outside the area of the bump. Whereas the specificity of the areawise test appears to be almost independent of the signal-to-noise ratio close to one, that of the pointwise test decreases for high background noise, as more and more spurious patches appear. At first sight, the difference between the two tests seems to be rather marginal, but taking into account the number of false-positive results, an obvious difference arises: The ratio $A_{FP}(pw)/A_{FP}(aw)$ between the two tests ranges from 1 for a high signal-to-noise ratio to $1/\alpha_{aw}=10$ for a low signal-to-noise ratio (the estimated values are corrupted by a high uncertainty, the order of the values rather than the values itself is interesting).

The specificity is—trivially—almost independent of the bump width as it considers the area off the bump. Also trivially, small bumps nearly free from background noise are detected almost totally. This occurs because the small bumps are shorter than the reproducing kernel and thus get enlarged by the estimation. For large bumps, the sensitivity is, in general, lower. However, the decrease of the sensitivity with noise is much larger for small bumps than for large bumps. That means that small patches get rather invisible as they get superimposed by strong noise.

[1] O. Rioul and P. Flandrin, IEEE Trans. Signal Process. **40**, 1746 (1992).
[2] D. Donoho, IEEE Trans. Inf. Theory **41**, 613 (1995).
[3] M. Holschneider, *Wavelets: An Analysis tool* (Oxford University Press, London, 1998).
[4] T.-H. Li, IEEE Trans. Inf. Theory **48**, 2922 (2002).
[5] D. Gu and S. Philander, J. Clim. **8**, 864 (1995).
[6] B. Grenfell, O. Bjœrnstad, and J. Kappey, Nature (London) **414**, 716 (2001).
[7] L. Hudgins, C. A. Friehe, and M. E. Mayer, Phys. Rev. Lett. **71**, 3279 (1993).
[8] C. Torrence and P. Webster, Q. J. R. Meteorol. Soc. **124**, 1985 (1998).
[9] C. Torrence and G. Compo, Bull. Am. Meteorol. Soc. **79**, 61 (1998).
[10] D. Maraun and J. Kurths, Nonlinear Processes Geophys. **11**, 505 (2004).
[11] I. Daubechies, *Ten Lectures on Wavelets* (SIAM, Philadelphia, 1992).
[12] R. Carmona, W.-L. Hwang, and B. Torresani, *Practical Time-Frequency Analysis. Gabor and Wavelet Transforms with an Implementation in S* (Academic Press, New York 1998).
[13] D. Percival and A. Walden, *Wavelet Methods for Time Series Analysis* (Cambridge University Press, Cambridge, England 2000).
[14] J. Timmer and M. König, Astron. Astrophys. **300**, 707 (1995).
[15] M. Jachan, G. Matz, and F. Hlawatsch, in *Proc. IEEE ICASSP-2005* (IEEE, New York, 2005), Vol. IV, pp. 301–304.
[16] G. Nason, R. von Sachs, and G. Kroisandt, J. R. Stat. Soc. Ser. B (Stat. Methodol.) **62**, 271 (2000).
[17] D. Percival and A. Walden, *Wavelet Methods for Time Series Analysis* (Cambridge University Press, Cambridge, England, 2000).
[18] For the frequency decomposition of a signal, one should utilize wavelets which are well localized in the frequency domain (e.g., no Haar wavelet) and complex (e.g., no real Mexican hat wavelet). For details see [3].
[19] More precisely, we use the term-reproducing kernel as a cer-

tain quantile containing a large fraction of the reproducing kernel's mass. This quantile will be defined in Sec. V B 2 and referred to as critical area. It is dilated to the scale under investigation.

[20] In reality, one usually observes only one realization of a certain process, i.e., one has no access to the expectation value as an ensemble average. For a local analysis, however, replacing the ensemble average by the time average is not valid.

[21] L. Qiu and M.-H. Er, Int. J. Electron. **79**, 665 (1995).

[22] We calculate wavelet coefficients at $a_i = a_0 2^{(i-1/(N_{\text{voice}})}$, $i = 1, \ldots, N_{\text{voice}} N_{\text{octave}} + 1$, $a_{\min} = a_0$ and $a_{\max} = a_0 2^{N_{\text{octave}}} a_0$ corresponds to the Nyquist frequency, N_{octave}: number of octaves, N_{voice} number of scales per octave.

[23] In case of the wavelet scalogram and an AR[1] background spectrum, the approximation of Torrence and Compo [9] may be used instead.

[24] K. Hasselmann, Tellus **28**, 473 (1976).

[25] As physicists, we would rather say coherent. We prefer to use the term stable to avoid confusion with the term coherence, which here refers to the interrelation between two processes.

[26] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Models* (Springer, Heidelberg, 1991).

[27] M. B. Priestley, *Spectral Analysis and Time Series* (Academic Press, New York, 1992).

[28] T. Schreiber and A. Schmitz, Physica D **142**, 346 (2000).

[29] A. Davison and D. Hinkley, *Bootstrap Methods and their Application*, Cambridge Series in Statistical and Probabilistic Mathematics (Cambridge University Press, Cambridge England, 1997).

[30] It can be downloaded from http://tocsy.agnld.uni-potsdam.de/wavelets.php

[31] J. Kiefer and J. Wolfowitz, Ann. Math. Stat. **23**, 462 (1952).