Extracting strong measurement noise from stochastic time series: Applications to empirical data

P. G. Lind,[1,2] M. Haase,[3] F. Böttcher,[4] J. Peinke,[4] D. Kleinhans,[5,6] and R. Friedrich[6]
[1]*Center for Theoretical and Computational Physics, University of Lisbon, Avenida Professor Gama Pinto 2, 1649-003 Lisbon, Portugal*
[2]*Departamento de Física, Faculdade de Ciências da Universidade de Lisboa, 1649-003 Lisboa, Portugal*
[3]*Institute for High Performance Computing, University of Stuttgart, Nobelstraße 19, D-70569 Stuttgart, Germany*
[4]*Institute of Physics, University of Oldenburg, D-26111 Oldenburg, Germany*
[5]*Institute for Marine Ecology, University of Gothenburg, Box 461, SE-405 30 Göteborg, Sweden*
[6]*Institute of Theoretical Physics, University of Münster, D-48149 Münster, Germany*
(Received 15 December 2009; revised manuscript received 30 March 2010; published 21 April 2010)

It is a big challenge in the analysis of experimental data to disentangle the unavoidable measurement noise from the intrinsic dynamical noise. Here we present a general operational method to extract measurement noise from stochastic time series even in the case when the amplitudes of measurement noise and uncontaminated signal are of the same order of magnitude. Our approach is based on a recently developed method for a nonparametric reconstruction of Langevin processes. Minimizing a proper non-negative function, the procedure is able to correctly extract strong measurement noise and to estimate drift and diffusion coefficients in the Langevin equation describing the evolution of the original uncorrupted signal. As input, the algorithm uses only the two first conditional moments extracted directly from the stochastic series and is therefore suitable for a broad panoply of different signals. To demonstrate the power of the method, we apply the algorithm to synthetic as well as climatological measurement data, namely, the daily North Atlantic Oscillation index, shedding light on the discussion of the nature of its underlying physical processes.

DOI: 10.1103/PhysRevE.81.041125 PACS number(s): 05.40.Ca, 02.50.Ey, 92.70.Gt

I. INTRODUCTION

Recently, much effort has been made to uncover the dynamical process underlying a given time series of scale and time dependent complex systems [1–3]. In many cases it is possible to describe such systems by a Langevin equation, extracted directly from the data, which separates the deterministic and stochastic processes inherent to the system [4]. Such an approach has already been carried out successfully, for instance, for data from turbulent fluid dynamics [5], financial data [6], climate indices [7,8], and for electroencephalographic recordings from epilepsy patients [9,10] and additional improvements were proposed to address the case of low sampling rates [11,12].

However, typically the signal is subject to noise due to experimental constraints or due to the measurement or discretization procedure leading to the data set to be studied. Such noise is not intrinsic to the system, differing from what is known as dynamical noise, and therefore one is interested to separate it from the stochastic process. We call such nonintrinsic noise measurement noise. To separate the measurement noise from the dynamics of the measured variable different predictor models or schemes for noise reduction may be used [1,3]. In this context, an alternative procedure has been proposed [13] to extract the intrinsic dynamics associated with Langevin processes strongly contaminated by measurement noise based solely on the two conditional moments directly calculated from the data [12,13].

In this paper we will revisit this nonparametric procedure, describing it in detail and explaining the main steps for its implementation, with the aim of applying it to empirical data sets. Let us consider a one-dimensional Langevin process $x(t)$ (an extension to more dimensions is straightforward) defined as

$$\frac{dx}{dt} = D_1(x) + \sqrt{D_2(x)}\Gamma_t, \tag{1}$$

where Γ_t represents a Gaussian δ-correlated white noise $\langle\Gamma_t\rangle = 0$ and $\langle\Gamma_t\Gamma_{t'}\rangle = \delta(t-t')$. Functions $D_1(x)$ and $D_2(x)$ are the drift and diffusion coefficients defined as

$$D_n(x) = \frac{1}{n!}\lim_{\tau\to 0}\frac{1}{\tau}M_n(x,\tau) \tag{2}$$

for $n=1,2$, where $M_n(x,\tau)$ denotes the nth order conditional moment of the data, as explained below. Further, we consider that $x(t)$ is "contaminated" by a Gaussian δ-correlated measurement white noise, which leads to the series of observations

$$y(t) = x(t) + \sigma\zeta(t), \tag{3}$$

where σ denotes the amplitude of the measurement noise.

When there is no measurement noise ($\sigma=0$), Eq. (3) yields the particular case $y(t)\equiv x(t)$, and the evolution equation underlying the signal can be extracted directly from the two conditional moments ($n=1,2$)

$$\hat{M}_n(y_i,\tau) = \langle[y(t+\tau) - y(t)]^n\rangle\big|_{y(t)=y_i} \tag{4}$$

as described in Refs. [4,6,8,13].

In the presence of measurement noise ($\sigma\neq 0$) the conditional moments depend on x, τ, and σ. Since generally the limit

$$\lim_{\tau\to 0}\hat{M}_n(x,\sigma\neq 0,\tau) \tag{5}$$

does not exist, Eq. (3) cannot be applied. The aim of this paper, however, is to explicitly derive a procedure which can transform the functional form of the "noisy conditional mo-

ments" $\hat{M}_1(x,\sigma,\tau)$ and $\hat{M}_2(x,\sigma,\tau)$ at small τ into the "true" coefficients $D_1(x)$ and $D_2(x)$ and simultaneously retrieve the amplitude σ of the associated measurement noise. For that, we show that $\hat{M}_n(y,\tau)$ for fixed y is typically linear in τ for a certain range $[\tau_1,\tau_2]$ of values (see Fig. 4 below). Therefore, even when $\sigma \neq 0$ one can estimate the quantities

$$D_n(y) = \frac{\hat{M}_n(y,\tau_2) - \hat{M}_n(y,\tau_1)}{n!(\tau_2 - \tau_1)}. \tag{6}$$

We start in Sec. II by briefly describing the procedure to extract Langevin equations from data sets and show how the drift and diffusion coefficients depend on the measurement noise strength σ. In particular, we will see that the proposed estimate [14] does not yield the correct value when the measurement noise is too strong. In Sec. III we then proceed to minimize a proper least square function using the Levenberg-Marquardt procedure [15]. By applying this algorithm to synthetic data we show that indeed this approach is able to reliably extract the noise amplitude even in cases where it is of the same order as the synthetic signal without noise. Furthermore, the procedure yields simultaneously more accurate estimates for the clean signal $x(t)$. Finally, in Sec. IV, we apply this framework to an empirical data set, namely, the North Atlantic Oscillation (NAO) daily index [16], giving some insight from the obtained results to the underlying system. Discussion and conclusions are given in Sec. V, where further possible applications are proposed. All details concerning the implementation of the minimization procedure to extract strong measurement noise are given as Appendixes A and C.

II. STOCHASTIC TIME SERIES WITH STRONG MEASUREMENT NOISE

We consider a time series generated by integrating Eq. (1) with drift and diffusion coefficient assumed to be linear and quadratic forms, respectively,

$$D_1(x) = d_{10} + d_{11}x, \tag{7a}$$

$$D_2(x) = d_{20} + d_{21}x + d_{22}x^2, \tag{7b}$$

and by adding separately to each data point the measurement noise term $\sigma\zeta(t)$ in Eq. (3). Though we concentrate on the particular expressions for D_1 and D_2 given above, it should be stressed that they comprehend a large collection of different processes, such as Ornstein-Uhlenbeck processes [13]. Further, some generalizations may be carried out as will be discussed in Sec. V. Using Eqs. (7a) and (7b), one has six parameters: five coefficients d_{ij} defining the evolution equation of the clean signal and a sixth parameter σ for the amplitude of the measurement noise.

Figure 1 illustrates this influence of noise for a particular choice of $D_1(x)$, $D_2(x)$. As shown in Fig. 1(a), for increasing σ one obtains broader probability density functions $P(y)$ as one intuitively expects. Quantitatively, the standard deviation θ of $P(y)$ varies quadratically with the measurement noise σ, while the mean value μ of $P(y)$ remains constant, as shown

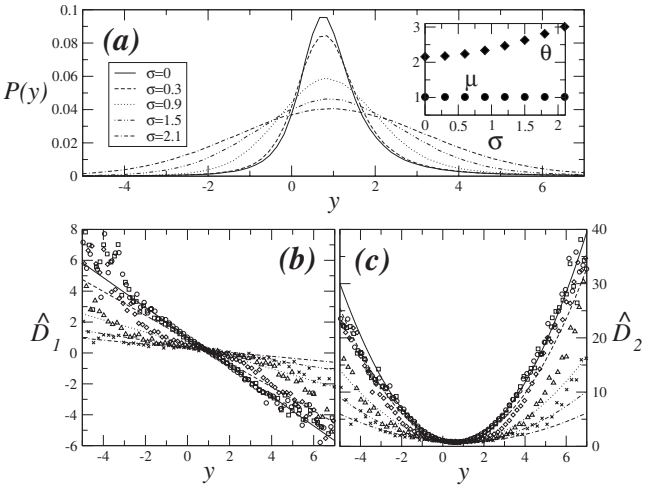

FIG. 1. Langevin time series with different measurement noise strengths. Here we show (a) the probability density function $P(y)$ of the series with noise [see Eq. (3)], with the corresponding mean value μ and standard deviation θ in the inset, and the corresponding functions (b) $\hat{D}_1(y)$ and (c) $\hat{D}_2(y)$, see Eq. (6). In all cases, the assumed time series $x(t)$ without measurement noise uses the functions $D_1(x)=1-x$ and $D_2(x)=1-x+x^2$.

in the inset of Fig. 1(a). The estimated functions $\hat{D}_1(y)$ and $\hat{D}_2(y)$ change significantly, as shown in Figs. 1(b) and 1(c), respectively. Assuming $\hat{D}_1(y)=\hat{d}_{10}+\hat{d}_{11}y$ and $\hat{D}_2(y)=\hat{d}_{20}+\hat{d}_{21}y+\hat{d}_{22}y^2$, Fig. 2 shows how the estimated parameters $\hat{d}_{ij}$ deviate from the true uncontaminated values d_{ij} in Eq. (7) when measurement noise increases. Notice that for $\sigma=0$—see left vertical axis in the plots of Fig. 2—the estimated parameter values are approximately correct.

To correctly derive the drift and diffusion coefficients $D_1(x)$ and $D_2(x)$ when σ is strong, we consider the measured conditional moments $\hat{M}_1(y_i,\tau)$ and $\hat{M}_2(y_i,\tau)$, as in Eq. (4), the hat indicating that they are calculated from the measured data $y(t)$ directly. Since this conditional moments depend in a

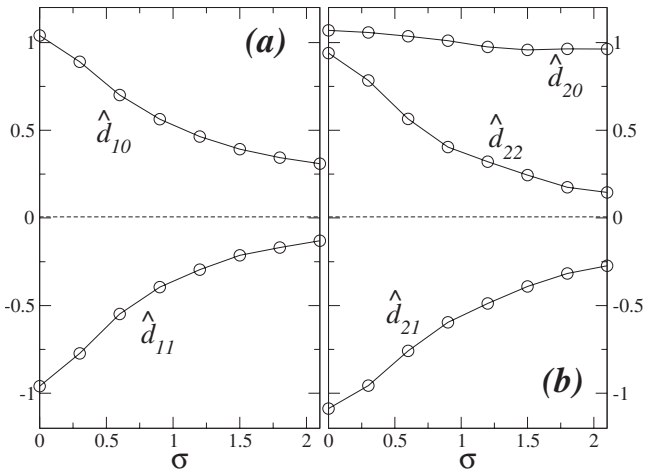

FIG. 2. Noise dependence of coefficients $\hat{d}_{ij}$ defining functions $\hat{D}_1(y)$ and $\hat{D}_2(y)$ [see text and Eq. (6)]. The underlying Langevin time series $x(t)$ without noise is the same as in Fig. 1.

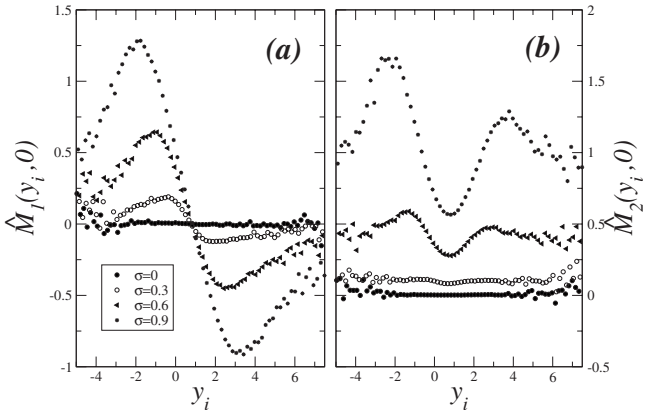

FIG. 3. Conditional moments $\hat{M}_1(y_i,\tau)$ and $\hat{M}_2(y_i,\tau)$ as a function of bin y_i, for $\tau=0$ and different measurement noise strengths. The asymmetry of $\hat{M}_2$ is due to $d_{21}\neq 0$ [see Eqs. (7)]. The same $x(t)$ as in Fig. 1 was used.

nontrivial way on both time τ and amplitude y_i, we approximate them up to first order on τ:

$$\hat{M}_1(y_i,\tau) = \langle y(t+\tau) - y(t)\rangle\big|_{y(t)=y_i} = \tau\hat{m}_1(y_i) + \hat{\gamma}_1(y_i) + \mathcal{O}(\tau^2), \tag{8a}$$

$$\hat{M}_2(y_i,\tau) = \langle (y(t+\tau) - y(t))^2\rangle\big|_{y(t)=y_i} = \tau\hat{m}_2(y_i) + \hat{\gamma}_2(y_i) + \sigma^2 + \mathcal{O}(\tau^2), \tag{8b}$$

where $y(t)$ is taken in the range $y_i \pm \Delta y/2$ for each bin i, and Δy depends on the binning considered. Appendix A gives the full derivation of Eqs. (8).

Figure 3 shows both conditional moments for $\tau=0$ and with different measurement noise strengths. Conversely, in Figs. 4(a) and 4(b) one sees that the conditional moments depend linearly on τ for a fixed amplitude y, which justifies the approximation assumed in Eqs. (8). Therefore, to study the dependence of the conditional moments on y we will consider the linear decompositions in Eqs. (8), as done in Fig. 5. Our simulations with synthetic data have shown that using a too large range of τ values yields results for D_1 and D_2 deviated from their true values. The best estimation for both D_1 and D_2 are obtained using the range $1 < \tau \lesssim 4$.

Notice that for sufficiently small measurement noise a good estimate of it is given by [13,14]

$$\sigma \approx \sqrt{\frac{\hat{M}_2(\mu,0)}{2}}, \tag{9}$$

where μ is the average value of $y(t)$ data points in the time series. For details see Appendix A. However, as shown in Fig. 4(c), this approximation is no longer valid for sufficiently high measurement noise, namely, when $\sigma \gtrsim 0.5$ [see inset of Fig. 4(c)] and even otherwise coefficients D_1 and D_2 are not correctly estimated (see Fig. 2). Therefore, a better algorithm to estimate such parameters is necessary.

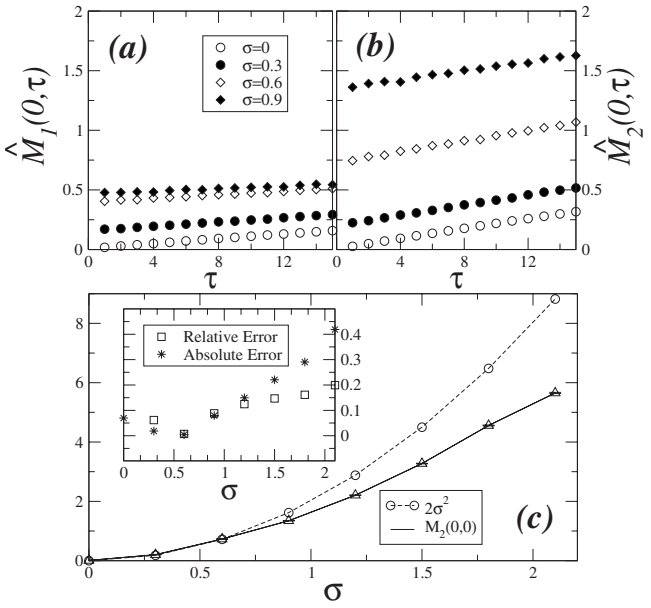

FIG. 4. Conditional moments (a) $\hat{M}_1(y_i,\tau)$ and (b) $\hat{M}_2(y_i,\tau)$ as a function of τ, for bin $y_i=0$ and different measurement noise strengths. In (c) one compares the true measurement noise with the approximation $\sigma_{app}=\hat{M}_2(0,0)\sim 2\sigma^2$ given in Eq. (9). In the inset the corresponding absolute and relative errors are given by $\zeta_a=|\sigma-\sigma_{app}|$ and $\zeta_r=\zeta_a/\sigma$, respectively. Errors for $\hat{M}_2$ are negligible. The same $x(t)$ as in Fig. 1 was used.

The heart of our procedure to correctly estimate measurement noise lies in the fact that while the functions $\hat{m}_i$ and $\hat{\gamma}_i$ $(i=1,2)$ are obtained explicitly for each bin value y_i, functions m_i and γ_i depend generally on the drift and diffusion coefficients as follows:

$$\gamma_1(y) = \int_{-\infty}^{+\infty} (x-y)\bar{f}_\sigma(x|y)dx, \tag{10a}$$

$$\gamma_2(y) = \int_{-\infty}^{+\infty} (x-y)^2\bar{f}_\sigma(x|y)dx, \tag{10b}$$

$$m_1(y) = \int_{-\infty}^{+\infty} D_1(x)\bar{f}_\sigma(x|y)dx, \tag{10c}$$

$$m_2(y) = 2\int_{-\infty}^{+\infty} [(x-y)D_1(x) + D_2(x)]\bar{f}_\sigma(x|y)dx, \tag{10d}$$

where $\bar{f}_\sigma(x|y)$ is the probability for the system to adopt the value x when a measured value y is observed. For details about the derivation of functions in Eqs. (10) see Appendix A and for the explicit expression of $\bar{f}_\sigma(x|y)$ see Appendix B.

In Fig. 5 we illustrate both the hat-functions in Eqs. (8) and their integral form in Eqs. (10). Due to the measurement noise fixed in this example at $\sigma=1$ the hat functions (symbols) are not properly fitted by the integral form in Eqs. (10) using the first estimate (dashed lines) of the parameters d_{ij}, taken from Fig. 2, and σ, computed from Eq. (9). If instead

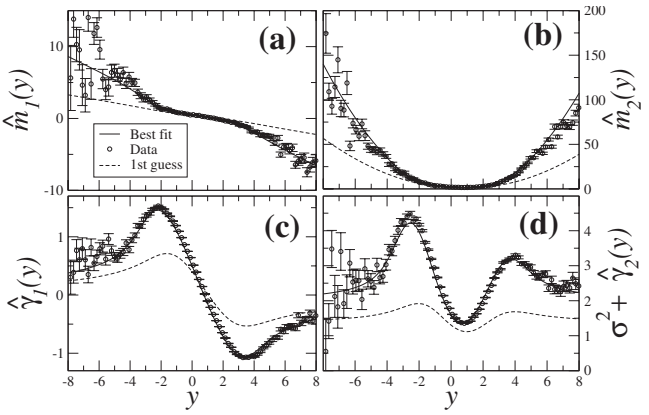

FIG. 5. Functions $\hat{m}_1$, $\hat{m}_2$, $\hat{\gamma}_1$, and $\hat{\gamma}_2$ (symbols) defining the conditional moments in Eqs. (8b). The underlying Langevin time series $x(t)$ without noise is characterized by a drift coefficient $D_1(x)=1-x$ and a diffusion coefficient $D_2(x)=1-x+x^2$. The measurement noise was fixed at $\sigma=1$. Each hat function is compared with the corresponding integral form in Eqs. (10) using the first estimate of parameters values (dashed lines) and the true values (solid lines).

we use the true parameter values in the integral forms of our m_i and γ_i functions a proper fit is obtained (solid lines).

Therefore, the problem we want to solve is to determine the parameters that minimize the function:

$$F = \frac{1}{M}\sum_{i=1}^{M}\left\{\frac{[\hat{\gamma}_1 - \gamma_1(y_i)]^2}{\sigma^2_{\hat{\gamma}_1}(y_i)} + \frac{[\hat{\gamma}_2 - \gamma_2(y_i) - \sigma^2]^2}{\sigma^2_{\hat{\gamma}_2}(y_i)} + \frac{[\hat{m}_1 - m_1(y_i)]^2}{\sigma^2_{\hat{m}_1}(y_i)} + \frac{[\hat{m}_2 - m_2(y_i)]^2}{\sigma^2_{\hat{m}_2}(y_i)}\right\}, \tag{11}$$

where the summation extends over all M bins, $\sigma_{\hat{\gamma}_1}(y_i)$ is the error associated to function $\hat{\gamma}_1$ at the value y_i and similarly for $\sigma_{\hat{\gamma}_2}$, $\sigma_{\hat{m}_1}$, and $\sigma_{\hat{m}_2}$. Notice that the values of such $\sigma_{\hat{\gamma}_i}$ and $\sigma_{\hat{m}_i}$ are taken directly from the data only. See Appendix A for details.

Taking again the example illustrated in Fig. 2 with $\sigma=1$ we plot in Fig. 6 function F in Eq. (11) as function of each one of the parameters keeping all others fixed at their true values. Evidently, the estimated values are near the minimum of F in each case. Further, the one-dimensional cuts of function F show only one minimum. One should note however that, for the entire six-dimensional parameter space, several local minima of F may appear. In fact, after minimizing F by varying one parameter, function F also changes as a function of the other parameters, i.e., its minimum as a function of the other parameter changes. In the next section we will see how to minimize function F, in order to find good estimates for the correct values for each parameter.

III. OPTIMIZATION PROCEDURE

After computing the functions $\hat{\gamma}_1$, $\hat{\gamma}_2$, $\hat{m}_1$, and $\hat{m}_2$ as well as the corresponding errors $\sigma_{\hat{\gamma}_1}$, etc., directly from the measured time series $y(t)$ and estimating the coefficients D_1 and

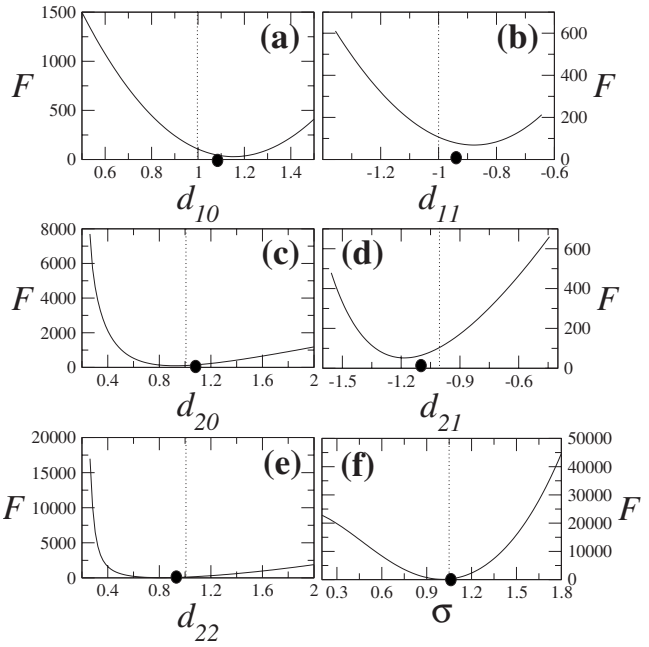

FIG. 6. Function F in Eq. (11) as a function of (a) d_{10}, (b) d_{11}, (c) d_{20}, (d) d_{21}, (e) d_{22}, and (f) σ. The same situation as in Fig. 2 is here chosen: $D_1(x)=1-x$, $D_2(x)=1-x+x^2$, and $\sigma=1$. Dashed lines indicate the true values used for generating the data series, while the bullet indicates the estimated values of functions D_1 and D_2 for $\sigma=0$. In each plot while varying one parameter, the remaining ones are fixed at their true values (see text).

D_2 given by the functional forms in Eqs. (7) there are several ways to minimize F. All of them start from the initially estimated set of values for the parameters and iteratively improve the solution by finding lower values of F until convergence is attained.

To proceed the following remark should be considered. Parameter d_{10} can always be eliminated with a simple transformation $x \to x' = x + d_{10}/d_{11}$. Alternatively and since we do not know beforehand the true values of d_{10} and d_{11}, we can consider also the fact that averaging Eq. (1) yields $d_{10} = -d_{11}\langle x \rangle$ and consider the transformation $x' = x - \langle x \rangle$. With these arguments, we henceforth disregard d_{10}, which reduces the dimension of parameter space by one. Parameter d_{10} is computed from the relations above, only after minimizing F. For simplicity the primes in x' will be omitted.

The simplest way is to minimize each term in F and repeat that a large number of times starting from different initial conditions for the parameters, in a sort of Monte Carlo procedure of random walks [17] or Lévy walks [18]. The Monte Carlo procedure assures that a substantial number of local minimal for F will be visited, and in the end, we take the minimum of all F values found. Simulations have shown however that a Monte Carlo procedure is too expensive in this case since there are different local minima and the choice of the minimum is strongly path dependent. We will therefore consider the Levenberg-Marquardt method [15].

For the Levenberg-Marquardt procedure one computes the first and second derivative of F. Symbolizing the parameters σ, d_{11}, d_{20}, d_{21}, and d_{22} by p_k with $k=1,\ldots,5$, respectively, these derivatives read as

$$\frac{\partial F}{\partial p_k} = -\frac{2}{M}\sum_{i=1}^{M}\left[\frac{\hat{\gamma}_1 - \gamma_1}{\sigma^2_{\hat{\gamma}_1}(i)}\frac{\partial \gamma_1}{\partial p_k} + \frac{\hat{\gamma}_2 - \gamma_2 - \sigma^2}{\sigma^2_{\hat{\gamma}_2}(i)}\frac{\partial(\gamma_2+\sigma^2)}{\partial p_k} + \frac{\hat{m}_1 - m_1}{\sigma^2_{\hat{m}_1}(i)}\frac{\partial m_1}{\partial p_k} + \frac{\hat{m}_2 - m_2}{\sigma^2_{\hat{m}_2}(i)}\frac{\partial m_2}{\partial p_k}\right], \tag{12}$$

$$\begin{aligned}\frac{\partial^2 F}{\partial p_k\,\partial p_\ell} = \frac{2}{M}\sum_{i=1}^{M}\Bigg[&\frac{1}{\sigma^2_{\hat{\gamma}_1}(i)}\frac{\partial \gamma_1}{\partial p_k}\frac{\partial \gamma_1}{\partial p_\ell} - \frac{\hat{\gamma}_1 - \gamma_1}{\sigma^2_{\hat{\gamma}_1}(i)}\frac{\partial^2 \gamma_1}{\partial p_k\,\partial p_\ell} + \frac{1}{\sigma^2_{\hat{\gamma}_2}(i)}\frac{\partial(\gamma_2+\sigma^2)}{\partial p_k}\frac{\partial(\gamma_2+\sigma^2)}{\partial p_\ell} - \frac{\hat{\gamma}_2 - \gamma_2 - \sigma^2}{\sigma^2_{\hat{\gamma}_2}(i)}\frac{\partial^2(\gamma_2+\sigma^2)}{\partial p_k\,\partial p_\ell}\\ &+ \frac{1}{\sigma^2_{\hat{m}_1}(i)}\frac{\partial m_1}{\partial p_k}\frac{\partial m_1}{\partial p_\ell} - \frac{\hat{m}_1 - m_1}{\sigma^2_{\hat{m}_1}(i)}\frac{\partial^2 m_1}{\partial p_k\,\partial p_\ell} + \frac{1}{\sigma^2_{\hat{m}_2}(i)}\frac{\partial m_2}{\partial p_k}\frac{\partial m_2}{\partial p_\ell} - \frac{\hat{m}_2 - m_2}{\sigma^2_{\hat{m}_2}(i)}\frac{\partial^2 m_2}{\partial p_k\,\partial p_\ell}\Bigg]\\ \sim \frac{2}{M}\sum_{i=1}^{M}\Bigg[&\frac{1}{\sigma^2_{\hat{\gamma}_1}(i)}\frac{\partial \gamma_1}{\partial p_k}\frac{\partial \gamma_1}{\partial p_\ell} + \frac{1}{\sigma^2_{\hat{\gamma}_2}(i)}\frac{\partial(\gamma_2+\sigma^2)}{\partial p_k}\frac{\partial(\gamma_2+\sigma^2)}{\partial p_\ell} + \frac{1}{\sigma^2_{\hat{m}_1}(i)}\frac{\partial m_1}{\partial p_k}\frac{\partial m_1}{\partial p_\ell} + \frac{1}{\sigma^2_{\hat{m}_2}(i)}\frac{\partial m_2}{\partial p_k}\frac{\partial m_2}{\partial p_\ell} - 2\delta_{\sigma p_k}\delta_{\sigma p_\ell}\frac{\hat{\gamma}_2 - \gamma_2 - \sigma^2}{\sigma^2_{\hat{\gamma}_2}(i)}\Bigg].\end{aligned} \tag{13}$$

In the right-hand side of Eq. (13) we neglect the terms containing second derivatives of γ and m functions. This last approximation of neglecting second derivatives is acceptable as far as the model is successful [15].

By symbolizing first and second derivatives as β_k and $\alpha_{k\ell}$, respectively, the iterative procedure computes the increments dp_k for each parameter p_k $(k=1,\ldots,5)$, which are the solutions of

$$\beta_k = -\sum_{\ell=1}^{5} \alpha_{k\ell} dp_\ell. \tag{14}$$

Furthermore, one assumes that $dp_\ell \propto \beta_\ell$, which considering dimensional analysis [15] can be written as

$$dp_\ell = \frac{\beta_\ell}{\lambda \alpha_{\ell\ell}}, \tag{15}$$

where typically $\lambda \gg 1$. For a given λ value, instead of the second derivatives α_{mn} one assumes $\alpha'_{mn} = \alpha_{mn}(1+\lambda)$ for $m=n$ and $\alpha'_{mn} = \alpha_{mn}$ otherwise and solves Eq. (14) for dp_k [15].

If $F(p_k+dp_k) < F(p_k)$, the parameter values are updated, $p_k \to p_k + dp_k$, and λ is typically decreased by 10%. Otherwise, if $F(p_k+dp_k) \geq F(p_k)$ one increases λ by 10% and determines new increments dp_k. The procedure stops after attaining the required convergence.

Using the same data as generated in Fig. 5 with $\sigma=1$, we now plot in Fig. 7 the functions $\hat{m}_i$ and $\hat{\gamma}_i$ for the data (symbols) and compare them with the integral forms of those functions for the first estimate of parameter values (dashed lines) and the optimized solution obtained with the Levenberg-Marquardt procedure (solid lines). Clearly, the optimized functions fit better the data and the minimum of F found is very close to its true value (see caption of Fig. 7).

Notice that the optimized values d'_{ik} are obtained for the transformed data $(x \to x' = x - \langle x \rangle)$, assuming $d'_{10}=0$. In practice one obtains $d'_{10} \sim 10^{-2}$, typically two orders of magnitude smaller than the other coefficients. Using $\langle x \rangle = -d_{10}/d_{11}$, one obtains the true coefficients according to $d_{10} = -d'_{11}\langle x \rangle$, $d_{11} = d'_{11}$, $d_{20} = d'_{20} - d'_{21}\langle x \rangle + d'_{22}\langle x \rangle^2$, $d_{21} = d'_{21} - 2d'_{22}\langle x \rangle$, and $d_{22} = d'_{22}$.

To show the power of the present procedure we next generate several synthetic data sets from Eq. (1) with different measurement noise amplitudes σ_I in the range [0,1.2]. The same $D_1(x)$ and $D_2(x)$ as in Fig. 2 is used. Results are shown in Fig. 8. The circles indicate the obtained parameter values for the first estimate, as in Fig. 2. The solid lines indicate the true values used to generate the data, while bullets indicate the value after optimization.

From Fig. 8(a) one sees that after optimization the value of σ_I is always correctly determined. Such finding is of major importance and shows the relevance of our approach for practical applications even for strong measurement noise since the uncontaminated series x typically lies within the range $[-2,2]$, having therefore values close to the amplitude σ_I of the measurement noise.

Figures 8(b) and 8(c) also show a very reliable estimate for the two parameters d_{10} and d_{11}, respectively, defining the

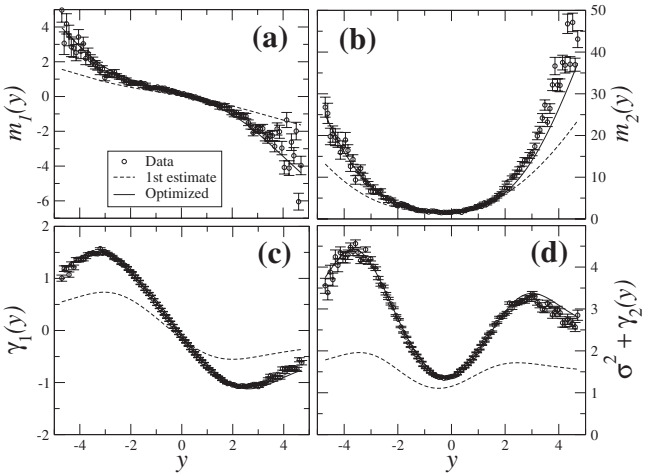

FIG. 7. Functions m_1, m_2, γ_1, and γ_2 for the Langevin process with $D_1(x)=1-x$, $D_2(x)=1-x+x^2$, and $\sigma=1$. Symbols indicate the functions obtained for the data, dashed line corresponds to the first estimate of the parameters and solid line corresponds to the parameter values obtained from the Levenberg-Marquardt procedure (see text). In this case, for the first estimate one has $F_0=3720$ while the final estimate retrieves $F_{LM}=33.1$. The true minimum is $F_m=29.2$.

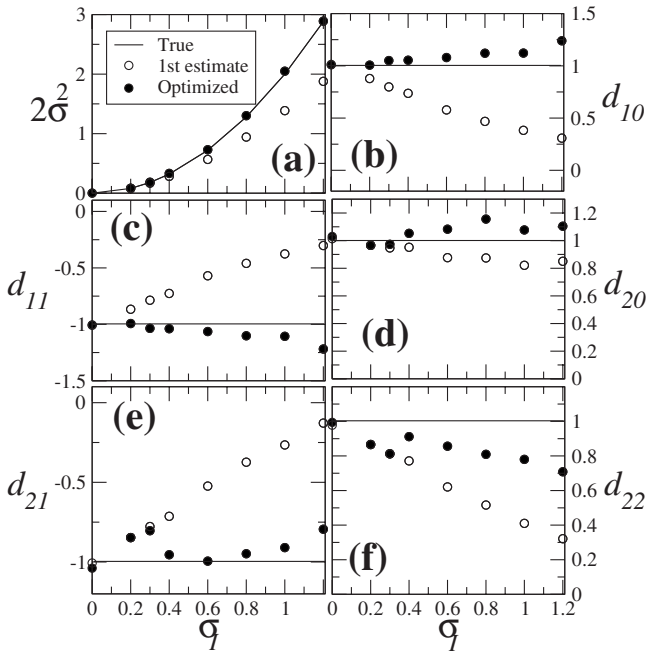

FIG. 8. Comparison of the optimized parameters values (bullet) with the first estimate and the true values for different input measurement noise strengths σ_I: (a) $2\sigma^2$, (b) d_{10}, (c) d_{11}, (d) d_{20}, (e) d_{21}, and (f) d_{22}. The measurement noise is correctly extracted as well as the parameters defining the drift coefficient $D_1(x)$ which controls the deterministic part of the underlying evolution equation (see text).

drift coefficient $D_1(x)$. Since this coefficient characterizes the deterministic part of the evolution equation for x, this accurate estimate should provide valuable insight into the dynamics of the underlying system.

As for the diffusion coefficient $D_2(x)$, Figs. 8(d)–8(f) show that the estimate of d_{22} is no longer as good as for the other parameters. Parameter d_{20} is reasonably estimated but the optimized estimate is as good as the first one.

For stronger measurement noise, namely, for $\sigma > 1.2$, one faces the problem that the optimization procedure is sometimes stuck in a local minimum of the function F leading to unreliable coefficients d_{ik}. This is in principle a shortcoming of the presently used minimization algorithm. In addition, the function F itself is based on estimated functions m and γ and therefore itself subject to errors. A forthcoming study will address the observed issues in the context of global optimization.

IV. NORTH ATLANTIC OSCILLATION: AN EMPIRICAL EXAMPLE

The NAO is a source of variability in the global atmosphere, describing a large-scale vacillation in atmospheric mass between the anticyclone near the Azores and the cyclone near Iceland [19]. The state of the NAO is usually measured by an index N defined as the normalized pressure difference between the high and the low poles, where the pressures are averaged over each, day, month, or year [8,19]. The NAO index and climate indices in general are receiving much attention due to their important role in climate change. Lately, evidences for the stochasticity of this index have been shown [7,8]. In this section we address the problem of estimating its measurement noise amplitude. In this section we apply our framework to the North Atlantic Oscillation daily index (Fig. 9), which presents data with strong measurement noise.

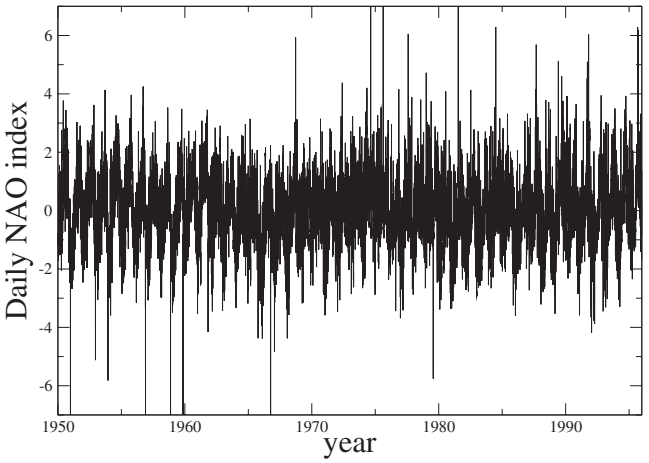

FIG. 9. Time-series of the NAO daily index, from 1950 until 1996. The data were extracted from [25]. More details in Ref. [16].

Figures 10(a) and 10(b) show the drift and diffusion coefficients, respectively, for the NAO daily index (bullets) and the corresponding fit (solid line). Table I summarizes the optimized values of d_{ij} for both D_1 and D_2 together with the amplitude of the measurement noise σ (column "NAO index").

To evaluate the reliability of considering the NAO index to be a Markov process described by Eq. (1) we also plot in Fig. 10 the results obtained when integrating such equation (circles) using the coefficient values in Table I including the amplitude of the measurement noise. The corresponding fit is represented with a dashed line and the parameter values are indicated in Table I for comparison (first column in "with measurement noise"). The values for each parameter are averaged over ten data sets, and the error is taken as the largest deviation from the average. Sample size and time increment are the same as for the empirical data.

As one sees from Table I, the very high level of measurement noise σ as well as the τ-independent parts γ_1 and γ_2 [see Eq. (8)] plotted in Figs. 10(e) and 10(f) are well represented, while the coefficient values d_{11} and d_{20} in our simulation deviate by a factor of two from the ones found for the NAO series. These deviations result in large scattering of the functions m_i and D_i, plotted in Figs. 10(a)–10(d) and are attributed to the extremely high σ value, at least three orders of magnitude larger than the d_{ij} parameters.

Moreover, the deviations appear also because of the small amount of data points (16 801 values) and of the too large time increment dt. To address the question concerning the size of the available series, we compare the results for sets with 10^4 points and time increments between successive measures $dt=1$ (same as for NAO series) with sets with

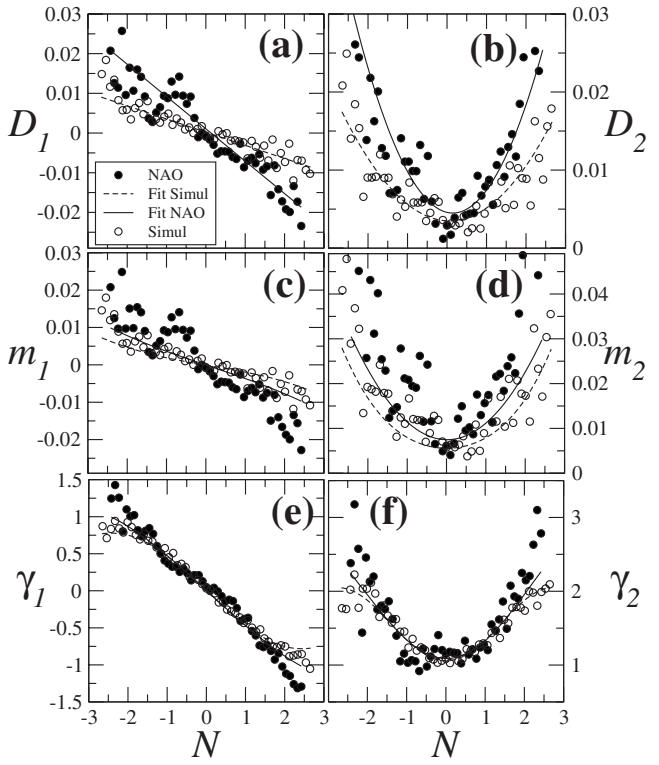

FIG. 10. [(a) and (b)] Estimate of the drift and diffusion coefficients $D_1(N)$ and $D_2(N)$ of the daily North Atlantic index N [16] (16 801 data points), together with the corresponding (c) $m_1(N)$, (d) $m_2(N)$, (e) $\gamma_1(N)$, and (f) $\gamma_2(N)$. Results for the empirical NAO index are represented with bullets whereas the synthetic data also with 16 801 data points and parameter values given by Table I is shown with circles for comparison. The corresponding fits of D_1 and D_2 and the analytical functions m_i and γ_i with the best parameter choices are given with solid and dashed lines, respectively.

different number N of points and intertime dt, remaining the time window $N\ dt$ constant, and also with sets where only dt is varied. Comparing these three different cases we see that for small dt with fixed $N\ dt$ the algorithm retrieves slightly better parameter values. See the three columns "with measurement noise." For all coefficients, the estimate is better in the case one uses the same time window $N\ dt$ but with a smaller time increment.

The mismatch between the empirical and synthetic series could raise the question if the NAO index is indeed suitably described by a Markovian stochastic process with a perceivable deterministic part. In fact, since one observes $\sigma \gg d_{ij}$ the series is approximately a pure white noise [i.e., $y(t)=\sigma\zeta(t)$ in Eq. (3)], which in fact also yields a linear drift and quadratic diffusion coefficients.

To address this problem we rerun our optimization procedure for synthetic data, for two additional situations: one where $\sigma=0$ and drift and diffusion coefficients are given by the NAO index and another one which simulates a pure white noise ($D_1=D_2=0$) with σ equal to the value found for the NAO series. The results are also given in Table I, columns "no measurement noise" and "only measurement noise," respectively.

For the pure white-noise process one obtains σ as the only nonzero parameter, apart from fluctuations. For $dt=1$, the white-noise amplitude is correctly estimated, while for the cases with smaller $dt=0.1$ only the first estimate is available, since the condition for optimizing iteratively the parameter values, namely, $d_{21}^2-4d_{20}d_{22}<0$, is eventually violated due to the numerically vanishing d_{ij} coefficients. These features raise difficulties in a proper minimum search for F and therefore most of the simulation trials stop at the first estimate of σ and d_{ij} (see last columns of Table I).

For the synthetic process with no noise, the order of magnitude for the parameters of D_1 and D_2 is correctly computed, whereas a nonzero measurement noise is re-

TABLE I. Optimized parameter values for the daily North Atlantic Oscillation index [16] compared with the average values for ten sets of synthetic data ("with measurement noise") using the same number of points and parameter values. In order to evaluate the reliability of our synthetic data we also run the optimization procedure for ten sets of synthetic data with the same D_1 and D_2 found in NAO series and $\sigma=0$ ("no measurement noise"). In the last columns we plot the results returned from the optimization procedure for synthetic data of pure measurement noise with amplitude $\sigma=0.847$, the one obtained for NAO series. In each case, we compare (i) results from sets of 10^4 points and $dt=1$ with results from sets of 10^5 and $dt=0.1$, i.e., same time window and different time increments dt between successive measures and (ii) results from sets of 10^5 and $dt=0.1$ with results from sets with 10^4 and $dt=0.1$, i.e., same time increment between measures and different time window.

		Simulations								
	NAO index 16801 pts	With measurement noise			No measurement noise			Only measurement noise		
Parameter	$dt=1$	10^4 pts $dt=1$	10^5 pts $dt=0.1$	10^4 pts $dt=0.1$	10^4 pts $dt=1$	10^5 pts $dt=0.1$	10^4 pts $dt=0.1$	10^4 pts $dt=1$	10^5 pts $dt=0.1$	10^4 pts $dt=0.1$
$\sigma\times10^3$	847	826 ± 90	790 ± 58	728 ± 110	52 ± 18	12 ± 7	13 ± 12	825 ± 90	598 ± 5	598 ± 16
$\mathbf{d_{10}}\times10^3$	−0.4	-0.3 ± 0.1	-0.3 ± 0.1	-1.1 ± 0.6	-0.7 ± 0.2	-0.62 ± 0.07	-3.1 ± 0.3	-0.3 ± 0.1	0 ± 10^{-3}	0 ± 0.01
$\mathbf{d_{11}}\times10^3$	−6.91	-3 ± 1	-3 ± 1	-5 ± 3	-6.0 ± 0.4	-5.9 ± 0.1	-13.5 ± 0.2	-2 ± 1	-0.1 ± 1	2 ± 2
$\mathbf{d_{20}}\times10^3$	5.9	3 ± 1	3 ± 2	2 ± 1	5.2 ± 0.2	5.81 ± 0.03	5.8 ± 0.2	2.8 ± 0.8	0.3 ± 2	-0.5 ± 3
$\mathbf{d_{21}}\times10^3$	−0.02	0.3 ± 0.7	-0.3 ± 0.5	1 ± 2	0.3 ± 0.4	-0.1 ± 0.1	1.7 ± 0.3	0.3 ± 0.6	-0.1 ± 4	2 ± 9
$\mathbf{d_{22}}\times10^3$	1.2	2.1 ± 0.7	0.9 ± 0.7	4 ± 5	1.4 ± 0.2	1.29 ± 0.03	2.2 ± 0.3	2.2 ± 0.7	0 ± 2	1 ± 4

trieved, being significantly better estimated for the cases with $dt=0.1$.

In this scope, our results point in the direction of previous arguments given by some authors [20]: differently from other climate indices such as the El Niño-Southern Oscillation (ENSO) index, the NAO index seems to be an almost pure white-noise process with only a minor contribution from a stochastic process governed by a Langevin-like equation. Alternative indices should be therefore considered and studied as recently suggested [8].

V. DISCUSSION AND CONCLUSIONS

We described in detail a nonparametric procedure to extract measurement noise in empirical stochastic series with strong measurement noise. The algorithm is able to accurately extract the strength of measurement noise and the values of the parameters defining the drift coefficient and to estimate with good accuracy the diffusion coefficient that fully describe the evolution equation for the measured quantity in the time series. This has been shown by synthetically generated data sets contaminated by increasing measurement noise. Additionally, the algorithm was applied to a set of measured data providing insight in the underlying systems. The data for the climate index shows a large scattering, probably due to the small amount of data points. Larger data sets for climate indices are not available up to our knowledge.

It should be noticed that the nonparametric reconstruction of the Langevin Eq. (1) from measured stationary data sets generally requires that the process exhibits Markovian properties and fulfils the Pawula theorem [8]. While the second constraint can be relaxed extending the analysis to a broader class of Langevin-like systems in which the Gaussian δ-correlated white-noise Langevin force is replaced by a more general Lévy noise [2,21], in general the Markov condition remains a crucial constraint.

Recently, it has been shown that processes corrupted from measurement noise may loose their Markov properties [22]. For this reason the proper analysis of data suffering from strong measurement noise in general is a complicated task. We, however, would like to point out, that the method presented here solely relies on Markov properties of the underlying undisturbed process $x(t)$. In case of δ-correlated measurement noise the method presents a general approach to access the process x and the noise amplitude σ at the same time.

Therefore, since the algorithm is general for a broad class of stochastic systems other applications can be proposed. Particularly in cases where the measurement procedure is subject to large measurement noise due to the distance between the location where the measure is taken and the location where the phenomena occurs. Two important applications in this context are seismographic data [2], where the epicenter cannot be predicted beforehand, and data from surface electroencephalogram (EEG) [9,10], which, though having stronger measurement noise, are much recommended instead of *in situ* measurements for the sake and comfort of the patient. A further application would be the analysis of sensors to which one has no access, for example, sensors being installed in remote systems showing more and more measurement noise due to aging effects. Here it should even be possible to know quite precisely the functional structure of the underlying process, an assumption of our analysis here.

Such applications however appeal for the extension of the present procedures to higher dimensions, i.e., more than one time series, which implies the consideration of different measurement noise sources and consequently noise mixing. To ascertain in which conditions and up to which point can we separate different measurement noise sources is an open question which we will address elsewhere.

In all simulations a linear function was assumed for the drift coefficient and a quadratic one for diffusion. Although such assumptions comprehend already a broad class of systems [2,8,13] our approach and all expressions may easily be extended to higher order polynomials for $D_1(x)$ and $D_2(x)$, as long as the number of parameters for modeling $D_1(x)$ and $D_2(x)$ is not too high. In this case the calculations presented in the appendices are valid if one considers proper higher powers in the integrand of integrals h_1 and h_2 [see Eqs. (C12) in Appendix C].

Furthermore, other possibilities for optimization are possible. For instance, though in this case we have shown that random Monte Carlo procedures are computationally expensive consuming, one could think of a nonlocal search procedure using, for example, bigger jumps such as the ones of a Lévy flight process [23]. Alternatively one may also study how good would be an optimization procedure that considers the minimization of a split cost function F. Preliminary results have shown that for a proper decomposition of F our optimization problem may be reduced to a cubic equation and a lower dimensional system of linear equations. Another possibility would be to use genetic algorithms [24]. These points will be addressed elsewhere.

ACKNOWLEDGMENTS

The authors thank the Wilhelm and Else Heraeus Foundation and Project DREBM/DAAD/03/2009 for the bilateral cooperation between Portugal and Germany. The authors also thank Matthias Wächter, Frank Raischel, Reza M. Baram, and Bibhu Biswal for useful discussions. P.G.L. thanks Fundação para a Ciência e a Tecnologia–Ciência 2007 for financial support.

APPENDIX A: THE CONDITIONAL MOMENTS OF AN ARBITRARY TIME SERIES AND THEIR LINEAR APPROXIMATIONS

Taking a series of measurements $y(t)$ as defined in Eq. (3), its nth order conditional moment reads as

$$\hat{M}_n(y_0,\tau) = \langle (y(t+\tau)-y(t))^n \rangle|_{y(t)=y_0}$$
$$= \int_{-\infty}^{+\infty} dx_0 \int_{-\infty}^{+\infty} dx \int_{-\infty}^{+\infty} dy (y-y_0)^n$$
$$\times f_\sigma(y|x) f_\tau(x|x_0) \bar{f}_\sigma(x_0|y_0), \tag{A1}$$

where $f_\sigma(y|x)$ is the probability to measure y in the presence of a measurement noise with variance σ^2, when the system (without noise) has the value x, $f_\tau(x|x_0)$ is the probability for the system to evolve from a value x_0 to a value x within a time interval τ and $\bar{f}_\sigma(x_0|y_0)$ has the inverse meaning of f_σ: it is the probability for the system to adopt the value x_0 when a measured value y_0 is observed. While f_τ is unknown, f_σ and $\bar{f}_\sigma$ are related with each other according to Bayes' theorem (see Appendix B).

From such assumptions one easily arrives to the identities

$$\int_{-\infty}^{+\infty} dy f_\sigma(y|x) = 1, \tag{A2a}$$

$$\int_{-\infty}^{+\infty} dy (y-x) f_\sigma(y|x) = 0, \tag{A2b}$$

$$\int_{-\infty}^{+\infty} dy (y-x)^2 f_\sigma(y|x) = \sigma^2, \tag{A2c}$$

and using these identities the general expression (A1) can be approximated up to first order assuming $\tau \ll 1$. More precisely, the first two moments $\hat{M}_1$ and $\hat{M}_2$ yield

$$\hat{M}_1(y_0,\tau) = \langle y(t+\tau)-y(t) \rangle|_{y(t)=y_0} = \int_{-\infty}^{+\infty} dx_0 \int_{-\infty}^{+\infty} dx \int_{-\infty}^{+\infty} dy (y-y_0) f_\sigma(y|x) f_\tau(x|x_0) \bar{f}_\sigma(x_0|y_0),$$
$$= \int_{-\infty}^{+\infty} dx_0 \int_{-\infty}^{+\infty} dx f_\tau(x|x_0) \bar{f}_\sigma(x_0|y_0) \int_{-\infty}^{+\infty} dy (y-x+x-y_0) f_\sigma(y|x)$$
$$= \int_{-\infty}^{+\infty} dx_0 \int_{-\infty}^{+\infty} dx f_\tau(x|x_0) \bar{f}_\sigma(x_0|y_0) \left(\int_{-\infty}^{+\infty} dy (x-y_0) f_\sigma(y|x) + \int_{-\infty}^{+\infty} dy (y-x) f_\sigma(y|x) \right)$$
$$= \int_{-\infty}^{+\infty} dx_0 \int_{-\infty}^{+\infty} dx f_\tau(x|x_0) \bar{f}_\sigma(x_0|y_0) \left((x-y_0) \int_{-\infty}^{+\infty} dy f_\sigma(y|x) + 0 \right)$$
$$= \int_{-\infty}^{+\infty} dx_0 \bar{f}_\sigma(x_0|y_0) \int_{-\infty}^{+\infty} dx (x-x_0+x_0-y_0) f_\tau(x|x_0)$$
$$= \int_{-\infty}^{+\infty} dx_0 \bar{f}_\sigma(x_0|y_0) \left(\int_{-\infty}^{+\infty} dx (x_0-y_0) f_\tau(x|x_0) + \int_{-\infty}^{+\infty} dx (x-x_0) f_\tau(x|x_0) \right)$$
$$= \int_{-\infty}^{+\infty} dx_0 \bar{f}_\sigma(x_0|y_0) ((x_0-y_0) + \tau D_1(x_0) + \mathcal{O}(\tau^2))$$
$$= \int_{-\infty}^{+\infty} dx_0 (x_0-y_0) \bar{f}_\sigma(x_0|y_0) + \tau \int_{-\infty}^{+\infty} dx_0 D_1(x_0) \bar{f}_\sigma(x_0|y_0) + \mathcal{O}(\tau^2) \equiv \hat{\gamma}_1(y_0) + \tau \hat{m}_1(y_0) + \mathcal{O}(\tau^2), \tag{A3}$$

$$\hat{M}_2(y_0,\tau) = \langle (y(t+\tau)-y(t))^2 \rangle|_{y(t)=y_0} = \int_{-\infty}^{+\infty} dx_0 \int_{-\infty}^{+\infty} dx \int_{-\infty}^{+\infty} dy (y-y_0)^2 f_\sigma(y|x) f_\tau(x|x_0) \bar{f}_\sigma(x_0|y_0),$$
$$= \int_{-\infty}^{+\infty} dx_0 \int_{-\infty}^{+\infty} dx f_\tau(x|x_0) \bar{f}_\sigma(x_0|y_0) \int_{-\infty}^{+\infty} dy (y-y_0)^2 f_\sigma(y|x)$$
$$= \int_{-\infty}^{+\infty} dx_0 \int_{-\infty}^{+\infty} dx f_\tau(x|x_0) \bar{f}_\sigma(x_0|y_0) \int_{-\infty}^{+\infty} dy (y-x+x-y_0)^2 f_\sigma(y|x)$$
$$= \int_{-\infty}^{+\infty} dx_0 \int_{-\infty}^{+\infty} dx f_\tau(x|x_0) \bar{f}_\sigma(x_0|y_0)$$
$$\times \left(\int_{-\infty}^{+\infty} dy (y-x)^2 f_\sigma(y|x) + 2(x-y_0) \int_{-\infty}^{+\infty} dy (y-x) f_\sigma(y|x) + (x-y_0)^2 \int_{-\infty}^{+\infty} dy f_\sigma(y|x) \right)$$

$$
\begin{aligned}
&= \int_{-\infty}^{+\infty} dx_0 \int_{-\infty}^{+\infty} dx f_\tau(x|x_0)\bar{f}_\sigma(x_0|y_0)(\sigma^2 + 0 + (x-y_0)^2) \\
&= \int_{-\infty}^{+\infty} dx_0 \bar{f}_\sigma(x_0|y_0) \int_{-\infty}^{+\infty} dx(\sigma^2 + (x-y_0)^2) f_\tau(x|x_0) \\
&= \int_{-\infty}^{+\infty} dx_0 \bar{f}_\sigma(x_0|y_0) \int_{-\infty}^{+\infty} dx(\sigma^2 + (x-x_0+x_0-y_0)^2) f_\tau(x|x_0) \\
&= \int_{-\infty}^{+\infty} dx_0 \bar{f}_\sigma(x_0|y_0) \left(\int_{-\infty}^{+\infty} dx(x-x_0)^2 f_\tau(x|x_0) + 2(x_0-y_0) \int_{-\infty}^{+\infty} dx(x-x_0) f_\tau(x|x_0)(\sigma^2 + (x_0-y_0)^2) \int_{-\infty}^{+\infty} dx f_\tau(x|x_0) \right) \\
&= \int_{-\infty}^{+\infty} dx_0 (2\tau D_2(x_0) + 2(x_0-y_0)\tau D_1(x_0) + \sigma^2 + (x_0-y_0)^2)\bar{f}_\sigma(x_0|y_0) + \mathcal{O}(\tau^2) \\
&= 2\tau \int_{-\infty}^{+\infty} dx_0 (D_2(x_0) + (x_0-y_0)D_1(x_0))\bar{f}_\sigma(x_0|y_0) + \sigma^2 + \int_{-\infty}^{+\infty} dx_0 (x_0-y_0)^2 \bar{f}_\sigma(x_0|y_0) + \mathcal{O}(\tau^2) \\
&\equiv \tau \hat{m}_2(y_0) + \sigma^2 + \hat{\gamma}_2(y_0) + \mathcal{O}(\tau^2).
\end{aligned} \tag{A4}
$$

From Eq. (A4) one has $\hat{M}(y_0,0) = \sigma^2 + \hat{\gamma}_2(y_0)$, where $\hat{\gamma}_2(y_0) = \int_{-\infty}^{+\infty} dx_0 (x_0-y_0)^2 \bar{f}_\sigma(x_0|y_0)$. Such observations justify the first estimate for the measurement noise stated in Eq. (9) since when σ is small enough, probability density function $\bar{f}_\sigma(x_0|y_0)$ is similar to $f_\sigma(y_0|x_0)$ [see Eq. (A2a)], and therefore, one can take as a first approximation $\bar{\gamma}_2(y_0) \sim \sigma^2$.

Notice that the last equalities in $\hat{M}_1$ and $\hat{M}_2$ yield first-order approximations under the assumption that $\tau \ll 1$. In Ref. [11] another approach is proposed for the estimation of drift and diffusion coefficients in the case of low sampling rates.

The errors for $\hat{\gamma}_1(y_0)$, $\hat{\gamma}_2(y_0)$, $\hat{m}_1(y_0)$, and $\hat{m}_2(y_0)$ are just given from the linear fit of $\hat{M}_1$ and $\hat{M}_2$ for each fixed y_0, given in Eqs. (8a) and (8b). The errors of $\hat{M}_1(y,\tau)$ and $\hat{M}_2(y,\tau)$ can also be directly computed from the data as

$$
\begin{aligned}
\sigma^2_{\hat{M}_1}(y,\tau) &= \langle [(y(t+\tau)-y(t)) - \langle y(t+\tau)-y(t)\rangle]^2 \rangle_{t\in\{t_1,\ldots,t_n\}} \\
&= \langle (y(t+\tau)-y(t))^2 + \hat{M}_1^2(y_0,\tau) - 2\hat{M}_1(y_0,\tau)(y(t+\tau)-y(t)) \rangle \\
&= \frac{1}{N_y}(\hat{M}_2(y_0,\tau) + \hat{M}_1^2(y_0,\tau) - 2\hat{M}_1^2(y_0,\tau)) \\
&= \frac{\hat{M}_2(y,\tau) - \hat{M}_1^2(y,\tau)}{N_y},
\end{aligned} \tag{A5a}
$$

$$
\begin{aligned}
\sigma^2_{\hat{M}_2}(y,\tau) &= \langle [(y(t+\tau)-y(t))^2 - \langle (y(t+\tau)-y(t))^2 \rangle]^2 \rangle_{t\in\{t_1,\ldots,t_n\}} \\
&= \langle (y(t+\tau)-y(t))^4 + \hat{M}_2^2(y_0,\tau) - 2\hat{M}_2(y_0,\tau)(y(t+\tau)-y(t))^2 \rangle \\
&= \frac{1}{N_y}(\hat{M}_4(y_0,\tau) + \hat{M}_2^2(y_0,\tau) - 2\hat{M}_2^2(y_0,\tau)) \\
&= \frac{\hat{M}_4(y,\tau) - \hat{M}_2^2(y,\tau)}{N_y},
\end{aligned} \tag{A5b}
$$

where N_y is the number of data points in bin y.

For the optimization procedure it is convenient to simplify the expressions for functions m_i and γ_i $(i=1,2)$. Namely, m_1 and m_2 can be written as expressions of γ_1 and γ_2. In fact, substituting Eqs. (7a) and (7b) into Eqs. (10c) and (10d) and adding and subtracting properly y yields

$$\begin{aligned} m_1(y) &= \int_{-\infty}^{+\infty} D_1(x)\bar{f}_\sigma(x|y)dx = \int_{-\infty}^{+\infty} (d_{10}+d_{11}x)\bar{f}_\sigma(x|y)dx = \int_{-\infty}^{+\infty} [d_{10}+d_{11}(x+y-y)]\bar{f}_\sigma(x|y)dx \\ &= d_{10}\int_{-\infty}^{+\infty} \bar{f}_\sigma(x|y)dx + d_{11}\int_{-\infty}^{+\infty} (x-y)\bar{f}_\sigma(x|y)dx + d_{11}y\int_{-\infty}^{+\infty} \bar{f}_\sigma(x|y)dx \\ &= d_{10}+d_{11}(y+\gamma_1(y)), \end{aligned} \tag{A6a}$$

$$\begin{aligned} m_2(y) &= 2\int_{-\infty}^{+\infty} [(x-y)D_1(x)+D_2(x)]\bar{f}_\sigma(x|y)dx = 2\int_{-\infty}^{+\infty} [(x-y)(d_{10}+d_{11}x)+d_{20}+d_{21}x+d_{22}x^2]\bar{f}_\sigma(x|y)dx \\ &= 2\int_{-\infty}^{+\infty} [(x-y)(d_{10}+d_{11}(x-y+y))+d_{20}+d_{21}(x-y+y)+d_{22}(x-y+y)^2]\bar{f}_\sigma(x|y)dx \\ &= 2[\gamma_1(y)d_{10}+(\gamma_2(y)+y\gamma_1(y))d_{11}+d_{20}+(\gamma_1(y)+y)d_{21}+(2y\gamma_1(y)+\gamma_2(y)+y^2)d_{22}]. \end{aligned} \tag{A6b}$$

Substituting Eqs. (A6a) and (A6b) into Eq. (11) yields F as a functional depending only on the integrals $\gamma_1(y)$ and $\gamma_2(y)$ defined in Eqs. (10a) and (10b), apart from the six parameters, σ and d_{jk}, we want to optimize.

APPENDIX B: THE PROBABILITY DENSITY FUNCTION $\bar{f}_\sigma(x|y)$

To solve the minimization problem we will need to explicitly write expressions for $\bar{f}_\sigma(x|y)$. This conditional probability density function appears in Eqs. (10a) and (10b) and according to Bayes theorem is given by

$$\bar{f}_\sigma(x|y) = \frac{f_\sigma(y|x)p(x)}{\int_{-\infty}^{+\infty} f_\sigma(y|x')p(x')dx'}, \tag{B1}$$

where $f_\sigma(y|x)$ is the probability density function of the measurement noise $\sigma\zeta_t$, i.e., a Gaussian function centered at y with variance σ^2,

$$f_\sigma(y|x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-(y-x)^2/2\sigma^2}, \tag{B2}$$

and $p(x)$ can be written, assuming that the process is stationary, as

$$p(x) = \frac{\mathcal{N}}{D_2(x)}e^{\Phi(x)}, \tag{B3}$$

where $\mathcal{N}$ is some normalized function such that $\int_{-\infty}^{\infty}p(x)dx=1$ and

$$\Phi(x) = \int_{-\infty}^{x} \frac{D_1(x')}{D_2(x')}dx'. \tag{B4}$$

For an Ornstein-Uhlenbeck process $D_1(x)=d_{10}+d_{11}x$ and $D_2(x)=d_{20}$ one finds

$$p_{OU}(x) = \sqrt{-\frac{d_{11}}{2d_{20}\pi}}e^{1/2d_{11}/d_{20}(x+d_{10}/d_{11})^2}, \tag{B5}$$

from which one easily sees that $d_{11}<0$ is a necessary condition to have a well-defined probability density function $p(x)$.

For the general case given by Eqs. (7) one has typically $D_2(x)>0$ with $d_{22}>0$, which yields $\Delta\equiv 4d_{20}d_{22}-d_{21}^2>0$. In these situations, $p(x)$ can also be integrated, yielding

$$p_G(x) = \mathcal{N}(D_2(x))^{d_{11}/2d_{22}-1}e^{(d_{10}-d_{21}d_{11}/2d_{22})h_0(x)}, \tag{B6}$$

with

$$h_0(x) = \frac{2}{\sqrt{\Delta}}\left[\arctan\left(\frac{2d_{22}x+d_{21}}{\sqrt{\Delta}}\right)+\frac{\pi}{2}\right]. \tag{B7}$$

APPENDIX C: THE DERIVATIVES OF γ_1, γ_2, m_1, AND m_2

The minimization problem needs also the expression of the derivatives for the γs and ms. To compute them one needs first to write the derivatives of function $\bar{f}_\sigma(x|y)$ defined in Eq. (B1).

Defining $g(x,y)\equiv f_\sigma(y|x)p(x)$ one has in general

$$\frac{\partial\bar{f}_\sigma(x|y)}{\partial X} = \frac{\frac{\partial g}{\partial X}\int_{-\infty}^{+\infty} g(x',y)dx' - g\int_{-\infty}^{+\infty}\frac{\partial g}{\partial X}dx'}{\left(\int_{-\infty}^{+\infty} g(x',y)dx'\right)^2}, \tag{C1}$$

where X is some variable on which $\bar{f}_\sigma$ depends. Since $p(x)$ depends only on parameters d_{ij} and $f_\sigma(y|x)$ depends only on σ, we have

$$\frac{\partial g(x,y)}{\partial \sigma} = \frac{\partial f_\sigma(y|x)}{\partial \sigma} p(x), \tag{C2a}$$

$$\frac{\partial g(x,y)}{\partial d_{ij}} = \frac{\partial p(x)}{\partial d_{ij}} f_\sigma(y|x), \tag{C2b}$$

where for $f_\sigma(y|x)$ we have

$$\frac{\partial f_\sigma(y|x)}{\partial \sigma} = f_\sigma(y|x)\frac{(x-y)^2}{\sigma^3} \tag{C3}$$

and for $p(x)$ we have

$$p(x) = \frac{\mathcal{N}}{D_2(x)} e^{\Phi(x)} \equiv \mathcal{N}\hat{p}(x), \tag{C4}$$

with

$$\mathcal{N} = \left(\int_{-\infty}^{+\infty} \hat{p}(x)dx \right)^{-1} \tag{C5}$$

and therefore

$$\frac{\partial p(x)}{\partial X} = \mathcal{N}\left(\frac{\partial \hat{p}(x)}{\partial X} - p(x)\int_{-\infty}^{+\infty} \frac{\partial \hat{p}(x')}{\partial X} dx' \right), \tag{C6}$$

with X one of the d parameters.

In the Ornstein-Uhlenbeck case

$$\frac{\partial \hat{p}_{OU}(x)}{\partial d_{10}} = \frac{1}{d_{20}}\left(x + \frac{d_{10}}{d_{11}} \right) p_{OU}(x), \tag{C7a}$$

$$\frac{\partial \hat{p}_{OU}(x)}{\partial d_{11}} = \left[\frac{1}{2d_{20}}\left(x^2 - \frac{d_{10}^2}{d_{11}^2} \right) + \frac{1}{2d_{11}} \right] p_{OU}(x), \tag{C7b}$$

$$\frac{\partial \hat{p}_{OU}(x)}{\partial d_{20}} = -\frac{1}{2d_{20}}\left[1 + \frac{d_{11}}{d_{20}}\left(x + \frac{d_{10}}{d_{11}} \right)^2 \right] p_{OU}(x), \tag{C7c}$$

and in the general case

$$\frac{\partial \hat{p}_G(x)}{\partial d_{10}} = h_0(x) p_G(x), \tag{C8a}$$

$$\frac{\partial \hat{p}_G(x)}{\partial d_{11}} = \left(\frac{1}{2d_{22}} \log D_2(x) - \frac{d_{21}}{2d_{22}} h_0(x) \right) p_G(x), \tag{C8b}$$

$$\frac{\partial \hat{p}_G(x)}{\partial d_{20}} = \left(\frac{1}{D_2(x)} + \frac{\partial h_0(x)}{\partial d_{20}} \right) p_G(x), \tag{C8c}$$

$$\frac{\partial \hat{p}_G(x)}{\partial d_{21}} = \left[\frac{x}{D_2(x)} - \frac{d_{11}}{2d_{22}} h_0(x) + \left(d_{10} - \frac{d_{21}d_{11}}{2d_{22}} \right) \frac{\partial h_0(x)}{\partial d_{21}} \right] p_G(x), \tag{C8d}$$

$$\frac{\partial \hat{p}_G(x)}{\partial d_{22}} = \left[\left(\frac{d_{11}}{2d_{22}} - 1 \right) \frac{x^2}{D_2(x)} - \frac{d_{11}}{2d_{22}^2} \log D_2(x) + \frac{d_{21}d_{11}}{2d_{22}^2} h_0(x) + \left(d_{10} - \frac{d_{21}d_{11}}{2d_{22}} \right) \frac{\partial h_0(x)}{\partial d_{22}} \right] p_G(x), \tag{C8e}$$

where

$$\frac{\partial h_0(x)}{\partial d_{20}} = -\frac{2d_{22}}{\Delta} h_0(x) - \frac{4d_{22}(2d_{22}x + d_{21})}{\Delta[\Delta + (2d_{22}x + d_{21})^2]}, \tag{C9a}$$

$$\frac{\partial h_0(x)}{\partial d_{21}} = \frac{d_{21}}{\Delta} h_0(x) + \frac{2}{\Delta}\frac{\Delta + d_{21}(2d_{22}x + d_{21})}{\Delta + (2d_{22}x + d_{21})^2}, \tag{C9b}$$

$$\frac{\partial h_0(x)}{\partial d_{22}} = -\frac{2d_{20}}{\Delta} h_0(x) + \frac{4}{\Delta}\frac{x\Delta - d_{20}(2d_{22}x + d_{21})}{\Delta + (2d_{22}x + d_{21})^2}. \tag{C9c}$$

So, neglecting the parameter d_{10} as explained in Sec. III, for the other parameters $\sigma, d_{11}, d_{20}, d_{21}, d_{22}$ we have

$$\frac{\partial \bar{f}_\sigma(x|y)}{\partial \sigma} = \frac{1}{\sigma^3}[(x-y)^2 - \gamma_2(y)]\bar{f}_\sigma(x|y), \tag{C10a}$$

$$\frac{\partial \bar{f}_\sigma(x|y)}{\partial d_{ij}} = \frac{e^{-(x-y)^2/2\sigma^2}\frac{\partial p(x)}{\partial d_{ij}} - \bar{f}_\sigma(x|y)\int_{-\infty}^{+\infty} e^{-(x'-y)^2/2\sigma^2}\frac{\partial p(x')}{\partial d_{ij}} dx'}{\int_{-\infty}^{+\infty} g(x',y)dx'} \tag{C10b}$$

and therefore considering Eqs. (10a) and (10b) that define functions $\gamma_1(y)$ and $\gamma_2(y)$ and also Eqs. (A6a) and (A6b) defining functions $m_1(y)$ and $m_2(y)$ it follows

$$\frac{\partial \gamma_1(y)}{\partial \sigma} = \frac{1}{\sigma^3}[h_1(y) - \gamma_1(y)\gamma_2(y)], \tag{C11a}$$

$$\frac{\partial \gamma_2(y)}{\partial \sigma} = \frac{1}{\sigma^3}[h_2(y) - \gamma_2^2(y)], \tag{C11b}$$

$$\frac{\partial m_1(y)}{\partial \sigma} = d_{11}\frac{\partial \gamma_1(y)}{\partial \sigma}, \tag{C11c}$$

$$\frac{\partial m_2(y)}{\partial \sigma} = 2\left\{ [d_{21} + y(d_{11} + 2d_{22})]\frac{\partial \gamma_1(y)}{\partial \sigma} + (d_{11} + d_{22})\frac{\partial \gamma_2(y)}{\partial \sigma} \right\}, \tag{C11d}$$

$$\frac{\partial \gamma_1(y)}{\partial d_{ij}} = \int_{-\infty}^{+\infty} (x' - y)\frac{\partial \bar{f}_\sigma(x'|y)}{\partial d_{ij}} dx', \tag{C11e}$$

$$\frac{\partial \gamma_2(y)}{\partial d_{ij}} = \int_{-\infty}^{+\infty} (x'-y)^2 \frac{\partial \bar{f}_\sigma(x'|y)}{\partial d_{ij}} dx', \tag{C11f}$$

$$\frac{\partial m_1(y)}{\partial d_{11}} = y + \gamma_1(y) + d_{11}\frac{\partial \gamma_1(y)}{\partial d_{11}}, \tag{C11g}$$

$$\frac{\partial m_1(y)}{\partial d_{2j}} = d_{11}\frac{\partial \gamma_1(y)}{\partial d_{2j}}, \tag{C11h}$$

$$\frac{\partial m_2(y)}{\partial d_{11}} = 2\left\{[d_{21} + y(d_{11}+2d_{22})]\frac{\partial \gamma_1(y)}{\partial d_{11}} + (d_{11}+d_{22})\frac{\partial \gamma_2(y)}{\partial d_{11}} + \gamma_2(y) + y\gamma_1(y)\right\}, \tag{C11i}$$

$$\frac{\partial m_2(y)}{\partial d_{20}} = 2\left\{[d_{21} + y(d_{11}+2d_{22})]\frac{\partial \gamma_1(y)}{\partial d_{20}} + (d_{11}+d_{22})\frac{\partial \gamma_2(y)}{\partial d_{20}} + 1\right\}, \tag{C11j}$$

$$\frac{\partial m_2(y)}{\partial d_{21}} = 2\left\{[d_{21} + y(d_{11}+2d_{22})]\frac{\partial \gamma_1(y)}{\partial d_{21}} + (d_{11}+d_{22})\frac{\partial \gamma_2(y)}{\partial d_{21}} + \gamma_1(y) + y\right\}, \tag{C11k}$$

$$\frac{\partial m_2(y)}{\partial d_{22}} = 2\left\{[d_{21} + y(d_{11}+2d_{22})]\frac{\partial \gamma_1(y)}{\partial d_{22}} + (d_{11}+d_{22})\frac{\partial \gamma_2(y)}{\partial d_{22}} + 2y\gamma_1(y) + \gamma_2(y) + y^2\right\}, \tag{C11l}$$

where

$$h_1(y) = \int_{-\infty}^{+\infty} (x'-y)^3 \bar{f}_\sigma(x'|y) dx', \tag{C12a}$$

$$h_2(y) = \int_{-\infty}^{+\infty} (x'-y)^4 \bar{f}_\sigma(x'|y) dx'. \tag{C12b}$$

[1] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis* (Cambridge University Press, Cambridge, England, 1997).
[2] R. Friedrich, J. Peinke, and M. R. R. Tabar, *Complexity in the View of Stochastic Processes in Springer Encyclopedia of Complexity and Systems Science* (Springer, Berlin, 2008).
[3] H. D. I. Abarbanel, R. Brown, J. J. Sidorowich, and L. S. Tsimring, Rev. Mod. Phys. **65**, 1331 (1993).
[4] R. Friedrich and J. Peinke, Phys. Rev. Lett. **78**, 863 (1997).
[5] A. P. Nawroth, J. Peinke, D. Kleinhans, and R. Friedrich, Phys. Rev. E **76**, 056102 (2007).
[6] R. Friedrich, J. Peinke, and Ch. Renner, Phys. Rev. Lett. **84**, 5224 (2000).
[7] C. Collette and M. Ausloos, Int. J. Mod. Phys. C **15**, 1353 (2004).
[8] P. G. Lind, A. Mora, J. A. C. Gallas, and M. Haase, Phys. Rev. E **72**, 056706 (2005).
[9] J. Prusseit and K. Lehnertz, Phys. Rev. E **77**, 041914 (2008).
[10] D. Lamouroux and K. Lehnertz, Phys. Lett. A **373**, 3507 (2009).
[11] D. Kleinhans, R. Friedrich, A. Nawroth, and J. Peinke, Phys. Lett. A **346**, 42 (2005).
[12] J. Gottschall and J. Peinke, New J. Phys. **10**, 083034 (2008).
[13] F. Bottcher, J. Peinke, D. Kleinhans, R. Friedrich, P. G. Lind, and M. Haase, Phys. Rev. Lett. **97**, 090603 (2006).
[14] M. Siefert, A. Kittel, R. Friedrich, and J. Peinke, Europhys. Lett. **61**, 466 (2003).
[15] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes* (Cambridge University Press, Cambridge, 1992).
[16] P. G. Lind, A. Mora, M. Haase, and J. A. C. Gallas, Int. J. Bifurcation Chaos Appl. Sci. Eng. **17**, 3461 (2007).
[17] J.-P. Bouchaud and A. Georges, Phys. Rep. **195**, 127 (1990).
[18] R. Metzler and J. Klafter, Phys. Rep. **339**, 1 (2000).
[19] J. W. Hurrell, Science **269**, 676 (1995).
[20] D. B. Stephenson, V. Pavan, and R. Bojariu, Int. J. Climatol. **20**, 1 (2000).
[21] S. Siegert and R. Friedrich, Phys. Rev. E **64**, 041107 (2001).
[22] D. Kleinhans, R. Friedrich, M. Wächter, and J. Peinke, Phys. Rev. E **76**, 041109 (2007).
[23] I. Pavlyukevich, J. Comput. Phys. **226**, 1830 (2007).
[24] D. Goldberg and K. Sastry, *Genetic Algorithms: The Design of Innovation* (Springer, Heidelberg, 2008).
[25] http://www.cru.uea.ac.uk/cru/data/nao.htm (2005).