Physics Letters A 384 (2020) 126307

Contents lists available at ScienceDirect

Physics Letters A

www.elsevier.com/locate/pla

Disentangling stochastic signals superposed on short localized oscillations

Gerrit Kampers [a,*], Matthias Wächter [a], Michael Hölling [a], Pedro G. Lind [b,c], Sílvio M.D. Queirós [d], Joachim Peinke [a]

[a] ForWind, Institute of Physics, University of Oldenburg, Küpkersweg 70, 26129 Oldenburg, Germany
[b] Department of Computer Science, Oslo Metropolitan University, P.O. Box 4 St. Olavs plass, N-0130 Oslo, Norway
[c] Instituto Universitário de Lisboa (ISCTE-IUL), ISTAR-IUL, Av. Forças Armadas, 1649-026 Lisboa, Portugal
[d] Centro Brasileiro de Pesquisas Físicas, Rua Dr. Xavier Sigaud 150, 22290-180 Rio de Janeiro, RJ, Brazil

ARTICLE INFO

Article history:
Received 28 August 2019
Received in revised form 16 January 2020
Accepted 2 February 2020
Available online 5 February 2020
Communicated by M. Perc

Keywords:
Empirical mode decomposition
Signal superposition
Stochastic modeling

ABSTRACT

We introduce a procedure for separating periodic oscillations superposed on a stochastic signal. The procedure combines empirical mode decomposition (EMD) of a signal with tools of data analysis based on stochastic differential equations, namely nonlinear Langevin equations. Taking the set of modes retrieved from the EMD of the signal, our procedure is able to separate them into two groups, one composing the periodic signal and another composing the stochastic signal. The framework is robust for a broad family of localized oscillations, in the range of large frequencies. In particular, we show that, in this context, the EMD method outperforms a low-pass filter and is robust for a wide interval of different frequency ranges and amplitudes of the periodic oscillation, as well as for a broad family of different non-linear Langevin processes.

1. Introduction

When studying natural phenomena one often needs to handle signals that result from the superposition of two or more processes, which should be properly resolved to enable assessing each process separately. A fundamental example is the superposition of measurement noise (also known as observation noise) when measuring a stochastic variable [19]. The measurement noise results from the measurement process and measurement devices and therefore should be distinguished from the noisy contributions belonging to the natural stochastic process. To this end, stochastic methods have been developed for resolving the superposition of natural stochastic processes and different measurement (additive) noise sources, namely uncorrelated [2] and correlated [9] with an extension for an arbitrary number of measurement noise sources [10,17]. In parallel, these methods have been extended to non-Markov processes, namely Langevin-type processes whose stochastic force is governed by a colored noise, e.g. an Orstein-Uhlenbeck noise [11].

Complementary to these situations, there are cases where the stochastic process under study is superposed on oscillatory signals. On one hand, extracting the periodic contributions of one process provides information e.g. the seasonality where one often regards the stochastic contributions as noise that should be filtered out. Typical tools in this scope are the Fourier filters [8] or the Discrete Wavelet Transform [1]. On the other hand, stochastic contributions are important when the property being analyzed reflects intrinsic features of stochastic dynamics. This is the case of several turbulence experiments, where periodic contributions in the measured signal often results from periodic signals that act as perturbations to the stochastic signal characterizing the turbulent behavior. For example, when addressing measurements of loads on airfoils, one usually considers their lift and drag component. Due to the inflow conditions, lift and drag forces may show a considerable amount of periodic oscillations that are taken as perturbations of the underlying signal and must be filtered out [12].

If the periodic perturbations are typically high-frequency oscillations, a low-pass filter would be a good tool for filtering them out. If they would be located at the low-frequency range of the signal's spectrum a high-pass filter would properly filter them out. However, many perturbations occur in the middle frequency range and often one needs to filter out the spectrally localized oscillations without affecting the original spectrum of the signal. An illustration of a such superposed signal with a stochastic and an oscillatory contribution is shown in Fig. 1. One often deals with such a superposed signal, Fig. 1(a), and to gain access to the stochastic contribution, Fig. 1(b) one needs to filter out the oscillatory contri-

* Corresponding author.
E-mail address: g.kampers@uni-oldenburg.de (G. Kampers).

https://doi.org/10.1016/j.physleta.2020.126307
0375-9601/

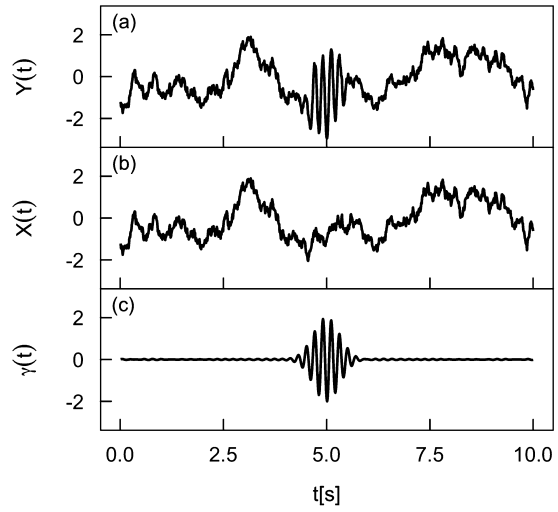

Fig. 1. Illustration of **(a)** a stochastic signal superposed with a periodic oscillation, **(b)** the stochastic signal alone and **(c)** the periodic oscillation resulting from the superposition of single-frequency signals with amplitudes A_i parameterized by a maximal value α and a bandwidth σ (see equations (5) and (6) and Fig. 2).

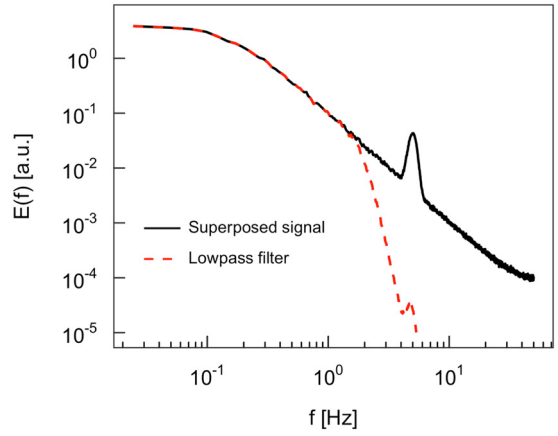

Fig. 2. Power spectrum of the stochastic signal superposed with periodic oscillations (black solid line) as illustrated in Fig. 1 and defined in equations (1), together with equations (2), (5) and (6). The effect of a low pass filter is on this signal is shown as red dashed line. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

bution, Fig. 1(c), or vice-versa. Fig. 2 shows the effect of a standard filtering procedure, namely a lowpass filter, using the power spectra of the superposed signal introduced in Fig. 1(a). As one sees, a standard filter such as a low-pass filter removes the full range of high frequencies, affecting also the stochastic process that cannot in this way be separated from the periodic oscillation. The contribution of the periodic oscillations is strongly attenuated but not removed from the signal.

In this study we introduce a method that is able to resolve the superposition of short localized oscillations from an underlying stochastic signal, retaining the power-law spectrum of the stochastic signal in the correct range of frequencies. The method combines the Empirical Mode Decomposition (EMD), introduced by Huang [7], and tools for deriving stochastic differential equations from series of data that were introduced by Siefert et al. [18] and developed till recently [3].

We start in section 2 by presenting the models considered for describing both the stochastic and the periodic components of a signal. In section 3 we introduce our framework that combines EMD and stochastic differential equations. The full framework is then applied to a specific case in section 4 and in section 5 we address the robustness of our framework by assessing its accuracy in resolving the superposition for other more general cases. Conclusions and outlook are given in section 6.

2. Model and assumptions

We consider the family of signals of the form

$$Y(t) = X(t) + \gamma(t), \tag{1}$$

where $X(t)$ and $\gamma(t)$ are two superposed components, yielding a measurable signal $Y(t)$. In the case both the components are stochastic, one falls into the case of additive measurement noise that was addressed in Refs. [19,2,9,10,17]. Here, we consider component $X(t)$ as a stochastic process and $\gamma(t)$ as a periodic component. The stochastic process is governed by a non-linear Langevin equation [4], namely

$$\frac{dX}{dt} = D^{(1)}(X) + \sqrt{D^{(2)}(X)}\Gamma(t), \tag{2}$$

with $D^{(1)}(X)$ and $D^{(2)}(X)$ being the drift and diffusion coefficients, typically low-order polynomials of X, and Γ_t being a Gaussian δ-correlated white noise fulfilling $\langle\Gamma_t\rangle = 0$ and $\langle\Gamma_t\Gamma_{t'}\rangle = 2\delta(t-t')$. Equation (2) is easily integrated using the Itô interpretation of the stochastic term [4]. Inversely, the Langevin equation can be directly extracted from a series of a Langevin process, by computing both the drift and diffusion coefficients [3]. Recently, an *R*-package was developed [15] that enables to extract drift and diffusion coefficients from series of measurements of a stochastic variable fulfilling equation (2)[1] Therefore we can consider the following parameterizations

$$D^{(1)}(X) = aX + bX^3\,, \tag{3}$$

$$D^{(2)}(X) = c + dX^2\,, \tag{4}$$

where $a < 0$ if $b = 0$, $a > 0$ if $b < 0$, and $c, d > 0$. All these cases will be considered for testing our framework.

Notice that, since $D^{(1)}(X)$ has units of $[X/t]$ and $D^{(2)}(X)$ has units of $[X^2/t]$, a and d have units of inverse time, $[t^{-1}]$, b has units of $[(tX^2)^{-1}]$ and c has units of $[X^2/t]$.

The oscillatory component is defined as the superposition of n single-frequency periodic modes

$$\gamma(t) = \sum_{i=1}^{n} A_i \sin(2\pi f_i t + \Phi_i)\,, \tag{5}$$

with amplitudes A_i, frequencies f_i and phases Φ_i. The amplitudes A_i are defined to follow a Gaussian function of the respective frequency, namely

$$A_i = \frac{\alpha}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(f_i-\mu)^2}{2\sigma^2}\right), \tag{6}$$

with the parameter σ scaling the amplitude distribution in frequency space. To relate the maximum amplitude $\gamma_{\max}$ of the oscillation to the standard deviation σ_X of $X(t)$,[2] we define the normalization parameter $\alpha = \gamma_{\max}/\sigma_X$. This parameter scales the

[1] The drift coefficient is typically a polynomial of odd order, whose coefficient of the highest power is negative. In this way, the outermost fixed points are stable. As for the diffusion coefficient, since it has to be positive semi-definite, one typically takes a constant (additive noise) or a quadratic polynomial in X (multiplicative noise).

[2] The value of σ_X can be derived from the stationary solution of the stochastic process for a given $D^{(1)}(X)$ and $D^{(2)}(X)$ as defined in equations (3) and (4).

amplitudes of the periodic oscillations with respect to the amplitude of the stochastic process and is addressed at a later point in the paper to evaluate the robustness of the proposed framework.

Taking both contributions together, $X(t)$ and $\gamma(t)$, one has six parameters, four for $X(t)$ (a, b, c and d in equations (3) and (4) and two for the periodic component (α and σ). Fig. 1(a) illustrates one superposed signal $Y(t)$ as defined in equation (1), where the stochastic part (Fig. 1(b)) is given by equation (2) together with (3) and (4) with $a = -1$, $b = 0$, $c = 1$ and $d = 0$ in proper units (see above), and the periodic component (Fig. 1(c)) is given by equations (5) and (6) with $\alpha = 2$ and $\sigma = 0.5$. We chose $n = 200$ and varied the frequencies f_i uniformly between 2 Hz and 8 Hz, having therefore a mean of $\mu = 5$ Hz. The phases Φ_i are randomly chosen following a normal distribution with a standard deviation of 0.1. These parameters result in a periodically occurring sweep in time and a broadband frequency peak in the spectrum. The signal is composed by 10^6 samples with a sampling frequency of 100 Hz. In Fig. 2 we show the power spectrum of the superposed signal plotted in Fig. 1(a) (solid line). The stochastic process is characterized by a power-law decay in the spectrum, whereas the periodic contribution $\gamma(t)$ is responsible for the peaked band around 5 Hz.

3. Method: the Huang-Langevin decomposition

The oscillatory disturbance is clearly visible in the superimposed time series illustrated in Fig. 1(a). In this section, we describe the method for resolving this superposition, which is composed of two steps. In step one, we apply a filter, based on EMD, to separate the signal into its stochastic and periodic components. After the filtering we extract the drift and diffusion coefficients for the stochastic contribution and compare the results to the known reference stochastic process $X(t)$ and the oscillatory component $\gamma(t)$.

3.1. The Huang-Langevin filter

The filter we introduce below is based on the fact that a stochastic process governed by equation (2) is typically Markovian, whereas periodic oscillations are not.

Therefore, the sum of the "non-markovian modes" shall reconstruct the periodic oscillation $\gamma(t)$, while the sum of all other modes shall retrieve the stochastic signal $X(t)$. It was recently shown [5] that EMD is helpful for reconstructing non-stationary contribution in stochastic signals. Therefore it is also appealing to use it for our purposes exposed above. For the mode decomposition of $Y(t)$ we use a variant of EMD. EMD is a model-free procedure that is able to decompose a process into its different intrinsic oscillation modes, so-called intrinsic mode functions (IMF), forming a complete and nearly orthogonal basis for the original signal [7]. Differently from the Fourier filter and the discrete wavelet transform, the EMD has no analytic definition. EMD expands one given process in its "natural" constitutive scales following a sifting algorithm. One starts by identifying of local extrema of the original data, and derive a cubic spline for the envelope connecting all maxima and the envelope connecting all minima. The upper and lower envelopes cover the full data set. The difference between the data set $Y(t)$ and the mean $m_1^{(1)}(t)$ of the maxima and minima envelopes yields a new signal $h_1^{(1)}(t)$ on which the same procedure can be applied to. By repeating iteratively this procedure one obtains a succession of signals $h_i^{(1)}$ for $i = 1, \ldots, k$. This sifting procedure is stopped by the S-number criterion proposed in [6], such that the number of zero crossings and extrema of $h_i^{(1)}$ differ at most by one for S consecutive iterations. We choose a value of k not larger than 50. The first IMF component of the data is then $C_1(t) = h_{k-1}^{(1)} - m_{k-1}^{(1)}$. The second IMF, $C_2(t)$, is then obtained by applying the same procedure to the signal $Y^{(1)}(t) = Y(t) - C_1(t)$. An improved version of this standard implementation, that we use in the following, is called Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) [20]. CEEMDAN applies the mode decomposition to an ensemble of white noise perturbed initial signals, for each IMF separately, yielding more meaningful IMFs while maintaining the completeness of the decomposition. Our CEEMDAN uses an R-package recently available at CRAN [13]. The stopping criterion is set to S-number = 4 and the noise strength is 0.1 standard deviations relative to $Y(t)$. Note that each IMF is more general than simple harmonic functions, since, instead of constant amplitude and frequency, an IMF can have variable amplitude and frequency along the time axis. Fig. 3 shows the first six IMFs of signal $Y(t)$ illustrated in Fig. 1(a). From the EMD of $Y(t)$, one obtains a set of IMFs, $C_i(t)$ with $i = 1, \ldots, m$ and a residual $R(t)$, such that

$$Y(t) = R(t) + \sum_{i=1}^{m} C_i(t)\,. \tag{7}$$

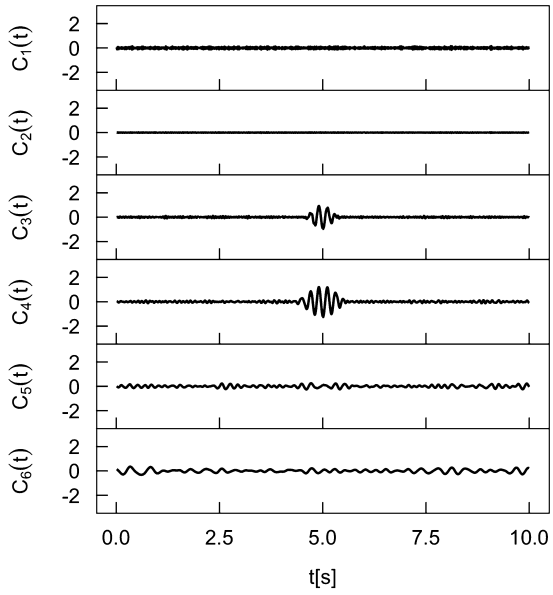

Fig. 3. Empirical mode decomposition (using variant CEEMDAN) of the stochastic signal superposed with a periodic oscillation $Y(t)$ shown in Fig. 1(a). Here only IMFs $C_1 - C_6$ are shown.

To determine how many periodic "modes" are reasonable to consider for a certain given process, the modes to be filtered out have been selected by direct inspection [5]. Below, we extend the methodology to automatically select the modes to be removed. To that end, we sort the IMFs $C_i(t)$ into two groups, one that adds up to recover the stochastic signal $X(t)$, and another that reconstructs $\gamma(t)$. The extraction of a full hierarchy of IMFs is done iteratively. Stochastic fluctuations are in a certain sense also oscillations, and can be easily confused with short-time periodic oscillations. Fig. 4 compares the spectra of signal $Y(t)$, with the signal after removing two illustrative IMFs, namely $C_4(t)$ and $C_6(t)$ (see Fig. 3). The effect of the removal of the IMF combination $C_3(t)$ and $C_4(t)$ is also shown and is discussed below. While removing $C_4(t)$ decreases significantly the peaked band, removing $C_6(t)$ has no effect on the peaked band. In this way, the EMD-based decomposition of a signal can be used as a band-pass filter. If the oscillations are typically in the range of large frequencies, one would expect a combination of

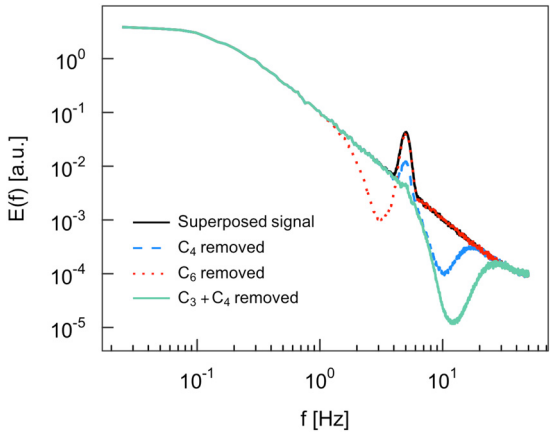

Fig. 4. Spectra of the signal after removing IMF $C_4(t)$ (dashed blue line), after removing IMF $C_6(t)$ (dotted red line) and after removing IMF $C_3(t) + C_4(t)$ (solid green line) compared with the spectrum of original signal $Y(t)$ (solid black line) illustrated in Fig. 1. The IMFs are plotted in Fig. 3.

IMF of low order to represent the oscillatory contribution and vice versa. We evaluate the best combination of IMF that represent the stochastic contribution, by analyzing the Markov property of the signal $Y^{(i)}(t) = Y(t) - C_i(t)$ for each function $C_i(t)$. The Markov property of $Y^{(i)}(t)$ on a time-scale τ is given by the equality of two conditional probability density functions (PDF), namely

$$\rho_\tau^{(i)}\left(y_0|y_1,\tau\right) = \eta_\tau^{(i)}\left(y_0|y_1,\tau;\ y_2,2\tau;\ldots;y_n,n\tau\right), \tag{8}$$

where y_n are specific values of the time shifted signal $Y^{(i)}(t-n\tau)$. Equation (8) is the mathematical definition of a Markov process. In practice such definition is not implementable, since it implies a comparison with all possible combination of previous values. Assuming that the dependency on previous values decays monotonically with the time-lag $n\tau$, one uses instead the simpler form

$$\rho_\tau^{(i)}\left(y_0|y_1,\tau\right) = \eta_\tau^{(i)}\left(y_0|y_1,\tau;\ y_2,2\tau\right), \tag{9}$$

with y_0 being a value in the range of observed values of the series $Y^{(i)}(t)$. Monotonic decrease of the dependency on previous steps ensures that if equation (9) is fulfilled, equation (8) also holds. The fulfillment of equation (8) is quantified by the so-called weighted mean square error in logarithmic space [14]:

$$\varepsilon_\tau^{(i)}(y_1,y_2) = \frac{\int \zeta_\tau^{(i)}\left(\ln(\rho_\tau^{(i)}) - \ln(\eta_\tau^{(i)})\right)^2 \mathrm{d}y_0}{\int \zeta_\tau^{(i)}\left(\ln^2(\rho_\tau^{(i)}) + \ln^2(\eta_\tau^{(i)})\right)\mathrm{d}y_0}, \tag{10}$$

where $\zeta_\tau^{(i)} = \rho_\tau^{(i)} + \eta_\tau^{(i)}$. The metric $\varepsilon_\tau^{(i)}$ is then averaged over y_1 and y_2

$$\varepsilon_\tau = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \varepsilon_\tau^{(i)}(y_1,y_2)w(y_1,y_2)\mathrm{d}y_1\mathrm{d}y_2, \tag{11}$$

where

$$w(y_1,y_2) = \frac{\sqrt{n(y_1,y_2)}}{\int\int\sqrt{n(y_1,y_2)}\mathrm{d}y_1\mathrm{d}y_2}, \tag{12}$$

where $n(y_1,y_2)$ denotes the number of counts in each conditional PDF $\eta_\tau^{(i)}$. The metric ε_τ is evaluated for l different time-scales of interest, and therefore we use the average

$$\varepsilon = \frac{1}{l}\sum_{\tau=1}^{l}\varepsilon_\tau \tag{13}$$

as the final measure to test for the Markov property.

Fig. 5 illustrates the comparison of conditional and double conditional PDFs for one time-scale $\tau = 0.1$ s, related to equation (8). The comparison is shown for (a) the superposed signal $Y(t)$ and (b) the superposed signal with C_4 removed $Y^{(4)}(t)$. For the superposed signal the PDFs clearly not match, as expected by the presence of the oscillatory component. The removal of C_4 only slightly improves the equality, as it contains a part of the oscillatory contribution (cf. Figs. 3 and 4). The visible improvement on this time-scale is only marginal in this representation, also implied by the values of ε_τ given in the figure. As other IMFs also might contain oscillatory contributions, we remove all possible combinations of IMFs and evaluate the value of ε by averaging ε_τ over different time-scales. The combination removal that minimizes ε in equation (13) is the selected combination that resolves the periodic oscillation and, simultaneously, all other modes sum up composing the stochastic signal.

3.2. Resolving the stochastic contribution when superposed with periodic oscillations

From the filtering procedure described previously, one obtains estimations of both $X(t)$ and $\gamma(t)$ components of signal $Y(t)$ (see equation (1)).

The periodic component $\gamma(t)$ is defined through the estimation of parameters α and σ (see equations (5) and (6)). Parameter σ can be estimated from the inspection of the power spectrum, since it scales with the width of the peaked band of frequencies

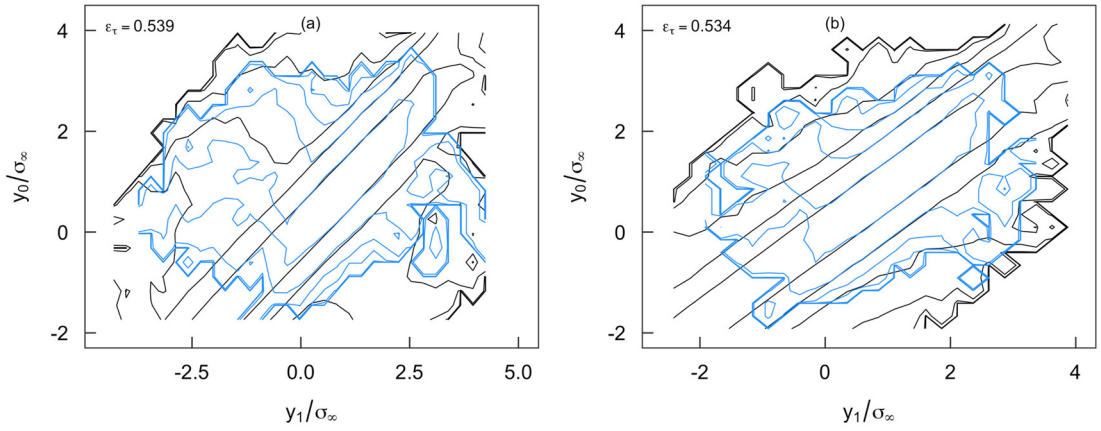

Fig. 5. (Color online) Contour plots of the conditional PDF (black) and double conditional PDF (blue) according to equation (8). (a) $Y(t)$, (b) $Y^{(4)}(t) = Y(t)) - C_4(t)$. The time-scale is $\tau = 0.1$ s. For clarity of presentation, only the plane at $y_2 = \sigma(y_2)$ is shown here.

(cf. Fig. 2). Parameter $\alpha = \gamma_{\max}/\sigma_X$ is given by the maximum amplitude of the periodic component $\gamma_{\max}$ and the standard deviation σ_X of the stationary solution of the stochastic component $X(t)$, whose PDF is given by

$$\rho_{\text{stat}}(x) = \frac{1}{D^{(2)}(x)} \exp\left(\int_x \frac{D^{(1)}(x')}{D^{(2)}(x')} dx' \right). \tag{14}$$

This stationary solution is obtained from the Fokker-Planck equation [16]

$$\frac{\partial \rho(x,t)}{\partial t} = -\frac{\partial}{\partial x}\left(D^{(1)}(x)\rho(x,t) \right) + \frac{\partial^2}{\partial x^2}\left(D^{(2)}(x)\rho(x,t) \right). \tag{15}$$

For the stationary solution one substitutes in equation (15) one has $\rho(x,t) \equiv \rho_{\text{stat}}(x)$ and therefore $\frac{\partial \rho_{\text{stat}}(x)}{\partial t} \equiv 0$, yielding $D^{(1)}(x)\rho(x,t) - \frac{\partial}{\partial x}\left(D^{(2)}(x)\rho(x,t) \right) = K$ being $K = 0$ since distribution $\rho(x,t)$ needs to be zero for $x \to \infty$. Dividing both terms by $D^{(2)}(x)\rho(x,t)$, integration leads straightforwardly to equation (14).

The stochastic component $X(t)$ is modeled by equation (2) and can be derived directly from the series of measurements following what has been called Langevin approach [15]: using the definition of the drift and diffusion coefficients as time derivatives of the first and second conditional moments, namely

$$\tilde{D}^{(n)}(x) = \frac{1}{n!} \lim_{\tau \to 0} \frac{1}{\tau} \tilde{M}^{(n)}(x,\tau), \tag{16}$$

with $\tilde{M}^{(n)}$ representing the estimated nth conditional moment defined as

$$\tilde{M}^{(n)}(x,\tau) = \left\langle (X(t+\tau) - X(t))^n \right\rangle_{X(t)=x}. \tag{17}$$

If the process is sufficiently sampled, obeys the Markov condition and is not spoiled by external sources of noise [9,10], one typically observes a linear dependence of the condition moments for the lowest range of τ values [15]. In practice one computes both conditional moments for the full range of observed values of X and for a few values of τ (typically between three and five times the inverse of the sampling frequency) and then solves the set of equations

$$M^{(1)}(X,\tau) = D^{(1)}(X)\tau, \tag{18}$$

$$M^{(2)}(X,\tau) = 2D^{(2)}(X)\tau + (D^{(1)}(X)\tau)^2, \tag{19}$$

in an autoregressive sense.

Details on the derivation of equation (2) from a stochastic process are described in [15] and a routine in R that implements this part of the procedure is available at https://cran.r-project.org/web/packages/Langevin/. Typically one finds drift and diffusion coefficients that are well fitted by polynomials as given in equations (3) and (4).

4. Application of the method to sample data

To test the method described in the previous sections, we start by considering the illustrative process introduced in section 2. Fig. 6 shows the measure ε_τ, defined in equation (10), for evaluating the Markovianity of a process. Each row corresponds to a time-scale τ, see equation (8), and each column to a combination of IMFs excluded in the EMD reconstruction. It is apparent, that the lowest values of ε_τ are achieved with different combinations of excluded modes for different τ. As the derivation of drift and diffusion coefficients following equations (16) and (17) involves a

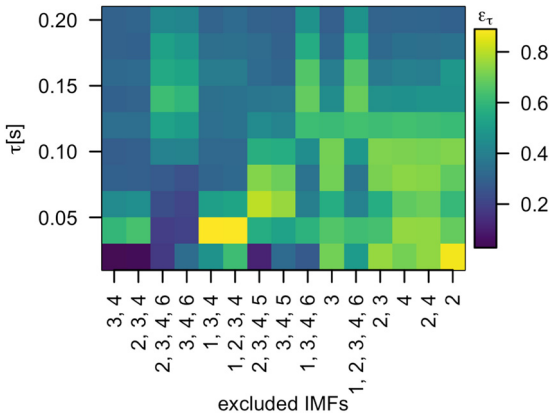

Fig. 6. Metric ε_τ, equation (10), for evaluating the Markovianity of a process after excluding specific IMFs. The horizontal line indicates the combination of IMFs that are filtered out, while the vertical line shows the value of τ used for computing the respective conditional PDFs, ρ_τ and η_τ, see equation (8). Here, τ was varied from 0.02 s to 0.2 s in steps of 0.02 s.

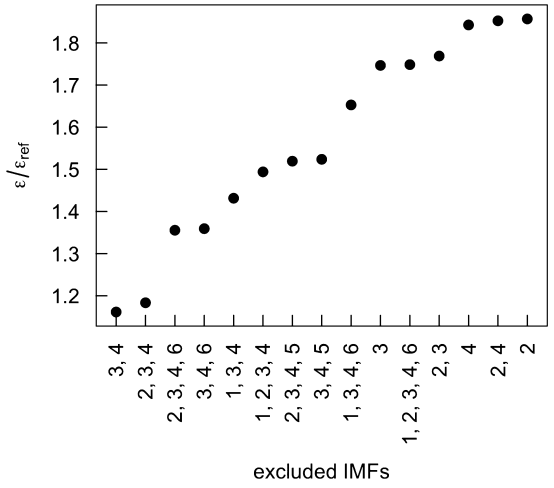

Fig. 7. Average distance ε of the metric ε_τ plotted in Fig. 6, see equation (13). The values are normalized by ε_{ref} of the unperturbed stochastic process $X(t)$.

range of τ, the combinations of excluded modes are sorted by increasing average values ε over the values of τ, as shown in Fig. 7, cf. equation (13). Here we normalized by $\varepsilon_{ref} \approx 0.24$, referred to the unperturbed stochastic process $X(t)$. For the perturbed and lowpass filtered signal (cf. Fig. 2), this measure amounts to $\varepsilon_L/\varepsilon_{ref} \approx 2.46$.

This representation clearly shows that the minimal value of ε is achieved by removing the third and the fourth IMFs from the signal. These modes spoil the Markov properties of our sample data $Y(t)$ the most and therefore are likely to contain most of the oscillatory part of the signal. Consequently, we separate our sample data into the extracted oscillatory contribution

$$\tilde{\gamma}(t) = C_3(t) + C_4(t)$$

and the stochastic contribution includes all other IMFs:

$$\tilde{X}(t) = C_1(t) + C_2(t) + C_5(t) + \ldots + C_m(t) + R(t).$$

Here we choose $m = 11$. Beyond the fifth mode, no significant oscillatory contribution is observed [cf. figure (3)], and therefore we can assume that all oscillatory contributions are included in modes three and four. This assumption is supported by the average peak height $\tilde{\alpha} = 2.043$ of the sweeps in $\tilde{\gamma}(t)$, which is in good agreement with the maximum amplitude $\alpha = 2$ of the peri-

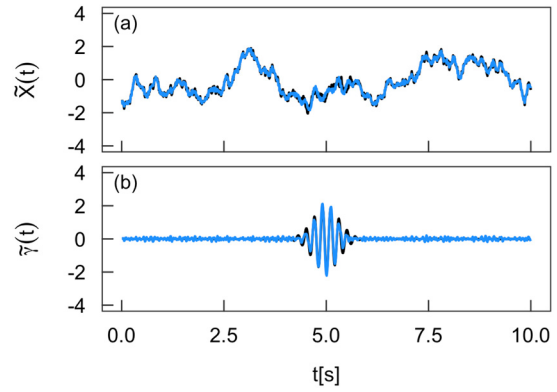

Fig. 8. (Color online) Comparison between segments of (a) the stochastic signal and (b) the periodic oscillations of the original superposition and the corresponding estimations after the decomposition.

odic oscillation $\gamma(t)$. Theoretically, one can take as much modes as needed. In our case, larger m lead to the same results.

Fig. 8 shows a sample of (a) the estimated stochastic contribution $\tilde{X}(t)$ and (b) the estimated periodic contribution $\tilde{\gamma}(t)$, resulting from the decomposition of the superposed process $Y(t)$. The stochastic signal $X(t)$ and the periodic oscillation $\gamma(t)$ (as shown in Fig. 1) are plotted as reference, clearly showing a very good match. The periodic contribution is almost completely removed from the estimated stochastic contribution, which is also apparent from the spectra (cf. Fig. 4).

Next we evaluate if this decomposition is also able to retain in the stochastic contribution the same statistical features of the original stochastic signal. After decomposing the original process into its estimated periodic and stochastic contributions, we perform the Langevin approach introduced in section 3.2 for modeling the stochastic contribution. For the quantitative comparison of the drift and diffusion coefficient estimates $\tilde{D}^{(n)}$ between the original stochastic signal and estimated stochastic contributions of the perturbed process, we use the weighted error

$$R = \sqrt{\left\langle \frac{\left(\tilde{D}^{(n)}(X) - \tilde{D}^{(n)}(\tilde{X})\right)^2}{\left(\frac{1}{2}\tilde{D}^{(n)}_{err}(X) + \frac{1}{2}\tilde{D}^{(n)}_{err}(\tilde{X})\right)^2} \right\rangle}, \tag{20}$$

where $D^{(n)}_{err}$ denotes the numerical errors of the coefficients, given by the previously mentioned R routine. If (in few bins) the numerical error becomes NaN, we cancel the denominator of the fraction for those bins, resulting in a standard RMS error. Fig. 9 shows the drift and diffusion coefficients of the stochastic contribution $X(t)$ of the original process (black lines), its estimate $\tilde{X}(t)$ (blue diamonds), the perturbed and lowpass filtered signal (red stars), as well as the original perturbed process $Y(t)$ (grey triangles). The estimated drift coefficient collapses very well with the original coefficients, showing that the deterministic term in the stochastic contribution is preserved under the decomposition of the original signal. As for the diffusion coefficients, one observes a small offset between the original stochastic contribution and its estimate, though agreeing in their functional shape. This offset in the estimated $D^{(2)}$ is consistent with the power spectra, since if one takes some modes out the integral of it get smaller than the one for the original time series (cf. equation (17)). The lowpass filter fails to recover the stochastic contribution from the superposed signal, which is supported by the significantly higher RMS errors. As expected, the estimated coefficients for the unfiltered superposed process $Y(t)$ (gray triangles) show a large deviation from the reference in both amplitude and functional shape, since the stochastic model is not suited to signals with oscillatory contributions.

Having both estimated coefficients, drift and diffusion, one is able to generate realizations of the estimated process $\tilde{X}(t)$, which can be taken as statistical reconstructions $X_{\mathrm{rec}}(t)$ of the stochastic contribution of the original process $X(t)$. Fig. 10(a) shows the PDF of the original stochastic contribution (solid black lines) and the respective estimate $\tilde{X}(t)$. Both PDFs overlap well. The same is observed for the power spectrum, as shown in Fig. 10(b), as well as for the two-point statistics, as illustrated in Fig. 10(c), where we further compare the distributions of the increments $\Delta_\tau X(t) = X(t+\tau) - X(t)$ and $\Delta_\tau \tilde{X}(t) = \tilde{X}(t+\tau) - \tilde{X}(t)$ for different time-lags, namely $\tau = 0.05$ s, 0.1 s, 0.5 s, 1 s and 10 s. All in all, the stochastic reconstruction shows a good quantitative and qualitative agreement with the original process, in terms of both one- and two-point statistics. The spectral agreement, previously shown in Fig. 2, further corroborates the accuracy of the method here proposed.

5. Robustness tests

In this section we address the robustness of the framework introduced in this paper, considering two separate cases: one where the oscillation parameters, α and σ are varied, and a second case where stochastic contributions with different dynamical features are considered.

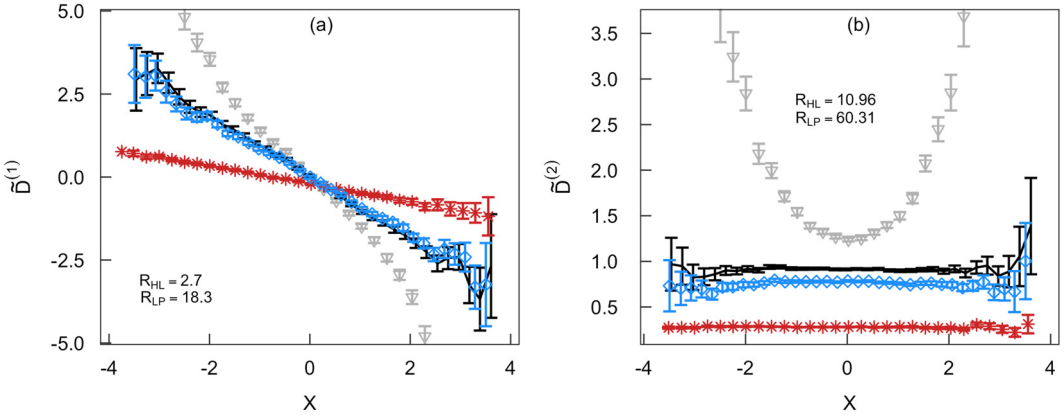

Fig. 9. Testing the efficiency of the Huang-Langevin decomposition for stochastic dynamics having a monostable drift and additive noise: $D^{(1)} = -X$ and $D^{(2)} = 1$. Comparison between the stochastic process $X(t)$ (black full line), the extracted stochastic contribution $\tilde{X}(t)$ (blue diamonds), the lowpass filtered signal (red stars) and the superposed signal $Y(t)$, with no filtering (gray triangles). (a) Estimated drift coefficient $\tilde{D}^{(1)}$ and (b) diffusion coefficient $\tilde{D}^{(2)}$. R_{HL} and R_{LP} denote the RMS errors for the Huang-Langevin and lowpass filters, respectively.

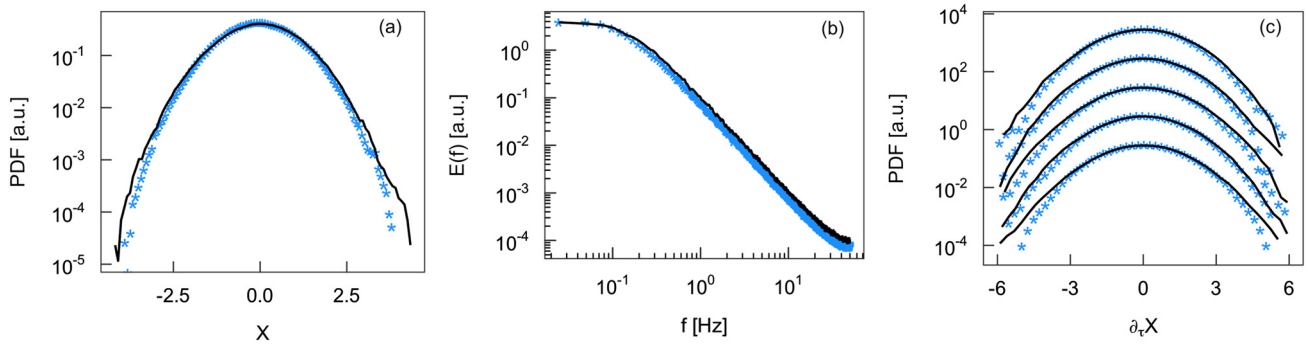

Fig. 10. Comparison of $X(t)$ (black solid lines) and the stochastic reconstruction $X_{\mathrm{rec}}(t)$ (blue stars), using the Langevin approach: (a) PDF, (b) spectra and (c) PDF of respective increments, $\Delta X_\tau(t) = X(t+\tau) - X(t)$ for $\tau = 0.05$ s, 0.1 s, 0.5 s, 1 s and 10 s.

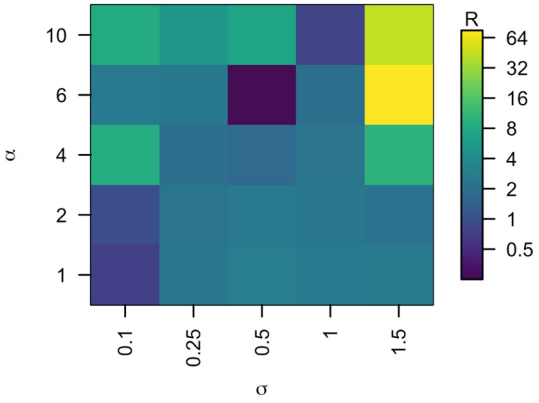

Fig. 11. Weighted error R (cf. equation (20)) between the estimated drift coefficients $\tilde{D}^{(1)}$ of the stochastic process $X(t)$ and the stochastic contribution of the perturbed and filtered process. The perturbation is varied by different oscillation parameters, σ and α (cf. equations (6) and (13)).

5.1. Varying the oscillation parameters

To check the robustness of our framework against the variation of the periodic contribution, we vary the amplitude α and standard deviation σ in the range $\alpha \in [1, 10]$ and $\sigma \in [0.1, 1.5]$ respectively, which corresponds to a wide range of different perturbations. The mean frequency $\mu = 5$ Hz is kept constant. For each parameter combination, the stochastic process $X(t)$, addressed above, (linear monostable drift and additive noise) is superimposed with the respective periodic contribution. The perturbed signals are then filtered following the procedure described previously, in section 4.

Here we take into account the series of IMFs $C_1, \ldots, C_6$ solely, from which we choose the combination of modes to represent the oscillatory contribution of the perturbed process. The perturbed processes are then separated into an oscillatory contribution (equation (5) and (6)) and the stochastic contribution (equation (2)). Time-scales τ_i are chosen as in section 4. The values of the error R (cf. equation (20)) between the estimated drift coefficients of the original and perturbed and filtered stochastic processes are shown in Fig. 11. For most combinations of the parameters σ and α, one observes small errors R, in the range of the previously shown case in Fig. 9 and lower, showing good agreement for the separated stochastic contribution. For combinations of small σ and large α, the estimated drift of the separated stochastic contribution tends to deviate more from the original (error bars do not overlap). For $\sigma = 1.5$ and $\alpha > 6$ the filtering method fails to recover the true stochastic contribution. Especially problematic are therefor cases where large values of the normalized amplitude, namely $\alpha > 4$, are combined with a wide frequency bandwidth. For small values of the normalized parameter $\alpha \leq 2$, the proposed filter allows for a good estimation of the deterministic dynamics of the perturbed stochastic process.

5.2. Variation of the stochastic dynamics of the stochastic signal

To evaluate the robustness of our framework against different stochastic dynamics, we use the periodic component introduced in section 2 ($\mu = 5$ Hz, $\alpha = 2$ and $\sigma = 0.5$) in two different cases: in one we superpose the periodic component on stochastic dynamics composed by a linear monostable drift ($D^{(1)}(X) = aX$) and in the other we consider a bistable drift ($D^{(1)}(X) = -X^3 + aX$). We vary a in $[-1, -10]$ for the monostable case, increasing the drift, and for the bistable case in $[1, 10]$, increasing the bistability of the drift. For both cases, the noise $D^{(2)}(X) = 1 + dX^2$ is varied with d in $[0, 1]$, covering additive and multiplicative noise with different noise levels. All superposed stochastic processes are filtered, following the procedure described in section 4 and the errors R (equation (16)) between the estimated drift coefficients of the original and perturbed and filtered stochastic processes are calculated for each filtered process, analogue to section 5.1. The results are plotted in Fig. 12 for the monostable cases (a) and the bistable cases (b). In the monostable case, the proposed filter succeeds to recover the stochastic contribution of the perturbed processes well, within overlapping error bars, for processes with additive or low multiplicative noise levels ($d \leq 0.1$) and weak drift ($a \leq 4$). For stronger drift and multiplicative noise levels, the recovery of the underlaying stochastic dynamics becomes poorer, as the periodic perturbation becomes hidden in stochastic noise.

In the bistable case, our framework succeeds to recover the stochastic contribution well for low to intermediate bistability ($a \leq 4$) for all tested noise levels. At higher levels of bistability, our method becomes less reliable and eventually fails to recover the stochastic contribution of some of the perturbed processes.

It should be noted, that the robustness tests covered in this section are solely related to the data intrinsic selection of IMFs covering the periodic component of the perturbed processes. Most of the perturbed processes for which our framework was found to be less reliable can still be treated with empirical mode decomposition. In these cases, the measure $\varepsilon/\varepsilon_{ref}$ often is very similar for several IMF combinations. One can consider different combinations of IMFs with low $\varepsilon/\varepsilon_{ref}$ values, paying attention also to spectra of the particular IMFs. This can yield plausible results, at the cost of requiring human interaction to the method.

6. Conclusions and outlook

In this paper we introduce a purely data driven framework for resolving the superposition of periodic oscillations in stochastic dynamics. The method is based on the estimate of the parameters describing each one of the components, the periodic perturbation and the stochastic process. To that end we apply the

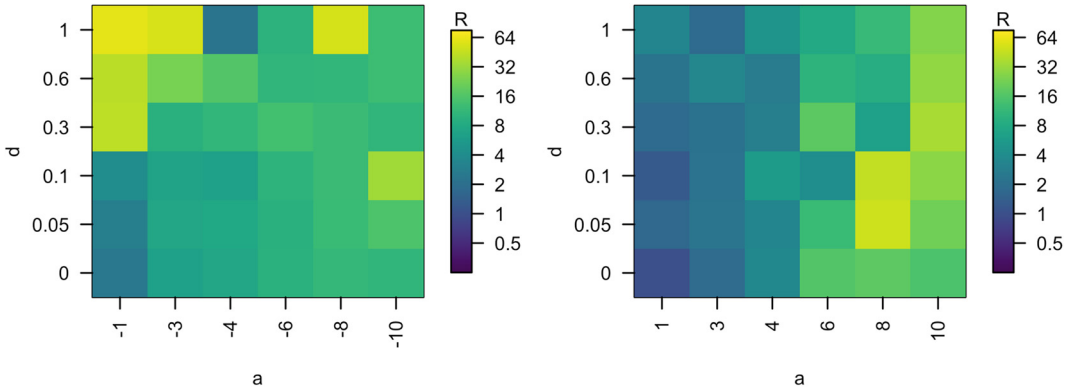

Fig. 12. Weighted error R (cf. equation (20)) between the estimated drift coefficients $\tilde{D}^{(1)}$ of different stochastic processes given by the parameters a, b, $c = 1$ and d (cf. equations (3) and (4)) and the related perturbed and filtered processes. (a) Monostable stochastic processes ($b = 0$) and (b) Bistable stochastic processes ($b = -1$). The perturbation is kept constant with $\sigma = 0.5$ and $\alpha = 2$ (cf. equations (6)).

empirical mode decomposition of the observed process and introduce a metric to automatically evaluate which of the intrinsic mode functions are more likely to be decoupled from the stochastic process. While the EMD can straightforwardly be applied to decompose a signal into a set of intrinsic modes, it does not provide a procedure to distinguish between its periodic and stochastic contributions. Our approach combines the Huang (EMD) decomposition with the Langevin modeling of the Markovian components, extending the applicability of typical procedures, namely EMD. The proposed framework retrieves good results for a wide range of frequency and amplitude in the periodic oscillation as well as for different stochastic processes (bistable/monostable dynamics, additive/multiplicative noise). A point for discussion at this stage concerns the apparent mismatch between the order of the mode and its relevance. In the cases considered above, the first mode of the empirical mode decomposition seems to be strongly coupled with the stochastic contribution in the measured signal. While at first this may seem counter-intuitive, one should notice that the intrinsic mode functions are not ranked by importance, e.g. according to their energy, as it would be the case in proper orthogonal decomposition. Empirical mode decomposition is data-driven: one calculates the modes by subtraction of average envelopes of the fluctuations, from high frequency to low frequency. Consequently, empirical mode decomposition behaves like a bandpass filter, i.e. each mode corresponds to a specific frequency band. In particular, if one removes something around 3-8 Hz in a 100 Hz signal, the first mode includes mostly higher frequency information and is not relevant to the perturbation contribution in the signal.

Concerning possible extensions of the present framework, we stress that the stochastic analysis used in this paper was already extended to series of values that, being non-Markovian in time, are Markovian in scale, i.e. have a corresponding series of increments for a given time-lag, that are Markovian through a hierarchy of different time-lags. See e.g. Ref. [3] and references therein.

Furthermore we have tested the presented framework on two stochastic processes with shot-noise, based on stochastic differential equations similar to the Langevin equation, namely $dv/dt = -av + h + \Gamma$, with $a = 0.5$, Γ being the Gaussian white noise, and the shot noise h having unitary rate and intensity. In one case, Γ was chosen to be 0 and in the second case to have a variance 0.01. Both processes were superposed with the periodic oscillations used in the main example in section 4. After filtering, the drift and diffusion coefficients $D^{(1)}$ and $D^{(2)}$ could be extracted from the data within overlapping error bars, in similar quality to the results shown in Fig. 9, with $D^{(4)} \approx 0$.

All in all, the present framework now enables to approach measurement data, spoiled with local oscillations at the middle frequency-range. Possible situations range from aerodynamical experiments with airfoils, e.g. when fluctuating lift and drag forces superposed with localized oscillation due to vibrations in the wind tunnel setup, to bacteria mobility in suspensions, when stochastic trajectories are suddenly perturbed by localized vibrations of the fluid.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work is part of the research focus “Wind Turbine Load Control under Realistic Turbulent In-Flow Conditions”, funded by the German Research Foundation, Ref. PE 478/15-1. Financial support from bilateral cooperation DFG-FAPERJ, Ref. LI-1599/3-1 is also acknowledged.

References

[1] P. Addison, The Illustrated Wavelet Transform Handbook, Institute of Physics, Bristol and Philadelphia, 2002.
[2] F. Böttcher, J. Peinke, D. Kleinhans, R. Friedrich, P. Lind, M. Haase, Reconstruction of complex dynamical systems affected by strong measurement noise, Phys. Rev. Lett. 97 (2006) 090603.
[3] R. Friedrich, J. Peinke, M. Sahimi, M.R.R. Tabar, Approaching complexity by stochastic methods: from biological systems to turbulence, Phys. Rep. 506 (2011) 87–162.
[4] C. Gardiner, Stochastic Methods, Springer-Verlag, Berlin, Heidelberg, 2009.
[5] A. Hadjihosseini, J. Peinke, N.P. Hoffmann, Stochastic analysis of ocean wave states with and without rogue waves, New J. Phys. 16 (2014), https://doi.org/10.1088/1367-2630/16/5/053037, arXiv:1402.4366.
[6] N.E. Huang, Z. Shen, S.R. Long, A New View of Nonlinear Water Waves: The Hilbert Spectrum 1, 1999.
[7] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, N.C. Yen, C.C. Tung, H.H. Liu, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, Proc. R. Soc. Lond., Ser. A, Math. Phys. Eng. Sci. 454 (1998) 903–995.
[8] T. Körner, Fourier Analysis, Cambridge University Press, Cambridge, 1989.
[9] B. Lehle, Analysis of stochastic time series in the presence of strong measurement noise, Phys. Rev. E 83 (2011) 021113.
[10] B. Lehle, Stochastic time series with strong, correlated measurement noise: Markov analysis in n dimensions, J. Stat. Phys. 152 (2013) 1145.
[11] B. Lehle, J. Peinke, Analyzing a stochastic process driven by Ornstein-Uhlenbeck noise, Phys. Rev. E 97 (2018) 012113.
[12] M.R. Luhur, J. Peinke, J. Schneemann, M. Wächter, Stochastic modeling of lift and drag dynamics under turbulent wind inflow conditions, Wind Energy (2014), https://doi.org/10.1002/we.1699.
[13] P.J.J. Luukko, J. Helske, E. Räsänen, Introducing libeemd: a program package, Comput. Stat. 31 (2016) 545–557, https://doi.org/10.1007/s00180-015-0603-9.

[14] A. Nawroth, J. Peinke, D. Kleinhans, R. Friedrich, Improved estimation of Fokker-Planck equations through optimization, Phys. Rev. E 76 (2007) 056102.
[15] P. Rinn, P.G. Lind, M. Wächter, J. Peinke, The Langevin approach: an R package for modeling Markov processes 4, https://doi.org/10.5334/jors.123, arXiv:1603.02036v1, 2016.
[16] H. Risken, The Fokker-Planck Equation, Springer-Verlag, Berlin, Heidelberg, 2011.
[17] T. Scholz, F. Raischel, V. Lopes, B. Lehle, M. Wächter, J. Peinke, P. Lind, Parameter-free resolution of the superposition of stochastic signals, Phys. Lett. A 381 (2017) 194–206.
[18] M. Siefert, R. Friedrich, J. Peinke, Analysis of data of stochastic systems, Phys. Lett. A 243 (1998) 275.
[19] M. Siefert, A. Kittel, R. Friedrich, J. Peinke, On a quantitative method to analyze dynamical and measurement noise, Europhys. Lett. 61 (2003) 466–472.
[20] M.E. Torres, M.A. Colominas, G. Schlotthauer, P. Flandrin, A Complete Ensemble Empirical Mode Decomposition, Laboratorio de Señales y Dinámicas no Lineales, Universidad Nacional de Entre Ríos, Laboratoire de Physique (UMR CNRS 5672), École Normale Supérieure de Lyon, France, 2011, pp. 4144–4147.