

**Nonlinear mode decomposition: A noise-robust, adaptive decomposition method**

Dmytro Iatsenko, Peter V. E. McClintock, and Aneta Stefanovska\*

*Department of Physics, Lancaster University, Lancaster LA1 4YB, United Kingdom*

(Received 1 April 2015; published 29 September 2015)

The signals emanating from complex systems are usually composed of a mixture of different oscillations which, for a reliable analysis, should be separated from each other and from the inevitable background of noise. Here we introduce an adaptive decomposition tool—*nonlinear mode decomposition* (NMD)—which decomposes a given signal into a set of physically meaningful oscillations for any wave form, simultaneously removing the noise. NMD is based on the powerful combination of time–frequency analysis techniques—which, together with the adaptive choice of their parameters, make it extremely noise robust—and surrogate data tests used to identify interdependent oscillations and to distinguish deterministic from random activity. We illustrate the application of NMD to both simulated and real signals and demonstrate its qualitative and quantitative superiority over other approaches, such as (ensemble) empirical mode decomposition, Karhunen-Loève expansion, and independent component analysis. We point out that NMD is likely to be applicable and useful in many different areas of research, such as geophysics, finance, and the life sciences. The necessary MATLAB codes for running NMD are freely available for download.

DOI: [10.1103/PhysRevE.92.032916](https://doi.org/10.1103/PhysRevE.92.032916)

PACS number(s): 05.45.Tp, 05.10.–a, 05.40.–a, 02.30.Zz

**I. INTRODUCTION**

Complex systems in real life are commonly studied by analysis of the signals that they generate. Such signals typically include many different superimposed oscillations, giving rise to complicated wave forms, and these oscillations are often characterized by time-varying amplitudes and frequencies. These oscillations, which are also referred to as modes, contain a great deal of valuable information about the originating system. Their properties can be used, e.g., to predict earthquakes from geophysical signals [1] or to assess health from signals generated by the human body [2,3]. However, to gain access to all of the properties of a particular mode, it should be first extracted from the signal. For a reliable analysis, therefore, one should decompose the signal, i.e., recover the individual oscillations that are present in it, separating them both from each other and from the inevitable background of noise.

How best to accomplish this decomposition is a problem of long standing, and many different approaches have been proposed. However, to the best of the authors' knowledge, each of the current methods has at least two of the following flaws.

(1) The method contains user-defined parameters and is quite sensitive to their choice. These parameters cannot be adaptively chosen and there are no more-or-less universal settings.

(2) The method is not noise robust.

(3) If an individual mode has a complex (nonsinusoidal) wave form, the method will decompose it into a few oscillations with simpler wave forms.

(4) The modes returned by the method are not always physically meaningful. For example, the method will decompose

even a random signal, such as Brownian noise, onto a set of oscillations.

To the best of the authors' knowledge, issue (4) has not been addressed previously, so that this drawback is common to all current methods (and thus will not be mentioned again in what follows). Next, the popular empirical mode decomposition (EMD) [4] method suffers additionally from disadvantage (2) [5], as well as from (3) (with the latter manifesting itself only in cases where the corresponding oscillation has a wave form with multiple peaks, as can easily be verified numerically).

To reduce the sensitivity of EMD to noise, Wu and Huang [5] recently proposed a variant called ensemble empirical mode decomposition (EEMD). The idea is to add independent realizations of white Gaussian noise to the signal and apply EMD each time, with the final modes being obtained as the ensemble averages of the modes for each noise realization. Even so, however, EEMD still cannot be regarded as noise-robust method (at least in comparison with other approaches), as seen in the examples below. Additionally, it contains a nonadaptive parameter [drawback (1)] controlling the standard deviation of the added noise, though its choice does not seem to exert a strong influence on the results [5].

For multivariate time series the most widely used decomposition methods are principal component analysis (PCA) [6,7] and independent component analysis (ICA) [8,9]. In the present work, however, we restrict our consideration to univariate signals, i.e., those represented by a single time series. To apply PCA and ICA in this case, one should first construct a multivariate signal from the original (univariate) one, which can be accomplished by using its time-shifted blocks. PCA and ICA can then be applied to this collection of original signal parts, in which case they are called single-channel PCA (also known as Karhunen-Loève expansion) and single-channel ICA [10], respectively. However, both of these methods are generally susceptible to the time shifts and the number of blocks used to create the multivariate signal from the original one (as was demonstrated, e.g., in [11] for the Karhunen-Loève expansion), thus suffering from the drawback (1).

\*aneta@lancaster.ac.uk

Published by the American Physical Society under the terms of the [Creative Commons Attribution 3.0 License](https://creativecommons.org/licenses/by/3.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

Another approach, which is now becoming increasingly popular, is to use a particular time-frequency representation (TFR) [12–17] of the signal, e.g., the wavelet transform, for decomposing it into separate modes. Qualitatively, the signal is first mapped onto the time-frequency plane, where its oscillatory components appear in the form of “curves,” formed by the sequences of amplitude peaks. Having identified these curves, one can then recover the corresponding oscillations, thereby decomposing the signal. Many different techniques for realizing this idea have been proposed [18–23], with the great advantage of such approaches being their high noise robustness [12,24]. However, any TFR-based decomposition also suffers from drawback (3), because oscillation with a nonsinusoidal wave form is typically represented by a number of curves in the time-frequency plane [25].

The flaws mentioned greatly restrict the applicability of the approaches currently in use, so that only for a very small class of signals can the decomposition be carried out successfully. To overcome these limitations, we now introduce nonlinear [26] mode decomposition (NMD). It is based on the powerful combination of the time-frequency analysis techniques reviewed and developed in [12,19], surrogate data tests [27–30], and the idea of harmonic identification [25]. NMD has proven to be extremely noise robust, and it returns only the physically meaningful oscillations with any wave forms, at the same time removing the noise. Furthermore, we develop a set of criteria in which almost all NMD settings can be adapted automatically to the signal, greatly improving its performance and making it a kind of superadaptive approach. The application of NMD and its advantages over the other existing approaches are demonstrated on both simulated and real data.

The structure of the work is as follows. In Sec. II we summarize the notation used and review the necessary background. The basic idea of the NMD and its implementation are considered in Sec. III. In Sec. IV we introduce some important upgrades to improve the NMD performance, while the choice of its parameters is considered in Sec. V. We demonstrate the power of the method by applying it to simulated signals in Sec. VI and to real signals in Sec. VII. We mention some other contexts where NMD can prove useful in Sec. VIII. The work is concluded in Sec. IX and the full NMD procedure is summarized in the Appendix.

## II. BACKGROUND AND TERMINOLOGY

### A. Main notation

Given some function  $f(t)$ , we denote its Fourier transform, positive frequency part, time average, and standard deviation as  $\hat{f}(\xi)$ ,  $f^+(t)$ ,  $\langle f(t) \rangle$ , and  $\text{std}[f(t)]$ , respectively,

$$\begin{aligned}\hat{f}(\xi) &\equiv \int_{-\infty}^{\infty} f(t)e^{-i\xi t} dt \Leftrightarrow f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}(\xi)e^{i\xi t} d\xi, \\ f^+(t) &\equiv \int_{0^+}^{\infty} \hat{f}(\xi)e^{i\xi t} d\xi, \quad \langle f(t) \rangle = \frac{1}{T} \int f(t) dt, \\ \text{std}[f(t)] &\equiv \sqrt{[\langle f(t)^2 \rangle] - [\langle f(t) \rangle]^2},\end{aligned}\quad (1)$$

where  $T$  denotes the overall time duration of  $f(t)$  (in theory,  $T \rightarrow \infty$ ).

Next, the *amplitude* and/or *frequency modulated component* (in short *AM* and/or *FM component*, or simply *component*) is defined as the function of time  $t$  having the form

$$x(t) = A(t) \cos \phi(t), \quad [A(t) > 0, \nu(t) = \phi'(t) > 0 \forall t], \quad (2)$$

where  $A(t)$ ,  $\phi(t)$ , and  $\nu(t)$  are respectively the instantaneous amplitude, phase, and frequency of the component. We restrict the consideration to components satisfying the analytic approximation

$$[A(t) \cos \phi(t)]^+ \approx \frac{1}{2} A(t) e^{i\phi(t)} \quad (3)$$

because, otherwise, the amplitude and phase are not uniquely defined and cannot, in principle, be recovered reliably by time-frequency analysis methods [12]. For a more detailed discussion of the definitions of  $A(t)$ ,  $\phi(t)$ ,  $\nu(t)$  and of some related issues, see, e.g., [31–33].

In real cases, a signal usually contains many components of the form (2). Moreover, some of them typically do not correspond to an independent activity, but arise due to a complex wave form of a particular mode to which they are related (see below). This is because real oscillations are rarely purely sinusoidal, but have more complicated shapes as the result of nonlinearities in the generating system and/or the measurement apparatus. For example, the AM and FM component (2) raised to the third power  $[A(t) \cos \phi(t)]^3 = \frac{3}{4} A^3(t) [\cos \phi(t) + \frac{1}{3} \cos 3\phi(t)]$  consists already of two components, although there is obviously only one meaningful oscillation.

It is therefore better to consider the full *nonlinear modes*, defined as the sum of all components corresponding to the same activity,

$$c(t) = A(t)v[\phi(t)] = A(t) \sum_h a_h \cos[h\phi(t) + \varphi_h], \quad (4)$$

where  $v[\phi(t)] = v[\phi(t) + 2\pi]$  is some periodic function of phase (also called the “wave-shape function” [34]), which, due to its periodicity, can always be expanded in the Fourier series (4). Without loss of generality, we fix the normalization of  $A(t)$  and  $\phi(t)$  in (4) by setting  $a_1 = 1$ ,  $\varphi_1 = 0$ . Then the instantaneous phase and frequency of the whole nonlinear mode (NM) are  $\phi(t)$  and  $\nu(t) = \phi'(t)$ , respectively [34]. In what follows, the components composing the NM are referred to as *harmonics*, with the  $h$ th harmonic being represented by a term  $\sim a_h \cos[h\phi(t) + \varphi_h]$  in (4).

We assume that the signal is composed of the NMs  $c_i(t)$  of form (4) plus some noise  $\eta(t)$  (the class of noise is considered later, in Sec. III E):

$$s(t) = \sum_i c_i(t) + \eta(t). \quad (5)$$

Our ultimate goal is then to extract all the NMs present in the signal and to find their characteristics, such as their amplitudes  $A(t)$ , phases  $\phi(t)$ , and frequencies  $\nu(t)$ , as well as the amplitude scaling factors  $a_h$  and phase shifts  $\varphi_h$  of the harmonics. It is important to note that the representation (5) is not unique (e.g., any harmonic of the NM can always be represented as a separate mode), but, among all possible choices, one wants to find the sparsest decomposition, i.e., the one characterized by the smallest number of modes subject to the constraints on the

amplitude and frequency modulation of the components (2) and the resolution restrictions of the time-frequency methods (see below).

### B. Time-frequency representations

The TFR of the signal is constructed by projecting it onto the time-frequency plane, which allows us to visualize and follow the evolution of the signal's spectral content in time. Depending on how the projection is constructed, there are many different types of TFRs. In our case, however, we need those from which the components present in the signal can be extracted and reconstructed in a straightforward way. TFRs meeting this criterion are the windowed Fourier transform (WFT)  $G_s(\omega, t)$  and the wavelet transform (WT)  $W_s(\omega, t)$ . All aspects of their use in practice, and our implementation of them, are discussed in [12] (see also the classical reviews and books [13–17]), to which we refer the reader for detail; here we just review briefly the minimal necessary background.

Given a real signal  $s(t)$ , its WFT and WT are constructed as

$$\begin{aligned} G_s(\omega, t) &\equiv \int_{-\infty}^{\infty} s^+(u)g(u-t)e^{-i\omega(u-t)}du \\ &= \frac{1}{2\pi} \int_0^{\infty} e^{i\xi t} \hat{s}(\xi) \hat{g}(\omega - \xi) d\xi, \\ W_s(\omega, t) &\equiv \int_{-\infty}^{\infty} s^+(u)\psi^* \left[ \frac{\omega(u-t)}{\omega_\psi} \right] \frac{\omega du}{\omega_\psi} \\ &= \frac{1}{2\pi} \int_0^{\infty} e^{i\xi t} \hat{s}(\xi) \hat{\psi}^* \left( \frac{\omega_\psi \xi}{\omega} \right) d\xi, \end{aligned} \quad (6)$$

where  $g(t)$  and  $\psi(t)$  are respectively the chosen window and wavelet functions, with the latter satisfying the admissibility condition  $\hat{\psi}(0) = 0$ ; the  $\omega_\psi = \arg\max |\hat{\psi}(\xi)|$  denotes wavelet peak frequency, while for the WFT we assume  $\hat{g}(\xi)$  to be centered around its peak:  $|\hat{g}(0)| = \max |\hat{g}(\xi)|$ . We choose a Gaussian window for the WFT and a log-normal wavelet for the WT,

$$\begin{aligned} \hat{g}(\xi) &= e^{-(f_0\xi)^2/2} \Leftrightarrow g(t) = \frac{1}{\sqrt{2\pi}f_0} e^{-(t/f_0)^2/2}, \\ \hat{\psi}(\xi) &= e^{-(2\pi f_0 \ln \xi)^2/2}, \quad \omega_\psi = 1, \end{aligned} \quad (7)$$

where  $f_0$  is the resolution parameter determining the trade-off between time and frequency resolution of the resultant transform (by default we are using  $f_0 = 1$ ). The forms (7) appear to be the most useful ones in terms of the joint time-frequency resolution and the reconstruction of components [12]. All of the considerations that follow, however, are applicable for any  $\hat{g}(\xi)$  and  $\hat{\psi}(\xi > 0)$ , which are real, positive, and unimodal around the single peak.

Note that the appropriate choice of window or wavelet function, which for the parametrization (7) is equivalent to the choice of  $f_0$ , is of paramount importance for time-frequency analysis based on the WFT/WT and therefore for any method that utilizes it. Thus, the window or wavelet should be well matched to the signal properties. The resultant representation is then able to resolve separate components and at the same time represent well their amplitude and frequency modulations. This is thoroughly discussed in [12] (see also Sec. V A).

The difference between the WFT and WT lies in the way they resolve the components present in the signal. The WFT has linear frequency resolution, i.e., distinguishes components on the basis of their frequency differences, whereas the WT has logarithmic frequency resolution, distinguishing components on the basis of ratios (or the differences between the logarithms) of their frequencies. As a result, the time resolution of the WT increases with frequency, so that the time variability of the higher frequency components is represented better than that for the components of lower frequency, while the time resolution of the WFT is the same at all frequencies.

In theory, the WFT and WT (6) depend on continuous  $\omega$ , but in practice they are calculated at particular discrete  $\omega_k$ . The latter can be chosen as  $\omega_k = (k - k_0)\Delta\omega$  for the WFT and  $\ln \omega_k = (k - k_0)\frac{\ln 2}{n_v}$  for the WT, reflecting the linear and logarithmic frequency resolution of the corresponding TFRs; the discretization parameters  $\Delta\omega$  and  $n_v$  can be selected using the criteria established in [12]. Due to the finite time length of real signals, all TFRs suffer from boundary distortions and, to reduce them, we use the predictive padding also suggested in [12].

### III. NONLINEAR MODE DECOMPOSITION: BASIC IDEA AND INGREDIENTS

The main goal of NMD is to decompose a given signal into a set of nonlinear modes (4). To do this, four steps are necessary:

- (a) extract the fundamental harmonic of an NM accurately from the signal's TFR;
- (b) find candidates for all its possible harmonics, based on its properties;
- (c) identify the true harmonics (i.e., corresponding to the same NM) among them;
- (d) subtract the resultant NM (obtained by summing together all the true harmonics) from the signal and iterate the procedure on the residual until a preset stopping criterion is met.

These individual subprocedures are explained in detail in the sections below, while the full NMD algorithm, including various improvements and related issues, is summarized in the Appendix.

Note that NMD can be based either on the WFT or on the WT, although in what follows (Sec. V B) we combine them to make the procedure more adaptive. Nevertheless, for the time being we assume that we have selected one of these TFR types and that we will use it for all operations.

#### A. Extraction and reconstruction of the main component

Provided that the TFR has sufficient resolution in time and frequency to resolve components that are close in frequency and, at the same time, to portray reliably their amplitude and frequency modulations, each component is represented by a unique sequence of TFR amplitude peaks (also called ridge points). We call such a sequence a *ridge curve* and denote it as  $\omega_p(t)$ ; an illustrative example showing the WFT and WT of a signal and the ridge curves of its components is presented in Fig. 1. Note that oscillations with a nonsinusoidal wave form

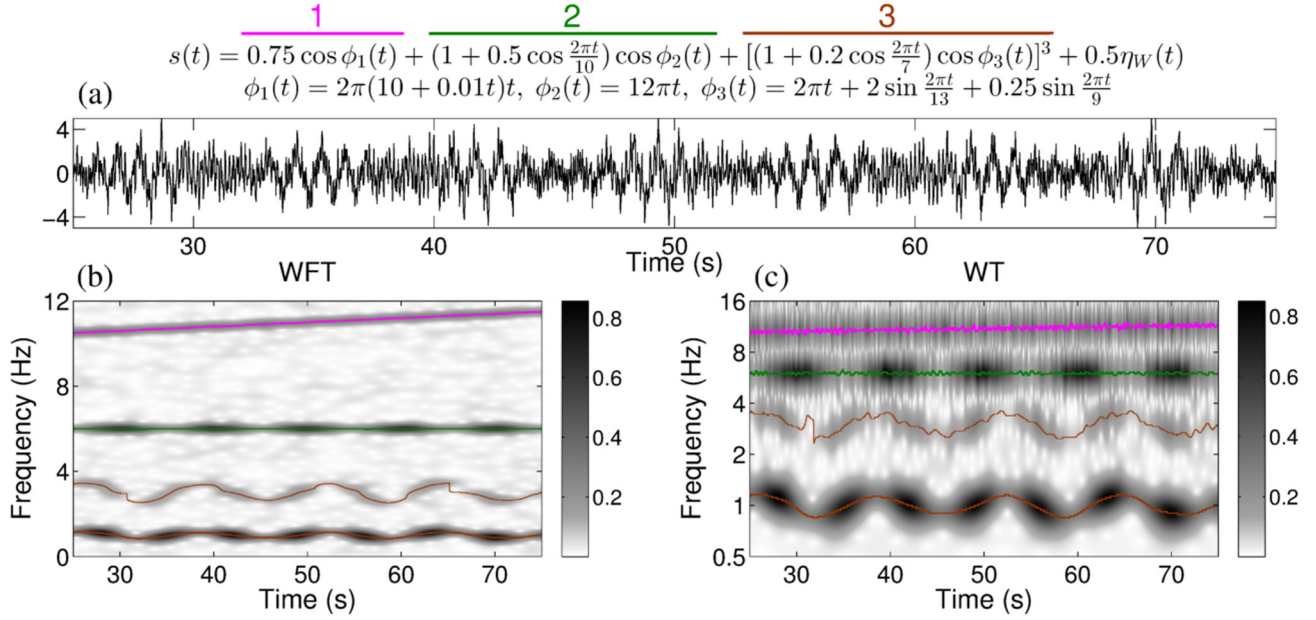


FIG. 1. (Color online) (a) The central 50-s section of the signal  $s(t)$  specified at the top, which is composed of three oscillations corrupted with white Gaussian noise of standard deviation equal to 0.5. The oscillations are (1) near 10 Hz, with frequency modulation (chirp); (2) near 6 Hz, with amplitude modulation; (3) near 1 Hz, with both modulations and a nonsinusoidal wave form. (b),(c) Respectively, the central parts of the WFT and WT of the signal  $s(t)$  shown in (a), where the ridge curves related to each oscillation are shown in each case by solid lines of the corresponding colors: 1, magenta (top line); 2, - green (second from the top); 3, brown (two bottom lines); the third oscillation is represented by two curves because of its complex wave form, while the apparent distortions in the ridge frequency profiles are due to noise; note also the logarithmic frequency scale for the WT, which is natural for it. The signal was sampled at 100 Hz for 100 s.

appear as two or more components in a TFR, consisting of the fundamental and its harmonics.

To extract a particular component from the TFR, one needs first to find its ridge curve. This is a nontrivial task, because in real cases it is not always clear which sequence of peaks corresponds to which component and which peaks are just noise-induced artifacts. The procedures for ridge curve extraction are discussed and developed in [19]. By default we use here scheme II(1,1) from the latter, as it appears to be very universal in applicability. In short, in the case of WFT we select the ridge curve  $\omega_p(t)$  as being that sequence of peaks (among all those that are possible) which maximizes the path functional

$$S[\omega_p(t)] \equiv \sum_{n=1}^N \left\{ \ln |G_s(\omega_p(t_n), t_n)| - \left| \frac{\Delta\omega_p(t_n) - m[\Delta\omega_p]}{s[\Delta\omega_p]} \right| - \left| \frac{\omega_p(t_n) - m[\omega_p]}{s[\omega_p]} \right| \right\}, \quad (8)$$

where  $\Delta\omega_p(t_n) \equiv \omega_p(t_n) - \omega_p(t_{n-1})$ , while  $m[\dots]$  and  $s[\dots]$  denote the median and 50% range, respectively. In the case of the WT, one uses the same functional (8), except that now all is taken on a logarithmic frequency scale ( $\omega_p \rightarrow \ln \omega_p$ ,  $\Delta\omega_p \rightarrow \Delta \ln \omega_p$ ). The approximate optimal  $\omega_p(t)$  in terms of (8) can be found in  $O(N \ln N)$  steps as described in [19].

Having found the ridge curve  $\omega_p(t)$ , the amplitude  $A(t)$ , phase  $\phi(t)$ , and frequency  $\nu(t) \equiv \phi'(t)$  of the corresponding component  $x(t) = A(t) \cos \phi(t)$  can be reconstructed by either of two methods, ridge and direct [12]. The ridge method

utilizes only the TFR values at the peak, and the corresponding formulas read

$$\begin{aligned} \text{ridge[WFT]:} \\ \nu(t) = \omega_p(t) + \delta\nu_d(t), \quad A(t)e^{i\phi(t)} &= \frac{2G_s(\omega_p(t), t)}{\hat{g}[\omega_p(t) - \nu(t)]}; \\ \text{ridge[WT]:} \\ \nu(t) = \omega_p(t)e^{\delta \ln \nu_d(t)}, \quad A(t)e^{i\phi(t)} &= \frac{2W_s(\omega_p(t), t)}{\hat{\psi}^*[\omega_p \nu(t)/\omega_p(t)]}; \end{aligned} \quad (9)$$

where  $\delta\nu_d(t)$  and  $\delta \ln \nu_d(t)$  are the corrections for discretization effects which can be found, e.g., by parabolic interpolation.

Another way is to reconstruct the component from the full time-frequency region into which it is mapped for a given TFR. We call this region the *time-frequency support*  $[\omega_-(t), \omega_+(t)]$  and estimate it as the widest region of unimodal and nonzero TFR amplitude around the ridge point  $\omega_p(t)$  at each time [12]. Based on estimated  $\omega_{\pm}(t)$ , the parameters of the component can be reconstructed as

$$\begin{aligned} \text{direct[WFT],} \\ A(t)e^{i\phi(t)} &= C_g^{-1} \int_{\omega_-(t)}^{\omega_+(t)} G_s(\omega, t) d\omega, \\ C_g &\equiv \frac{1}{2} \int_{-\infty}^{\infty} \hat{g}(\xi) d\xi = \pi g(0), \\ \nu(t) &= \text{Re} \left[ \frac{\int_{\omega_-(t)}^{\omega_+(t)} \omega G_s(\omega, t) d\omega}{\int_{\omega_-(t)}^{\omega_+(t)} G_s(\omega, t) d\omega} - \bar{\omega}_g \right], \end{aligned}$$

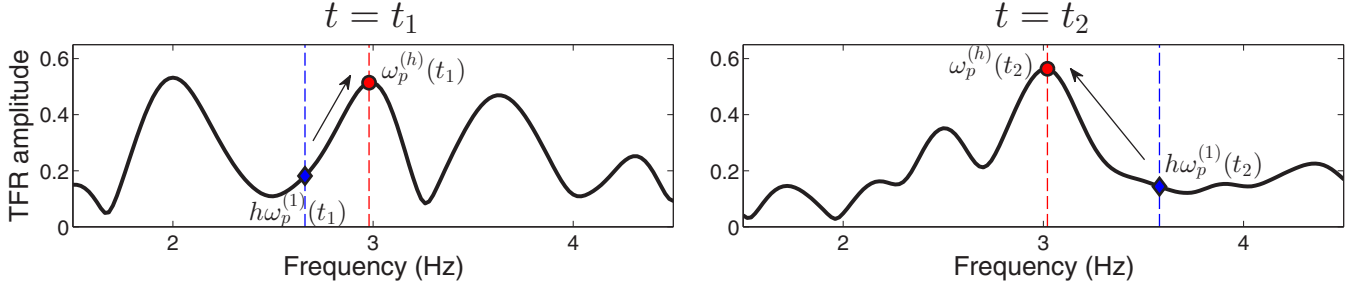


FIG. 2. (Color online) Illustration of the extraction of the  $h$ th harmonic ridge curve  $\omega_p^{(h)}(t)$  based on the fundamental ridge frequency  $\omega_p^{(1)}(t)$ . At each time, starting from the expected ridge frequency  $h\omega_p^{(1)}(t)$  of the harmonic (blue diamonds), one climbs (i.e., follows in the direction of TFR amplitude increase) to the nearest peak, which is then assigned to  $\omega_p^{(h)}(t)$  (red circles).

$$\bar{\omega}_g \equiv \frac{C_g^{-1}}{2} \int_{-\infty}^{\infty} \xi \hat{g}(\xi) d\xi$$

$[\bar{\omega}_g = 0$  for symmetric  $\hat{g}(\xi)$ , e.g., Gaussian window]; (10)  
direct[WT]:

$$A(t)e^{i\phi(t)} = C_\psi^{-1} \int_{\omega_-(t)}^{\omega_+(t)} W_s(\omega, t) \frac{d\omega}{\omega},$$

$$C_\psi \equiv \frac{1}{2} \int_0^\infty \hat{\psi}^*(\xi) \frac{d\xi}{\xi},$$

$$v(t) = \text{Re} \frac{D_\psi^{-1} \int_{\omega_-(t)}^{\omega_+(t)} \omega W_s(\omega, t) \frac{d\omega}{\omega}}{C_\psi^{-1} \int_{\omega_-(t)}^{\omega_+(t)} W_s(\omega, t) \frac{d\omega}{\omega}},$$

$$D_\psi \equiv \frac{\omega_\psi}{2} \int_0^\infty \frac{1}{\xi} \hat{\psi}^*(\xi) \frac{d\xi}{\xi}.$$

The difference and relative performance of the two reconstruction methods were studied in [12]. In brief, the direct method better recovers the time variations of the component's amplitude and frequency but is less robust to noise and interferences than the ridge method. Assume for now that we reconstruct everything by a particular method, e.g., ridge; its adaptive choice is discussed in Sec. V.

### B. Finding possible harmonics

The component extracted in the previous step will generally represent a harmonic of some NM. For simplicity, however, in this and the next section we assume that it corresponds to a fundamental, i.e., the first, harmonic (we get rid of this assumption in Sec. III D). Then, given the component's ridge frequency  $\omega_p^{(1)}(t)$ , its  $h$ th harmonic is expected to lie near  $h\omega_p^{(1)}(t)$ . Therefore, the ridge curve  $\omega_p^{(h)}(t)$  corresponding to the  $h$ th harmonic can be extracted simply as the sequence of peaks which are located in the same time-frequency support (region of unimodal TFR amplitude at each time) as  $h\omega_p^{(1)}(t)$ ; or, in other words, the sequence of peaks nearest to  $h\omega_p^{(1)}(t)$  in the direction of TFR amplitude increase. This is illustrated in Fig. 2. Having found  $\omega_p^{(h)}(t)$ , the parameters of the harmonic  $A^{(h)}(t), \phi^{(h)}(t), v^{(h)}(t)$  can be reconstructed as discussed in the previous section.

### C. Distinguishing true harmonics

In general, the procedure of the previous section yields what is not necessarily a genuine harmonic, but just a candidate for one. Thus, even if the NM does not contain a particular harmonic, one will still obtain some signal for it, consisting of noise or components lying near the frequency of the putative harmonic. Hence, having extracted the  $h$ th harmonic candidate, we must next determine whether it is a true harmonic or not. To tackle this problem, we use the method of surrogate data [27,28], testing against the null hypothesis of independence between the first harmonic and the extracted harmonic candidate. Thus, we first select a measure to quantify the degree of dependence between the dynamics of two harmonics, and we then calculate it for the original harmonic and for a collection of surrogates, a specially constructed time series consistent with the null hypothesis being tested. If the original value of the selected measure lies outside the distribution of its surrogate values, this indicates genuine interdependence and the harmonic is regarded as true; otherwise, it is discarded as false.

The amplitude, phase, and frequency dynamics of a true harmonic should depend on those for the first harmonic (fundamental) as  $A^{(h)}(t)/A^{(1)}(t) \equiv a_h = \text{const}$ ,  $\phi^{(h)}(t) - h\phi^{(1)}(t) \equiv \phi_h - h\phi_1 = \text{const}$ ,  $v^{(h)}(t) = hv^{(1)}(t)$ . One can introduce measures  $q_{A, \phi, v}^{(h)} \in [0, 1]$  quantifying the degree of consistency with these laws (0, no consistency; 1, full consistency), i.e., the dependence of the parameters of the  $h$ th harmonic on the corresponding parameters of the first one:

$$\begin{aligned} q_A^{(h)} &\equiv \exp \left\{ -\frac{\sqrt{\langle [A^{(h)}(t)A^{(1)}(t)] - A^{(1)}(t)A^{(h)}(t) \rangle^2}}{\langle A^{(1)}(t)A^{(h)}(t) \rangle} \right\}, \\ q_\phi^{(h)} &\equiv a \left| \langle \exp \{ i[\phi^{(h)}(t) - h\phi^{(1)}(t)] \} \rangle \right|, \\ q_v^{(h)} &\equiv \exp \left\{ -\frac{\sqrt{\langle [v^{(h)}(t) - hv^{(1)}(t)]^2 \rangle}}{\langle v^{(h)}(t) \rangle} \right\}. \end{aligned} \quad (11)$$

The overall measure of interdependence between the harmonics can then be taken as

$$\rho^{(h)}(w_A, w_\phi, w_v) = (q_A^{(h)})^{w_A} (q_\phi^{(h)})^{w_\phi} (q_v^{(h)})^{w_v}, \quad (12)$$

with the parameters  $w_{A, \phi, v}$  giving weights to each of the consistencies  $q_{A, \phi, v}^{(h)}$ . By default, we use  $\rho^{(h)} \equiv \rho^{(h)}(1, 1, 0)$ , giving equal weights to the amplitude and phase consistencies and no weight to the frequency consistency. The latter is

because the procedure of harmonic extraction (see Sec. III B), being based on the instantaneous frequency of the first harmonic, in itself introduces a dependence of  $v^{(h)}(t)$  on  $v^{(1)}(t)$ , so it is better not to base any conclusions on it.

Ideally, for true harmonics one should have  $q_{A,\phi,v}^{(h)} = 1$ , but noise and interference with the other components introduce errors, making the consistencies smaller than unity. Hence, one cannot identify harmonics based only on the value of  $\rho^{(h)}$  but also needs to use the surrogate test. We employ the idea of time-shifted surrogates [29,30], using as a first harmonic surrogate its time-shifted version and as the other harmonic surrogate the corresponding candidate harmonic extracted from the time-shifted TFR. Such time delay destroys any temporal correlations between the signals while preserving all other their properties, thus creating surrogates consistent with the null hypothesis of independence.

Given the maximal time-shift (in samples)  $M$ , the surrogate parameters for the first harmonic are taken as shifted backwards by  $\Delta T_d/2$ ,

$$\begin{aligned} A_d^{(1)}(\tau) &= A^{(1)}(\tau - \Delta T_d/2), \\ \phi_d^{(1)}(\tau) &= \phi^{(1)}(\tau - \Delta T_d/2), \\ v_d^{(1)}(\tau) &= v^{(1)}(\tau - \Delta T_d/2), \\ \tau &= \{t_{i=1+M/2,\dots,N-M/2}\}, \\ \Delta T_{d=1,\dots,N_d} &= M(1 - 2d/N_d)/(2f_s), \end{aligned} \quad (13)$$

where  $f_s$  is the signal sampling frequency and  $N$  is its total length in samples (note that the length of the surrogate time series is smaller than the original ones, being  $N - M$ ). Using  $v_d^{(1)}(\tau)$  (13) as a reference profile, the surrogate  $h$ th harmonic and its parameters  $A_d^{(h)}(\tau), \phi_d^{(h)}(\tau), v_d^{(h)}(\tau)$  are extracted in the same way as described in Sec. III B, but from the signal's TFR shifted on  $\Delta T_d/2$  forward in time [ $G(\omega, \tau + \Delta T_d/2)$  or  $W(\omega, \tau + \Delta T_d/2)$ ]. The procedure can greatly be accelerated by initial preprocessing of the TFR, constructing its ‘‘skeleton’’ [12] at the beginning and utilizing it for each surrogate.

Summarizing, we calculate the amplitude-phase consistencies  $\rho_{d=1,\dots,N_d}^{(h)}(1,1,0)$  (12) for the surrogate parameters (13) and compare them with the value  $\rho_0^{(h)}(1,1,0)$  calculated in the same way but for the zero time shift  $\Delta T_0 = 0$ . The probability measure (although mathematically not the true probability) that the extracted  $h$ th harmonic curve is a true harmonic of the main one is then quantified by the significance of the surrogate test, i.e., by the relative part of surrogates for which  $\rho_d^{(h)} < \rho_0^{(h)}$ . For example, if we found 1000 surrogate amplitude-phase consistencies  $\rho_{d=1,\dots,1000}^{(h)}$  and 792 of them are smaller than the original value  $\rho_0^{(h)}$ , then the rough probability that the extracted curve represents a true harmonic is 79.2%. We regard a harmonic as true if the probability calculated in this way is  $\geq 95\%$ . By default we use  $N_d = 100$  surrogates and a maximum time shift that equals a quarter length of the signal  $M = N/4$ , so that the surrogates are of  $N - M = 3N/4$  length each.

Note that the significance of the surrogate test does not depend on the magnitude of  $\rho^{(h)}$  (12). Thus, there might be an independent component located at the frequency of a possible harmonic, so that the amplitude-phase consistency would be

high but, because it does not fully adjust its amplitude and phase to that of the fundamental harmonic, the surrogate test will not reject the null hypothesis of independence (e.g., see below, Table I). Thus, the possibility of picking up a spurious component that is nearby in frequency is largely eliminated.

It should be noted that, since we test many harmonic candidates, it is natural to expect that sometimes we will encounter false positives, i.e., the surrogate test will regard as true a harmonic which is actually false. Thus, for some noise realizations the parameters of the TFR around the expected harmonic frequency might indeed appear to be correlated with the parameters of the fundamental harmonic. However, in such cases the extracted harmonic candidate will usually have quite small amplitude-phase consistency  $\rho^{(h)}$  (12). To reduce the probability of false positives, therefore, we introduce some threshold  $\rho_{\min}$  and regard the harmonic as true only if it both passes the surrogate test and at the same time is characterized by  $\rho^{(h)} \geq \rho_{\min}$ . Empirically, we set this threshold as

$$\rho_{\min} = 0.5^{w_A + w_\phi}, \quad (14)$$

where  $w_A, w_\phi$  are the weightings used in (12) (as mentioned, we use  $w_A = w_\phi = 1$ ); the value (14) was found to be quite universal and to work well in most cases. Note that for a true harmonic one can also have  $\rho^{(h)} < \rho_{\min}$ , but this will usually mean that it is badly damaged by noise or other influences and cannot be recovered without large errors.

#### D. Practical issues of harmonics extraction

*Extracting in order.* To improve the accuracy of reconstruction, each harmonic which is identified as true should be subtracted from the signal before extracting and testing the next one. This decreases the errors related to interference between the harmonics and makes all procedures more accurate. The same consideration applies to the first harmonic which, after being found by the methods described in Sec. III A, should be subtracted from the signal before extraction of any of the other harmonics.

*How many harmonics to extract?* Clearly, the maximum number of harmonics one can extract, in principle, is  $h_{\max} = (f_s/2)/\langle v^{(1)}(t)/2\pi \rangle$ , where  $f_s$  is the sampling frequency of the signal (so that  $f_s/2$  is the maximally achievable, Nyquist frequency) and  $v^{(1)}(t)$  denotes the extracted instantaneous frequency of the first harmonic. However, checking all harmonics might be computationally very expensive and is often not needed. In practice, it is better to stop the search after some chosen number of sequential harmonics has been identified as false, making it likely that there are no more true harmonics. We choose this number to be  $S = 3$  by default.

*What if the extracted component is not the first harmonic?* Although intuitively the first harmonic should have the highest amplitude (and will therefore be extracted as the dominant curve), for some complicated wave forms (e.g., with sharp peaks) this might be untrue. Therefore, before starting extraction of the harmonics, one should first ensure that one is starting with the first harmonic and, if not, then find it and switch to it. To do this, we apply the same procedure described for harmonic extraction and identification, but in the reverse direction, i.e., using  $h = 1/2, 1/3, 1/4, \dots$ . Then if some of these are identified as true, we switch to the one with the

TABLE I. Summary of the results of NMD applied to the signal shown in Fig. 9. For each NM the significance of the surrogate test against noise (see Sec. III E) was based on 100 surrogates. The columns left to right provide information for the  $h$ th harmonic: the value of the consistency  $\rho^{(h)}(1, 1, 0)$  (12); the significance of the test against independence (see Sec. III C); the method chosen for reconstruction of the harmonic (see Sec. V C); the resolution parameter  $f_0^{(h)}$  adapted for the harmonic considered (see Sec. IV A); the true amplitude ratio  $a_h \equiv A^{(h)}(t)/A^{(1)}(t)$ ; the extracted amplitude ratio; the true phase shift  $\varphi_h \equiv \phi^{(h)}(t) - h\phi^{(1)}(t)$ ; the extracted phase shift. Harmonic was identified as true if both  $\rho^{(h)} \geq \rho_{\min} = 0.25$  (14) and the significance level is  $\geq 95\%$ , with the corresponding rows being shaded with gray. The numbers of real harmonics are in bold type, and it can be seen that only they were identified as being true. For simplicity, the results for  $h = 1/n$  harmonics (which are tested to check whether the extracted component is the first harmonic) are not shown; in each case all three consecutive  $h = 1/2, 1/3, 1/4$  harmonics were identified as false.

Harmonic No. $h$	$\rho^{(h)}$	Significance level (%)	Reconstruction method	$f_0^{(h)}$	Amplitude ratio $a_h$		Phase shift $\varphi_h/\pi$	
					True	Extracted	True	Extracted
First NM (significance against noise = 100%)								
<b>1</b>			Ridge	1.00	1	1	0	0
2	0.41	50	Ridge	2.85				
<b>3</b>	0.75	100	Ridge	2.08	0.75	0.74	-0.20	-0.20
4	0.29	69	Ridge	1.77				
<b>5</b>	0.63	98	Ridge	2.01	0.50	0.51	0.25	0.26
6	0.11	52	Ridge	1.77				
7	0.04	67	Direct	1.21				
8	0.06	67	Ridge	0.41				
Second NM (significance against noise = 100%)								
<b>1</b>			Ridge	1.00	1	1	0	0
<b>2</b>	0.68	98	Ridge	2.85	0.50	0.45	0.50	0.51
<b>3</b>	0.40	95	Ridge	1.71	0.25	0.29	-0.33	-0.33
4	0.09	79	Ridge	1.95				
5	0.11	91	Ridge	0.64				
6	0.07	17	Ridge	1.00				
Residual (significance against noise = 37%)								

smallest frequency and start from there. The minimal  $h$  one can go for can be set as  $h_{\min} = (1/T)/\langle v^{(1)}(t)/2\pi \rangle$ , although the statistics will be bad already for  $h < 5h_{\min}$ , when the related oscillation has less than five cycles over the whole signal time duration [12]. Nevertheless, it is better to stop after  $S_0 = 3$  consecutive  $1/n$  harmonics have been identified as false, in the same manner as is done for the usual harmonics.

### E. Criterion for stopping the decomposition

Once all harmonics are identified and reconstructed, they are summed up into the NM, which is then subtracted from the signal and the procedure is repeated on the residual. The natural question arises, therefore, of how many NMs to extract, i.e., when to stop the decomposition. Obviously, decomposition of any noise, e.g., white, Brownian, any other correlated or not, does not make much sense, so the reasonable goal is to extract all oscillatory components present in the signal and leave the noise and trends as the residual. Therefore, after extraction of each NM one needs to decide whether what is left contains any more meaningful oscillations, in which case one continues the decomposition, or whether it just represents noise, in which case one should stop. The problem is thus reduced to distinguishing deterministic from random dynamics.

To solve it, one can use the surrogate test against the null hypothesis of linear noise [27,28], which includes white and colored noises (e.g., Brownian). The surrogates

for this task, called Fourier transform (FT) surrogates, are constructed by taking the inverse FT of the signal's FT with randomized phases of the Fourier coefficients:  $s_s(t) = (2\pi)^{-1} \int [\hat{s}(\xi) e^{i\varphi_s(\xi)}] e^{i\xi t} d\xi$ , where  $\varphi_s(\xi) = -\varphi_s(-\xi)$  denote the phases taken at random uniformly on  $[0, 2\pi]$  for each frequency  $\xi > 0$ . The reason for this is that any linear noise [an Autoregressive Moving Average or ARMA process  $x(t_n) = a_0 + b_0\eta_W(t_n) + \sum_{p=1}^P a_p x(t_{n-p}) + \sum_{m=1}^M b_m \eta_W(t_{n-p})$ , where  $\eta_W(t_n)$  denotes Gaussian white noise of unit variance] is characterized only by the amplitudes of the Fourier coefficients. Randomization of the Fourier phases preserves the power spectrum, so that the surrogate time series will represent another realization of the same random process if the original time series is noise, thus being consistent with the tested null hypothesis. On the other hand, if some meaningful NMs are present in the signal, the randomization of the phases will destroy the particular phase relationships responsible for the amplitude and frequency modulations, making their behavior less deterministic [35].

One now needs to select the discriminating statistics, which is calculated for the original signal and the surrogates, so that the null hypothesis of linear noise is accepted if the original value lies within the surrogate values and rejected otherwise. The commonly used statistics involve one of the generalized Renyi dimensions [37–39], with the correlation dimension calculated by Grassberger-Procaccia approach [40–42] remaining the most popular choice. However, we find that the surrogate test based on such measures is extremely

sensitive to noise, prohibiting their use in the NMD which is intended to be noise robust.

Therefore, we need to devise another discriminating statistics. There are virtually no restrictions on its choice [43], so that almost any measure can be used. The only question to be asked is how powerful it is, i.e., how good in distinguishing the deterministic dynamics from noise. Given that NMD is based on the WFT/WT, it seems reasonable to select statistics based on the properties of the TFR obtained. Namely, since at the first step of the NMD process we extract the component from the signal's TFR (see Sec. III A), we can base our statistics on the properties of the components extracted (in the same way) from the original signal and from its FT surrogates. Thus, if the original component is true (and not just formed from noise peaks picked in the time-frequency plane), then it is expected to have more deterministic amplitude and frequency modulation than the surrogate components, which should be more stochastic; otherwise, there will be no difference.

The degree of order in the extracted amplitude  $A(t)$  or frequency  $\nu(t)$  can be quantified by their spectral entropies  $Q[\hat{A}(\xi)]$ ,  $Q[\hat{\nu}(\xi)]$ , so that the discriminating statistics  $D$  for the surrogate test can be taken as their combination,

$$D(\alpha_A, \alpha_\nu) \equiv \alpha_A Q[\hat{A}(\xi)] + \alpha_\nu Q[\hat{\nu}(\xi)], \quad (15)$$

$$Q[f(x)] \equiv - \int \frac{|f(x)|^2}{\int |f(x)|^2 dx} \ln \frac{|f(x)|^2}{\int |f(x)|^2 dx} dx.$$

Note that, in practice, due to the finite sampling frequency  $f_s$  and sample length  $N$  of real signals, the integrals over frequency  $\xi$  in  $Q[\hat{A}(\xi)]$ ,  $Q[\hat{\nu}(\xi)]$  (15) are discretized into sums over the discrete FT frequencies  $\xi_n = (n/N - 1/2)f_s$ ,  $n = 1, \dots, N$ .

In the present context, we have found the statistics  $D(\alpha_A, \alpha_\nu)$  (15) to be more meaningful and much more powerful than other choices (e.g., the popular correlation dimension [40–42]). This statistics is directly related to the quality of the representation of component in the signal's TFR, so that the significance of the surrogate test based on it reflects the proportion of the “deterministic” part in the extracted amplitude and frequency dynamics. Thus, if the residual does not pass the surrogate test (null hypothesis is not rejected), this might mean either that the residual is indeed noise, or that the component simply cannot be reliably extracted from the TFR (e.g., because resolution characteristics of the TFR are not appropriate to reliably represent the related amplitude and frequency modulations and/or to segregate the component from noise).

The power of  $D(\alpha_A, \alpha_\nu)$ , i.e., its ability to distinguish deterministic from random dynamics, depends strongly on the complexity of the component's amplitude and frequency modulations: The lower the original spectral entropies  $Q[\hat{A}(\xi)]$ ,  $Q[\hat{\nu}(\xi)]$  are, the more powerful is the test. However, even in the (quite unrealistic) case when the signal contains a meaningful component without any amplitude or frequency modulation, i.e., a pure tone  $A \cos \nu t$ , due to numerical issues [44,45] the surrogate test will still be quite powerful in rejecting the null hypothesis (unless this tone has an integer number of cycles over the time-length of the signal). The power of the test is also inversely proportional to the spread

of  $\hat{A}(\xi)$ ,  $\hat{\nu}(\xi)$ : The more concentrated they are, the narrower is the frequency band that the component  $A(t) \cos \phi(t)$  occupies, so that the less is the noise power that is contained in it.

As to the choice of  $\alpha_A, \alpha_\nu$  in (15), we have found that the powers of  $D(1,0) = Q[\hat{A}(\xi)]$  and  $D(0,1) = Q[\hat{\nu}(\xi)]$  are often inversely proportional to the strengths of the amplitude and frequency modulation, respectively. Thus,  $D(1,0)$  is preferable for components with relatively small amplitude variability and considerable frequency variability, while  $D(0,1)$  is better otherwise. Although the procedure is not fully rigorous mathematically, we therefore perform three tests, using  $D(1,0)$ ,  $D(0,1)$ , and  $D(1,1)$ , and then select the significance as the maximum among them. It remains to be established, however, whether some better statistics not having the drawback mentioned can be found.

Summarizing, we extract the components from the TFR of the original signal and compute the corresponding  $D_0(\alpha_A, \alpha_\nu)$  (15); then we create  $N_s$  FT surrogates of the signal, for each of them calculate the corresponding TFR, extract the component from it, and compute the respective  $D_{s=1, \dots, N_s}(\alpha_A, \alpha_\nu)$ . We use  $N_s = 40$  surrogates and set the significance level to 95%, rejecting the tested null hypothesis of noise if the number of surrogates with  $D_s > D_0$  is equal or higher than  $0.95 \times 40 = 38$ . The test is performed for three different  $(\alpha_A, \alpha_\nu)$  in (15), using  $D(1,1)$ ,  $D(0,1)$ , and  $D(1,0)$  as a discriminating statistic; if at least for one of them the null hypothesis is rejected, we regard the signal as not noise and continue the decomposition.

Note that, at the very start of each NMD iteration, the signal should be detrended, which we do by subtracting a third order polynomial fit from it. This is especially important for the surrogate test, as trends might greatly affect its performance. In particular, the FT surrogates should be constructed using an already detrended signal as a base, because otherwise one might end up with testing the trend against noise rather than the oscillatory component of interest. Additionally, to guarantee that the boundary distortions are of the same nature in the original TFR and those for the surrogates, it is recommended to use padding with zeros [12]; the original  $D_0$  should thus be recalculated using parameters of the component extracted from the TFR of the zero-padded signal. This issue, however, concerns only the surrogate test, while in all other cases we use the more accurate (but at the same time more computationally expensive) predictive padding [12].

#### IV. ADDITIONAL IMPROVEMENTS

The NMD as described in the previous section already represents a very powerful decomposition tool. However, it may be made even better with the help of the upgrades outlined below, although at the expense of greatly increased computational cost for some of them [though it still remains  $O(N \ln N)$ ].

##### A. Adaptive representation of the harmonics

Even if the first harmonic of some NM is well represented in the TFR and can be accurately extracted and reconstructed, it does not mean that the same applies to all the other harmonics, too. For example, harmonics of an NM with only amplitude



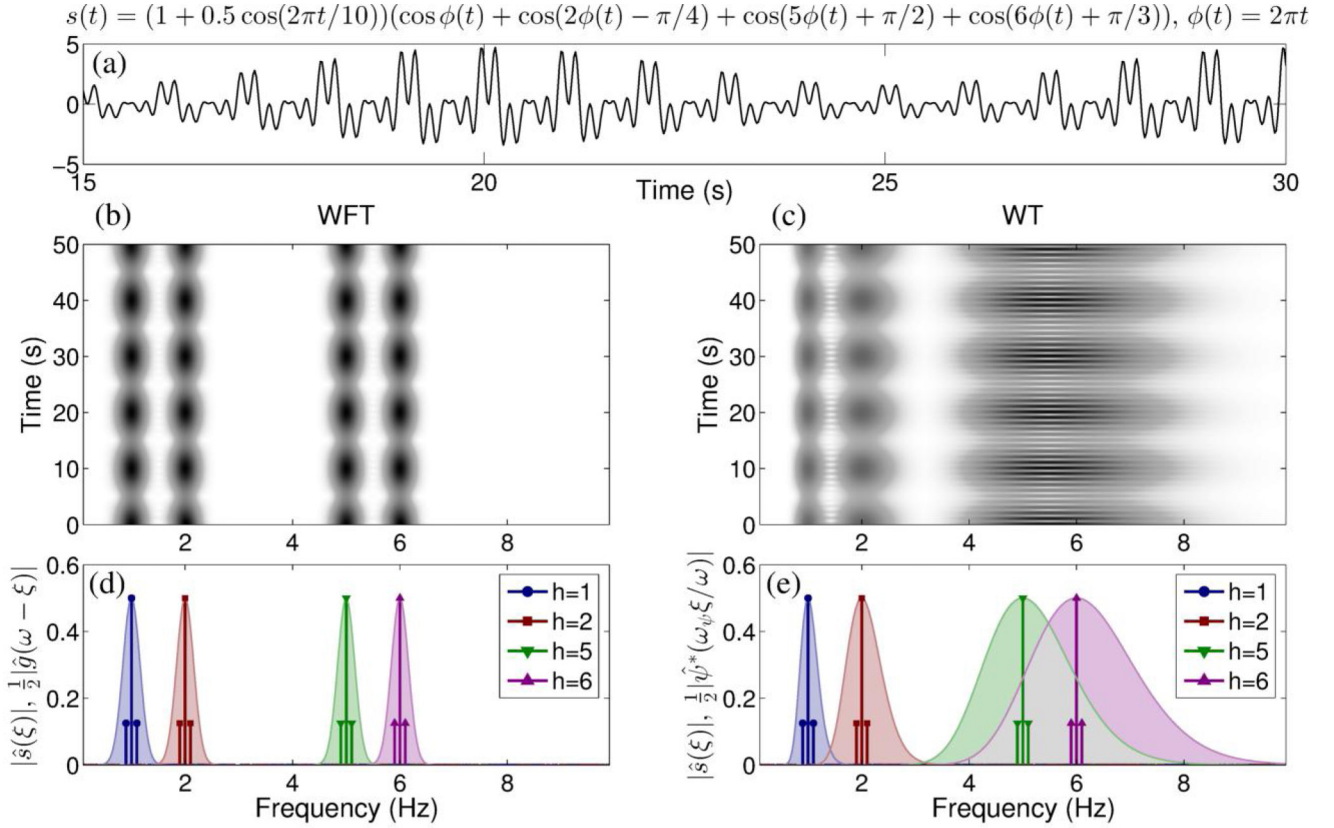


FIG. 3. (Color online) (a) Nonlinear mode with amplitude modulation, as specified by the equation for  $s(t)$  above the box. (b),(c) The corresponding WFT and WT amplitudes, respectively. (d),(e) The signal's FT, integrated over one frequency bin (vertical lines), with the parts responsible for different harmonics shown using different colors and markers; the shaded areas show the absolute values of the window functions  $\hat{g}(\omega - \xi)$  or wavelet functions  $\hat{\psi}^*(\omega_\psi \xi / \omega)$  centered at the mean frequencies of the harmonics  $\omega = 2\pi h$  and rescaled to half of the harmonics' mean amplitudes. The signal was sampled at 50 Hz for 50 s.

modulation require the same time and frequency resolution, so that the WFT can represent them all well, while for the WT, where time and frequency resolution scale with frequency, one will need to adjust the resolution parameter  $f_0$  for each harmonic. Thus, consider an NM with simple sinusoidal amplitude modulation:

$$\begin{aligned}
 s(t) &= [1 + r_a \cos(v_a t + \varphi_a)] \sum_{h=1}^{\infty} a_h \cos(hv t + \varphi_h) \\
 &= \sum_{h=1}^{\infty} a_h \left\{ \cos(hv t + \varphi_h) + \frac{r_a}{2} \cos[(hv + v_a)t + (\varphi_h + \varphi_a)] + \frac{r_a}{2} \cos[(hv - v_a)t + (\varphi_h - \varphi_a)] \right\} \\
 &\equiv \sum_{h=1}^{\infty} x^{(h)}(t). \tag{16}
 \end{aligned}$$

Clearly, all harmonics have the same amplitude variability (in relative terms), so that the TFR time resolution should also be the same for them. This is reflected in the fact that each harmonic  $x^{(h)}(t)$  is composed of three Fourier terms, which all have identical amplitude ratios, frequency differences, and phase relationships for each  $h$ . Furthermore, the frequency distance between two consecutive harmonics remains the

same, meaning that the frequency resolution also should not be changed for harmonics.

Therefore, the WFT, having constant time and frequency resolution, will be a perfect match for this case. This means, first, that if the amplitude modulation of the first harmonic is represented reliably in the WFT, then the same will also apply to all other harmonics and, second, that if two first harmonics do not interfere in the WFT, then any two harmonics will also be well-separated. For the WT, the former is also true, as the time resolution, i.e., the ability to reflect changes in time, increases with frequency for this type of TFR. However, the frequency resolution of the WT progressively worsens with the increase of frequency, so that the higher are the harmonics, the harder it is to resolve them.

This is illustrated in Fig. 3, where from panels (b) and (c) it is clear that all harmonics can be well represented in the WFT, but for the WT higher harmonics begin to interfere. This issue is explained schematically in Figs. 3(d) and 3(e). Thus, both the WFT and the WT can be seen as convolutions in the frequency domain of the signal with a window  $\hat{g}(\omega - \xi)$  and wavelet  $\hat{\psi}^*(\omega_\psi \xi / \omega)$ , as seen from (6). Figures 3(d) and 3(e) show the signal's discrete FT  $\hat{s}(\xi)$  together with the (rescaled and centered at the mean frequencies of the harmonics  $\omega/2\pi = 1, 2, 5, 6$  Hz) window and wavelet FTs  $\hat{g}(\omega - \xi)$  and  $\hat{\psi}^*(\omega_\psi \xi / \omega)$  (7), with which  $\hat{s}(\xi)$  is convoluted while constructing the WFT and WT, respectively.

Roughly speaking, the quality of the representation of time variability (the amplitude modulation in the present case) for each harmonic can be estimated based from the proportion of the corresponding Fourier coefficients lying below the shown window or wavelet FTs. For both the WFT and the WT, each harmonic has three Fourier coefficients (16), all of which lie in the correspondingly shaded areas, meaning that the time resolution in each case is sufficient to represent amplitude modulation appropriately. The degree of interference between the harmonics in the WFT and WT can be estimated from the area of overlap between the corresponding window/wavelet FTs. For the WFT [Fig. 3(d)] the interference between harmonics does not depend on their frequency, while for the WT [Fig. 3(e)] it increases for the higher harmonics, which results in the fifth and sixth harmonics being represented as a single component [Fig. 3(c)]. This is because the minimal frequency difference between two harmonics is equal to the frequency of the first harmonic, thus being defined on a linear scale (which is natural for the WFT), but not on a logarithmic one (natural for the WT).

The situation changes when there is frequency modulation. In this case neither the WFT nor the WT provide optimal representation of the harmonics; in addition, some harmonics cannot, in principle, be reliably resolved with time-frequency analysis methods. Thus, consider an NM with simple sinusoidal frequency modulation,

$$\begin{aligned} s(t) &= \sum_{h=1}^{\infty} a_h \cos[h\phi(t) + \varphi_h] \equiv \sum_{h=1}^{\infty} x^{(h)}(t), \\ \phi(t) &\equiv vt + r_b \sin(v_b t + \varphi_b), \\ \Rightarrow x^{(h)}(t) &= a_h \cos[hvt + hr_b \sin(v_b t + \varphi_b) + \varphi_h] \quad (17) \\ &= a_h \operatorname{Re} \left[ \sum_{n=-\infty}^{\infty} J_n(hr_b) e^{i(\varphi_h + n\varphi_b)} e^{i(hv + nv_b)t} \right], \end{aligned}$$

where we have used the formula  $e^{ia \sin \phi} = \sum_{n=-\infty}^{\infty} J_n(a) e^{in\phi}$ , with  $J_n(a) = (-1)^n J_{-n}(a)$  denoting a Bessel function of the first kind. Since after some  $|n|$  all terms in (17) become negligible, one can, in practice, restrict the summation to  $|n| \leq n_J(ha)$ , with the maximum non-negligible order  $n_J$  being determined as

$$n_J(a) : \frac{\sum_{|n| > n_J(a)} |J_n(a)|^2}{\sum_{n=-\infty}^{\infty} |J_n(a)|^2} < \epsilon_p, \quad (18)$$

where  $\epsilon_p$  denote a chosen accuracy threshold. The value of  $n_J(a)$  increases with  $a$ , being (for  $\epsilon_p = 0.001$ ):  $n_J(0 \lesssim a \lesssim 0.3) = 2$ ,  $n_J(0.3 \lesssim a \lesssim 0.65) = 3$ ,  $n_J(0.65 \lesssim a \lesssim 1.13) = 4$ , etc. As a result, the higher the harmonic, the larger the frequency range it occupies, i.e., the larger is the number of non-negligible terms in (17).

Consequently, to reflect reliably the frequency modulation of the higher harmonics, one needs higher time resolution for them, a requirement that is satisfied by the WT, but not by the WFT. This issue is illustrated in Fig. 4, where it can be seen that the WT can represent reliably both the first and the third harmonics, while the WFT reflects appropriately only the first one. However, the increased time-resolution of the WT is provided at the expense of decreased frequency resolution, leading to stronger interference between harmonics, as seen

from the previous case (Fig. 3). Figure 4 also shows that in some cases it might, in principle, be impossible to represent reliably two harmonics in the TFR. Thus, as can be seen from Figs. 4(d) and 4(e), the FTs  $\hat{x}^{(h)}(\xi)$  of the sixth and seventh harmonics are ‘‘entangled,’’ i.e., the frequency regions in which they are contained overlap. Therefore, unless specifically designed for this case, any window or wavelet function which picks the Fourier coefficients of one harmonic will inevitably also pick those corresponding to the other one, too [46].

Summarizing, for the general case where both amplitude and frequency modulation are present in the NM, accurate representation of higher harmonics requires higher time resolution, but the same frequency resolution. However, an increase in time resolution inevitably leads to a decrease of frequency resolution, so usually one needs to search for some compromise. Due to this, it is often the case that neither the WFT nor the WT with a constant resolution parameter  $f_0$  in (7) can represent all harmonics reliably.

To tackle this problem, one can adaptively adjust  $f_0$  for each harmonic individually. Assuming that the extracted first harmonic is reconstructed well, the quality of the representation of the  $h$ th harmonic can be assessed through its consistency  $\rho^{(h)}$  (12) with the first one. Therefore, the optimal resolution parameter for the  $h$ th harmonic,  $f_0^{(h)}$ , can be selected as that for which  $\rho^{(h)}$  (12) achieves its maximum and at the same time the harmonic passes the surrogate test, i.e., is identified as a true harmonic.

The remaining question is the region within which to search for  $f_0^{(h)}$ . To find it, we first need to understand in what frequency band  $[\omega_f^{(h)} - \Delta\omega_f^{(h)}/2, \omega_f^{(h)} + \Delta\omega_f^{(h)}/2]$  the  $h$ th harmonic FT  $\hat{x}^{(h)}(\xi)$  is concentrated, given the frequency band of the first harmonic. We define  $\omega_f^{(h)}$  and  $\Delta\omega_f^{(h)}$  as

$$\begin{aligned} \omega_f^{(h)}, \Delta\omega_f^{(h)} : \int_0^{\omega_f^{(h)} - \Delta\omega_f^{(h)}/2} |\hat{x}^{(h)}(\xi)|^2 d\xi &= \frac{\epsilon_p}{2} \frac{E_{\text{tot}}^{(h)}}{2}, \\ \int_{\omega_f^{(h)} + \Delta\omega_f^{(h)}/2}^{\infty} |\hat{x}^{(h)}(\xi)|^2 d\xi &= \frac{\epsilon_p}{2} \frac{E_{\text{tot}}^{(h)}}{2}, \end{aligned} \quad (19)$$

where  $E_{\text{tot}}^{(h)} \equiv \int |\hat{x}^{(h)}(\xi)|^2 d\xi = \int |x^{(h)}(t)|^2 dt$  is the total energy of the harmonic and  $\epsilon_p$  denotes the chosen accuracy threshold [note that, as everywhere else, we assume that the analytic approximation (3) holds for  $x^{(h)}(t) = A^{(h)}(t) \cos\phi^{(h)}(t)$ ]. Then it can be shown that, given  $\omega_f^{(1)}, \Delta\omega_f^{(1)}$ , the frequency range for the  $h$ th harmonic will be  $[\omega_f^{(h)} - \Delta\omega_f^{(h)}/2, \omega_f^{(h)} + \Delta\omega_f^{(h)}/2]$ , with

$$\omega_f^{(h)} = h\omega_f^{(1)}, \quad \Delta\omega_f^{(1)} \lesssim \Delta\omega_f^{(h)} \lesssim h\Delta\omega_f^{(1)}. \quad (20)$$

It is clear from the preceding discussion that, for the simple AM mode (16), one has  $\omega_f^{(h)} = h\omega_f^{(1)}$ ,  $\Delta\omega_f^{(h)} = \Delta\omega_f^{(1)}$ . The same is true for any AM mode, as can be shown by representing the amplitude modulation through the Fourier series  $A(t) = \sum_n r_n \cos(vnt + \varphi_n)$  and expanding each harmonic in the same way as in (16). For the FM mode (17), the expression (20) follows from the fact that, at least for small enough  $\epsilon_p$  (e.g., 0.001) in the definition (18), one has

$$n_J(a) \lesssim n_J(ha) \lesssim hn_J(a), \quad (21)$$

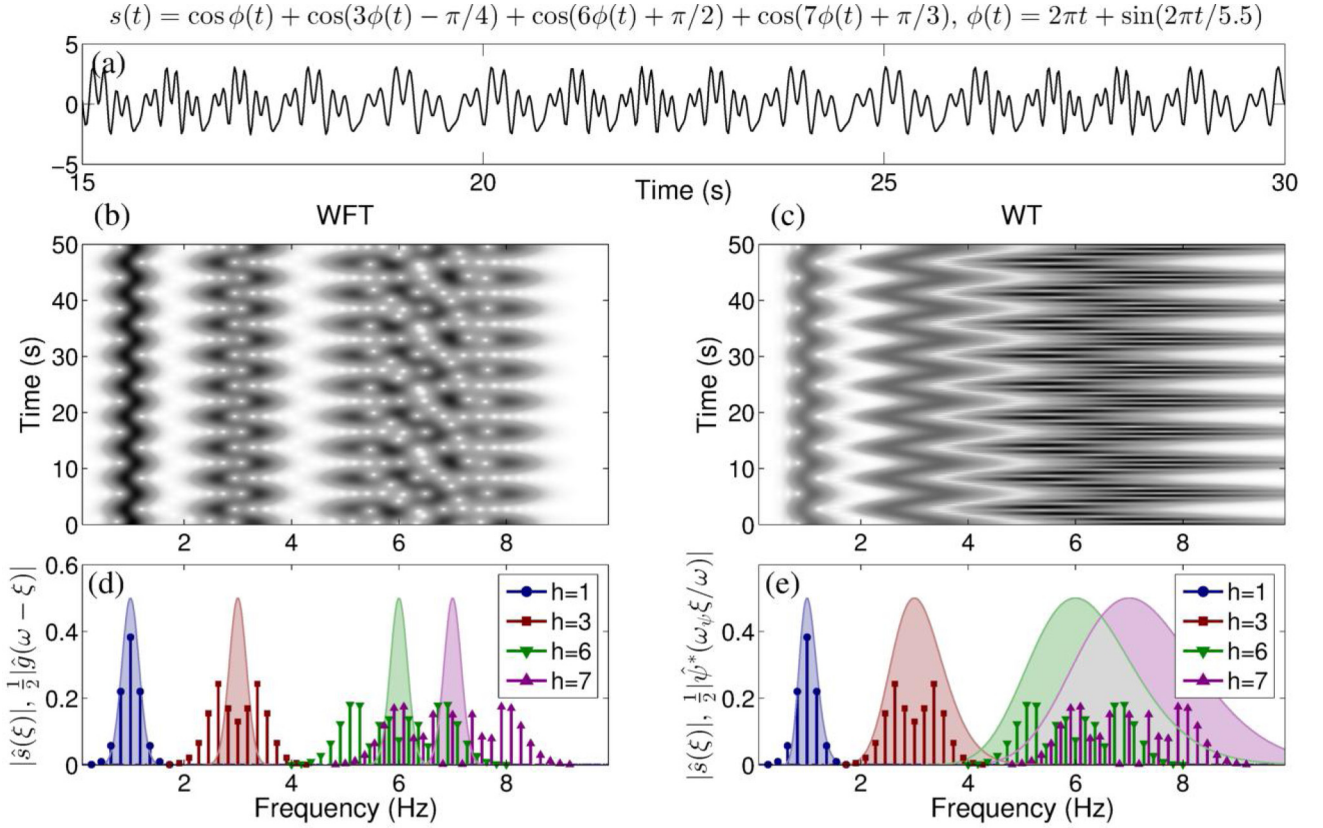


FIG. 4. (Color online) (a) Nonlinear mode with frequency modulation, as specified by the equation for  $s(t)$  above the box. (b),(c) The corresponding WFT and WT amplitudes, respectively. (d),(e) The signal's FT, integrated over one frequency bin (vertical lines), with the parts responsible for different harmonics shown using different colors and markers; the shaded areas show the absolute values of the window functions  $\hat{g}(\omega - \xi)$  or wavelet functions  $\hat{\psi}^*(\omega_\psi \xi / \omega)$  centered at the mean frequencies of the harmonics  $\omega = 2\pi h$  and rescaled to half of the harmonics' mean amplitudes. The signal was sampled at 50 Hz for 55 s.

for most  $a$ , as can be confirmed numerically. Then, taking into account (21) in the expansion (17), one obtains (20). For the  $h$ th harmonic of the NM with a more complicated double-sinusoidal frequency modulation, one has

$$\begin{aligned}
 x^{(h)}(t) &= A \cos[h\nu t + \varphi + hr_1 \sin(\nu_1 t + \varphi_1) \\
 &\quad + hr_2 \sin(\nu_2 t + \varphi_2)] \\
 &= A \operatorname{Re} \left[ e^{i(h\nu t + \varphi)} \sum_{n=-\infty}^{\infty} J_n(hr_1) e^{i(n\nu_1 t + n\varphi_1)} \right. \\
 &\quad \left. \times \sum_{m=-\infty}^{\infty} J_m(hr_2) e^{i(m\nu_2 t + m\varphi_2)} \right] \\
 &= \sum_{n=-\infty}^{\infty} A J_n(hr_1) \operatorname{Re} \left[ e^{i(h\nu t + n\nu_1 t + \varphi + n\varphi_1)} \right. \\
 &\quad \left. \times \sum_{m=-\infty}^{\infty} J_m(hr_2) e^{i(m\nu_2 t + m\varphi_2)} \right].
 \end{aligned} \tag{22}$$

Clearly, it can be viewed as a sum of components with single-sinusoidal frequency modulation, but amplitudes  $A J_n(hr_1)$  and frequencies  $(h\nu + n\nu_1)$ . The frequency range of each of these components, and the number of them with a non-negligible amplitudes, both scale with harmonic number  $h$  according

to (20) and (21). Therefore, the frequency range of the whole  $x^{(h)}(t)$  (22) satisfies (20). Since frequency modulation can always be expanded in a Fourier series, using the same trick as with (22), it can be shown by induction that (20) holds for any FM mode. The generalization to the case of any joint amplitude and frequency modulation is then straightforward.

Based on (20) and the scaling properties of the WFT/WT, one can now determine the region within which the optimal resolution parameter  $f_0^{(h)}$  for the  $h$ th harmonic lies. Given that the first harmonic was extracted from the TFR calculated with resolution parameter  $f_0^{(1)}$ , and assuming that the corresponding time and frequency resolutions are appropriate for this first harmonic, one has

$$\text{WFT: } f_0^{(h)} \in [f_0^{(1)}/h, f_0^{(1)}]; \quad \text{WT: } f_0^{(h)} \in [f_0^{(1)}, hf_0^{(1)}]. \tag{23}$$

The optimal  $f_0^{(h)}$  is searched for by first breaking the region (23) into  $N_r$  values  $f_{r=1, \dots, N_r}^{(h)}$  (we use  $N_r = 10$ ). For each of them, one calculates the TFR, extracts the harmonic from it (see Sec. III B), estimates the corresponding consistency  $\rho_r^{(h)}$  (12), and tests the harmonic for being true (see Sec. III C). Among the  $f_r^{(h)}$  for which the harmonic is identified as true, the one characterized by the highest  $\rho_r^{(h)}$  is selected. It is then used as the initial value for an iterative

golden section search of the optimal  $f_0^{(h)}$  (with default accuracy being  $\epsilon_f = 0.01$ ), maximizing the consistency  $\rho^{(h)}$  (12).

Note that (23) does not take into account the interference between harmonics and with the other components which might lie nearby in frequency, so that in some cases the upper bound on  $f_0^{(h)}$  might be higher than (23); the same consideration applies to the lower bound. Therefore, if near the minimum or maximum of the tested  $f_0^{(h)}$  the consistency  $\rho^{(h)}$  is found to grow, we continue the search in that direction until the peak appears. We perform an identical procedure when we check whether the extracted component represents a first or a higher harmonic by extracting and testing its  $h = 1/2, 1/3, \dots$  harmonic candidates (see Sec. III D). The formulas (20) and (23) remain valid in this case, but the upper and lower bounds for  $\Delta\omega_f^{(h)}$  in (20) and  $f_0^{(h)}$  in (23) change places.

$$\begin{aligned}\tilde{A}^{(h)}(t) &= \langle A^{(h)}(t) \rangle \frac{\sum_{h'} A^{(h')}(t)}{\sum_{h'} \langle A^{(h')}(t) \rangle}, \\ \tilde{\phi}^{(h)}(t) &= \arg \left( \sum_{h'} \min(1, h'/h) \langle A^{(h')}(t) \rangle \exp \left[ i \frac{h\phi^{(h')}(t) - \Delta\phi_{h',h}}{h'} \right] \exp \left\{ -\frac{2\pi i}{h'} I \left[ \frac{h\phi^{(h')}(t) - h'\phi^{(h)}(t) - \Delta\phi_{h',h}}{2\pi} \right] \right\} \right), \\ \tilde{\nu}^{(h)}(t) &= \frac{\sum_{h'} \min(1, h'/h) \langle A^{(h')}(t) \rangle h\nu^{(h')}/h'}{\sum_{h'} \min(1, h'/h) \langle A^{(h')}(t) \rangle},\end{aligned}\quad (24)$$

where  $\Delta\phi_{h',h} \equiv \arg [e^{i[h\phi^{(h')}(t) - h'\phi^{(h)}(t)]] \in [-\pi, \pi]$ , and  $I[\dots]$  denotes rounding to the nearest integer, so that  $I[0.8] = 1$ ,  $I[-0.6] = -1$  (the corresponding term is needed to eliminate possible noise-induced phase slips, i.e., the rapid growth by  $2\pi$  in the phase differences). Note also the multiplier  $\min(1, h'/h)$ , appearing for phase and frequency refinement in (24). It is needed to account for the scaling of phase and frequency errors of lower harmonics when they are mapped to higher ones. Thus, if  $\nu^{(1)}(t)$  has an error  $\epsilon(t)$ , then  $h\nu^{(1)}(t)$  will have error  $h\epsilon(t)$ .

The refinement (24) not only makes the reconstructed NM much more accurate, solving the problem of picking up the cumulative noise of all the harmonics, but also reduces the noise level in each harmonic separately. Thus, noise-induced variations in the parameters of different harmonics are expected to be mutually independent, and so they average out being added together. As a result, the more harmonics are contained in the NM, the more accurately it can be reconstructed. While NMD is generally noise robust, due to being based on time-frequency methods for which only the spectral power of noise in the vicinity of components' frequencies matters, NM reconstruction by (24) raises this robustness to an extreme extent.

### C. Speed improvement

To extract a particular harmonic candidate, one does not need to calculate the TFR over the whole frequency range, which would be computationally expensive. Rather, one can calculate it only in the range where the possible harmonic is expected to lie. As discussed previously, given the frequency

### B. Improved reconstruction of nonlinear modes

Given the reconstructed amplitudes  $A^{(h)}(t)$ , phases  $\phi^{(h)}(t)$ , and frequencies  $\nu^{(h)}(t)$  of all the true harmonics, the most straightforward way to reconstruct the full NM is just to add all  $A^{(h)} \cos \phi^{(h)}(t)$  together. However, in this way the NM picks up noise contributions from all harmonics, which can make it quite inaccurate.

Fortunately, there is a more clever way to perform the reconstruction, also yielding more accurate parameters for the individual harmonics. Thus, one can utilize the theoretical amplitude, phase, and frequency relationships between the harmonics, i.e.,  $A^{(h)} = a_h A^{(1)}(t)$ ,  $\phi^{(h)} - h\phi^{(1)} = \varphi_h$ , and  $\nu^{(h)}(t) = h\nu^{(1)}(t)$ . Then, because the components with the higher amplitudes are expected to be less noise corrupted, one can refine the parameters of each harmonic by weighted averaging over the parameters of all harmonics:

range  $[\omega_{\min}^{(1)}, \omega_{\max}^{(1)}] \equiv [\bar{\omega}_c^{(1)} - \Delta\omega_c^{(1)}, \bar{\omega}_c^{(1)} + \Delta\omega_c^{(1)}]$ , where the first harmonic is concentrated in TFR, it is clear from (20) that the appropriate frequency range for the  $h$ th harmonic will be characterized by  $\omega_{\min, \max}^{(h)} = h\bar{\omega}_c^{(1)} \mp \max(1, h)\Delta\omega_c^{(1)}$ . The multiplier  $\max(1, h)$  takes into account that  $h$  can be smaller than unity, since we use  $h = 1/2, 1/3, \dots$  when checking whether the extracted component is first (or higher) harmonic (see Sec. III D); the frequency range should not be squeezed for  $h < 1$  because, e.g., for NM with only amplitude modulation all harmonics will have the same bandwidth (see Sec. IV A).

Given the extracted time-frequency support of the first harmonic  $[\omega_-(t), \omega_+(t)]$ , the required frequency range for it to be fully contained in the TFR can be estimated as  $[\omega_{\min}^{(1)}, \omega_{\max}^{(1)}] = [\min \omega_-(t), \max \omega_+(t)]$ . However, the maximum value of the  $\omega_+(t)$  might be excessively large (it is, in theory, infinite, e.g., if the signal is represented by a single tone); the same applies to minimum  $\omega_-(t)$ . Therefore, for determination of the harmonic frequency range it is better to use a support  $[\tilde{\omega}_-(t), \tilde{\omega}_+(t)]$  that is narrower in frequency, determined as the maximum region where the TFR amplitude of the first harmonic is non-negligible. Mathematically, it can be defined as

$$\begin{aligned}\tilde{\omega}_-(t) &: |H_s\{\omega \in [\omega_-(t), \tilde{\omega}_-(t)], t\}| < \epsilon_s |H_s(\omega_p(t), t)|, \\ \tilde{\omega}_+(t) &: |H_s\{\omega \in [\tilde{\omega}_+(t), \omega_+(t)], t\}| < \epsilon_s |H_s(\omega_p(t), t)|,\end{aligned}\quad (25)$$

where  $\omega_p(t) \in [\omega_-(t), \omega_+(t)]$  is the position of the peak, as always. If  $\tilde{\omega}_-(t)$  (25) does not exist, e.g., if the  $|H_s(\omega_-(t), t)| \geq \epsilon_s |H_s(\omega_p(t), t)|$ , we assign  $\tilde{\omega}_-(t) = \omega_-(t)$ ; we do the same for  $\tilde{\omega}_+(t)$  in a similar case. In this work we set  $\epsilon_s = 0.001$ .

Using (25), the frequency range for the  $h$ th harmonic can be estimated as

$$\begin{aligned}\omega_{\min, \max}^{(h)} &= h\langle\omega_p^{(1)}(t)\rangle + \max(1, h)[\omega_{\min, \max}^{(1)} - \langle\omega_p^{(1)}(t)\rangle], \\ \omega_{\min}^{(1)} &= \min \left\{ \text{perc}_{0.05}[\tilde{\omega}_-(t)], \min[\omega_p(t)] \right\}, \\ \omega_{\max}^{(1)} &= \max \left\{ \text{perc}_{0.95}[\tilde{\omega}_+(t)], \max[\omega_p(t)] \right\},\end{aligned}\quad (26)$$

where  $\text{perc}_p$  denotes the  $p$ th percentage largest value of the argument (with  $\text{perc}_0$  and  $\text{perc}_1$  corresponding to usual minimum and maximum, respectively). The 95% largest value of  $\tilde{\omega}_+(t)$  instead of the overall maximum is taken in (26) to prevent selection of an excessively wide region, because, due to noise,  $\tilde{\omega}_+(t)$  might sometimes be too large; the same applies to  $\tilde{\omega}_-(t)$ . In the process of harmonic extraction for  $h > 1$  ( $h < 1$ ) it is also useful to restrict  $\omega_{\min}^{(h)}$  ( $\omega_{\max}^{(h)}$ ) to be higher than the 5% (smaller than the 95%) value of the instantaneous frequency of the last true harmonic being extracted.

The range  $[\omega_{\min}^{(h)}, \omega_{\max}^{(h)}]$  can, however, be more compressed or stretched, depending on the window or wavelet resolution parameter  $f_0^{(h)}$  used for the analysis of the  $h$ th harmonic. The expression (26) is optimal only for

$$\begin{aligned}\text{WFT: } f_0^{(h)} &= h^{-1} f_0^{(1)} \min(1, h), \\ \text{WT: } f_0^{(h)} &= f_0^{(1)} \min(1, h).\end{aligned}\quad (27)$$

Indeed, as discussed previously, if the FT of the first harmonic is contained in the band  $[\omega_1^{(1)}, \omega_2^{(1)}]$ , the largest band where  $h$ th harmonic can be contained is the one within  $\omega_{1,2}^{(h)} = h \frac{\omega_1^{(1)} + \omega_2^{(1)}}{2} \mp \max(1, h) \frac{\omega_2^{(1)} - \omega_1^{(1)}}{2}$ . However, the range  $[\omega_{\min}^{(h)}, \omega_{\max}^{(h)}]$  where harmonic lies in the TFR is larger due to the finite frequency resolution, being  $[\omega_1^{(h)} - \Delta/f_0, \omega_2^{(h)} + \Delta/f_0]$  for the WFT with a Gaussian window (7) and  $[\omega_1^{(h)} e^{-\Delta/f_0}, \omega_2^{(h)} e^{\Delta/f_0}]$  for the WT with a log-normal wavelet (7), where  $\Delta$  is a particular constant. As a result, for  $h f_0^{(h)} > f_0^{(1)} \min(1, h)$  (WFT) or  $f_0^{(h)} > f_0^{(1)} \min(1, h)$  (WT), the range (26) is too large; although one can, in principle, “squeeze” it in this case, due to noise, the appropriate squeezing cannot always be estimated well, so it is better to leave all as it is. On the other hand, for  $h f_0^{(h)} < f_0^{(1)} \min(1, h)$  (WFT), or  $f_0^{(h)} < f_0^{(1)} \min(1, h)$  (WT), one might need a wider range than (26), so it should be stretched. Therefore, if (27) is not the case, the values (26) should be updated as

$$\begin{aligned}\text{WFT:} \\ \omega_{\min, \max}^{(h)} &= \bar{\omega}^{(h)} + [\omega_{\min, \max}^{(h)} - \bar{\omega}^{(h)}] \max \left[ 1, \frac{\min(1, h) f_0^{(1)}}{h f_0^{(h)}} \right]; \\ \text{WT:} \\ \omega_{\min, \max}^{(h)} &= \bar{\omega}^{(h)} (\omega_{\min, \max}^{(h)} / \bar{\omega}^{(h)})^{\max(1, f_0^{(1)} \min(1, h) / f_0^{(h)})},\end{aligned}\quad (28)$$

where  $\bar{\omega}^{(h)} \equiv h\langle\omega_p^{(1)}(t)\rangle$ , and  $\omega_{\min, \max}^{(h)}$  are the estimates given by (26). The expressions (26) and (28) together give the frequency range in which to calculate the WFT/WT [based on the window or wavelet (7) with the resolution parameter  $f_0^{(h)}$ ] for the  $h$ th harmonic.

To reduce the computational cost of the surrogate test against noise, utilized for the stopping criterion (see Sec. III E), one can use the same trick as above, calculating TFRs for surrogates in the frequency range where the original component resides. After extracting the component from the original TFR, this range can be estimated by (26) with  $h = 1$ .

## V. METHOD PARAMETERS AND THEIR CHOICE

There are different settings that can be used while applying NMD. However, many of the parameters are either set at well-established values, or can be chosen adaptively, thus removing the ambiguity and making NMD a kind of superadaptive method. The settings and their choice are discussed below.

### A. TFR resolution parameter $f_0$

The resolution parameter  $f_0$  of the window or wavelet (7) determines the time and frequency resolution of the WFT/WT; i.e., how fast time changes can be reflected and how close in frequency components can be resolved, respectively. Time and frequency resolutions are not independent, being inversely proportional to each other, and an increase of  $f_0$  increases the frequency resolution but decreases the time resolution. This inability to locate the component precisely in both time and frequency is a general issue, which manifests itself in nearly every aspect of signal analysis and represents the time-frequency analog of the Heisenberg uncertainty principle in quantum mechanics [13, 15, 49, 50].

Although we have discussed how to choose  $f_0$  for harmonics given the extracted component, it is still unclear how to choose it at the first step, when we extract the main component. Generally, one needs to select  $f_0$  based on a compromise between better reflecting time variability and better resolving components in frequency, so the optimal choice depends on the characteristics of the components contained in the signal. Unfortunately, at present there does not seem to be any universal approach enabling one to choose  $f_0$  appropriately for any given signal (see [12, 13] for a discussion of this issue and the effects of different choices), so it remains the only important parameter of NMD that cannot be adaptively selected. Its choice, however, seems to be slightly more universal for the WT, as the latter adjusts its resolution in relation to the frequency band being studied; for it, one usually uses  $f_0 = 1$ , setting some standard limits on the allowable relative amplitude and frequency modulations of the components and the frequency distances between them.

Nevertheless, the very possibility of adjusting the time and frequency resolution is a great advantage of TFR-based methods. For example, as discussed previously in Ref. [46], the (E)EMD has time and frequency resolution properties similar to WT but, in contrast to the latter, its resolutions are set around some implicit values and cannot be changed. The choice of  $f_0$  therefore gives NMD more flexibility in comparison to methods not possessing such a parameter.

### B. TFR type: WFT or WT?

The difference between the WFT and the WT lies in the type of frequency resolution, which is linear for the former and logarithmic for the latter. Thus, the ability of the WT to resolve

components in frequency worsens with increasing frequency, while its ability to reflect time variations improves; in contrast, the WFT does not discriminate between components based on their characteristic frequencies. Preference for one type of TFR over the other therefore depends on the signal structure. The WT is especially suitable if the higher frequency components contained in the signal are more distant in frequency and have higher time variability than those at lower frequencies, which is the case for some types of signals; otherwise, the WFT is preferable.

Without some *a priori* knowledge, it is hard to select adaptively (i.e., automatically) the most appropriate type, especially given the associated problems related to the choice of the resolution parameter, as discussed above. However, after one component has been extracted (even roughly), selection of the optimal resolution type can be made based on its properties. Thus, if the time variability of the component's parameters increases with frequency, then the WT is the most suitable, whereas otherwise one should prefer the WFT. Given the initially extracted component's amplitude  $A(t)$ , phase  $\phi(t)$ , and frequency  $\nu(t)$ , an empirical condition for selecting the most appropriate TFR type can be stated as

$$\begin{aligned}
 R < 1 &\Rightarrow \text{use WFT}, \quad R > 1 \Rightarrow \text{use WT}, \\
 R &\equiv \{1 + V[\partial_t \nu(t), \nu(t)]\}^{-1} \\
 &\quad + \{1 + V[\partial_t A(t), \nu(t)]\}^{-1}, \quad (29) \\
 V[x(t), y(t)] &\equiv \frac{\text{std}[|x(t)/y(t)|^+]}{\text{std}[|x(t)/\langle y(t) \rangle|^+]}, \quad [|f(t)|^+ \equiv |f^+(t)|].
 \end{aligned}$$

Thus, the values of  $V[\partial_t A(t), \nu(t)]$  and  $V[\partial_t \nu(t), \nu(t)]$  quantify whether the amplitude and frequency modulation of the component become stronger with increasing frequency ( $V < 1$ ) or not ( $V > 1$ ). In the former case a reliable representation of the component requires higher time resolution for higher frequencies, so the WT is preferred, while for the latter case the WFT should be used. For example, for linear  $[\nu(t) = \nu_0 + at]$  and hyperbolic  $[\nu(t) = \exp(at)]$  chirps one has  $V[\partial_t \nu(t), \nu(t)] = \infty$  and  $V[\partial_t A(t), \nu(t)] = 0$ , reflecting the well-known [13] fact that the WFT and WT are most appropriate for their representation, respectively.

The derivatives  $\partial_t \nu(t)$  and  $\partial_t A(t)$  in (29) can be estimated by numerical differentiation. However, when noise is present in  $A(t)$  and  $\nu(t)$ , it will be greatly amplified in such estimates, so they will be generally quite noisy. Consequently, instead of standard deviations  $\text{std}[x(t)]$  in (29), it is better to use 75 percentiles, i.e., the width of the range within which 75% of the values of  $x(t)$  are contained. Alternatively, one can reconstruct  $\partial_t \nu(t)$  and  $\partial_t A(t)$  by deriving the direct reconstruction formulas for them as explained in [12].

The remaining question is which TFR type to use for the preliminary signal exploration, i.e., extraction of the component's parameters to be used in (29). As was discussed, the answer depends on the signal structure, and there is no universal criterion. However, since the WT is usually faster to calculate (due to its logarithmic scale) and generally has a more universal choice of the resolution parameter, we use it. Additionally, the threshold for  $R$  in (29) is changed from 1 to 1.1 (with this value being established empirically); such a change is done to account for the artificial correlation between

$\nu(t)$  and  $\partial_t \nu(t), \partial_t A(t)$  estimated from the WT, which appears as a result of the logarithmic frequency resolution of the latter.

Summarizing, we calculate the WT of the signal and extract from it the component and its parameters. Then we utilize the criterion (29) and determine which is the best type of TFR to use in the given case. If this is the WT, we retain the component already extracted; otherwise, we calculate the WFT of the signal in the corresponding frequency range (26) and re-extract all the parameters from it. To preserve the time and frequency resolution properties for which the component was extracted from the WT (which we assume to be appropriate for it), the resolution parameter for the WFT  $f_0^{(WFT)}$  should be adjusted accordingly. If the WFT and WT are calculated using a Gaussian window and log-normal wavelet (7), the rule is

$$f_0^{(WFT)} \approx 2\pi f_0^{(WT)} / \langle \nu(t) \rangle. \quad (30)$$

It follows from the fact that, for window and wavelet (7) and not too small  $f_0^{(WT)}$ , the frequency resolution of the WT around  $\omega = 2\pi$  (if "linearized") is similar to the frequency resolution of the WFT for the same resolution parameter [12]; taking into account the scaling nature of the WT, i.e., its frequency-dependent resolution, one then obtains (30). Since for all harmonics, by definition, one would have identical  $V[\partial_t \nu(t), \nu(t)]$  and  $V[\partial_t A(t), \nu(t)]$  in (29), the same type of frequency resolution (linear or logarithmic) is appropriate for all of them. Hence, for the extraction of harmonics we use the same TFR type as was selected for the original component.

### C. Reconstruction method: Direct or ridge?

As mentioned in Sec. III A, one can use either of two alternative methods for reconstruction of the components from the WFT/WT: direct (10) or ridge (9). The differences and errors of each method were studied in detail in [12]. It was found that the ridge method is more noise robust, but that the direct method allows the time variability in the component's parameters at low noise levels to be followed more accurately.

For some parts of the NMD procedure it is inherently better to use a particular reconstruction method. Thus, in the criterion for selecting the TFR type (29) we always use the parameters reconstructed by the ridge method due to the noise robustness of such estimates. Additionally, while testing the signal against noise in the stopping criterion (Sec. III E) we also use the ridge estimates for calculating the discriminating statistics  $D$  (15). This seems natural because the curve extraction is based on the amplitudes and frequencies of the peaks (see Sec. III A), and we have found such an approach to be more stable than using the direct estimates; the noise-robustness of ridge reconstruction is advantageous here as well.

For the other parts of the procedure, the reconstruction method can be chosen adaptively. To choose it at the first step, when we have extracted the time-frequency support  $[\omega_-(t), \omega_+(t)]$  of some component and need to reconstruct its parameters, one can use the approach suggested in [12]. Thus, we first reconstruct the parameters by both methods, getting  $A^{(d,r)}(t)$ ,  $\phi^{(d,r)}(t)$ ,  $\nu^{(d,r)}(t)$ , where "d" and "r" denote direct and ridge estimates, respectively. Then we compute the TFR (using the same parameters as originally) of the signal  $s^{(d)}(t) = A^{(d)} \cos \phi^{(d)}$ , extract the ridge curve from it [taking simple maxima  $\omega_p(t) = \arg \max_{\omega} |H_s(\omega, t)|$  is sufficient here],

and reconstruct by direct method the “refined” parameters  $\tilde{A}^{(d)}$ ,  $\tilde{\phi}^{(d)}$ ,  $\tilde{\nu}^{(d)}$ . We do the same for the “ridge” signal  $s^{(r)}(t) = A^{(r)} \cos \phi^{(r)}$ , now using ridge method to reconstruct the refined estimates. Then we compute inconsistencies  $\varepsilon_{a,\phi,\nu}^{(d,r)}$  between the original and refined estimates for each method,

$$\begin{aligned}\varepsilon_A^{(d,r)} &\equiv \kappa_A^{(d,r)} \sqrt{[\tilde{A}^{(d,r)}(t) - A^{(d,r)}(t)]^2}, \\ \varepsilon_\phi^{(d,r)} &\equiv \kappa_\phi^{(d,r)} \sqrt{1 - |\langle e^{i[\tilde{\phi}^{(d,r)}(t) - \phi^{(d,r)}(t)]} \rangle|^2}, \\ \varepsilon_\nu^{(d,r)} &\equiv \kappa_\nu^{(d,r)} \sqrt{[\tilde{\nu}^{(d,r)}(t) - \nu^{(d,r)}(t)]^2},\end{aligned}\quad (31)$$

where  $\kappa_{A,\phi,\nu}^{(d,r)}$  are the coefficients to tune the performance of the approach (we empirically found  $\kappa_{A,\phi,\nu}^{(d)} = \{3, 4, 2\}$ ,  $\kappa_{A,\phi,\nu}^{(r)} = 1$ ). Obviously, for the exact estimates one has  $\varepsilon_{A,\phi,\nu} = 0$ , so it is natural to assess the relative performance of the reconstruction methods based on their inconsistencies (31). Therefore, we use the direct amplitude estimate if  $\varepsilon_A^{(d)} < \varepsilon_A^{(r)}$ , and the ridge amplitude estimate otherwise; the same applies to the phase and frequency estimates. Although empirical, this approach works very well in practice.

For harmonics the appropriate reconstruction method can be determined simply as that giving the highest consistency  $\rho^{(h)}$  (12). Therefore, while adapting the resolution parameter for the harmonic, at each  $f_0$  we reconstruct its parameters by both methods and at the end select the one characterized by highest  $\rho^{(h)}$ .

An adaptive choice of reconstruction method, as outlined above, ensures that in most cases we get the best possible estimates: When the amplitude and frequency modulations are low and noise is high, the noise-robust ridge reconstruction method is used, while in the opposite case the direct method is applied, giving more accurate estimates of amplitude and frequency variations. This approach further increases the noise robustness and the accuracy of the NMD.

#### D. Other parameters

All the other parameters can be partitioned into two groups: some prefixed settings and the parameters of numerical accuracy. The former group includes the significance levels for the two surrogate tests (used for identification of the harmonics, Sec. III C, and for the stopping criterion, Sec. III E), the minimal consistency  $\rho_{\min}$  (14), etc. Each of these parameters is set to a well-established value, either corresponding to a standard convention (such as 95% significance of the surrogate tests [27,28]) or found empirically [e.g., the expression (14) for  $\rho_{\min}$ ].

The second group includes the accuracy with which to determine the optimal  $f_0$  while adapting it to the harmonics, the precision  $\varepsilon_s$  to use for determining the minimal support (25), etc. Here one faces the usual trade-off between accuracy and computational cost. The default values, however, are sufficient in most cases and further increase of precision leads to only a slight improvement in the accuracy of the method.

## VI. SIMULATED EXAMPLES

Having described all parts of the NMD procedure, we now illustrate the method by consideration of some specific examples. In this section we consider simulated signals whose composition is precisely known, so that one can assess reliably the performance of NMD and compare it to that of the other methods.

#### A. Example 1: One simple mode in Brownian noise

As a first, simple, and illustrative example, we take the signal depicted in Fig. 5(a). Its WFT and WT are shown in Figs. 5(b) and 5(c), respectively. From the figure it is immediately evident that, while in the WFT the harmonics with frequencies around 3 and 4 Hz are distinguishable, in the

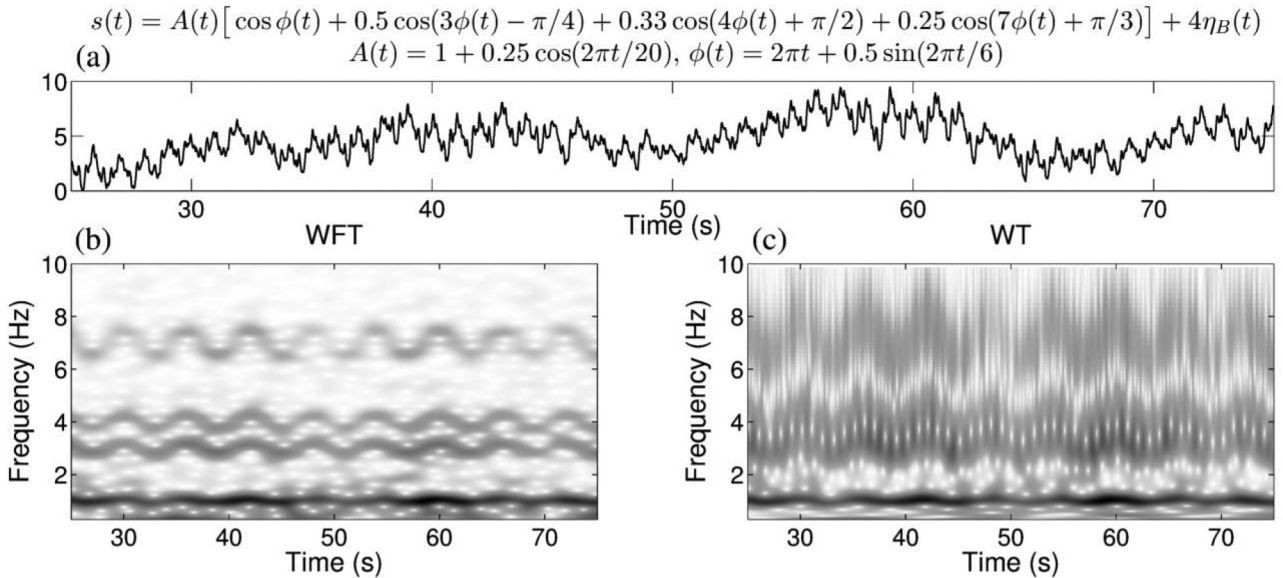


FIG. 5. (a) The central 50-s section of the signal  $s(t)$  specified by the equation at the top, which represents a single NM corrupted by Brownian noise of standard deviation equal to 4 [Brownian noise of unit deviation  $\eta_B(t_n) \sim \sum_{m=1}^n \eta_W(t_{m-1}) \sqrt{t_m - t_{m-1}}$  is obtained as the normalized cumulative sum of the Gaussian white noise signal  $\eta_W(t)$ ]. (b),(c) Central parts of the WFT and WT, respectively, of the signal  $s(t)$  shown in (a). The signal was sampled at 100 Hz for 100 s.

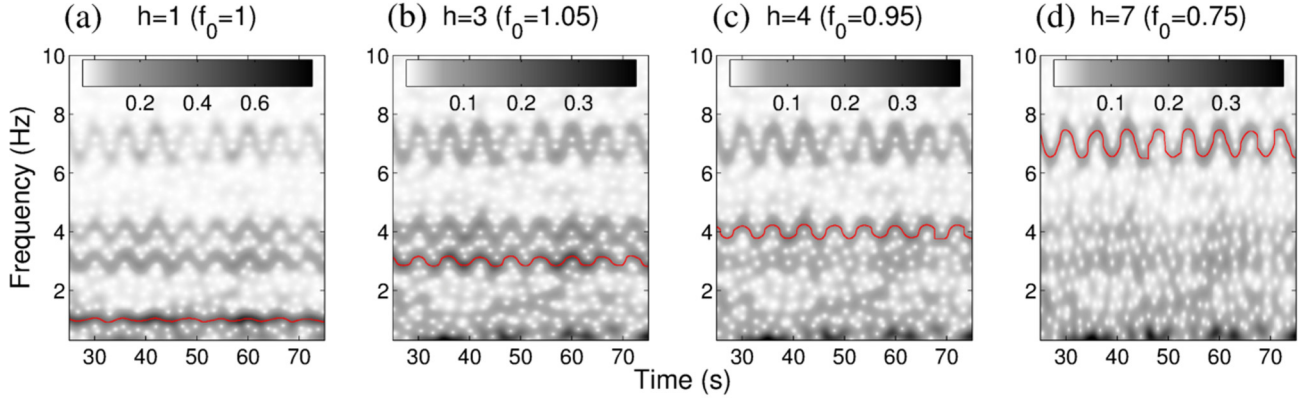


FIG. 6. (Color online) The WTFs from which each true harmonic is reconstructed, with the corresponding extracted ridge curves being shown by solid red lines. The window resolution parameter  $f_0$  is adjusted individually for each harmonic. After the harmonic is reconstructed, it is subtracted from the signal, so that, e.g., the first harmonic no longer appears in the WTFs of (b)–(d). Note that the color scaling in (b)–(d) covers half the amplitude range of that in (a).

WT they interfere strongly and cannot be resolved; in contrast, the WT has high enough time resolution at 7 Hz to represent the frequency modulation of the corresponding harmonic, while in the WTF this highest harmonic self-interferes (there appear the “bridges” between consecutive frequency modulation cycles), indicating that the time resolution is insufficient. Therefore, for the present case one cannot appropriately extract all harmonics using either WTF or WT with constant  $f_0$ . However, adaptive representation of the harmonics, as discussed in Sec. IV A, solves the problem.

The NMD proceeds as follows. First, the WT of the signal is calculated, and the dominant component is extracted from it, as described in Sec. III A; in result one obtains the first harmonic of the NM (located around 1 Hz). The extracted component is then tested with the surrogates against noise (see Sec. III E); in the present case, it passes the test (100% significance) and is therefore regarded as genuine. By application of the criteria (29), the WTF is determined to be more suitable for the representation of this component than the WT and is used in what follows. Thus, the component is reextracted from the corresponding signal’s WTF (see Sec. V B), and its parameters are reconstructed by both the direct (10) and the ridge (9) methods. Using (31), it is established that the ridge estimates of amplitude, phase, and frequency seem to be the more accurate in the present case and are therefore taken as the ones to be used.

Next we test whether the extracted component is the first harmonic by checking its possible  $1/h$  harmonics. Thus, first the  $1/2$  harmonic is extracted and tested for being true (see Sec. III C) using different resolution parameters  $f_0$  within the range (23); among the direct and ridge estimates obtained for  $f_0$  at which the harmonic was identified as true, we choose those maximizing the consistency (12). In the present case, the  $1/2$  harmonic is identified as false for all tested  $f_0$ , so it is discarded. We do the same for  $1/3$  and  $1/4$  harmonics, which are both identified as false. Since there are  $S_0 = 3$  consecutive false harmonics ( $1/2$ ,  $1/3$ , and  $1/4$ ), we stop the procedure and correctly conclude that the extracted component is first (and not higher) harmonic of the NM.

We then extract and test the higher harmonics  $h = 2, 3, \dots$  in qualitatively the same way as was done for  $h = 1/2, 1/3, \dots$ . If some harmonic is identified as true, we subtract it from the

signal to remove its interference with higher harmonics. As a result of the procedure, all genuine harmonics  $h = 3, 4, 7$  were correctly identified as true, and all the others as false. The resolution parameters were adapted for each harmonic so as to optimally represent and reconstruct it, as discussed in Sec. IV A and illustrated in Fig. 6 for the present case. Harmonic extraction was stopped when  $S = 3$  consecutive harmonics  $h = 8, 9, 10$  were identified as false.

We then refine the harmonics’ parameters using (24) and reconstruct the full NM. This NM is then subtracted from the signal, and the procedure is repeated. However, when we extract the next component, it does not pass the surrogate test against noise (see Sec. III E), and therefore NMD is stopped, with one NM being extracted and the residual correctly identified as noise (in the present case Brownian). The result of NMD is shown in Fig. 7, from which one can see that even the residual Brownian noise is well recovered.

For example, the results of EMD [4] and EEMD [5] procedures are shown in Fig. 8. In contrast to NMD, (E)EMD produces 13 distinct components, with only the first harmonic of the NM being more-or-less reliably recovered in one of them. Thus, in Fig. 8 C4 for EMD and C5 for EEMD represent the first harmonic, C3 for EMD and C4 for EEMD is the noise-spoiled mix of the 3 and 4 harmonics, C2 for EMD, and C3 for EEMD is the badly corrupted seventh harmonic (with influence from harmonics 3 and 4 in the case of EEMD), while none of the other “components” has any physical meaning at all.

### B. Example 2: Two modes with interchanging harmonics in white noise

To demonstrate the exceptional noise robustness of NMD and the power of the surrogate test in distinguishing true from false harmonics, we consider the signal shown together with its WTF and WT in Fig. 9. As can be seen, the harmonics of the second NM are located exactly at the places where the harmonics of the first NM are expected to be, so that they can easily be confused with them. Moreover, it is very hard to distinguish true from false harmonics in the present case because each NM has constant amplitudes and only very small frequency modulation (the absolute deviation between



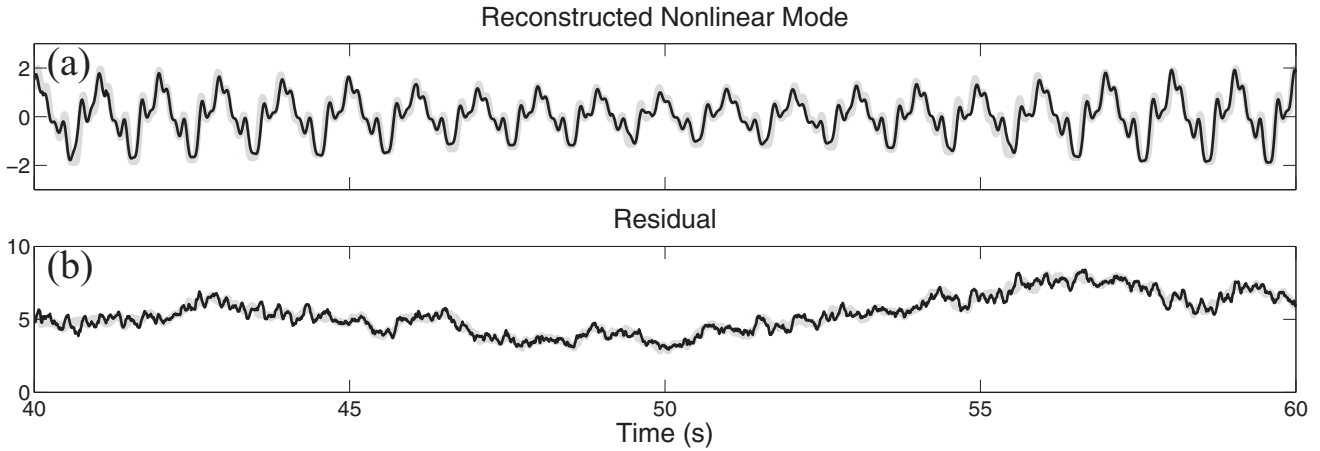


FIG. 7. The result of applying NMD to the signal shown in Fig. 5(a). (a) The reconstructed NM (black line) is compared with the true NM (gray background line); (b) the residual provided by NMD (black line) is compared with the actual background noise (gray background line).

the expected frequency of the second harmonic of the first NM and the frequency of the first harmonic of the second NM is only  $|2\nu_1^{(1)}(t) - \nu_2^{(1)}|/2\pi = 0.016 \pm 0.014$  Hz). Furthermore, the noise is exceedingly strong, in standard deviation being 1.5 times that of the full noise-free signal, 1.8 times that of the first NM, 2.7 times that of the second NM, and 12.2 times that of the smallest (third) harmonic of the second NM, located at around 6 Hz [the latter is buried under the noise and not even visible in the WFT shown in Fig. 9(b)].

Because the noise is white in the present case, its power in the WT grows with frequency, as seen from Fig. 9 (note, however, that such a situation rarely arises for real signals). Consequently, instead of using the WT and then adaptively selecting the appropriate type of TFR, for the present case we use the WFT from the very beginning. Nonlinear mode decomposition then proceeds as usual. It first extracts the

dominant component and tests it against noise. If it passes the test, NMD extracts its harmonics and identifies the true ones. Then the full NM is reconstructed and subtracted from the signal, after which the procedure is repeated on the residual until it does not pass the surrogate test against noise.

The relevant information about the NMD outcome is summarized in Table I. As can be seen, NMD correctly identifies all harmonics and joins them into two distinct modes. Moreover, the amplitude ratios  $a_h$  and phase shifts  $\varphi_h$  for each NM [see (4)], calculated from the reconstructed amplitude and phases as  $a_h = \langle A^{(h)} \rangle / \langle A^{(1)} \rangle$  and  $\varphi_h \equiv \arg [e^{i(\phi^{(h)} - h\phi^{(1)})}]$ , are very close to their actual values; this is true even for the (buried in noise) third harmonic of the second NM. The ridge method is automatically selected for reconstructing of all harmonics, which is indeed the preferable choice due to its noise robustness [12].

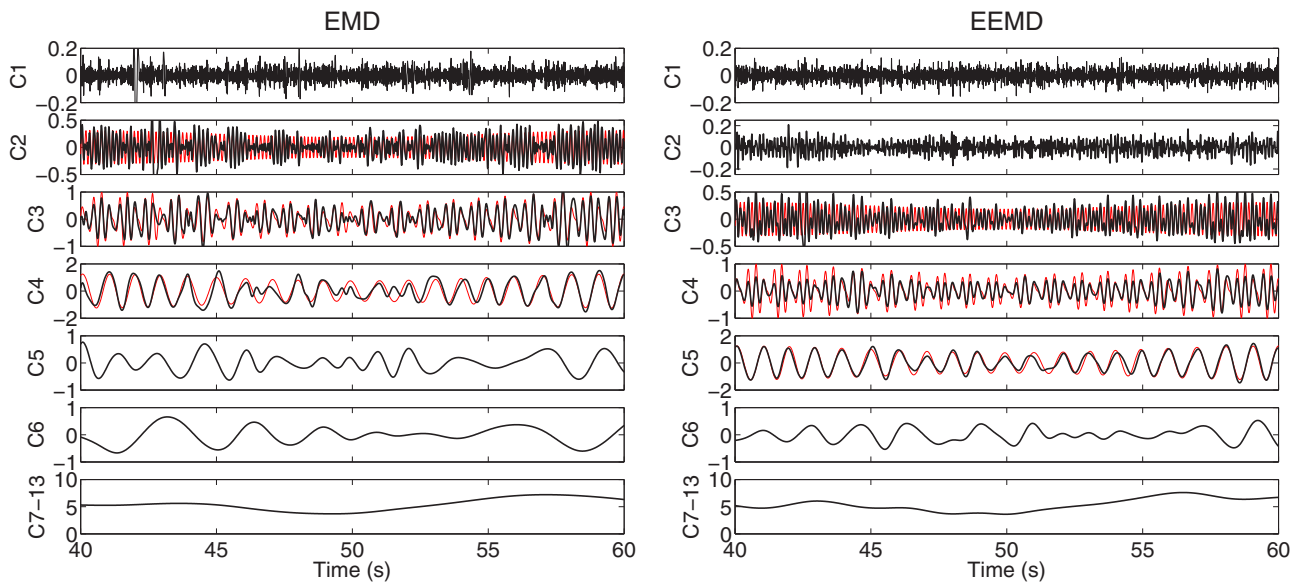


FIG. 8. (Color online) The result of applying EMD (left panels) and EEMD (right panels) to the signal shown in Fig. 5. Thin red lines, where present, show the real first harmonic (in C4 for EMD and C5 for EEMD), sum of the third and fourth harmonics (in C3 for EMD and C4 for EEMD), and the seventh harmonic (in C2 for EMD and C3 for EEMD). The bottom panels show the sum of components 7 to 13. For EEMD, we have used 1000 noise realizations with standard deviations ten times smaller than that of the signal.

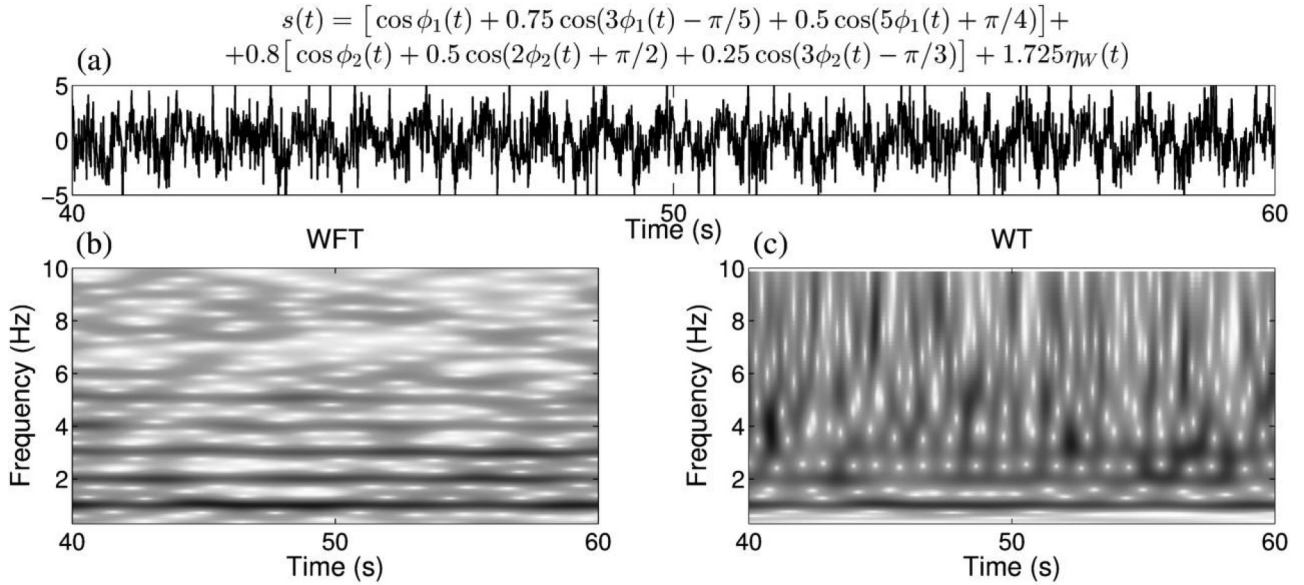


FIG. 9. (a) The central 20-s section of the signal  $s(t)$  specified by the equation at the top, representing the sum of two NMs corrupted by Gaussian white noise of standard deviation equal to 1.725; the phases of the NMs  $\phi_{1,2}(t)$  were obtained as  $\phi_{1,2}(t) = \int_0^t v_{1,2}(\tau) d\tau$ , with  $v_{1,2}(t)/2\pi = (1,2) + 0.01\tilde{\eta}_{B;1,2}(t)$ , where  $\tilde{\eta}_{B;1,2}(t)$  are two independent realizations of the unit deviation Brownian noise filtered in the range  $[0.01, 0.2]$  Hz. (b),(c) Central parts of the WFT and WT of the signal  $s(t)$  shown in (a). The signal was sampled at 100 Hz for 100 s.

From Table I it can be noted that, for all true harmonics, the resolution parameter  $f_0$  used is higher than the original. This is because the higher the  $f_0$ , the easier it is to segregate the component from noise [12]. However, increasing  $f_0$  at the same time worsens the accuracy of representation of amplitude or frequency modulation of the component [12], so its choice is determined by a compromise between reflecting well the time variability of component's parameters and suppressing the noise; the adaptive scheme

that we use, described in Sec. IV A, effectively implements this criterion.

The final results of NMD are shown in Fig. 10. Both NMs are reconstructed with great accuracy, which is like a miracle given such strong noise corruption of the original signal (see Fig. 9); even the residual noise is recovered almost exactly. Such performance is unachievable with the other existing methods; e.g., (E)EMD in the present case produces 13 components, and none of them reliably represent any harmonic

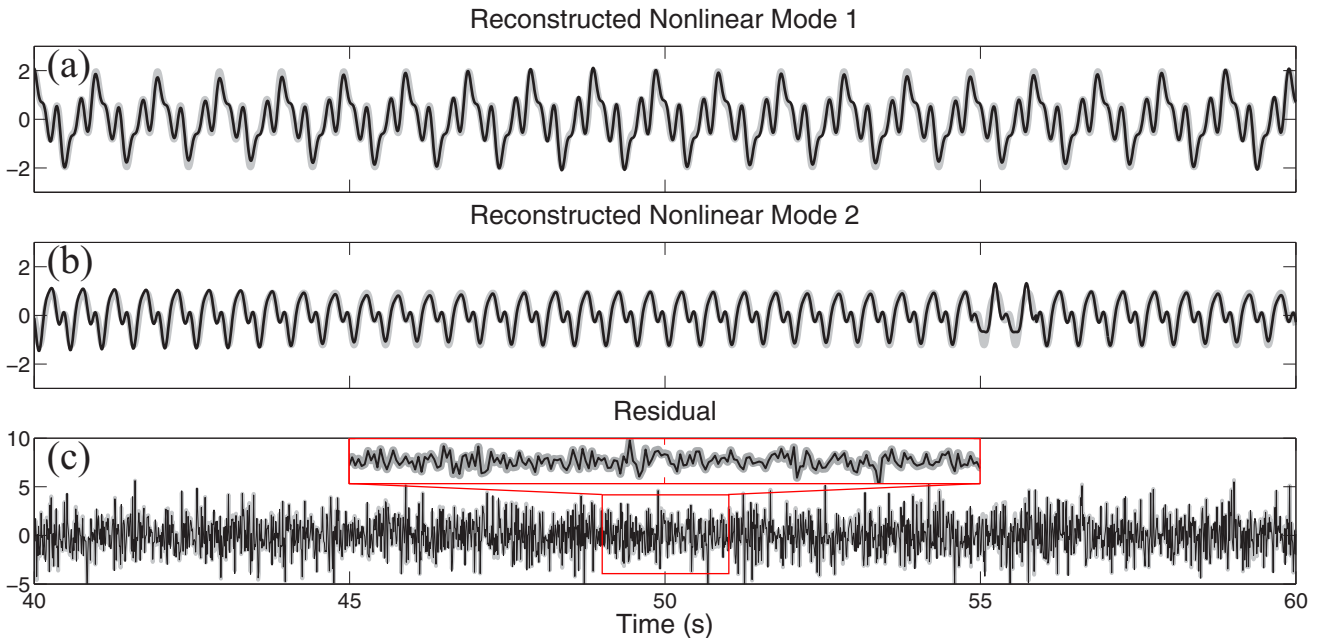


FIG. 10. (Color online) The result of applying NMD to the signal shown in Fig. 9. In (a) and (b) black lines indicate the two reconstructed NMs and the gray background lines show the true NMs for comparison. (c) Similarly, the black and gray lines show the residual returned by NMD and true background noise, respectively.

(not shown). Note that NMD can produce even more accurate results if the resolution parameter is adjusted from the very beginning, i.e., for the first harmonics (and not only the higher ones). However, as mentioned in Sec. V A, at present there does not seem to be a good and universal way of doing this.

## VII. REAL LIFE EXAMPLES

After demonstrating its success in the analysis of simulated signals, we now apply NMD to real data.

### A. Decomposing human blood flow signals

The decomposition of skin blood flow signals, measured noninvasively by laser-Doppler flowmetry (LDF) [51], is a very tough task, with no method at present being able to do it well. This was demonstrated in [11] for the example of Karhunen-Loève decomposition, while (E)EMD usually also fails. Thus, blood flow signals contain a large amount of “physiological” noise, as well as having components with amplitude or frequency modulation whose strength and speed change with time. Nonetheless, as will be seen, NMD can often tackle even such complicated cases.

Blood flow signals exhibit oscillatory activity at multiple frequencies, and the WT has been found especially useful in studies of their structure [52]. Six oscillations have been identified in blood flow and attributed to different physiological mechanisms [3,52–54], with characteristic frequency ranges of (approximately): 0.6–2 Hz (I), 0.15–0.6 Hz (II), 0.04–0.15 Hz (III), 0.02–0.04 Hz (IV), 0.01–0.02 Hz (V), and 0.005–0.01 Hz (VI). Range I corresponds to cardiac oscillations, which originate from rhythmical pumping of the heart. Range II corresponds to respiratory oscillations, which are the consequence of the mechanical influence of respiration on the cardiac output and, to a smaller extent, the respiratory modulation of the heart rate [55,56]. The mechanism underlying range III oscillations, which are present in most hemodynamic signals, is not generally agreed: In studies of cardiac activity and blood pressure variability they are usually attributed to the sympathetic nerve activity, being regarded as a result of time delays in the baroreflex feedback loop [57,58], while the many authors studying microvascular blood flow relate these oscillations to myogenic activity of the smooth muscle cells [3,52,53]. Finally, the oscillations in the IV, V, and VI ranges were attributed to the neurogenic, NO-dependent endothelial, and NO-independent endothelial activity [3,52–54], respectively.

Armed with this knowledge, we now apply NMD. To better utilize the prior information, we apply the procedure to each of the above-mentioned physiological frequency ranges individually, starting from the first one. Thus, for a given range we extract the dominant component and test it against noise. If it passes the test, we then extract its harmonics, reconstruct the NM and subtract it from the signal. We then repeat the procedure for the same range until the next extracted component does not pass the test against noise.

The results of the procedure are shown in Figs. 11 and 12 for the examples of two blood flows. Clearly, NMD is able to decompose these signals into a physically meaningful oscillations with complex wave forms (and it also returns their amplitudes, phases, and frequencies). In both cases, we

were able to extract the cardiac component (around 1 Hz), while activity in ranges IV, V, and VI did not pass the test against noise. However, in the example of Fig. 11 there are strong myogenic oscillations (around 0.1 Hz), which were extracted, while we see no activity in the frequency range II (respiratory) [Fig. 11(b)]. In contrast, for the example of Fig. 11 the respiratory oscillations are present and the myogenic oscillations are buried under noise. The wave forms of the cardiac oscillations are also different in two cases.

In general, the properties and presence of the oscillations in blood flow vary from subject to subject, being influenced by many factors, such as the state of the microvasculature (which might be influenced by age and gender), properties of the skin, etc. As was demonstrated, NMD can be very useful for the study and classification of these effects; e.g., it can be used to investigate the relationship between the cardiac wave form (as well as other blood flow properties) and the health conditions.

It is important to note that, as discussed in Sec. III E, even if the component extracted from a particular frequency range does not pass the surrogate test against noise (as, e.g., for ranges IV, V, and VI in the above examples), this does not necessarily mean that there is no physiologically meaningful activity there. Thus, the underlying oscillations might be simply very small so that they are easily masked by noise, which is often the case for the respiratory oscillations. The other possibility is that the resolution parameter used is not appropriate to represent reliably the component of interest. This is often the case for myogenic oscillations, which might change their amplitude and/or frequency very rapidly at certain times. In fact, the best choice in such situations would be probably some time-varying  $f_0(t)$ , but its form is generally very hard to choose.

We have found that, in practice, NMD is almost always able to extract the cardiac component accurately from the blood flow signal using the default settings, though, to improve the accuracy and speed of the method, it is advisable to filter the signal in the corresponding frequency range before applying NMD. The respiratory component, on the other hand, can be extracted only in a limited number of cases on account of often being very small in amplitude compared to the physiological noise inherent in blood flow signals. Thus, one can always obtain “some” component from the corresponding frequency range, but it is often very inaccurate and consequently fails the surrogate test. Nevertheless, if the respiratory signal is also available, then one can use it as a reference signal for extracting the corresponding oscillations accurately from the blood flow signal even where they are very weak. An approach for doing this is discussed in the next section; it can also be used to improve the extraction of cardiac oscillations if the corresponding ECG, from which cardiac phase can be extracted more accurately than from the blood flow signal, is available. Finally, although we were able to extract myogenic oscillations (range III) in some cases, they are, in general, very hard to extract, as are also those at even lower frequencies (ranges IV–VI).

### B. Removing cardiac artifacts from a human EEG signal

Nonlinear mode decomposition can also be used to filter the signal from an extraneous oscillatory activity, provided that there is an associated signal from which the phase and

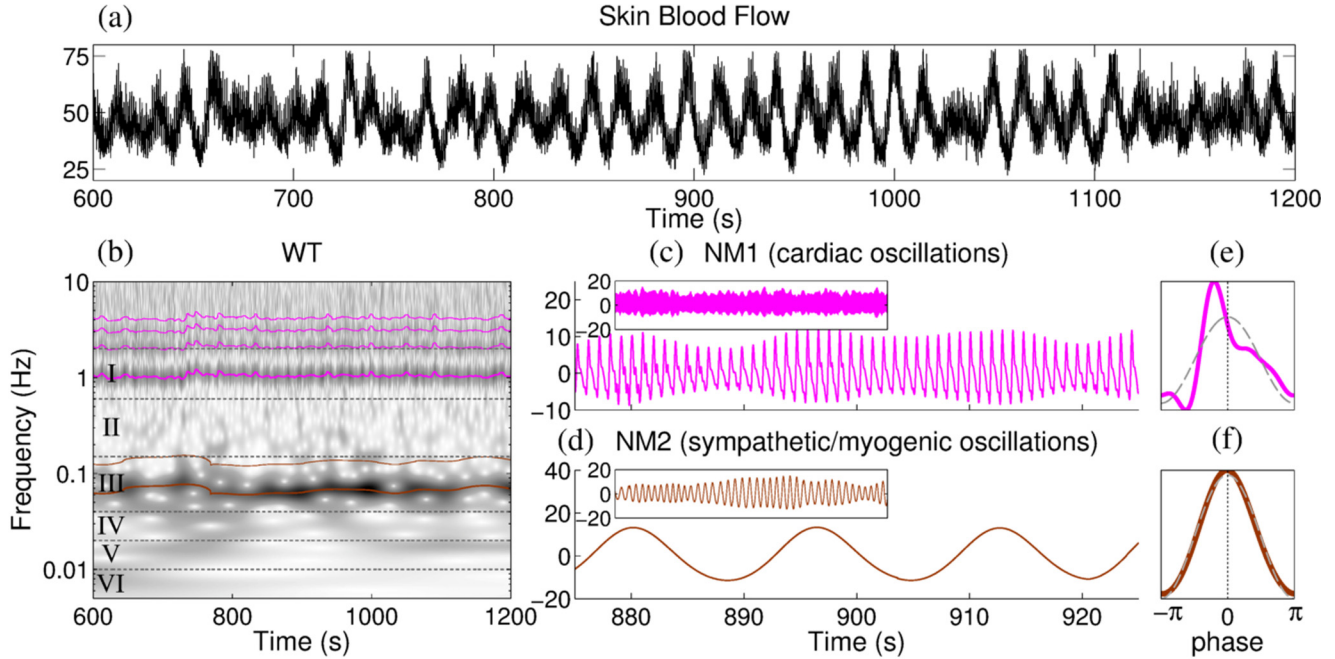


FIG. 11. (Color online) (a) An example of the blood flow signal, measured by LDF with the probe over the right wrist *caput ulna* (for more details, see [3,53]); the signal is sampled at 40 Hz for 1800 s, and the panel shows its central 600 s part. (b) The WT of the signal shown in (a). Gray dotted lines partition the frequency axis into the regions within which the physiologically meaningful oscillations are located (according to [3,52–54]; see the text). Bold colored lines show those extracted components which pass the surrogate test against noise, with thinner lines of the same color showing their higher harmonics. (c),(d) The reconstructed NMs corresponding to four upper magenta lines and two bottom brown lines on (b), respectively; the main graphs show the modes during 50 s and small insets during 600 s [as in (a)]. (e),(f) The wave forms of the oscillations shown in (c) and (d), respectively; the gray dashed line shows a pure sinusoidal wave form of amplitude equal to that of the first harmonic and is provided as a guide for the eye. The cardiac wave form (e) has four harmonics  $h = 1, 2, 3, 4$ , with  $a_h = [1, 0.52, 0.37, 0.16]$  and  $\varphi_h/\pi = [0, 0.4, 0.66, -0.97]$  [in the notation of (4)]; the wave form in (f) is characterized by  $h = 1, 2$ ,  $a_h = [1, 0.08]$ , and  $\varphi_h/\pi = [0, 0.07]$ .

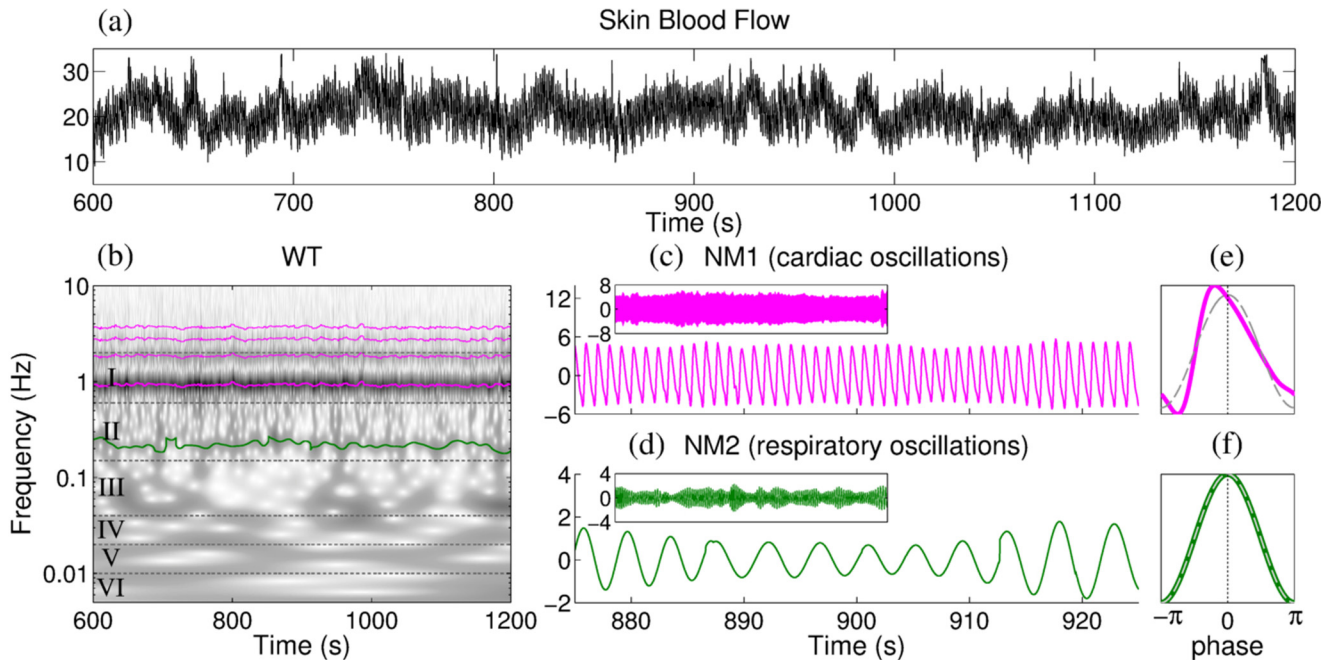


FIG. 12. (Color online) Same as in Fig. 11, but for the blood flow measured from a different subject. The nonlinear modes shown in (c) and (d) correspond, respectively, to the four upper lines (magenta) and the bottom (green) line in (b). The cardiac wave form (e) has four harmonics  $h = 1, 2, 3, 4$ , with  $a_h = [1, 0.32, 0.15, 0.04]$  and  $\varphi_h/\pi = [0, 0.41, 0.95, -0.44]$  [in the notation of (4)]; the respiratory oscillations have only one harmonic, so that the wave form in (f) is a pure sinusoid.

frequency of the latter can accurately be extracted. Thus, using the NMD harmonic extraction procedure, the fundamental harmonic of the extraneous mode can be extracted as the  $h = 1$  harmonic of the reference component based on its (reference) phase and frequency. In this way the resolution parameter is adjusted from the start, allowing one to represent this fundamental oscillation well, even if it is strongly corrupted by noise or other influences. We illustrate this alternative use of NMD through the example of removing cardiac artifacts from the human electroencephalogram (EEG) recording, using the electrocardiogram (ECG) as the reference signal.

The EEG often contains artifacts related to heart activity, which arise due to blood flow pulsations underneath the probes (the so-called ballistocardiogram (BCG) artifacts [61]) and possibly also due to direct pick up of heart electrical activity. The BCG artifacts are extremely prominent in the EEG measured in a magnetic field (e.g., simultaneously with a magnetic resonance imaging scan), in which case they are usually filtered out by ICA [8,9]. However, ICA requires as many simultaneous EEG measurements as possible, with cardiac artifacts being prominent and taking a similar form in most of them, which is not always the case in practice. It will obviously fail to remove the artifacts if only one EEG and one ECG signal are available, because the form taken by cardiac activity in the EEG is completely different from its form in the ECG signal (see below), contradicting the assumption of linear mixing on which ICA is based. In fact, we have found that ICA fails even given four EEGs, which is probably because the artifacts in them are relatively small, being hard to distinguish (although still capable of affecting some time-frequency measures), and because the form and magnitude of these artifacts might be different in different EEGs (perhaps dependent on probe position). The (E)EMD method, too, fails to provide meaningful results in the present case.

Simultaneously measured EEG and ECG time series for the same subject are presented in Figs. 13(a) and 13(b), with their WFTs for the default resolution parameter  $f_0 = 1$  being shown in (c) and (d), respectively. Clearly, for  $f_0 = 1$  the cardiac harmonics are not well distinguishable in the EEG's WFT [Fig. 13(c)], so that the extracted curve might not be very accurate. However, the cardiac phase  $\phi^{(1)}(t)$  and frequency  $\nu^{(1)}(t)$  can, of course, be extracted *directly* from the WFT of the ECG [Fig. 13(d)]. They should be the same for the first cardiac harmonic in the EEG and in ECG, because both activities obviously have the same rhythmicity; however, the corresponding amplitudes might be different and perhaps not even proportional to each other. This is because, depending on the measurement and the environment, the same activity might undergo various transformations that can change its amplitude dynamics and the corresponding wave form, but leave phase dynamics largely unaltered. For example, NMs  $c(t) = A(t)[\cos \phi(t) + a \cos 2\phi(t)]$  and  $\exp[c(t)]$  will have the same fundamental phase and frequency, but different amplitude dynamics (i.e., the ratio between the amplitudes of the corresponding fundamental components will be time varying) and different relationships between the harmonics.

Therefore, we use the cardiac phase  $\phi^{(1)}(t)$  and frequency  $\nu^{(1)}(t)$  extracted from the ECG as the parameters of the reference component and extract the main cardiac component from the EEG as its  $h = 1$  harmonic. This is done in the same

way as discussed in Sec. III B, i.e., selecting the peaks nearest to the expected frequency  $\nu^{(1)}(t)$ . We test this harmonic for being true (Sec. III C) and adapt the resolution parameter for its best representation (Sec. IV A) in the usual way, except that now we use everywhere the phase consistency  $\rho^{(1)}(0, 1, 0)$  (12) instead of the usual amplitude-phase consistency  $\rho^{(1)}(1, 1, 0)$ . This is because of the possible discrepancy between the amplitude dynamics of the components mentioned above [note that the threshold (14) also becomes  $\rho_{\min} = 0.5$ ]. If the extracted first harmonic is regarded as true, we use its parameters to extract the higher cardiac harmonics from EEG (now using unmodified procedures) and reconstruct the corresponding NM.

The EEG's WFT adapted [by maximizing phase consistency  $\rho^{(1)}(0, 1, 0)$ ] for representation of the cardiac component is shown in Fig. 13(e). Clearly, the corresponding ridge curves became much more visible than in the default WFT presented in Fig. 13(c). The cardiac artifacts extracted from the EEG are shown in Fig. 13(f). Their wave form, presented in Fig. 13(g), very much resembles the shape of the cardiac waves in the blood flow (cf. Fig. 12), but not that of the ECG. This is an indication of the BCG mechanism by which these cardiac artifacts are generated, which is also supported by the fact that their strength (and even their shape) might be different in EEGs from different probes for the same subject. Note that, depending on the particular EEG measurement, the artifacts might be inverted.

There are many other possible applications of NMD-based filtering. Thus, as discussed in the previous section, one can use this approach to extract the cardiac and respiratory components from the blood flow signal more accurately by utilizing reference phases obtained from the corresponding ECG and respiration signals. NMD can also be used to check whether a given signal contains oscillatory activity related to another one: If there is no oscillation originating from the same source as the reference signal, then the corresponding first harmonic extracted from a given signal based on the reference phase and frequency will be regarded as false. For example, we have not found any respiratory-related activity in the EEG, implying that the measurement process is almost unaffected by breathing.

In general, the situation when one signal contains components related to other signals is ubiquitous in real life, so that the NMD-based filtering is expected to be very useful in many situations. A great advantage of this approach is that it does not require the related oscillations in different signals to be of the same form (as is the case of, e.g., ICA), but only to have the same phase dynamics.

## VIII. OTHER APPLICATIONS

There are many other contexts where NMD can prove useful. For example, it can serve as an initial preprocessing which needs to be performed prior to applying any of the numerous existing methods for studying monocomponent signals or their sets (for a review, see [62]). Thus, having first decomposed the original signal into its constituent NMs, one can then investigate the latter using one or some of the huge diversity of available techniques. For example, the structure of the interactions between different modes can be recovered by

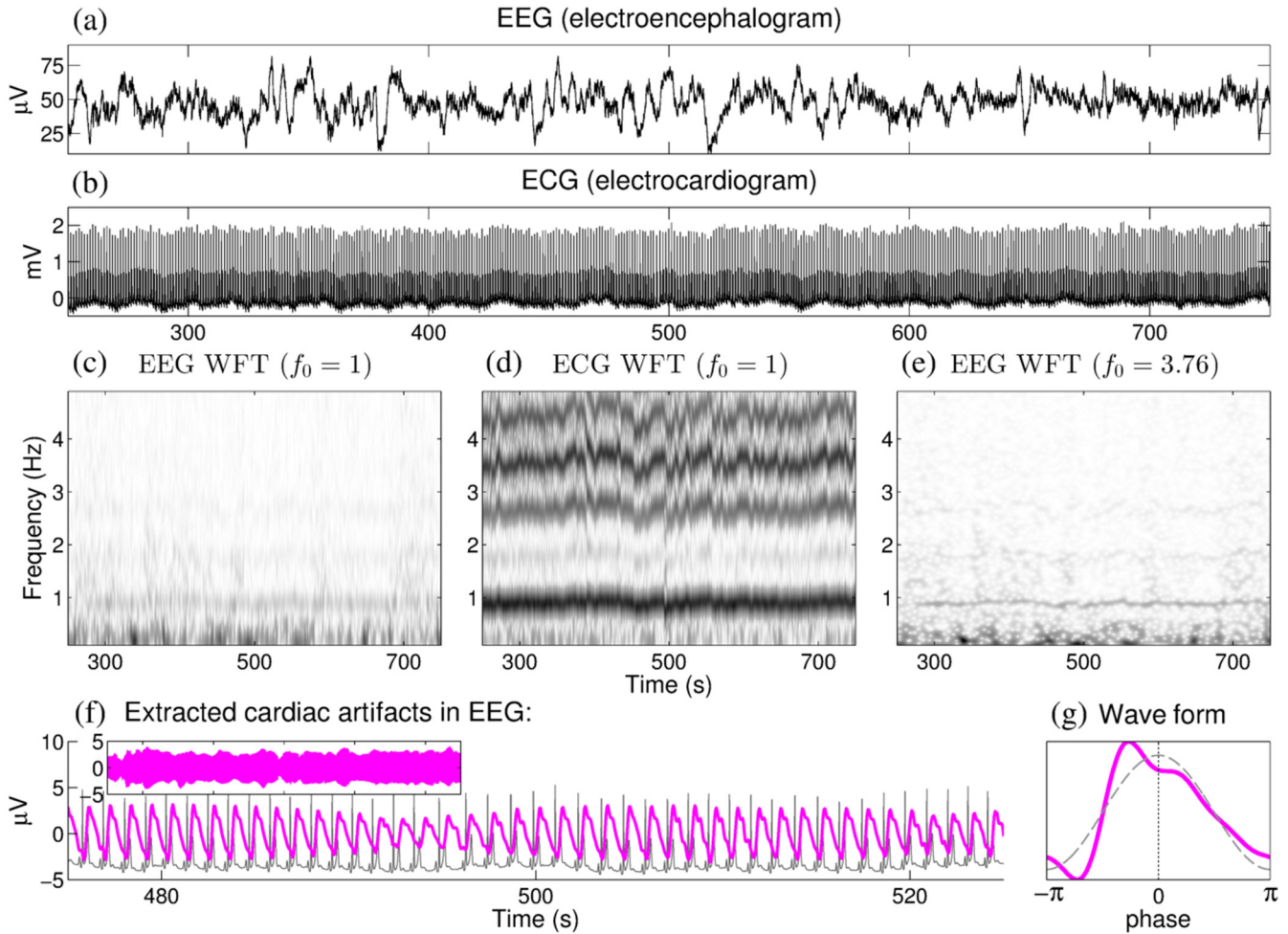


FIG. 13. (Color online) (a) The EEG signal [measured using a BIS<sup>TM</sup> 4-Electrode Sensor (Covidien-Medtronic, Dublin, Ireland and Fridley, Minnesota) placed on the forehead and a purpose-built signal conditioning unit (Jožef Stefan Institute, Slovenia), as described in [59]]. (b) The ECG signal (three-lead, with electrodes on shoulders and the lowest left rib, see, e.g., [60]). (c),(d) The WFTs of the EEG and ECG signals shown in (a) and (b), respectively, calculated using the default resolution parameter  $f_0 = 1$ . (e) The WFT of the EEG signal calculated using the adaptively adjusted resolution parameter  $f_0 = 3.76$ , due to which the cardiac component becomes much more visible than in (c). (f) The cardiac artifacts extracted from the EEG, with an inset showing them over the same 500 s as presented in (a),(b); gray background lines show the ECG scaled to the dimensions of the plot. (g) The wave form of the cardiac artifacts, which has four harmonics with  $a_h = [1, 0.33, 0.24, 0.11]$  and  $\phi_h/\pi = [0, 0.49, 1, -0.59]$  [in the notation of (4)]; the gray dashed line represents a pure sinusoid with the amplitude of the first harmonic, for comparison. The full signals are of duration 20 min and are sampled at 80 Hz; the EEG was actually measured under anaesthesia, but the same artifacts arise under all conditions.

applying Bayesian inference to the extracted phases [63,64], as was done in [2] to reconstruct the cardiorespiratory interaction; in this application, the high accuracy of the phases returned by NMD is especially advantageous.

Another important application is the classification of the oscillations present in the signal, which is of particular interest because it might yield valuable insights into the nature and properties of the underlying phenomena. Following the recent introduction of chronotaxic systems [65,66], it is clearly desirable to be able to determine whether the originating systems generating the different modes are chronotaxic or not. Roughly speaking (see [65,66] for a more detailed definition), the system is chronotaxic if it is (a) oscillatory (i.e., characterized by a limit cycle) and (b) its phase  $\phi(t)$  does not just move freely along the cycle, as in self-sustained oscillators, but is attracted to a particular deterministically time-varying fixed point  $\phi_A(t)$ ,

conferring an ability to resist external perturbations. This is exactly what one often observes in living systems, which are able to maintain their activity within physiological ranges even when strongly perturbed, so that chronotaxic behavior is expected to be abundant within the life sciences.

To establish which oscillations contained in the signal are chronotaxic, one needs to study them separately; i.e., the signal should be first decomposed into its NMs. This can conveniently and accurately be achieved with NMD. Subsequently, for each mode one can apply the approach suggested by Clemson *et al.* [66]. Basically, by applying different types of filters to the extracted phase, one estimates the expected difference  $\phi(t) - \phi_A(t)$  and uses detrended fluctuation analysis [3,67] to analyze the associated correlations, which are expected to differ between chronotaxic and nonchronotaxic systems. We have applied this method to the NMs of both of the

examples in Figs. 11 and 12, but have not found clear evidence of chronotaxis in any of the corresponding oscillations. However, since the method being used is based on a particular set of assumptions, these oscillations could, in principle, still be chronotaxic though falling outside the model considered in [66]; note also that the respiratory modulation of the cardiac frequency was shown to demonstrate signs of chronotaxis [66].

## IX. CONCLUSIONS

The decomposition method that we have introduced—NMD—is based on the time-frequency analysis techniques [12,19], surrogate data tests [27–30], and the idea of identification of the time-variable harmonics [25]. NMD consists of a number of ingredients, as described in Sec. III, each of which is a useful technique in its own right. For example, one can use the approach discussed in Sec. III E for testing the signal against noise (within a large class; see Sec. III E), which proves to be more noise robust than the conventional tests based on the correlation dimension or other nonlinearity measures [27,28,68].

We have successfully applied NMD to both simulated and real data and demonstrated its advantages over the existing approaches. Unlike many previous methods, NMD recovers the full underlying oscillations of any wave form. It is also extremely noise robust and in a sense superadaptive, meaning that most of its settings are automatically adapted to the properties of the particular signal under investigation. Finally, in contrast to the other methods, NMD returns only physically meaningful modes, stopping the decomposition when the residual is just noise.

The area of applicability of NMD is extremely wide. For example, it would now be reasonable to reconsider all those cases where (E)EMD has been applied to date (e.g., see references in [5,69]). Furthermore, as demonstrated above, the exceptional noise robustness of NMD and its other advantages allow one to apply it even in cases where all the other methods fail completely. Thus, it can be applied to almost all multicomponent signals of the kind that are ubiquitous in the life sciences, geophysics, astrophysics, econometrics, etc. We therefore expect that, with time, NMD will become a standard method in the field.

The latest MATLAB codes needed for running NMD are freely available [70], together with detailed instructions and examples, in both text and video formats.

## ACKNOWLEDGMENTS

We are grateful to Philip Clemson for valuable discussions and to Valentina Ticcinelli for testing the codes and for useful comments. The work was supported by the Engineering

and Physical Sciences Research Council (UK) [Grant No. EP/100999X1].

## APPENDIX: SUMMARY OF THE NMD ALGORITHM

The algorithm for performing NMD, including all the steps and improvements described above, can be summarized as follows.

(1) Calculate the WT of the given signal using the wavelet (7) with the chosen resolution parameter  $f_0$  (see Sec. V A).

(2) Extract the dominant component (in what follows, the *reference component*) from the signal's WT (see Sec. III A), and reconstruct its characteristics (see Secs. III A and V C).

(3) Test the reference component against noise using surrogates, as described in Sec. III E. Stop the decomposition if it does not pass this test.

(4) Check whether the WT is the “right” representation for the reference component and, if not, switch to using the WFT and reextract the component from it (see Sec. V B). In what follows, TFR denotes the chosen representation: WT or WFT.

(5) For  $h = 1/2, 1/3, \dots$  do the following.

(a) Calculate the signal's TFR within the frequency range determined by (28) using different values of the resolution parameter  $f_0^{(h)}$  [see (23) and related discussion], for each of which

(i) extract the  $h$ th harmonic of the reference component from the TFR and reconstruct its amplitude, phase, and frequency (see Secs. III B and V C);

(ii) test the current harmonic for being true as described in Sec. III C.

(b) If for some  $f_0^{(h)}$  the harmonic was identified as true, then set its characteristics to those reconstructed for the  $f_0^{(h)}$  that is characterized by the highest consistency (11) with the reference component among  $f_0^{(h)}$  for which the harmonic was identified as true.

(c) Stop when a prechosen number (default = 3) of consequent harmonics were identified as false for all tested  $f_0^{(h)}$ .

(6) If some harmonic was identified as true in the previous step, take the true harmonic with the smallest  $h$  as the reference component. The reference component is therefore guaranteed to be the first harmonic of the corresponding NM now.

(7) Perform step 5 for  $h = 1, 2, \dots$ , remembering the reconstructed parameters of the harmonics.

(8) Using the parameters of the harmonics that were identified as true in the previous step, reconstruct the full NM as described in Sec. IV B.

(9) Subtract the reconstructed NM from the signal, and repeat steps 1–8 on the residual.

- [1] G. C. Beroza and T. H. Jordan, Searching for slow and silent earthquakes using free oscillations, *J. Geophys. Res.* **95**, 2485 (1990).
- [2] D. Iatsenko, A. Bernjak, T. Stankovski, Y. Shiogai, P. J. Owen-Lynch, P. B. M. Clarkson, P. V. E. McClintock, and A. Stefanovska, Evolution of cardio-respiratory interac-

tions with age, *Philos. Trans. R. Soc., A* **371**, 20110622 (2013).

- [3] Y. Shiogai, A. Stefanovska, and P. V. E. McClintock, Nonlinear dynamics of cardiovascular ageing, *Phys. Rep.* **488**, 51 (2010).
- [4] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N. Yen, C. C. Tung, and H. H. Liu, The empirical

- mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, *Proc. R. Soc. London, Ser. A* **454**, 903 (1998).
- [5] Z. Wu and N. E. Huang, Ensemble empirical mode decomposition: a noise-assisted data analysis method, *Adv. Adapt. Data Anal.* **01**, 1 (2009).
- [6] I. T. Jolliffe, *Principal Component Analysis*, 2nd ed. (Springer-Verlag, Berlin, 2002).
- [7] H. Abdi and L. J. Williams, Principal component analysis, *Wiley Interdiscip. Rev. Comput. Stat.* **2**, 433 (2010).
- [8] P. Comon, Independent component analysis, A new concept?, *Signal Proc.* **36**, 287 (1994).
- [9] A. Hyvärinen and E. Oja, Independent component analysis: algorithms and applications, *Neural Networks* **13**, 411 (2000).
- [10] M. E. Davies and C. J. James, Source separation using single channel ICA, *Signal Proc.* **87**, 1819 (2007).
- [11] M. Hožič and A. Stefanovska, Karhunen—Loève decomposition of peripheral blood flow signal, *Phys. A (Amsterdam, Neth.)* **280**, 587 (2000).
- [12] D. Iatsenko, P. V. E. McClintock, and Aneta Stefanovska, Linear and synchrosqueezed time-frequency representations revisited: Overview, standards of use, reconstruction, resolution, concentration, and algorithms, *Digital Signal Proc.* **42**, 1 (2015).
- [13] S. Mallat, *A Wavelet Tour of Signal Processing*, 3rd ed. (Academic Press, Burlington, MA, 2008).
- [14] P. S. Addison, *The Illustrated Wavelet Transform Handbook: Introductory Theory and Applications in Science, Engineering, Medicine and Finance* (IOP Publishing, Bristol, UK, 2002).
- [15] L. Cohen, *Time-Frequency Analysis* (Prentice Hall PTR, New Jersey, 1995), Vol. 778.
- [16] F. Hlawatsch and G. F. Boudreaux-Bartels, Linear and quadratic time-frequency signal representations, *IEEE Signal Proc. Mag.* **9**, 21 (1992).
- [17] I. Daubechies, *Ten Lectures on Wavelets*, in CBMS-NSF Regional Conference Series in Applied Mathematics, Vol. 61 (SIAM, Philadelphia, 1992).
- [18] N. Delprat, B. Escudie, P. Guillemain, R. Kronland-Martinet, P. Tchamitchian, and B. Torrèsani, Asymptotic wavelet and Gabor analysis: Extraction of instantaneous frequencies, *IEEE Trans. Inf. Theor.* **38**, 644 (1992).
- [19] D. Iatsenko, P. V. E. McClintock, and A. Stefanovska, On the extraction of instantaneous frequencies from ridges in time-frequency representations of signals, [arXiv:1310.7276](https://arxiv.org/abs/1310.7276).
- [20] R. A. Carmona, W. L. Hwang, and B. Torresani, Multiridge detection and time-frequency reconstruction, *IEEE Trans. Signal Proc.* **47**, 480 (1999).
- [21] R. A. Carmona, W. L. Hwang, and B. Torresani, Characterization of signals by the ridges of their wavelet transforms, *IEEE Trans. Signal Proc.* **45**, 2586 (1997).
- [22] I. Daubechies, J. Lu, and H.-T. Wu, Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool, *Appl. Comput. Harmonic Anal.* **30**, 243 (2011).
- [23] G. Thakur, E. Brevdo, N. S. Fuckar, and H.-T. Wu, The Synchrosqueezing algorithm for time-varying spectral analysis: Robustness properties and new paleoclimate applications, *Sig. Process.* **93**, 1079 (2013).
- [24] Y.-C. Chen, M.-Y. Cheng, and H.-T. Wu, Nonparametric and adaptive modeling of dynamic periodicity and trend with heteroscedastic and dependent errors, *J. R. Stat. Soc. B* **76**, 651 (2014).
- [25] L. W. Sheppard, A. Stefanovska, and P. V. E. McClintock, Detecting the harmonics of oscillations with time-variable frequencies, *Phys. Rev. E* **83**, 016206 (2011).
- [26] The word “Nonlinear” in NMD does not have any relation to classical nonlinear analysis, but simply reflects the fact that the method decomposes the signal into oscillations with complex wave forms (a feature that commonly appears due to nonlinearities in either the system or the measurement apparatus); see also (4) and related discussion.
- [27] T. Schreiber and A. Schmitz, Surrogate time series, *Phys. D (Amsterdam, Neth.)* **142**, 346 (2000).
- [28] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. Doyne Farmer, Testing for nonlinearity in time series: the method of surrogate data, *Phys. D (Amsterdam, Neth.)* **58**, 77 (1992).
- [29] R. G. Andrzejak, A. Kraskov, H. Stögbauer, F. Mormann, and T. Kreuz, Bivariate surrogate techniques: Necessity, strengths, and caveats, *Phys. Rev. E* **68**, 066202 (2003).
- [30] R. Q. Quiroga, A. Kraskov, T. Kreuz, and P. Grassberger, Performance of different synchronization measures in real data: a case study on electroencephalographic signals, *Phys. Rev. E* **65**, 041903 (2002).
- [31] B. Picinbono, On instantaneous amplitude and phase of signals, *IEEE Trans. Signal Proc.* **45**, 552 (1997).
- [32] B. Boashash, Estimating and interpreting the instantaneous frequency of a signal. I. Fundamentals, *Proc. IEEE* **80**, 520 (1992).
- [33] B. Boashash, Estimating and interpreting the instantaneous frequency of a signal. II. algorithms and applications, *Proc. IEEE* **80**, 540 (1992).
- [34] H.-T. Wu, Instantaneous frequency and wave shape functions (i), *Appl. Comput. Harmonic Anal.* **35**, 181 (2013).
- [35] Note that, in place of the FT surrogates used here, one can instead utilize the more advanced IAAFT surrogates [36]. For this particular task, however, we have found that the performance of both types is the same. In fact, the outcome of the test seems to be determined primarily by the “goodness” of the component representation in TFR, which is affected only by the noise power and its distribution in the time-frequency region around the component’s instantaneous frequency.
- [36] T. Schreiber and A. Schmitz, Improved surrogate data for nonlinearity tests, *Phys. Rev. Lett.* **77**, 635 (1996).
- [37] H. G. E. Hentschel and I. Procaccia, The infinite number of generalized dimensions of fractals and strange attractors, *Phys. D (Amsterdam, Neth.)* **8**, 435 (1983).
- [38] P. Grassberger, Generalized dimensions of strange attractors, *Phys. Lett. A* **97**, 227 (1983).
- [39] J. D. Farmer, E. Ott, and J. A. Yorke, The dimension of chaotic attractors, *Phys. D (Amsterdam, Neth.)* **7**, 153 (1983).
- [40] P. Grassberger and I. Procaccia, Measuring the strangeness of strange attractors, *Phys. D (Amsterdam, Neth.)* **9**, 189 (1983).
- [41] P. Grassberger and I. Procaccia, Characterization of strange attractors, *Phys. Rev. Lett.* **50**, 346 (1983).
- [42] J. Theiler, Efficient algorithm for estimating the correlation dimension from a set of discrete points, *Phys. Rev. A* **36**, 4456 (1987).
- [43] J. Theiler and D. Prichard, Constrained-realization Monte-Carlo method for hypothesis testing, *Phys. D (Amsterdam, Neth.)* **94**, 221 (1996).



- [44] J. Theiler and T. Lookman, Statistical error in a chord estimator of correlation dimension: the “rule of five”, *Int. J. Bifurcation Chaos Appl. Sci. Eng.* **03**, 765 (1993).
- [45] C. J. Stam, J. P. M. Pijn, and W. S. Pritchard, Reliable detection of nonlinearity in experimental time series with strong periodic components, *Phys. D (Amsterdam, Neth.)* **112**, 361 (1998).
- [46] Interestingly, the (E)EMD [4, 5] has logarithmic frequency resolution [47, 48] and therefore suffers from the same drawbacks as the WT in relation to representation of harmonics. Note also that, while for the WT the logarithmic frequency resolution can be tuned by changing the resolution parameter  $f_0$  in (6), for (E)EMD the resolution is fixed around some implicit value [47, 48].
- [47] M. Feldman, Analytical basics of the EMD: Two harmonics decomposition, *Mech. Syst. Signal Proc.* **23**, 2059 (2009).
- [48] G. Rilling and P. Flandrin, One or two frequencies? The empirical mode decomposition answers, *IEEE Trans. Signal Proc.* **56**, 85 (2008).
- [49] G. B. Folland and A. Sitaram, The uncertainty principle: a mathematical survey, *J. Fourier Anal. Appl.* **3**, 207 (1997).
- [50] G. Kaiser, *A Friendly Guide to Wavelets* (Birkhäuser, Boston, 1994).
- [51] G. E. Nilsson, T. Tenland, and P. A. Oberg, Evaluation of a laser Doppler flowmeter for measurement of tissue blood flow, *IEEE Trans. Biomed. Eng.* **BME-27**, 597 (1980).
- [52] A. Stefanovska and M. Bračič, Physics of the human cardiovascular system, *Contemp. Phys.* **40**, 31 (1999).
- [53] A. Stefanovska, M. Bračič, and H. D. Kvernmo, Wavelet analysis of oscillations in the peripheral blood circulation measured by laser Doppler technique, *IEEE Trans. Biol. Med. Eng.* **46**, 1230 (1999).
- [54] P. Kvandal, S. A. Landsverk, A. Bernjak, A. Stefanovska, H. D. Kvernmo, and K. A. Kirkeboen, Low-frequency oscillations of the laser Doppler perfusion signal in human skin, *Microvasc. Res.* **72**, 120 (2006).
- [55] J. P. Saul, R. D. Berger, P. Albrecht, S. P. Stein, M. H. Chen, and R. J. Cohen, Transfer function analysis of the circulation: Unique insights into cardiovascular regulation, *Am. J. Physiol.: Heart Circ. Physiol.* **261**, H1231 (1991).
- [56] N. Japundžić, M.-L. Grichois, P. Zitoun, D. Laude, and J.-L. Elghozi, Spectral analysis of blood pressure and heart rate in conscious rats: effects of autonomic blockers, *J. Auton. Nerv. Syst.* **30**, 91 (1990).
- [57] S. C. Malpas, Neural influences on cardiovascular variability: possibilities and pitfalls, *Am. J. Physiol.: Heart. Circ. Physiol.* **282**, H6 (2002).
- [58] C. Julien, The enigma of Mayer waves: Facts and models, *Cardiovasc. Res.* **70**, 12 (2006).
- [59] A. Stefanovska, P. V. E. McClintock, J. Ræder, and A. F. Smith (eds.), *Nonlinear Dynamics of Anaesthesia: From Theory to Clinical Practice* (unpublished).
- [60] M. Bračič Lotrič, A. Stefanovska, D. Štajer, and V. Urbančič-Rovan, Spectral components of heart rate variability determined by wavelet analysis, *Physiol. Meas.* **21**, 441 (2000).
- [61] G. Srivastava, S. Crottaz-Herbette, K. M. Lau, G. H. Glover, and V. Menon, ICA-based procedures for removing ballistocardiogram artifacts from EEG data acquired in the MRI scanner, *Neuroimage* **24**, 50 (2005).
- [62] P. T. Clemson and A. Stefanovska, Discerning non-autonomous dynamics, *Phys. Rep.* **542**, 297 (2014).
- [63] T. Stankovski, A. Duggento, P. V. E. McClintock, and A. Stefanovska, Inference of time-evolving coupled dynamical systems in the presence of noise, *Phys. Rev. Lett.* **109**, 024101 (2012).
- [64] A. Duggento, T. Stankovski, P. V. E. McClintock, and A. Stefanovska, Dynamical Bayesian inference of time-evolving interactions: From a pair of coupled oscillators to networks of oscillators, *Phys. Rev. E* **86**, 061126 (2012).
- [65] Y. F. Suprunenko, P. T. Clemson, and A. Stefanovska, Chronotaxic systems: a new class of self-sustained non-autonomous oscillators, *Phys. Rev. Lett.* **111**, 024101 (2013).
- [66] Y. F. Suprunenko, P. T. Clemson, and A. Stefanovska, Chronotaxic systems with separable amplitude and phase dynamics, *Phys. Rev. E* **89**, 012922 (2014).
- [67] C.-K. Peng, Sergey V. Buldyrev, Shlomo Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger, Mosaic organization of DNA nucleotides, *Phys. Rev. E* **49**, 1685 (1994).
- [68] T. Schreiber and A. Schmitz, Discrimination power of measures for nonlinearity in a time series, *Phys. Rev. E* **55**, 5443 (1997).
- [69] N. E. Huang and Z. Wu, A review on Hilbert-Huang transform: method and its applications to geophysical studies, *Rev. Geophys.* **46**, RG2006 (2008).
- [70] The NMD Toolbox for MATLAB, which contains all the needed codes, is freely available at <http://www.physics.lancs.ac.uk/research/nbmphysics/diats/nmd/>.