Analyzing a stochastic time series obeying a second-order differential equation

B. Lehle and J. Peinke

Institute of Physics, University of Oldenburg, D-2611 Oldenburg, Germany

(Received 6 December 2014; published 10 June 2015)

The stochastic properties of a Langevin-type Markov process can be extracted from a given time series by a Markov analysis. Also processes that obey a stochastically forced second-order differential equation can be analyzed this way by employing a particular embedding approach: To obtain a Markovian process in $2N$ dimensions from a non-Markovian signal in N dimensions, the system is described in a phase space that is extended by the temporal derivative of the signal. For a discrete time series, however, this derivative can only be calculated by a differencing scheme, which introduces an error. If the effects of this error are not accounted for, this leads to systematic errors in the estimation of the drift and diffusion functions of the process. In this paper we will analyze these errors and we will propose an approach that correctly accounts for them. This approach allows an accurate parameter estimation and, additionally, is able to cope with weak measurement noise, which may be superimposed to a given time series.

DOI: 10.1103/PhysRevE.91.062113 PACS number(s): 02.50.Ey, 02.50.Ga, 05.40.Ca

I. INTRODUCTION

Many dynamical systems can be modelled as continuous-time Markov process $\mathbf{Y}(t)$ that is driven by Gaussian white noise $\boldsymbol{\xi}(t)$ with $\langle \xi_i(t) \rangle = 0$ and $\langle \xi_i(t)\xi_j(t') \rangle = \delta_{ij}\delta(t-t')$. Such a process is commonly referred to as diffusion process. Its temporal evolution obeys a Langevin equation—a first-order ordinary differential equation (ODE) that is stochastically forced,

$$\dot{\mathbf{Y}} = \mathbf{a}(\mathbf{Y}) + \mathbf{b}(\mathbf{Y})\,\boldsymbol{\xi}(t). \tag{1}$$

Here and in the following Itô's definition of a stochastic integral is used [1]. Furthermore, a stationary stochastic process is looked at, whereas in general $\mathbf{a}$ and $\mathbf{b}$ may depend on time.

The Kramers-Moyal coefficients of the Fokker-Planck equation corresponding to Eq. (1) are denoted by $\mathbf{D}^{(1)}$ and $\mathbf{D}^{(2)}$ and commonly referred to as drift and diffusion functions, respectively [2]. These functions uniquely define the stochastic process and are related to $\mathbf{a}$ and $\mathbf{b}$ by

$$\mathbf{D}^{(1)}(\mathbf{y}) = \mathbf{a}(\mathbf{y}), \quad \mathbf{D}^{(2)}(\mathbf{y}) \;=\; \mathbf{b}(\mathbf{y})\mathbf{b}^T(\mathbf{y}). \tag{2}$$

There are several approaches that allow us to estimate these functions from experimental data. A method based on spectral data is presented in [3,4]. Based on the spectrum of the finite-increment transition probability matrix of a given time series, a reference spectrum for the generator of the diffusion process is calculated. An estimate for the generator (and thus for the Kramers-Moyal coefficients) then is obtained by minimizing an object function that measures the distance between reference and generator spectrum. Other approaches use Bayesian frameworks, where the process parameters are estimated by maximizing their likelihood [5–8]. Such approaches also allow the analysis of partially observed processes and processes that are spoilt by observation errors.

In the following an approach called Markov analysis will be looked at. This technique, also denoted as the direct estimation method, has been introduced in the late 1990s [9–12]. As it provides a very simple and direct way to estimate the process parameters from a given time series, it has found widespread use since then. Reviews on Markov analysis and its applications can be found, e.g., in [13,14].

The method is based on the fact that the moments $\mathbf{M}^{(k)}$ of the conditional process increments of $\mathbf{Y}$ can be expressed in terms of the Kramers-Moyal coefficients,

$$\begin{aligned}\mathbf{M}^{(k)}(\mathbf{y},\tau) :&= \langle [\mathbf{Y}(t+\tau) - \mathbf{Y}(t)]^k \rangle|_{\mathbf{Y}(t)=\mathbf{y}} \\ &= \tau \mathbf{D}^{(k)}(\mathbf{y}) + O(\tau^2), \quad k \;=\; 1,2.\end{aligned} \tag{3}$$

Here and in the following the kth power of a vector denotes a k-fold dyadic product. The time argument t of $\mathbf{M}^{(k)}$ is suppressed here because a stationary process is assumed. This assumption also allows a moment estimation from a single time series—ensemble averages can be replaced by time averages then (tacitly assuming ergodicity). For a nonstationary process an ensemble of time series would be needed (alternatively a windowing strategy could be applied, assuming a slowly varying time dependence).

The moments $\mathbf{M}^{(k)}$ [Eq. (3)] can be expressed in terms of moments $\mathbf{m}^{(k)}$ of the two-point probability density function (PDF) of $\mathbf{Y}$ at times t and $t+\tau$. These moments $\mathbf{m}^{(k)}$ are defined as

$$\mathbf{m}^{(k)}(\mathbf{y},\tau) := \int_{\mathbf{s}} (\mathbf{s}-\mathbf{y})^k\, p(\mathbf{y},t;\mathbf{s},t+\tau)\, \boldsymbol{ds}, \tag{4}$$

where again the time argument t is suppressed because of the assumption of stationarity. Using the well known relations $p(a;b) = p(b)p(a|b)$ and $\int_a f(a)p(a|b) = \langle f(A)|b \rangle$ leads to

$$\mathbf{m}^{(k)}(\mathbf{y},\tau) = p(\mathbf{y},t)\, \mathbf{M}^{(k)}(\mathbf{y},\tau). \tag{5}$$

For $k=0$, this yields $m^{(0)}(\mathbf{y}) = p(\mathbf{y},t)$ (suppressing the unneeded argument τ and taking into account the scalar nature of $m^{(0)}$). Consequently one can write $\mathbf{M}^{(k)} = \mathbf{m}^{(k)}/m^{(0)}$ and one obtains

$$\frac{\mathbf{m}^{(k)}(\mathbf{y},\tau)}{m^{(0)}(\mathbf{y})} = \tau \mathbf{D}^{(k)}(\mathbf{y}) + O(\tau^2), \quad k = 1,2. \tag{6}$$

The moments $\mathbf{m}^{(k)}(\mathbf{y},\tau)$ can directly be estimated from a given time series. In practice, this is usually done by applying a binning approach. Estimating the moments for a number of time increments τ then allows us to solve Eq. (6) for $\mathbf{D}^{(k)}(\mathbf{y})$

in a least square sense. Usually a low order polynomial in τ is used for a fit of the right hand side, as the higher order terms in the above equation are known to be powers of τ [15,16]. This strategy will be denoted as standard Markov analysis (SMA) in the following.

However, not every problem can be modelled by Eq. (1). This applies, e.g., to the large class of physical and engineering systems that are described by force equations. Such systems can only be modelled by a first-order equation if they are dominated by damping, i.e., when inertial forces can be neglected. In general, they need to be modelled by a process $\mathbf{X}(t)$ obeying a *second-order* ODE,

$$\ddot{\mathbf{X}} = \mathbf{f}(\mathbf{X},\dot{\mathbf{X}}) + \mathbf{g}(\mathbf{X},\dot{\mathbf{X}})\,\boldsymbol{\xi}(t), \quad \mathbf{X} \in \mathbb{R}^N, \tag{7}$$

where $\boldsymbol{\xi}(t)$ denotes Gaussian white noise again. Another example would be a Langevin-like process $\mathbf{X}(t)$ that is not driven by white noise but by an independent Langevin process, say $\mathbf{Z}(t)$, which itself is driven by white noise. Due to the continuity of $\mathbf{Z}$ the dynamic of $\mathbf{X}$ then can be captured by an equation of the above form.

As Eq. (7) is a second-order ODE, such a process is not Markovian, i.e., the statistics of its increments do not only depend on the value of $\mathbf{X}$ but also on its derivative. In an extended phase space, however, consisting of the values of $\mathbf{X}$ and $\dot{\mathbf{X}}$, the dynamic becomes Markovian. With the definitions

$$\mathbf{Y}_1(t) := \mathbf{X}(t), \quad \mathbf{Y}_2(t) := \dot{\mathbf{X}}(t), \tag{8}$$

Eq. (7) can be written as a system of first-order equations that define a Langevin process $\mathbf{Y}^T(t) := [\mathbf{Y}_1^T(t),\mathbf{Y}_2^T(t)]$ in $2N$ dimensions,

$$\dot{\mathbf{Y}} = \begin{bmatrix}\dot{\mathbf{Y}}_1\\ \dot{\mathbf{Y}}_2\end{bmatrix} = \begin{bmatrix}\mathbf{Y}_2\\ \mathbf{f}(\mathbf{Y}) + \mathbf{g}(\mathbf{Y})\,\boldsymbol{\xi}(t)\end{bmatrix}. \tag{9}$$

The Kramers-Moyal coefficients of the corresponding Fokker-Planck equation are simpler than in the general $2N$-dimensional case, as they are given by

$$\mathbf{D}^{(1)}(\mathbf{y}) = \begin{bmatrix}\mathbf{y}_2\\ \mathbf{f}(\mathbf{y})\end{bmatrix}, \quad \mathbf{D}^{(2)}(\mathbf{y}) = \begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{g}\mathbf{g}^T(\mathbf{y})\end{bmatrix}. \tag{10}$$

Of course, these coefficients can be estimated from a given time series of $\mathbf{Y}(t)$ by the above mentioned SMA. But for that the values of $\mathbf{X}$ *and* $\dot{\mathbf{X}}$ must be given [Eq. (8)]. For real word data this will not always be the case. Frequently only a series of "positions" $\mathbf{Y}_1(t) \equiv \mathbf{X}(t)$ will be given for a second-order process obeying Eq. (7), while the corresponding "velocities" $\mathbf{Y}_2(t) \equiv \dot{\mathbf{X}}(t)$ are missing. It may, for example, be hard to accurately measure the velocities in a given experimental setup. Or it may not have been realized in advance that $\mathbf{X}(t)$ needs to be modelled as a second-order process. Or it may simply have been assumed that a highly resolved series of position values will provide sufficiently accurate information on the velocities.

If $\dot{\mathbf{X}}$ is missing, these velocity values need to be estimated numerically. The errors associated with this estimation will be called "reconstruction errors" in the following (after all, we try to reconstruct the missing series $\dot{\mathbf{X}}$). These errors seem to impose no major problem as $\mathbf{X}(t)$ is a continuously differentiable function. Its derivative can be estimated by a discrete differencing scheme with arbitrary accuracy—provided the step size of the scheme (here and in the following denoted by θ) can be chosen small enough. So for a "sufficiently" fine sampled series of positions the reconstruction errors will become negligible. The standard approach for an analysis, therefore, goes like this: Choose some small step size θ and estimate the series $\mathbf{Y}_2$ using the given series $\mathbf{Y}_1$. Then apply a SMA to the resulting series $\mathbf{Y}$. This strategy will be denoted as the standard embedding approach (SEA) in the following.

This approach, however, has its flaws. For a Markov analysis, the moments of process *increments* will be looked at [see Eq. (3)]. For these quantities the reconstruction errors will show to be of importance unless the step size θ (used for velocity estimation) can be chosen much smaller than the time increment τ (used for increment calculation). At the same time, however, τ needs to be small compared to the characteristic time scale T of the process under investigation. Otherwise the higher order terms in Eq. (6) can no longer be approximated by a low order polynomial. The requirement $\theta \ll \tau \ll T$ will only rarely be fulfilled in practice as it requires data with a very high temporal resolution (compared to the characteristic time scale T). Further discussion on the effects of the choice of τ in the presence of multiple time scales can, for example, be found in [17].

Also another source of errors has to be considered for real data: Virtually all experimental data will be superimposed by additional observation errors, in the following referred to as measurement noise. This is a well known problem and there are several approaches to deal with it [18–23]. In the given context, we face a special variant of this problem. We will apply a differencing scheme to the given time series $\mathbf{Y}_1$. So any measurement noise that afflicts the values of $\mathbf{Y}_1$ will lead to an additional error in the estimation of $\mathbf{Y}_2$. For a differencing scheme with step size θ, this error will be proportional to θ^{-1}, as will be seen later (assuming uncorrelated measurement noise). So even if the measurement noise is very small, and thus negligible for $\mathbf{Y}_1$ itself, it may become important in the estimation of $\mathbf{Y}_2$ for small values of θ.

The above considerations imply that for real data neither the values of $\mathbf{Y}_1(t)$ nor that of $\mathbf{Y}_2(t)$ are known accurately. The values actually at hand will be denoted by $\mathbf{Y}^*(t)$ and referred to as "noisy" values in the following. Here the term noisy does *not* refer to the dynamic noise of the stochastic process but to the errors caused by measurement noise and differencing scheme.

The aim of this paper is to provide a modified embedding approach (MEA) that accounts for these errors. As a by-product also a quantitative description of the errors of the SEA will be found. However, only *weak* measurement noise can be accounted for. This restriction is a consequence of the perturbative approach that will be used. The requirements on the noise will be given later, but, roughly speaking, the noise must be negligible for the position values and its effect on the velocity increments may at most be of the same order as the effects of the driving stochastic force $\boldsymbol{\xi}$.

We want to emphasize that our direct approach does not focus on any smoothing methods (like, for example, different denoising methods do, which average out different noise contributions). Instead, we use the unprocessed, noisy data and try to extract from these directly the spoiling measurement noise, the involved dynamical Langevin noise, and finally the underlying deterministic dynamics.

This paper is organized as follows: In Sec. II the observable moments $\mathbf{m}^{*(k)}$ of the noisy time series will be expressed in terms of moments of the *noisy* values $\mathbf{Y}^*(t)$ and $\mathbf{Y}^*(t+\tau)$ conditioned on the *true* value $\mathbf{Y}(t)$. Subsequently, based on a Taylor-Itô expansion, these conditional noisy values will be expressed in terms of process parameters, measurement noise, and stochastic integrals of $\boldsymbol{\xi}$ in Sec. III. The resulting expressions, together with an assumption on the magnitude of the measurement noise, will lead to an explicit description of $\mathbf{m}^{*(k)}$ in Sec. IV then. This description will serve two purposes. First, the effects of the reconstruction errors of a SEA can be quantified (Sec. V). Second, a MEA can be specified that allows an accurate estimation of the Kramers-Moyal coefficients and the properties of the measurement noise (Sec. VI). Subsequently a numerical test case will be specified in Sec. VII, which will be used to compare the results of SEA and MEA with and without measurement noise (Secs. VIII and IX).

II. MOMENTS OF THE NOISY VALUES

For a series of noisy values $\mathbf{Y}^*$ only the noisy counterparts $\mathbf{m}^{*(k)}$ of the moments $\mathbf{m}^{(k)}$ can be estimated. In analogy to Eq. (4) they can be defined as

$$\mathbf{m}^{*(k)}(\mathbf{y}^*) := \int_{\mathbf{s}} (\mathbf{s}-\mathbf{y}^*)^k\, p(\mathbf{Y}^*=\mathbf{y}^*;\mathbf{Y}^*_\tau=\mathbf{s})\, d\mathbf{s}. \quad (11)$$

Here and in the following, the time arguments t and τ are omitted to allow for a more compact notation. Stochastic variables implicitly refer to time t now, and the shortcut $\mathbf{Y}^*_\tau$ is used to denote $\mathbf{Y}^*(t+\tau)$.

Next the moments $\mathbf{m}^{*(k)}$ need to be related to the process parameters and the properties of the measurement noise. As outlined in Sec. I, the first step will be to express the moments $\mathbf{m}^{*(k)}$ in terms of moments of the conditional noisy values $\mathbf{Y}^*|_{\mathbf{Y}=\mathbf{y}}$ and $\mathbf{Y}^*_\tau|_{\mathbf{Y}=\mathbf{y}}$. This can be done as follows: First, the PDF in Eq. (11) is rewritten as

$$p(\mathbf{Y}^*=\mathbf{y}^*;\mathbf{Y}^*_\tau=\mathbf{s}) \equiv \int_{\mathbf{y}} \rho(\mathbf{y})p(\mathbf{Y}^*=\mathbf{y}^*;\mathbf{Y}^*_\tau=\mathbf{s}|\mathbf{Y}=\mathbf{y})\, d\mathbf{y}, \quad (12)$$

where $\rho(\mathbf{y}) := p(\mathbf{Y}=\mathbf{y})$ denotes the PDF of $\mathbf{Y}$. Inserting Eq. (12) and interchanging the order of integration thus allows us to write the moments $\mathbf{m}^{*(k)}$ in the form

$$\mathbf{m}^{*(k)}(\mathbf{y}^*) = \int_{\mathbf{y}} \rho(\mathbf{y})\, \mathbf{F}^{(k)}(\mathbf{y}^*,\mathbf{y})\, d\mathbf{y} \quad (13)$$

with

$$\mathbf{F}^{(k)}(\mathbf{y}^*,\mathbf{y}) = \int_{\mathbf{s}} (\mathbf{s}-\mathbf{y}^*)^k\, p(\mathbf{Y}^*=\mathbf{y}^*;\mathbf{Y}^*_\tau=\mathbf{s}|\mathbf{Y}=\mathbf{y})\, d\mathbf{s}. \quad (14)$$

Expressing the integral in Eq. (13) by a moment expansion yields (using summation convention)

$$m^{*(k)}_{i_1,\ldots,i_k}(\mathbf{y}^*) = \sum_{\nu=0}^{\infty} \frac{(-1)^\nu}{\nu!} \frac{\partial}{\partial y^*_{j_1}} \cdots \frac{\partial}{\partial y^*_{j_\nu}} \times \left[\rho(\mathbf{y}^*) M^{(k,\nu)}_{i_1,\ldots,i_k,j_1,\ldots,j_\nu}(\mathbf{y}^*)\right], \quad (15)$$

where the moments are defined as

$$\mathbf{M}^{(k,\nu)}(\mathbf{y}^*) := \int_{\mathbf{z}} \mathbf{F}^{(k)}(\mathbf{z},\mathbf{y}^*) \otimes (\mathbf{z}-\mathbf{y}^*)^\nu\, d\mathbf{z}. \quad (16)$$

Here $\otimes$ denotes a dyadic product. Inserting the definition of $\mathbf{F}^{(k)}$ first leads to

$$\mathbf{M}^{(k,\nu)}(\mathbf{y}^*) = \int_{\mathbf{s},\mathbf{z}} (\mathbf{s}-\mathbf{z})^k \otimes (\mathbf{z}-\mathbf{y}^*)^\nu \times p(\mathbf{Y}^*=\mathbf{z};\mathbf{Y}^*_\tau=\mathbf{s}|\mathbf{Y}=\mathbf{y}^*)\, d\mathbf{s}\, d\mathbf{z}. \quad (17)$$

Using the relation $\int_a f(a)p(a|b) = \langle f(A)|b\rangle$ then gives

$$\mathbf{M}^{(k,\nu)}(\mathbf{y}^*) = \langle(\mathbf{Y}^*_\tau-\mathbf{Y}^*)^k \otimes (\mathbf{Y}^*-\mathbf{Y})^\nu\rangle|_{\mathbf{Y}=\mathbf{y}^*}. \quad (18)$$

The general form of the observable moments $\mathbf{m}^{*(k)}$ therefore reads (dropping the asterisk on the parameter $\mathbf{y}$)

$$m^{*(k)}_{i_1,\ldots,i_k}(\mathbf{y}) = \sum_{\nu=0}^{\infty} \frac{(-1)^\nu}{\nu!} \frac{\partial}{\partial y_{j_1}} \cdots \frac{\partial}{\partial y_{j_\nu}} \times \left\{\rho(\mathbf{y})\left\langle[\mathbf{A}^k(\mathbf{y})]_{i_1,\ldots,i_k}[\mathbf{B}^\nu(\mathbf{y})]_{j_1,\ldots,j_\nu}\right\rangle\right\} \quad (19)$$

with

$$\mathbf{A}(\mathbf{y}) : = \mathbf{Y}^*_\tau|_{\mathbf{Y}=\mathbf{y}} - \mathbf{Y}^*|_{\mathbf{Y}=\mathbf{y}}\ , \quad (20a)$$

$$\mathbf{B}(\mathbf{y}) : = \mathbf{Y}^*|_{\mathbf{Y}=\mathbf{y}} - \mathbf{y}. \quad (20b)$$

This is a quite general result—no information on how $\mathbf{Y}^*(t)$ and $\mathbf{Y}(t)$ are related is used so far. This will be done in the next section, where the conditional values of $\mathbf{Y}^*$ and $\mathbf{Y}^*_\tau$ will be expressed explicitly.

III. CONDITIONAL VALUES OF Y*

In this section we will specify the assumptions on the measurement noise together with the details of the differencing scheme. This will allow us to express the conditional values of $\mathbf{Y}^*$ and $\mathbf{Y}^*_\tau$ in terms of measurement noise and conditional values of $\mathbf{Y}_1$. Based on a Taylor-Itô expansion, these conditional values $\mathbf{Y}_1$ can then be expressed in terms of the driving stochastic force and process parameters.

To avoid confusion, time arguments will be given explicitly again in the following. However, the shortcut $(\cdots)|_{\mathbf{y}}$ will be used to indicate conditioning on $\mathbf{Y}(t)=\mathbf{y}$.

The given values $\mathbf{Y}^*_1(t)$ are assumed to be spoilt by additive, Gaussian distributed, and temporally uncorrelated measurement noise $\boldsymbol{\Gamma}(t)$ with an expectation value of zero and covariance matrix $\mathbf{V}$

$$\mathbf{Y}^*_1(t) := \mathbf{Y}_1(t) + \boldsymbol{\Gamma}(t), \quad (21)$$

with

$$\langle\boldsymbol{\Gamma}(t)\rangle = \mathbf{0}, \quad (22a)$$

$$\langle\boldsymbol{\Gamma}(t)\boldsymbol{\Gamma}^T(t')\rangle = \delta_{t,t'}\mathbf{V}, \quad \delta_{t,t'} := \begin{cases}1, & t=t'\\ 0, & t\neq t'\end{cases}. \quad (22b)$$

The noise is also assumed to be independent of $\boldsymbol{\xi}$ and $\mathbf{Y}$ [implying $\boldsymbol{\Gamma}(t)|_{\mathbf{y}} \equiv \boldsymbol{\Gamma}(t)$]. The conditional values $\mathbf{Y}^*_1|_{\mathbf{y}}$ are thus given by

$$\mathbf{Y}^*_1(t+\Delta)|_{\mathbf{y}} := \mathbf{Y}_1(t+\Delta)|_{\mathbf{y}} + \boldsymbol{\Gamma}(t+\Delta). \quad (23)$$

Here the symbol Δ is used to denote some arbitrary, non-negative time increment. For the reconstruction of $\mathbf{Y}_2$ a first-order forward differencing scheme with a step size of θ, applied to the observable values $\mathbf{Y}_1^*$, will be used in the following:

$$\mathbf{Y}_2^*(t+\Delta,\theta) := \frac{1}{\theta}[\mathbf{Y}_1^*(t+\Delta+\theta) - \mathbf{Y}_1^*(t+\Delta)]. \quad (24)$$

The conditional values $\mathbf{Y}_2^*|_{\mathbf{y}}$ therefore are given by

$$\mathbf{Y}_2^*(t+\Delta,\theta)|_{\mathbf{y}} := \frac{1}{\theta}[\mathbf{Y}_1(t+\Delta+\theta)|_{\mathbf{y}} - \mathbf{Y}_1(t+\Delta)\big|_{\mathbf{y}}] + \frac{1}{\theta}[\mathbf{\Gamma}(t+\Delta+\theta) - \mathbf{\Gamma}(t+\Delta)]. \quad (25)$$

The values $\mathbf{Y}_1|_{\mathbf{y}}$ at time $t+\Delta$ can be expressed by a Taylor-Itô expansion (see Appendix A)

$$\mathbf{Y}_1(t+\Delta)|_{\mathbf{y}} = \mathbf{y}_1 + \mathbf{y}_2\,\Delta + \mathbf{f}(\mathbf{y})\frac{\Delta^2}{2} + \mathbf{g}(\mathbf{y})\mathbf{I}^{t,\Delta} + \mathbf{R}^{t,\Delta}(\mathbf{y}). \quad (26)$$

Here $\mathbf{I}^{t,\Delta}$ denotes a vector of stochastic integrals that only depend on the realization of $\boldsymbol{\xi}$ in the interval $[t,t+\Delta)$. The components of this vector are of magnitude $O(\Delta^{3/2})$ and have an expectation value of zero. All other expansion terms are summarized in the remainder $\mathbf{R}^{t,\Delta}(\mathbf{y})$ with a magnitude of $O(\Delta^2)$ and an expectation value of $O(\Delta^3)$.

In summary, the above results lead to the following expressions for $\mathbf{Y}^*|_{\mathbf{y}}$:

$$\mathbf{Y}_1^*(t)|_{\mathbf{y}} = \mathbf{y}_1 + \mathbf{\Gamma}(t), \quad (27a)$$

$$\mathbf{Y}_1^*(t+\tau)|_{\mathbf{y}} = \mathbf{y}_1 + \mathbf{y}_2\,\tau + \mathbf{f}(\mathbf{y})\frac{\tau^2}{2} + \mathbf{g}(\mathbf{y})\mathbf{I}^{t,\tau} + \mathbf{\Gamma}(t+\tau) + \mathbf{R}^{t,\tau}(\mathbf{y}), \quad (27b)$$

$$\mathbf{Y}_2^*(t,\theta)|_{\mathbf{y}} = \mathbf{y}_2 + \mathbf{f}(\mathbf{y})\frac{\theta}{2} + \mathbf{g}(\mathbf{y})\frac{\mathbf{I}^{t,\theta}}{\theta} + \frac{\mathbf{\Gamma}(t+\theta)-\mathbf{\Gamma}(t)}{\theta} + \frac{\mathbf{R}^{t,\theta}(\mathbf{y})}{\theta}, \quad (27c)$$

$$\mathbf{Y}_2^*(t+\tau,\theta)|_{\mathbf{y}} = \mathbf{y}_2 + \mathbf{f}(\mathbf{y})\left(\tau+\frac{\theta}{2}\right) + \mathbf{g}(\mathbf{y})\frac{\mathbf{I}^{t,\tau+\theta}-\mathbf{I}^{t,\tau}}{\theta} + \frac{\mathbf{\Gamma}(t+\tau+\theta)-\mathbf{\Gamma}(t+\tau)}{\theta} + \frac{\mathbf{R}^{t,\tau+\theta}(\mathbf{y})-\mathbf{R}^{t,\tau}(\mathbf{y})}{\theta}. \quad (27d)$$

IV. MOMENTS $\mathbf{M}^{(k,\nu)}$

Now the moments $\mathbf{M}^{(k,\nu)}$ can be attacked. For a calculation of $\mathbf{M}^{(k,\nu)}$ explicit expressions for the vectors $\mathbf{A}$ and $\mathbf{B}$, as defined in Eq. (20), are needed. Using the results from the previous section [Eq. (27)] one finds

$$\mathbf{A}(\mathbf{y},\tau,\theta) = \begin{bmatrix} \mathbf{y}_2\,\tau + \mathbf{g}(\mathbf{y})\mathbf{I}^{t,\tau} + \mathbf{\Gamma}(t+\tau) - \mathbf{\Gamma}(t) + \mathbf{f}(\mathbf{y})\frac{\tau^2}{2} + \mathbf{R}^{t,\tau}(\mathbf{y}) \\ \mathbf{f}(\mathbf{y})\tau + \mathbf{g}(\mathbf{y})\frac{\mathbf{I}^{t,\tau+\theta}-\mathbf{I}^{t,\tau}-\mathbf{I}^{t,\theta}}{\theta} + \frac{\mathbf{\Gamma}(t+\tau+\theta)-\mathbf{\Gamma}(t+\tau)-\mathbf{\Gamma}(t+\theta)+\mathbf{\Gamma}(t)}{\theta} + \frac{\mathbf{R}^{t,\tau+\theta}(\mathbf{y})-\mathbf{R}^{t,\tau}(\mathbf{y})-\mathbf{R}^{t,\theta}(\mathbf{y})}{\theta} \end{bmatrix}, \quad (28a)$$

$$\mathbf{B}(\mathbf{y},\tau,\theta) = \begin{bmatrix} \mathbf{\Gamma}(t) \\ \mathbf{f}(\mathbf{y})\frac{\theta}{2} + \mathbf{g}(\mathbf{y})\frac{\mathbf{I}^{t,\theta}}{\theta} + \frac{\mathbf{\Gamma}(t+\theta)-\mathbf{\Gamma}(t)}{\theta} + \frac{\mathbf{R}^{t,\theta}(\mathbf{y})}{\theta} \end{bmatrix}. \quad (28b)$$

These expressions contain infinitely many terms, summarized in the remainders $\mathbf{R}$. To allow for a series truncation, a quantity denoted by ε is introduced now. This quantity is assumed to be small and its only purpose is to express the magnitude of terms [it is tacitly assumed here that the problem is described in dimensionless form with $\mathbf{f}$ and $\mathbf{g}$ being of order $O(1)$]. It will be assumed that τ and θ are of the same order of magnitude as ε and that the measurement noise Γ_i is of the same order as $\varepsilon^{3/2}$

$$\tau \overset{!}{=} O(\varepsilon), \quad \theta \overset{!}{=} O(\varepsilon), \quad V_{ij} \overset{!}{=} O(\varepsilon^3). \quad (29)$$

In a strict sense, the use of the Landau symbols here is not appropriate, because there is no functional relation between ε and, e.g., τ. The above notation is rather used to express the assumptions that, first, τ, ε, and Γ_i are small quantities, which allows us to sort powers by magnitude (like, e.g., $\tau^2 \ll \tau$). Second, it is assumed that τ, θ, and $|\Gamma_i|^{2/3}$ are of "comparable size," where comparable size means that, when restricting to small exponents, also powers of different quantities can be sorted by size (like, e.g., $\tau^3 \ll \theta^2$ or $|V_{ij}| \ll \tau^2$). This will be sufficient for appropriate low order approximations.

With this assumption the lowest order terms in $\mathbf{A}$ and $\mathbf{B}$ are of order $O(\varepsilon^{1/2})$. The magnitude of a moment $\mathbf{M}^{(k,\nu)}$, therefore, is given by (omitting arguments)

$$\mathbf{M}^{(k,\nu)} = \langle \mathbf{A}^k \otimes \mathbf{B}^\nu \rangle \;=\; O(\varepsilon^{(k+\nu)/2}). \quad (30)$$

For a first-order description of the moments $\mathbf{m}^{*(k)}$ thus only moments $\mathbf{M}^{(k,\nu)}$ with $k+\nu \leqslant 2$ need to be taken into account. Using Eqs. (28), (30), (A12) and the properties of $\mathbf{\Gamma}$, one finds

$$\mathbf{M}^{(0,0)} = 1, \quad (31a)$$

$$\mathbf{M}^{(0,1)} = \begin{bmatrix} \mathbf{0} \\ \frac{1}{2}\theta\mathbf{f} \end{bmatrix} + O(\varepsilon^2), \quad (31b)$$

$$\mathbf{M}^{(0,2)} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{1}{3}\theta\mathbf{g}\mathbf{g}^T + 2\frac{\mathbf{V}}{\theta^2} \end{bmatrix} + O(\varepsilon^2), \quad (31c)$$

$$\mathbf{M}^{(1,0)} = \begin{bmatrix} \tau\mathbf{y}_2 \\ \tau\mathbf{f} \end{bmatrix} + O(\varepsilon^2), \quad (32a)$$

$$\mathbf{M}^{(1,1)} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{\tau-\psi}{2}\mathbf{g}\mathbf{g}^T - (2+\delta_{\tau,\theta})\frac{\mathbf{V}}{\theta^2} \end{bmatrix} + O(\varepsilon^2), \quad (32b)$$

$$\mathbf{M}^{(2,0)} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \psi\,\mathbf{g}\mathbf{g}^T + 2(2+\delta_{\tau,\theta})\frac{\mathbf{V}}{\theta^2} \end{bmatrix} + O(\varepsilon^2), \quad (33)$$

with

$$\psi := \begin{cases} \tau^2/\theta - \frac{1}{3}\tau^3/\theta^2, & \tau < \theta \\ \tau - \frac{1}{3}\theta, & \tau \geqslant \theta \end{cases}. \tag{34}$$

Inserting these expressions into Eq. (15), finally, yields a first-order description of the moments $\mathbf{m}^{*(k)}$ in terms of ρ, $\mathbf{f}$, $\mathbf{g}$, and $\mathbf{V}$. It turns out that derivatives with respect to components of $\mathbf{y}_1$ do not appear in the terms up to order $O(\varepsilon)$—so for a first-order description only the derivatives with respect to the components of $\mathbf{y}_2$ need to be considered. It also turns out that only the upper half of the vector $\mathbf{m}^{*(1)}$ and the upper quarter of the matrix $\mathbf{m}^{*(2)}$ need to be looked at (those components that correspond to moments of the increments of $\mathbf{Y}_2^*$). To take (syntactical) advantage of this reduction in dimensionality the notations

$$\hat{\partial}_i := \frac{\partial}{\partial y_{N+i}}, \quad \hat{m}^{(0)} := m^{*(0)}, \tag{35a}$$

$$\hat{m}_i^{(1)} := m_{N+i}^{*(1)}, \quad \hat{m}_{ij}^{(2)} := m_{N+i,N+j}^{*(2)}, \tag{35b}$$

are introduced, where i and j are in the range $1,\dots,N$. The relevant equations can now be written compactly as

$$\hat{m}^{(0)}(\mathbf{y},\theta) = \rho - \frac{\theta}{2}\,\hat{\partial}_i[\rho f_i] + \frac{\theta}{6}\,\hat{\partial}_i\hat{\partial}_j[\rho\,(\mathbf{g}\mathbf{g}^T)_{ij}] + \frac{V_{ij}}{\theta^2}\,\hat{\partial}_i\hat{\partial}_j\rho + O(\varepsilon^2), \tag{36a}$$

$$\hat{m}_i^{(1)}(\mathbf{y},\tau,\theta) = \tau\rho f_i - \frac{1}{2}(\tau-\psi)\,\hat{\partial}_j[\rho\,(\mathbf{g}\mathbf{g}^T)_{ij}] + (2+\delta_{\tau,\theta})\frac{V_{ij}}{\theta^2}\,\hat{\partial}_j\rho + O(\varepsilon^2), \tag{36b}$$

$$\hat{m}_{ij}^{(2)}(\mathbf{y},\tau,\theta) = \psi\rho\,(\mathbf{g}\mathbf{g}^T)_{ij} + 2(2+\delta_{\tau,\theta})\frac{V_{ij}}{\theta^2}\,\rho + O(\varepsilon^2). \tag{36c}$$

These equations directly relate the unknown quantities ρ, $\mathbf{f}$, $\mathbf{g}\mathbf{g}^T$, and $\mathbf{V}$ and the observable quantities $\hat{\mathbf{m}}^{(k)}$. The function argument of ρ, $\mathbf{f}$, and $\mathbf{g}$ is given by $\mathbf{y}$. The function ψ depends on τ and θ and has a piecewise definition only, Eq. (34).

V. SYSTEMATIC ERRORS OF THE STANDARD EMBEDDING APPROACH (SEA)

Next the SEA will be analyzed, using the final result of the previous section [Eq. (36)]. Only the case without measurement noise, i.e., $\mathbf{V} \equiv \mathbf{0}$, will be looked at. This will show the "pure" effects of the reconstruction errors caused by the numerical estimation of $\mathbf{Y}_2$.

For a time series, where $\mathbf{Y}_2$ has been reconstructed by a first-order forward differencing scheme with step size θ, the observable moments $\hat{\mathbf{m}}^{(k)}$ are described by Eq. (36). Ignoring this result and attempting a Markov analysis as outlined in Sec. I will put the focus on the terms $\hat{\mathbf{m}}^{(k)}(\mathbf{y},\tau,\theta)/[\tau\hat{m}^{(0)}(\mathbf{y},\theta)]$. According to Eq. (6), these terms should be finite-increment estimates of $\mathbf{f}$ and $\mathbf{g}\mathbf{g}^T$ ($k = 1$, respectively $k = 2$). In fact, however, the terms evaluate to

$$\frac{\hat{m}_i^{(1)}(\mathbf{y},\tau,\theta)}{\tau\hat{m}^{(0)}(\mathbf{y},\theta)} = f_i - \frac{1-\psi(\tau,\theta)/\tau}{2\hat{m}^{(0)}}\,\hat{\partial}_j[\hat{m}^{(0)}\,(\mathbf{g}\mathbf{g}^T)_{ij}] + O(\varepsilon), \tag{37a}$$

$$\frac{\hat{m}_{ij}^{(2)}(\mathbf{y},\tau,\theta)}{\tau\hat{m}^{(0)}(\mathbf{y},\theta)} = \frac{\psi(\tau,\theta)}{\tau}(\mathbf{g}\mathbf{g}^T)_{ij} + O(\varepsilon). \tag{37b}$$

Trying to extrapolate these estimates to $\tau = 0$ then becomes problematic. Instead of being approximately constant, as expected from Eq. (6), the values will show nonlinear behavior caused by the function ψ/τ. For fixed θ this function starts linear with a value of zero at $\tau/\theta = 0$, passes through $2/3$ at $\tau/\theta = 1$, and approaches a value of 1 for $\tau/\theta \to \infty$. Simply fitting a low order polynomial to all estimates up to some maximum increment $\tau_{\max}$ will thus, in general, underestimate $\mathbf{g}\mathbf{g}^T$ (because of $|\psi/\tau| < 1$). An error of comparable size (although with arbitrary sign) will occur when estimating $\mathbf{f}$.

In principle, however, the estimates for large τ, i.e., where $\psi/\tau \approx 1$, could be used for a fit. On the other hand also the influence of higher order terms becomes stronger for large increments. Unless a time series is sampled with a very small time step, such an approach will also fail to provide accurate estimates for f_i and $(\mathbf{g}\mathbf{g}^T)_{ij}$.

VI. MODIFIED EMBEDDING APPROACH (MEA)

Based on Eq. (36), we now will propose a modified approach that takes into account the effects of the differencing scheme as well as the effects of measurement noise. An important point in this approach will be to keep the ratio of τ and θ fixed. This provides an easy way to avoid problems caused by the nonlinear term $\psi(\tau,\theta)$. In the following $\theta \equiv \tau$ is chosen. From a practical point of view this means that we no longer use a fixed step size to estimate all velocities. Instead, to calculate a velocity increment $\mathbf{Y}_2(t+\tau) - \mathbf{Y}_2(t)$, the involved velocities will be estimated using a step size $\theta = \tau$. Equation (36) then reads

$$\hat{m}^{(0)}(\mathbf{y},\tau) = \rho - \frac{\tau}{2}\,\hat{\partial}_i[\rho f_i] + \frac{\tau}{6}\,\hat{\partial}_i\hat{\partial}_j[\rho\,(\mathbf{g}\mathbf{g}^T)_{ij}] + \frac{V_{ij}}{\tau^2}\,\hat{\partial}_i\hat{\partial}_j\rho + O(\varepsilon^2), \tag{38a}$$

$$\hat{m}_i^{(1)}(\mathbf{y},\tau,\tau) = \tau\rho f_i - \frac{\tau}{6}\,\hat{\partial}_j[\rho\,(\mathbf{g}\mathbf{g}^T)_{ij}] + 3\frac{V_{ij}}{\tau^2}\,\hat{\partial}_j\rho + O(\varepsilon^2), \tag{38b}$$

$$\hat{m}_{ij}^{(2)}(\mathbf{y},\tau,\tau) = \tau\frac{2}{3}\rho\,(\mathbf{g}\mathbf{g}^T)_{ij} + 6\frac{V_{ij}}{\tau^2}\,\rho + O(\varepsilon^2). \tag{38c}$$

A fixed ratio of τ and θ also leads to a simpler form of the higher order terms (see Appendix B). Each term of order $O(\varepsilon^n)$ on the right hand side of Eq. (38) has the form

$$Q^{(n)} = c(\mathbf{y})\,\tau^a(\varepsilon^3/\tau^2)^b \tag{39}$$

with

$$0 \leqslant a \leqslant n, \quad b = n - a, \quad c = O(1). \tag{40}$$

Here the symbol $Q^{(n)}$ is used to denote such a term and ε^3 accounts for the assumption on the magnitude of $\mathbf{V}$. The functional form of $Q^{(n)}$ (with respect to τ) can thus be described by a function-base $\mathcal{B}^{(n)}$ that consists of $n+1$ functions τ^{a-2b}. As noted in Appendix B, this implies $\mathcal{B}^{(n)} \subset \mathcal{B}^{(n+3)}$ and thus puts a limit on the accuracy that can be achieved in least square fits of $\hat{\mathbf{m}}^{(k)}$. For example, it is not possible to distinguish a first-order term $c\tau$ and a fourth-order term $c'\tau^3(\varepsilon^3/\tau^2)$ by their functional form.

In the following Eq. (38a) will be used in the form $\hat{m}^{(0)} = \rho + O(\varepsilon)$, i.e., the explicit results for the first-order terms will not be used. This avoids the need to numerically calculate the derivatives that appear within these terms. Next, Eqs. (38b) and (38c) are divided by ρ. Replacing ρ by $\hat{m}^{(0)}$ in the resulting left hand sides will only result in additional terms of order $O(\varepsilon^2)$ and higher for the right hand sides. One finds (omitting function arguments again)

$$\frac{\hat{m}_i^{(1)}}{\hat{m}^{(0)}} = \tau \tilde{f}_i + \frac{3V_{ij}\hat{\partial}_j\rho}{\tau^2\rho} + O(\varepsilon^2), \tag{41a}$$

$$\frac{\hat{m}_{ij}^{(2)}}{\hat{m}^{(0)}} = \tau\frac{2}{3}(\mathbf{gg}^T)_{ij} + \frac{6V_{ij}}{\tau^2} + O(\varepsilon^2), \tag{41b}$$

with the shortcut

$$\tilde{f}_i := f_i - \frac{\hat{\partial}_j[\rho\,(\mathbf{gg}^T)_{ij}]}{6\rho}. \tag{42}$$

The term $3V_{ij}(\hat{\partial}_j\rho)/(\tau^2\rho)$ in Eq. (41a) will now be expressed as $\tau^{-2}c_i$, where c_i is an unknown constant [of order $O(\varepsilon^3)$]. Finally, it will be assumed that $\mathbf{V}$ is known. This assumption is not mandatory—$\mathbf{V}$ could be estimated using Eq. (41b)—but this quantity can be estimated more easily in advance by, for example, analyzing the autocovariance of $\mathbf{Y}_1^*$ (see Appendix C). The final set of equations now reads

$$\hat{m}^{(0)} = \rho + O(\varepsilon), \tag{43a}$$

$$\frac{\hat{m}_i^{(1)}}{\hat{m}^{(0)}} = \tau\tilde{f}_i + \frac{1}{\tau^2}c_i + O(\varepsilon^2), \tag{43b}$$

$$\frac{\hat{m}_{ij}^{(2)}}{\hat{m}^{(0)}} - \frac{6V_{ij}}{\tau^2} = \tau\frac{2}{3}(\mathbf{gg}^T)_{ij} + O(\varepsilon^2). \tag{43c}$$

The terms on the left hand sides can be estimated for different values of τ from a given time series. Choosing appropriate sets of regression functions thus allows us to estimate ρ, $\tilde{\mathbf{f}}$, and $\mathbf{gg}^T$ by a linear regression analysis. Once these quantities have been estimated, Eq. (42) can be used to finally calculate $\mathbf{f}$ [the derivative that appears in Eq. (42) can, e.g., be calculated using a density-weighted local polynomial fit of $\rho\mathbf{gg}^T$].

The functional form of the higher order terms can be shown to still obey Eq. (39). Therefore $\{1,\tau,\tau^{-2}\}$, $\{\tau,\tau^{-2}\}$, and $\{\tau\}$ are appropriate function sets for Eq. (43a), (43b), and (43c) if terms up to order $O(\varepsilon)$ shall be taken into account. To also take into account second-order terms, the functions $\{\tau^2,\tau^{-1},\tau^{-4}\}$ must be added to the sets. In principle, also third-order terms can be accounted for in Eqs. (43b) and (43c) by also adding the functions $\{\tau^3,1,\tau^{-3},\tau^{-6}\}$. In practice, however, a large number of regression functions and also large negative exponents lead to numerical problems. As a compromise, the terms can partially be accounted for. In the numerical example given later, e.g., only τ^3 is used as a regression function for third-order terms.

VII. NUMERICAL TEST CASE

To check the analytical results and to compare the different embedding approaches, a numerical example is investigated now. As test case a scalar process $X(t)$ is chosen that obeys the second-order ODE,

$$\ddot{X} = f(X,\dot{X}) + g(X,\dot{X})\xi(t), \tag{44}$$

where f and g are defined as

$$f(X,\dot{X}) := -X - 3\dot{X}, \quad g(X,\dot{X}) := 1. \tag{45}$$

Again $\xi(t)$ denotes Gaussian white noise with $\langle\xi(t)\xi(t')\rangle = \delta(t-t')$. The above ODE can be rewritten as a system of first-order ODEs for a two-dimensional (2D) process $\mathbf{Y}(t)$, the components of which are given by position $Y_1 \equiv X$ and velocity $Y_2 \equiv \dot{X}$ of the 1D process $X(t)$,

$$\dot{Y}_1 = Y_2, \tag{46a}$$

$$\dot{Y}_2 = -Y_1 - 3Y_2 + \xi. \tag{46b}$$

These equations describe an Ornstein-Uhlenbeck process in two dimensions and can be solved analytically. The characteristic time scales of the autocovariance of $\mathbf{Y}$ are found to be $(3+\sqrt{5})/2 \approx 2.618$ and $(3-\sqrt{5})/2 \approx 0.382$. The values of $\mathbf{Y}$ are Gaussian distributed and have a variance of $\langle\mathbf{YY}^T\rangle = \mathbf{Id}/6$.

For this process a time series of $\mathbf{Y}$, consisting of 10^7 values, sampled with a time increment $\Delta t = 0.01$, is generated. Excerpts of the resulting series for Y_1 and Y_2 are shown in Figs. 1 and 2. Here also a basic problem of the SEA can be seen, which was noted in Sec. (I) and quantified in Sec. (V): Even if a series is sampled sufficiently fine to allow an "accurate" estimation of Y_2 by a numerical differencing scheme, the velocity *increments* (for time increment τ) will still show notable errors for small τ. This error depends on the ratio θ/τ (here $\theta = \Delta t$) and its effects can be quantified by the function ψ in Eq. (37).

To obtain a baseline for the accuracy that can be achieved with the given data, a SMA is applied to the true 2D series $\mathbf{Y}$ first. Here and for subsequent analyses a binning approach is used, where the region $[-1,1]\times[-1,1]$ of the (y_1,y_2) plane is

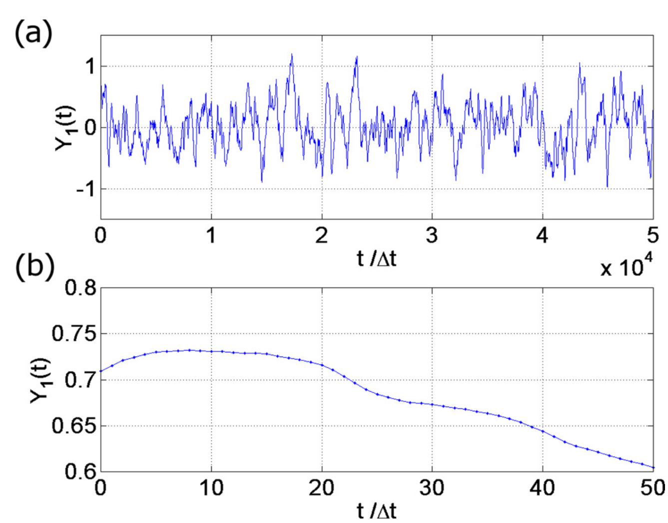

FIG. 1. (Color online) Excerpt of the generated series of position values Y_1 (a). A zoomed-in view (b) shows that the signal in fact is smooth and thus allows us to numerically estimate its derivative if the sampling time step Δt is sufficiently small.

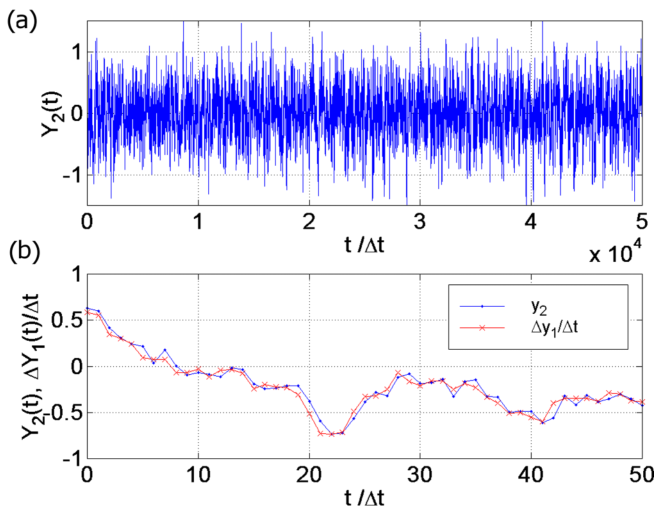

FIG. 2. (Color online) Excerpt of the generated series of velocity values Y_2 (a). In the zoomed-in view (b) additionally the numerically estimated derivative of Y_1 is shown. Although the values of both series (true and estimated) quite accurately match, there are notable differences for the small scale increments.

covered by 30 × 30 bins. For one of these bins the estimated moments of the conditional velocity increments are shown in Fig. 3.

Actually the moments in Fig. 3 are scaled by τ^{-1}, as is usually done for their visual presentation. This allows us to interpret the estimation of f and g^2 as "extrapolating the scaled moments to $\tau = 0$." Later on, however, when measurement noise enters the scene, a more general interpretation will be needed, where f and g^2 are found by a linear regression strategy. Of course this interpretation is also valid in the given setup. The values of f and g^2 are given by the coefficients of

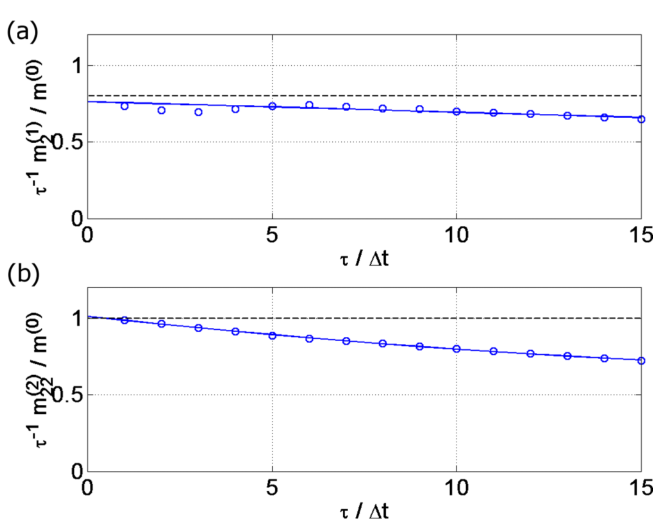

FIG. 3. (Color online) First (a) and second moment (b) of the conditional velocity increments (obtained by a SMA). The estimated values (circles) are scaled by τ^{-1}. The corresponding fits are shown as solid curves. Estimates are taken at $(y_1, y_2) = (-0.1, -0.2333)$. Here f and g^2 have values of 0.8 and 1.0 respectively (dashed lines).

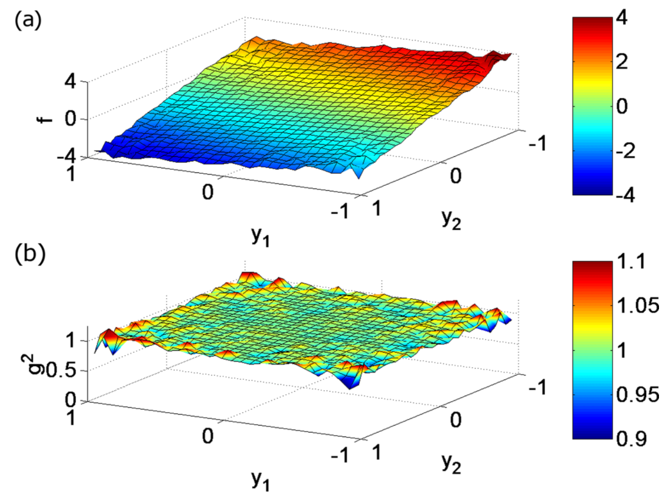

FIG. 4. (Color online) Estimates for f and g^2, (a) and (b), obtained by a SMA.

the linear part (in τ) of the conditional moments $m_2^{(1)}/m^{(0)}$ and $m_{22}^{(2)}/m^{(0)}$ respectively [see Eq. (6)].

In the following, the regression functions $\{\tau, \tau^2\}$ and $\{\tau, \tau^2, \tau^3\}$ are used to fit the estimated first and second conditional moments [this corresponds to fitting a linear function to the values in Fig. 3(a) and a quadratic function to those in Fig. 3(b)]. The maximum increment that is used for these fits is chosen as $\tau_{\max} = 15\Delta t$. The resulting estimates for f and g^2 are shown in Fig. 4. In Fig. 5 the absolute errors δf and δg^2 of these estimates are shown. In regions with low density (as noted above, the PDF of $\mathbf{Y}$ is a symmetric Gaussian with a standard deviation of ≈0.408) fluctuations become larger but there is no obvious bias of the results.

VIII. EMBEDDING APPROACHES WITHOUT MEASUREMENT NOISE

Next the results of the different embedding approaches are looked at. First a SEA is used to perform an analysis solely

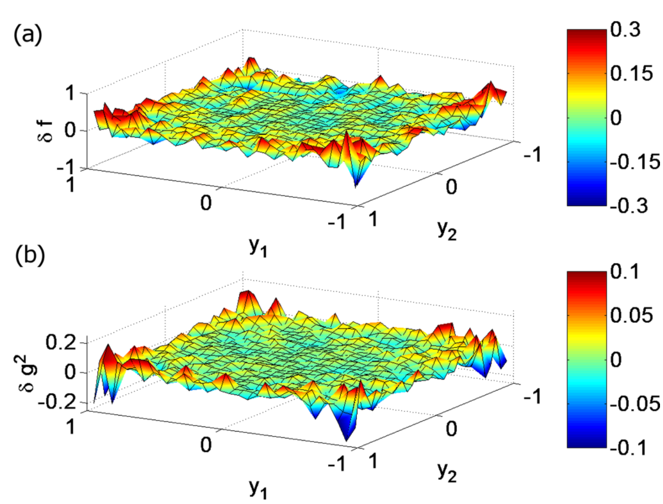

FIG. 5. (Color online) Absolute errors of the estimates for f and g^2, (a) and (b), obtained by a SMA.

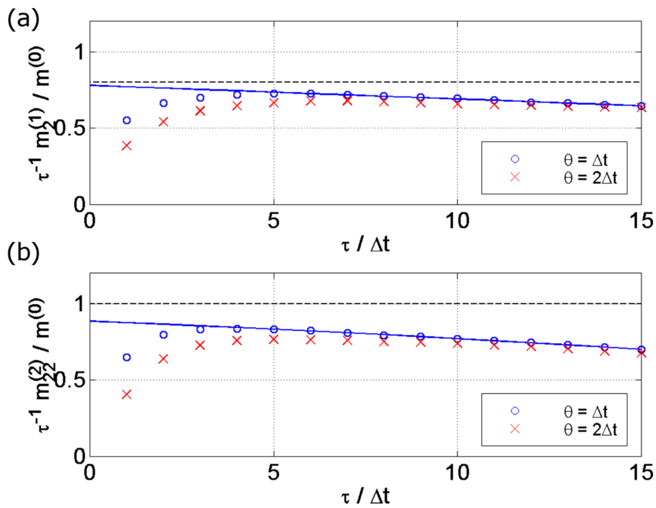

FIG. 6. (Color online) First (a) and second moment (b) of the conditional velocity increments (obtained by a SEA with $\theta = \Delta t$). The estimated values (circles) are scaled by τ^{-1}. The corresponding fits are shown as solid curves. Additionally, the estimates obtained by a SEA with $\theta = 2\Delta t$ are shown (crosses). Estimates are taken at $(y_1, y_2) = (-0.1, -0.2333)$. Here f and g^2 have values of 0.8 and 1.0 respectively (dashed lines).

based on the 1D series of positions Y_1. The corresponding velocity values Y_2 are estimated by a first-order forward differencing scheme with a step size of $\theta = \Delta t$ and the resulting 2D series then is analyzed by a SMA.

As is shown in Fig. 6, the estimated moments of the conditional velocity increments behave quite differently compared to those obtained from the true 2D series (shown in Fig. 3). As expected from Eq. (37), the moments show strongly nonlinear behavior for small increments τ. For an estimation of f and g^2, therefore, only increments with $5 \leqslant \tau/\Delta t \leqslant 15$ are used. Least square fits are performed using the same sets of regression functions as in the previous section. The absolute errors δf and δg^2 of the resulting estimates are shown in Fig. 7. Of course, the fluctuations become larger now as fewer

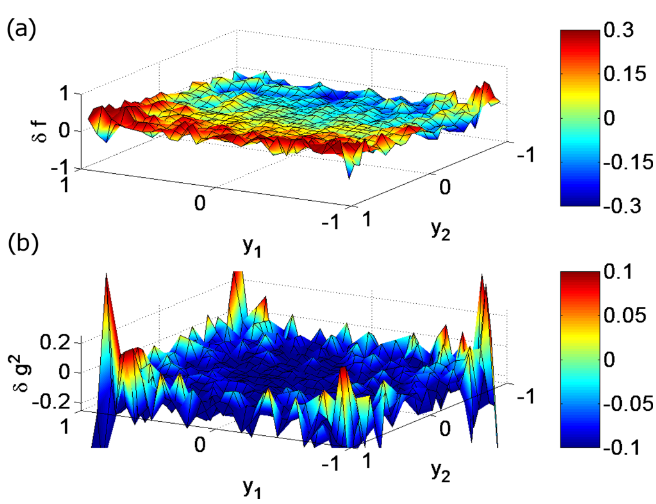

FIG. 7. (Color online) Absolute errors of the estimates for f and g^2, (a) and (b), obtained by a SEA.

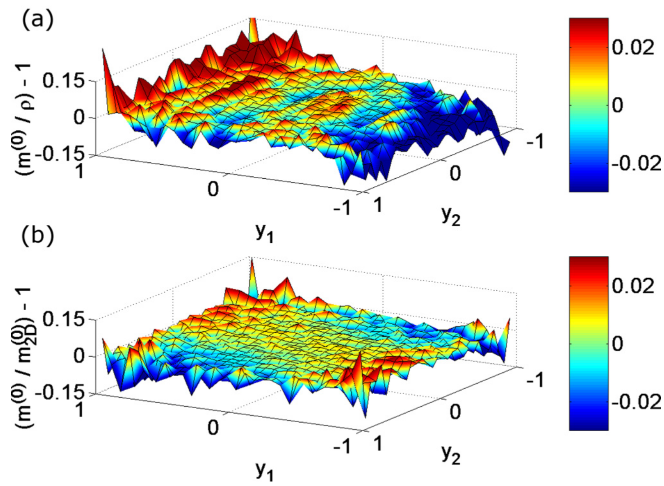

FIG. 8. (Color online) Relative errors of the estimated density values, obtained by a SEA. In (a) errors relative to the true density ρ are shown. In (b) errors are relative to the binned density of the 2D series.

increments are used for the fits. But, more importantly, it is obvious that g^2 is systematically underestimated. And also the estimates for f clearly show a significant bias that is approximately linear in y_2.

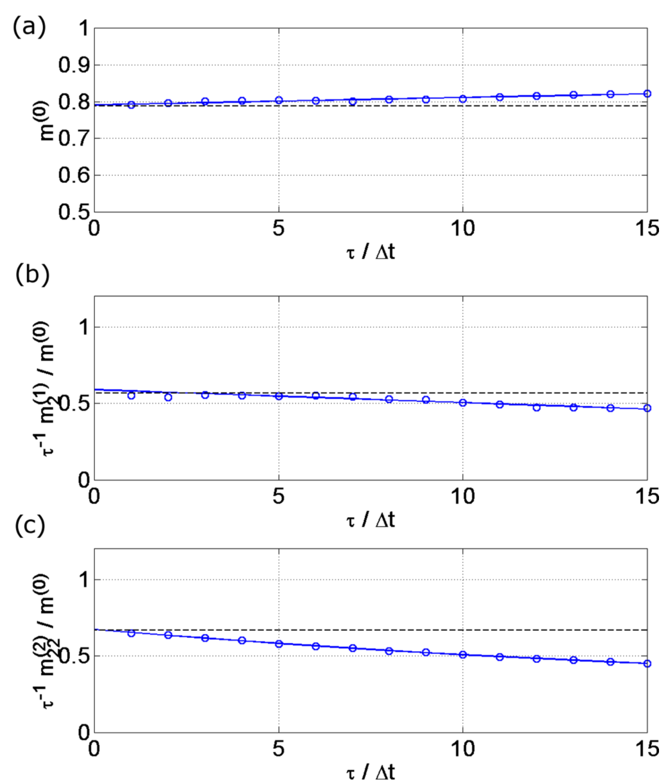

FIG. 9. (Color online) Estimated densities (a) and estimated moments (scaled by τ^{-1}) of the conditional velocity increments (b), (c). The estimates (circles) have been obtained by a MEA. The corresponding fits are shown as solid curves. Estimates are taken at $(y_1, y_2) = (-0.1, -0.2333)$. Here $\tilde{f}$ [see Eq. (42)] and $2g^2/3$ have values of 0.5667 and 0.6667 respectively and the binned density of the 2D series has a value of 0.7873 (dashed lines).

Using a SEA also affects the estimates for the process density ρ. It would be misleading, however, to compare the estimates $m^{(0)}$ to the true density ρ, as is done in Fig. 8(a). To a large extent the observed errors are caused by finite size effects and not by the reconstruction approach (the binned density of the true 2D series would show very similar errors). To assess the errors that are introduced by the embedding approach, the estimates $m^{(0)}$ thus should be compared to the binned density of the 2D series. This is done in Fig. 8(b), where the errors are found to be biased by a hyperbolic function in y_1 and y_2.

Now a MEA, as proposed in Sec. VI, is applied. Again the analysis is purely based on the 1D series of positions Y_1. Opposed to a SEA, however, velocities are no longer estimated by a differencing scheme with a fixed step size. Instead, velocities and velocity increments for time increment τ are estimated using the step size $\theta = \tau$. Using a binning approach, it is not much more effort than for a SEA to implement the calculation of the density $m^{(0)}$ and of the conditional moments $m_1^{(1)}/m^{(0)}$ and $m_{22}^{(2)}/m^{(0)}$. In pseudocode this reads

```
fori = 1 : n − kmax
fork = 1 : kmax
pos = x[i]
velo = (x[i + k] − x[i])/k/dt
dvelo = (x[i + 2 ∗ k] − 2 ∗ x[i + k] + x[i])/k/dt
idx = getBinIndex(pos,velo)
if(isValid(idx))
m0[idx][k]+ = 1
m1[idx][k]+ = dvelo
m2[idx][k]+ = dvelo ∗ dvelo
end
end
end
foridx = 1 : idxmax
fork = 1 : kmax
m1[idx][k]/ = m0[idx][k]
m2[idx][k]/ = m0[idx][k]
m0[idx][k]/ = (n − kmax) ∗ binSize
end
end
```

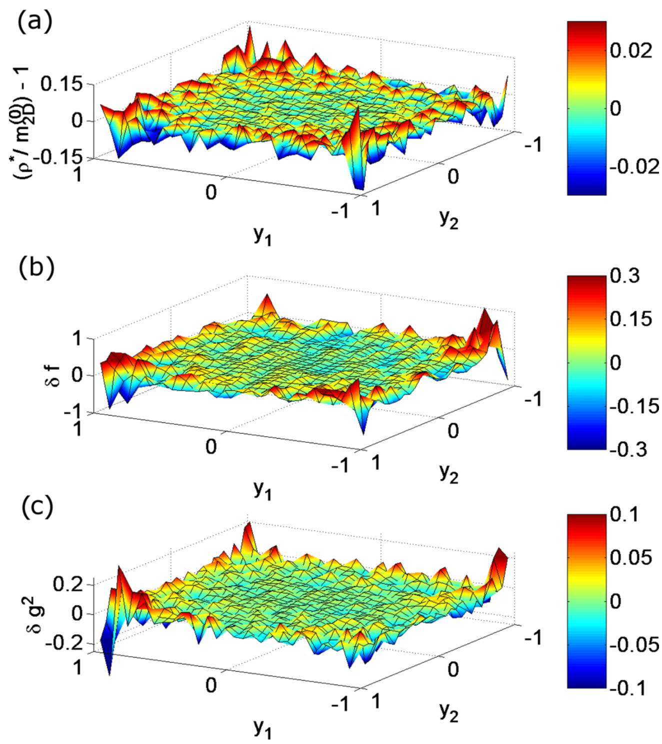

FIG. 10. (Color online) Relative error of the estimated density ρ (a) and absolute errors of the estimates for f and g^2, (b) and (c), obtained by a MEA. Errors in (a) are relative to the binned density of the 2D series.

Estimates for density and conditional moments obtained by a MEA are shown in Fig. 9. As expected from Eq. (41), the scaled moments now approach $\tilde{f}$ and $2g^2/3$ respectively for $\tau \to 0$. Also the density $m^{(0)}$ now depends on τ and approaches the density of the 2D series.

All increments up to $\tau_{\max} = 15\Delta t$ are used for the least square fits. The regression functions $\{1,\tau\}$ are used to fit the density estimates. For the fits of the estimated first and second conditional moments again the functions $\{\tau,\tau^2\}$ respectively $\{\tau,\tau^2,\tau^3\}$ are used. There is no need to add functions like τ^{-2}, as still a case without measurement noise is looked at. The errors of the resulting estimates for ρ, f, and g^2 are shown in Fig. 10. Opposed to a SEA, shown in Fig. 7, no obvious biasing of the estimates can be observed and the fluctuations of δf and δg^2 are comparable to those observed in an analysis of the 2D series using a SMA.

IX. EMBEDDING APPROACHES WITH MEASUREMENT NOISE

So far, only data without measurement noise as been analyzed. Next, a series of "noisy" values Y_1^* is generated by adding Gaussian, uncorrelated noise with a variance of $V = 1.6667 \times 10^{-7}$ (this corresponds to a noise-to-signal amplitude ratio of 10^{-3}) to the series Y_1. This noisy series Y_1^* is then analyzed—first by applying a SEA and next by applying a MEA.

Moments obtained by a SEA are shown in Fig. 11. Due to the measurement noise the scaled moments now diverge

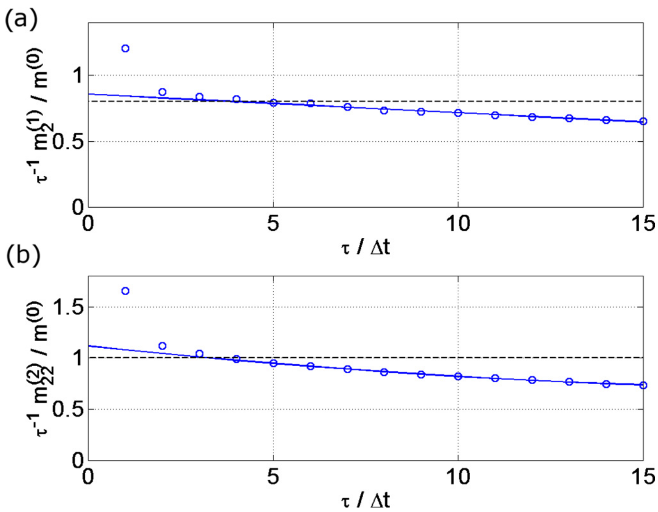

FIG. 11. (Color online) First (a) and second moment (b) of the conditional velocity increments of the noisy series Y_1^* (obtained by a SEA with $\theta = \Delta t$). The estimated values (circles) are scaled by τ^{-1}. The corresponding polynomial fits are shown as solid curves. Estimates are taken at $(y_1, y_2) = (-0.1, -0.2333)$. Here f and g^2 have values of 0.8 and 1.0 respectively (dashed lines).

for $\tau \to 0$. For an estimation of f and g^2, therefore, again only increments with $5 \leqslant \tau/\Delta t \leqslant 15$ are used. Least square fits again are performed using the functions $\{\tau, \tau^2\}$ and $\{\tau, \tau^2, \tau^3\}$ respectively. The absolute errors δf and δg^2 of the resulting estimates are shown in Fig. 12. It turns out that g^2 is systematically *overestimated* now. The estimates for f still show a significant bias that is approximately linear in y_2—although the bias now has switched sign.

Finally our proposed MEA is applied to the series Y_1^*, which leads to estimates for density and conditional moments as shown in Fig. 13. For $\tau \to 0$ the moments are diverging because of the terms proportional to τ^{-2} (and other higher order terms proportional to negative powers of τ), as described by Eq. (41). It is thus necessary now to add appropriate regression functions that account for these terms: For density

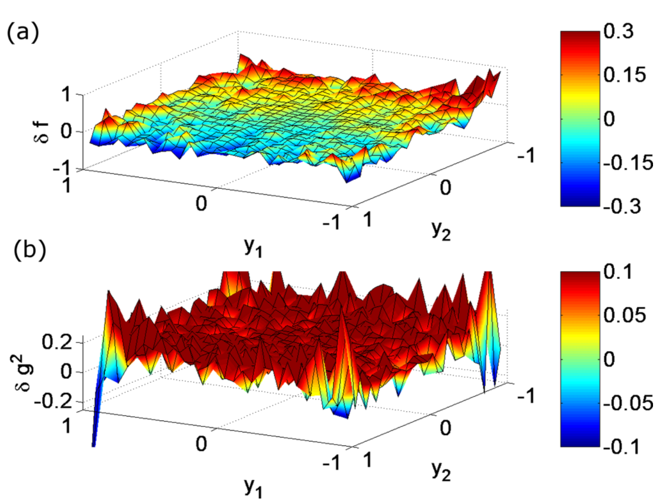

FIG. 12. (Color online) Absolute errors of the estimates for f and g^2, (a) and (b), obtained by using a SEA for the noisy series Y_1^*.

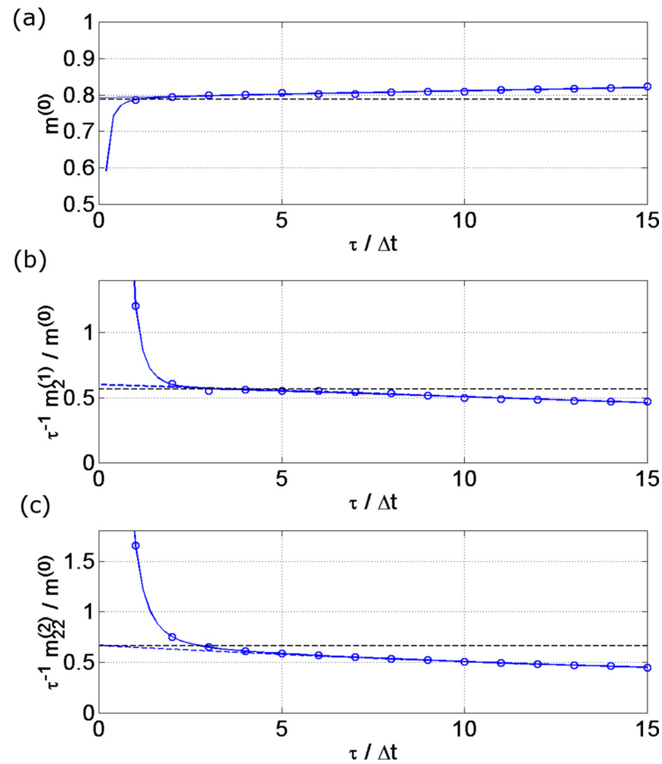

FIG. 13. (Color online) Estimated densities (a) and estimated moments (scaled by τ^{-1}) of the conditional velocity increments (b) and (c) of the noisy series Y_1^*. The estimates (circles) have been obtained by a MEA. The corresponding fits are shown as solid curves. The nondiverging parts of these fits are shown as dashed curves. Estimates are taken at $(y_1, y_2) = (-0.1, -0.2333)$. Here $\tilde{f}$ [see Eq. (42)] and $2g^2/3$ have values of 0.5667 and 0.6667 respectively and the binned density of the 2D series has a value of 0.7873 (dashed lines).

estimation, all terms up to order $O(\varepsilon)$ are accounted for by using the functions $\{1, \tau, \tau^{-2}\}$. Fits of the first conditional moments (yielding an estimate for $\tilde{f}$) are performed using the regression functions $\{\tau, \tau^{-2}, \tau^2, \tau^{-1}, \tau^{-4}\}$, i.e., considering terms up to order $O(\varepsilon^2)$. Fits of the second conditional moments, finally (yielding an estimate for $2g^2/3$), are performed using the regression functions $\{\tau, \tau^2, \tau^{-1}, \tau^{-4}, \tau^3\}$. This choice needs some explanation. First, only τ^3 is present to account for third-order terms. This is a compromise for numerical reasons—it reduces the number of regression functions and avoids numerical problems with large negative powers of τ. Second, the first-order term $6V/\tau^2$ is not accounted for by any regression function. This term is assumed to be known and thus does not need to be estimated. The value of V is estimated in advance by extrapolating the autocovariance function of Y_1^* to $\tau = 0$ and then taking the difference to $\langle Y_1^{*2} \rangle$ (see Appendix C). For the given example a fifth-order polynomial without linear term has been used to extrapolate the autocovariance values calculated at $\tau/\Delta t = 1, \ldots, 5$. Estimates for V that are obtained this way are accurate within about five percent, as has been checked numerically.

Using the above sets of regression functions and all increments up to $\tau_{\max} = 15\Delta t$ then leads to estimates for f and g^2, the absolute errors of which are shown in Fig. 14. The

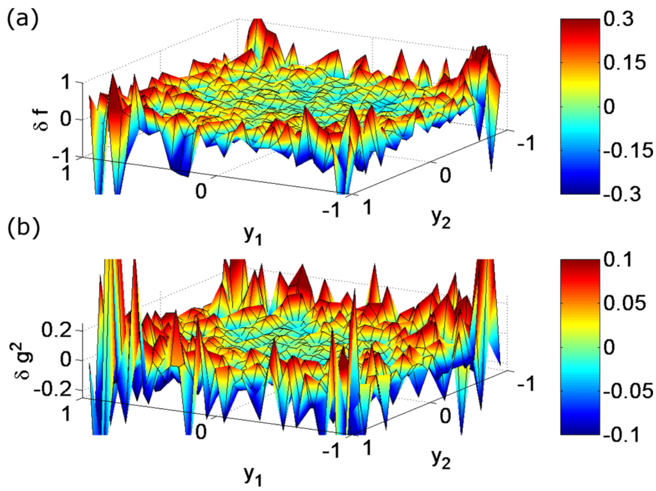

FIG. 14. (Color online) Absolute errors of the estimates for f and g^2, (a) and (b), obtained by using a MEA for the noisy series Y_1^*.

estimates are quite heavily fluctuating now. But—at least to the bare eye—the results seem not to be biased.

X. SUMMARY OF NUMERICAL RESULTS

Solid quantitative results for biases of the results of the different analyses that have been performed would require an averaging over a large number of analyses of independent realizations of $\mathbf{Y}$. Instead, a simpler approach is chosen to numerically compare the results. A polynomial P with

$$P = a + b_1 y_1 + b_2 y_2 + c_{11} y_1^2 + c_{12} y_1 y_2 + c_{22} y_2^2 \quad (47)$$

is fitted to the results for f and g^2 using a density weighted least square fit. According to Eq. (45) the only nonzero coefficients for f should be $b_1 = -1$ and $b_2 = -3$. For g^2 only $a = 1$ should be nonzero. Defining rms as the root of the density weighted mean of the squared differences between the actual estimates and P allows us to also assess the fluctuations. The results for f and g^2 are given in Table I.

TABLE I. Polynomial coefficients and mean errors of a fit of the estimates for f and g^2 respectively. Here SEA* and MEA* denote results for the noisy series Y_1^*. Bold values are discussed in the text.

f	a	b_1	b_2	c_{11}	c_{12}	c_{22}	rms
Exact	0.00	−1.00	−3.00	0.00	0.00	0.00	
SMA	0.00	−1.00	−2.97	−0.05	0.01	0.02	0.06
SEA	0.00	−1.00	**−2.77**	−0.05	0.01	0.02	0.06
MEA	0.00	−0.99	−2.97	−0.05	0.01	0.02	0.06
SEA*	0.00	−1.00	**−3.16**	−0.05	0.01	0.01	0.06
MEA*	0.00	−0.97	−2.98	−0.05	0.00	0.00	**0.13**
g^2	a	b_1	b_2	c_{11}	c_{12}	c_{22}	rms
Exact	1.00	0.00	0.00	0.00	0.00	0.00	
SMA	1.00	0.00	0.00	0.00	0.00	0.01	0.02
SEA	**0.89**	0.00	0.00	0.00	0.00	−0.04	0.04
MEA	1.00	0.00	0.00	0.00	0.01	0.00	0.02
SEA*	**1.11**	0.00	0.00	0.00	0.04	0.08	0.04
MEA*	1.00	0.01	0.00	0.01	0.02	0.00	**0.06**

The most pronounced effects of a SEA can be observed for the coefficient b_2, when estimating f, respectively for the coefficient a, when estimating g^2. These coefficients are also strongest affected by the presence of measurement noise. Applying a MEA, however, yields results that are comparable to those obtained by an analysis of the 2D series—at least if no measurement noise is present. When analyzing noisy data, the coefficients still are quite accurate, which indicates the absence of systematical estimation errors. However, the fluctuations, rms, are quite large then. Eventually this is caused by the metric of the regression functions. Roughly speaking, rms becomes large, when the angle between the function $\{\tau\}$ (the coefficient of which we want to estimate) and the subspace spanned by the other regression functions becomes small. For above analyses, this angle is smallest for the set of regression functions required by a MEA able to analyze noisy data.

XI. CONCLUSIONS

For a time series analysis of a process $\mathbf{X}$ that is described by a stochastically forced second-order ODE, frequently an embedding strategy as outlined in Sec. I is used: First the temporal derivative $\dot{\mathbf{X}}$ is estimated for each point in time by a numerical differencing scheme, and a new series $\mathbf{Y}^T := (\mathbf{X}^T, \dot{\mathbf{X}}^T)$ is built. Then a Markov analysis is applied to the series $\mathbf{Y}$ in order to estimate its drift and diffusion functions. However, the errors that are caused by the differencing scheme lead to notably biased estimates for these functions. Additionally, even a very small amount of measurement noise has strong influence on the results.

The errors of the above "standard" approach have been studied analytically and a modified approach has been proposed. This approach allows for an accurate estimation of the drift and diffusion functions. Additionally, this approach can be adapted to also deal with weak measurement noise. This has been verified for a numerical test case.

In this numerical test it also could be seen that measurement noise is a bigger problem than one might think intuitively. Already measurement noise with an noise-to-signal amplitude ratio of only 10^{-3} had a severe influence. For the standard approach, it introduces an additional, notable bias to the results. When using the "noise-aware" version of our modified approach, still unbiased results are obtained. However, the additional ability to also deal with measurement noise has to be paid for with much stronger fluctuations of the estimates.

The implementation of the presented approach is easily done and straightforward. The algorithm is not demanding with respect to memory or CPU power. All calculations have been performed on a standard desktop PC, where each analysis took less than one minute.

Compared to the standard approach, our modified embedding approach performs much better at comparable costs. It, therefore, should be the method of choice in the given setup.

APPENDIX A: TAYLOR–ITÔ EXPANSION OF $\mathbf{Y}_1$

A Taylor-Itô expansion of $\mathbf{Y}_1(t)$ provides a stochastic description of the values $\mathbf{Y}_1(t+\Delta)$ for given $\mathbf{Y}(t)$. Assuming smooth functions $\mathbf{f}$ and $\mathbf{g}$, the expansion can be written as an infinite sum of deterministic and stochastic integrals that only depend on Δ and $\boldsymbol{\xi}$ and that are weighted by coefficient functions. These functions only depend on the values and derivatives of $\mathbf{f}$ and $\mathbf{g}$, evaluated at $\mathbf{Y}(t)$. In the following, some properties of the integrals will shortly be summarized. A detailed description of the Taylor-Itô expansion and the properties of the stochastic integrals can be found, e.g., in [1].

Using a multi-index $\boldsymbol{\alpha}$, the expansion of $\mathbf{Y}_1$ can be written quite compactly,

$$\mathbf{Y}_1(t+\Delta)|_{\mathbf{y}} = \mathbf{y}_1 + \sum_{\boldsymbol{\alpha}} \mathbf{c}_{\boldsymbol{\alpha}}(\mathbf{y}) I^{t,\Delta}_{\boldsymbol{\alpha}}(\boldsymbol{\xi}), \tag{A1}$$

with

$$\boldsymbol{\alpha} := (\alpha_1,\ldots,\alpha_n), \quad n \in \mathbb{N}, \tag{A2}$$

$$\alpha_i \in \{0,\ldots,N\}. \tag{A3}$$

Here $\mathbf{c}_{\boldsymbol{\alpha}}$ denotes vectors of the above mentioned coefficient functions. The multiple integrals $I_{\boldsymbol{\alpha}}$ may contain integrations with respect to time as well as integrations with respect to components of the Wiener process $\mathbf{W}(t)$, associated with the Gaussian noise $\boldsymbol{\xi}(t)$. The structure of each integral is determined by its multi-index $\boldsymbol{\alpha}$,

$$I^{t,\Delta}_{\boldsymbol{\alpha}} := \int_{s_n=t}^{t+\Delta} \int_{s_{n-1}=t}^{s_n} \cdots \int_{s_1=t}^{s_2} dZ_1 \cdots dZ_n, \tag{A4}$$

with

$$dZ_i := \begin{cases} ds_i, & \alpha_i = 0 \\ dW_{\alpha_i}(s_i), & \alpha_i \neq 0 \end{cases}. \tag{A5}$$

The ith integration, therefore, will be performed with respect to s_i if the ith component of $\boldsymbol{\alpha}$ is zero. Otherwise, if $\alpha_i = j \neq 0$, the integration will be performed with respect to the jth component of the vector $\mathbf{W}(s_i)$. The multi-index also determines the order of magnitude of the integral

$$I^{t,\Delta}_{\boldsymbol{\alpha}} = O(\Delta^{m(\boldsymbol{\alpha})}), \tag{A6}$$

with

$$m(\boldsymbol{\alpha}) := \sum_{\alpha_i=0} 1 + \sum_{\alpha_i\neq 0} \frac{1}{2}. \tag{A7}$$

Because of Itô's definition of the stochastic integral, the expectation value of $I_{\boldsymbol{\alpha}}$ will be zero if it contains any integration with respect to a Wiener process, i.e., if there are any nonzero components in its index vector. Otherwise, when all components are zero, the integral becomes purely deterministic and evaluates to $(\Delta^n)/n!$, where n indicates the length of $\boldsymbol{\alpha}$.

In Appendix B expectation values of multiple products of integrals will be of interest. These values in general are given by multivariant polynomials in the increments Δ_i of the involved integrals, where for each monomial the powers sum up to a value r, which is determined by the index vectors of the integrals,

$$\left\langle \prod_{i=1}^{k} I^{t,\Delta_i}_{\boldsymbol{\alpha}_i} \right\rangle = \begin{cases} P^{(r)}(\Delta_1,\ldots,\Delta_k) \\ 0 \end{cases}, \tag{A8}$$

with

$$r := \sum_{i=1}^{k} m(\boldsymbol{\alpha}_i). \tag{A9}$$

Here a sufficient (but not necessary) condition for a vanishing expectation value is an odd total number of nonzero entries in the index vectors, i.e., a nonintegral value of r. Consequently, for nonvanishing expectation values, r will always be integral. The coefficients of the above mentioned polynomials, in general, will not have a uniform definition but will depend on size relations between the increments. To give an example (and also to provide a result needed for the calculation of the moments $\mathbf{M}^{(k,\nu)}$ in Sec. IV) one such expectation value is calculated explicitly now. We are interested in

$$E := \left\langle I^{t,\Delta_1}_{(i,0)} I^{t,\Delta_2}_{(j,0)} \right\rangle. \tag{A10}$$

Using the (somewhat loose) notation $dW_i(s) = \xi_i(s)\,ds$, one first finds

$$\begin{aligned} I^{t,\Delta_1}_{(i,0)} &= \int_{s_2=t}^{t+\Delta_1} \int_{s_1=t}^{s_2} \xi_i(s_1)\,ds_1\,ds_2 \\ &= \int_{s_1=t}^{t+\Delta_1} \int_{s_2=s_1}^{t+\Delta_1} \xi_i(s_1)\,ds_2\,ds_1 \\ &= \int_{s_1=t}^{t+\Delta_1} (t+\Delta_1-s_1)\xi_i(s_1)\,ds_1. \end{aligned} \tag{A11}$$

With $\langle \xi_i(s)\xi_j(s')\rangle = \delta_{ij}\delta(s-s')$ this leads to

$$\begin{aligned} E &= \delta_{ij} \int_{s=0}^{\Delta_1} \int_{s'=0}^{\Delta_2} (\Delta_1 - s)(\Delta_2 - s')\delta(s-s')\,ds\,ds' \\ &= \delta_{ij} \int_{s=0}^{\min(\Delta_1,\Delta_2)} (\Delta_1 - s)(\Delta_2 - s)\,ds \\ &= \delta_{ij} \begin{cases} \frac{1}{2}\Delta_1\Delta_2^2 - \frac{1}{6}\Delta_2^3, & \Delta_2 \leqslant \Delta_1 \\ \frac{1}{2}\Delta_1^2\Delta_2 - \frac{1}{6}\Delta_1^3, & \Delta_2 > \Delta_1 \end{cases}. \end{aligned} \tag{A12}$$

In Appendix B a more restrictive case is looked at, where the ratios Δ_i/Δ_j of the increments are kept fix. This can be expressed by $\Delta_i = \lambda_i\Delta$, where Δ denotes some reference increment and the factors λ_i are constants. The expectation values then become proportional to the rth power of Δ,

$$\Delta_i \stackrel{!}{=} \lambda_i\Delta \quad \Rightarrow \quad \left\langle \prod_{i=1}^{k} I^{t,\Delta_i}_{\boldsymbol{\alpha}_i} \right\rangle = \begin{cases} C\,\Delta^r \\ 0 \end{cases}. \tag{A13}$$

Next the actual expansion will be given. There is one special point in the expansion of $\mathbf{Y}_1$: If the last entry of an index vector is nonzero, the corresponding coefficient function $\mathbf{c}_{\boldsymbol{\alpha}}$ will be vanishing [this is due to the fact that $\mathbf{Y}_1$ is not directly driven by noise; see Eq. (9)]. The remaining integrals will thus all be at least of order $O(\Delta)$,

$$\mathbf{Y}_1(t+\Delta)|_{\mathbf{y}} = \mathbf{y}_1 + \mathbf{y}_2\,\Delta + \mathbf{f}(\mathbf{y})\frac{\Delta^2}{2} + \mathbf{g}(\mathbf{y})\mathbf{I}^{t,\Delta} + \mathbf{R}^{t,\Delta}(\mathbf{y}), \tag{A14}$$

with

$$I^{t,\Delta}_i := I^{t,\Delta}_{(i,0)}. \tag{A15}$$

The remainder $\mathbf{R}$ is used to summarize all remaining expansion terms. Its lowest order stochastic terms are given by $\mathbf{c}_{(j,k,0)} I^{t,\Delta}_{(j,k,0)}$ and its lowest order deterministic term by $\mathbf{c}_{(0,0,0)} I^{t,\Delta}_{(0,0,0)}$. Thus $\mathbf{R}$ is a term of order $O(\Delta^2)$ with the statistical properties

$$\left\langle R_i^{t,\Delta}\right\rangle = O(\Delta^3), \tag{A16}$$

$$\left\langle R_i^{t,\Delta} R_j^{t,\Delta}\right\rangle = O(\Delta^4). \tag{A17}$$

APPENDIX B: FUNCTIONAL FORM OF HIGHER ORDER TERMS

Equation (36) is accurate up to first order only. The "classical" Markov analysis, as sketched in Sec. I, faces the same problem: Equation (6), the relation between the moments $\mathbf{m}^{(k)}$ and the Kramers-Moyal coefficients, is accurate up to order $O(\tau)$ only. However, for Eq. (6) the functional form (with respect to τ) of the higher order terms is known—terms of order $O(\tau^n)$ simply are proportional to τ^n. Performing a linear regression with a function base $\{\tau,\tau^2,\ldots,\tau^n\}$ will thus allow parameter estimations with an accuracy of $O(\tau^n)$ (of course, there are practical limitations for n).

For higher order estimations in the given setup, the functional form (with respect to τ and θ) of the higher order terms of $\hat{\mathbf{m}}^{(k)}$ is needed. Because the functional form of *all* terms of $\hat{\mathbf{m}}^{(k)}$ is dictated by the form of the moments $\mathbf{M}^{(k,\nu)} = \langle \mathbf{A}^k \otimes \mathbf{B}^\nu\rangle$, the starting point will be the vectors $\mathbf{A}$ and $\mathbf{B}$.

According to Eq. (28) the components of both vectors can be expressed as linear combination of terms that either stem from the Taylor-Itô expansion or from the measurement noise. Denoting the former by q^ξ and the later by q^γ, the terms can be expressed as [using $m(\boldsymbol{\alpha})$ as defined in Eq. (A7)]

$$q^\xi \in \left\{I_{\boldsymbol{\alpha}}^{t,\Delta}, \theta^{-1} I_{\boldsymbol{\beta}}^{t,\Delta}\right\}, \tag{B1}$$

$$\Delta \in \{\theta,\tau,\tau+\theta\},\ \ m(\boldsymbol{\alpha}) \geqslant 1,\ \ m(\boldsymbol{\beta}) \geqslant 3/2, \tag{B2}$$

and

$$q^\gamma \in \{\Gamma_i(t+\Delta), \theta^{-1}\Gamma_i(t+\Delta)\}, \tag{B3}$$

$$\Delta \in \{0,\theta,\tau,\tau+\theta\}. \tag{B4}$$

A component of $\mathbf{M}^{(k,\nu)}$, therefore, can be expressed as a linear combination of expectation values of $k+\nu$ factors q. Because $\boldsymbol{\Gamma}$ is assumed to be external noise, each expectation value, denoted by Q, can be factorized,

$$Q := \left\langle \prod_{i=1}^{n_1} q_i^\xi \prod_{j=1}^{n_2} q_j^\gamma \right\rangle = \left\langle \prod_{i=1}^{n_1} q_i^\xi \right\rangle \left\langle \prod_{j=1}^{n_2} q_j^\gamma \right\rangle, \tag{B5}$$

with

$$n_1 + n_2 = k + \nu. \tag{B6}$$

The components Γ_i have been assumed to be Gaussian noise with a magnitude of $O(\varepsilon^{3/2})$. A nonvanishing expectation value of a product of n factors $\Gamma_i(t+\Delta_i)$ will thus be given by $C^\gamma \varepsilon^{3n/2}$, where C^γ in general depends on whether or not τ equals θ. As q^γ either denotes a factor Γ or a factor $\theta^{-1}\Gamma$, one finds

$$\left\langle \prod_{j=1}^{n_2} q_j^\gamma \right\rangle = C^\gamma \theta^{n_2 - n_2'} (\varepsilon^3/\theta^2)^{n_2/2}, \tag{B7}$$

with

$$0 \leqslant n_2' \leqslant n_2. \tag{B8}$$

The expectation value of a product of integrals $I_{\boldsymbol{\alpha}}$ will be a polynomial P in τ and θ, where the coefficients in general will depend on whether or not τ is smaller than θ. For each monomial the powers of τ and θ will sum up to a value n_1'', determined by the index vectors of the integrals. As q^ξ either denotes a factor $I_{\boldsymbol{\alpha}}$ or a factor $\theta^{-1} I_{\boldsymbol{\beta}}$, one finds

$$\left\langle \prod_{i=1}^{n_1} q_i^\xi \right\rangle = \theta^{-n_1'} P^{(n_1'')}(\tau,\theta), \tag{B9}$$

with

$$0 \leqslant n_1' \leqslant n_1, \quad n_1'' \geqslant n_1 + n_1'/2. \tag{B10}$$

The expectation values Q can thus be written as a linear combination of terms Q', as defined below. Here it has been used that odd moments of Γ are vanishing, i.e., only even values n_2 have to be considered,

$$Q' = C\tau^a\theta^b(\varepsilon^3/\theta^2)^c, \tag{B11}$$

with

$$a \geqslant 0, \quad a+b \geqslant 0, \quad c \geqslant 0. \tag{B12}$$

The value of C depends on whether τ is smaller, equal or larger than θ. Keeping the ratio of τ and θ fixed, therefore, leads to a constant factor C. The functional form of the terms Q' (and thus of all terms in the moments $\hat{\mathbf{m}}^{(k)}$) is then given by

$$\tau/\theta \overset{!}{=} \text{const} \quad \Rightarrow \quad Q' \ \sim\ \tau^a(\varepsilon^3/\tau^2)^b, \tag{B13}$$

with

$$a \geqslant 0, \quad b \geqslant 0. \tag{B14}$$

Because τ is assumed to be of order $O(\varepsilon)$, the term Q' is of order $O(\varepsilon^{a+b})$. The function base of terms of order $O(\varepsilon^n)$, denoted by $\mathcal{B}^{(n)}$, thus consists of the τ-dependent parts of all terms Q' with $a+b=n$,

$$\mathcal{B}^{(0)} = \{1\}, \tag{B15}$$

$$\mathcal{B}^{(1)} = \{\tau,\tau^{-2}\}, \tag{B16}$$

$$\vdots \tag{B17}$$

$$\mathcal{B}^{(n)} = \{\tau^n,\tau^{n-3},\ldots,\tau^{-2n}\}. \tag{B18}$$

Unfortunately this means $\mathcal{B}^{(n)} \subset \mathcal{B}^{(n+3)}$, which puts a limit on the accuracy that can be achieved. It is, for example, not possible to distinguish some of the terms of order $O(\varepsilon^4)$ from the terms of order $O(\varepsilon)$. At most, therefore, an accuracy of order 3 can be achieved [if no $O(1)$ terms are present].

APPENDIX C: ESTIMATING V FROM THE AUTOCOVARIANCE OF $\mathbf{Y}_1^*$

Before attacking the problem of estimating $\mathbf{V}$, the autocovariance $\mathbf{a}(\tau)$ of $\mathbf{Y}(t)$ will be looked at,

$$\mathbf{a}(\tau) := \langle \mathbf{Y}(t)\mathbf{Y}^T(t+\tau)\rangle. \tag{C1}$$

As $\mathbf{Y}(t)$ is a Langevin process, some useful properties of $\mathbf{a}(\tau)$ can be derived from a Taylor–Itô expansion. First $\mathbf{Y}(t+\tau)$ is expressed as

$$\mathbf{Y}(t+\tau) = \mathbf{Y}(t) + \mathbf{h}(\mathbf{Y}(t),\tau), \tag{C2}$$

with the conditional process increment

$$\mathbf{h}(\mathbf{Y}(t),\tau) := \mathbf{Y}(t+\tau)|_{\mathbf{Y}(t)} - \mathbf{Y}(t). \tag{C3}$$

This leads to

$$\begin{aligned} a_{ij}(\tau) &= \langle Y_i(t)Y_j(t)\rangle + \langle Y_i(t)h_j(t)\rangle \\ &= a_{ij}(0) + \int_{\mathbf{y}} p(\mathbf{y})\, y_i \langle h_j(\mathbf{y},\tau)\rangle\, \boldsymbol{d}y. \end{aligned} \tag{C4}$$

From the Taylor-Itô expansion the form of $\langle \mathbf{h}(\mathbf{y},\tau)\rangle$ is known to be given by

$$\langle \mathbf{h}(\mathbf{y},\tau)\rangle = \sum_{k=1}^{\infty} \mathbf{c}^{(k)}(\mathbf{y})\,\tau^k, \qquad \tau \geqslant 0. \tag{C5}$$

So the general form of $\mathbf{a}(\tau)$ is given by

$$\mathbf{a}(\tau) = \mathbf{a}(0) + \sum_{k=1}^{\infty} \mathbf{a}^{(k)}\tau^k, \quad \tau \geqslant 0. \tag{C6}$$

For small increments τ, therefore, the components of $\mathbf{a}(\tau)$ can be approximated by polynomials in τ. It should be noted, however, that in general the derivatives of $\mathbf{a}(\tau)$ are discontinuous at $\tau = 0$. A polynomial approximation, therefore, should be restricted to either positive or negative increments.

Next, the autocovariance of the given time series will be looked at. As $\mathbf{Y}$ is partitioned into the subvectors $\mathbf{Y}_1$ and $\mathbf{Y}_2$, the matrices $\mathbf{a}(\tau)$ and $\mathbf{a}^{(k)}$ can accordingly be partitioned into submatrices $\mathbf{a}_{ij}(\tau)$ and $\mathbf{a}_{ij}^{(k)}$ with $i,j = 1,2$. The autocovariance of the true values $\mathbf{Y}_1$, therefore, is given by

$$\mathbf{a}_{11}(\tau) := \left\langle \mathbf{Y}_1(t)\mathbf{Y}_1^T(t+\tau)\right\rangle = \mathbf{a}_{11}(0) + \sum_{k=1}^{\infty} \mathbf{a}_{11}^{(k)}\tau^k. \tag{C7}$$

However, the given time series $\mathbf{Y}_1^*(t)$ is spoilt by measurement noise $\mathbf{\Gamma}(t)$. This noise is assumed to have vanishing mean and to be independent of $\mathbf{Y}$ [see Eqs. (21) and (22)]. The covariance of $\mathbf{Y}_1^*$, therefore, is given by

$$\begin{aligned} \mathbf{a}_{11}^*(\tau) &:= \left\langle \mathbf{Y}_1^*(t)\mathbf{Y}_1^{*T}(t+\tau)\right\rangle \\ &= \left\langle \mathbf{Y}_1(t)\mathbf{Y}_1^T(t+\tau)\right\rangle + \langle \mathbf{\Gamma}(t)\mathbf{\Gamma}^T(t+\tau)\rangle. \end{aligned} \tag{C8}$$

As the noise is also assumed to be uncorrelated, one finds

$$\mathbf{a}_{11}^*(\tau) = \begin{cases} \mathbf{a}_{11}(0) + \mathbf{V}, & \tau = 0 \\ \mathbf{a}_{11}(\tau), & \tau \neq 0 \end{cases}. \tag{C9}$$

For all values but $\tau = 0$ we thus can estimate the true covariance $\mathbf{a}_{11}$ from the given time series $\mathbf{Y}_1^*$. According to Eq. (C7), therefore, the value at $\tau = 0$ can be estimated by fitting a polynomial to a number of estimates for $\tau > 0$ and subsequently evaluating the polynomial at $\tau = 0$. Subtracting this estimate for $\mathbf{a}_{11}(0)$ from the value $\mathbf{a}_{11}^*(0)$ (calculated from the series $\mathbf{Y}_1^*$) then gives an estimate for $\mathbf{V}$.

Even though this approach works, it will show to be useful to have a closer look at the coefficient $\mathbf{a}^{(1)}$ of the linear term in Eq. (C6). As from the Taylor-Itô expansion the coefficient $\mathbf{c}^{(1)}$ in Eq. (C5) is known to be given by $\mathbf{D}^{(1)}$, one first finds

$$a_{ij}^{(1)} = \int_{\mathbf{y}} p(\mathbf{y})\, y_i D_j^{(1)}(\mathbf{y})\, \boldsymbol{d}y. \tag{C10}$$

As we are looking at a stationary process, the Fokker-Planck equation yields the relation (using summation convention)

$$\frac{\partial}{\partial y_k}\left[p(\mathbf{y})D_k^{(1)}(\mathbf{y})\right] = \frac{1}{2}\frac{\partial^2}{\partial y_k \partial y_l}\left[p(\mathbf{y})D_{kl}^{(2)}(\mathbf{y})\right]. \tag{C11}$$

Multiplying by $y_i y_j$ and applying integration by parts then yields (omitting the function argument $\mathbf{y}$)

$$\int_{\mathbf{y}} p\left[y_i D_j^{(1)} + y_j D_i^{(1)}\right] \boldsymbol{d}y = -\int_{\mathbf{y}} p\, D_{ij}^{(2)}\, \boldsymbol{d}y \tag{C12}$$

and therefore

$$\mathbf{a}^{(1)} + \mathbf{a}^{(1)T} = -\int_{\mathbf{y}} p(\mathbf{y})\,\mathbf{D}^{(2)}(\mathbf{y})\, \boldsymbol{d}y. \tag{C13}$$

Partitioning $\mathbf{D}^{(2)}$ into submatrices $\mathbf{D}_{ij}^{(2)}$ and noting that $\mathbf{D}_{11}^{(2)}$ equals zero [see Eq. (10)] first gives

$$\mathbf{a}_{11}^{(1)} + \mathbf{a}_{11}^{(1)T} = -\int_{\mathbf{y}} p(\mathbf{y})\,\mathbf{D}_{11}^{(2)}(\mathbf{y})\, \boldsymbol{d}y = \mathbf{0}. \tag{C14}$$

Additionally taking into account the symmetry of $\mathbf{V}$ then finally leads to (note that the sum now starts with $k = 2$)

$$\tilde{\mathbf{a}}_{11}(\tau) = \tilde{\mathbf{a}}_{11}(0) + \sum_{k=2}^{\infty} \tilde{\mathbf{a}}_{11}^{(k)}\tau^k, \tag{C15}$$

$$\tilde{\mathbf{a}}_{11}^*(\tau) = \begin{cases} \tilde{\mathbf{a}}_{11}(0) + \mathbf{V}, & \tau = 0 \\ \tilde{\mathbf{a}}_{11}(\tau), & \tau \neq 0 \end{cases}, \tag{C16}$$

with the symmetric matrices

$$\tilde{\mathbf{a}}_{11}(\tau) := \tfrac{1}{2}\left[\mathbf{a}_{11}(\tau) + \mathbf{a}_{11}^T(\tau)\right], \tag{C17}$$

$$\tilde{\mathbf{a}}_{11}^{(k)} := \tfrac{1}{2}\left(\mathbf{a}_{11}^{(k)} + \mathbf{a}_{11}^{(k)T}\right), \tag{C18}$$

$$\tilde{\mathbf{a}}_{11}^*(\tau) := \tfrac{1}{2}\left[\mathbf{a}_{11}^*(\tau) + \mathbf{a}_{11}^{*T}(\tau)\right]. \tag{C19}$$

The components of $\tilde{\mathbf{a}}_{11}(\tau)$, therefore, can be approximated by polynomials in τ *without* a linear part. Using such polynomials to fit the values $\tilde{\mathbf{a}}_{11}^*(\tau)$, therefore, should lead to more accurate estimates of $\mathbf{V}$ due to the reduced number of unknown polynomial coefficients.

[1] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations* (Springer, New York, 1999).

[2] H. Risken, *The Fokker-Planck Equation* (Springer, New York, 1989).

[3] D. Crommelin and E. Vanden-Eijnden, J. Comput. Phys. **217**, 782 (2006).

[4] D. Crommelin and E. Vanden-Eijnden, Commun. Math. Sci. **4**, 651 (2006).

[5] A. Dembo and O. Zeitoni, Stoch. Proc. Appl. **23**, 91 (1986).

[6] F. Campillo, F. Le Gland, Stoch. Proc. Appl. **33**, 245 (1989).

[7] A. Golightly and D. J. Wilkinson, Comput. Stat. Data Anal. **52**, 1674 (2008).

[8] D. Peavoy, C. Franzke, and G. O. Roberts, Comput. Stat. Data Anal. **83**, 182 (2014).
[9] R. Friedrich and J. Peinke, Phys. Rev. Lett. **78**, 863 (1997).
[10] S. Siegert, R. Friedrich, and J. Peinke, Phys. Lett. A **243**, 275 (1998).
[11] R. Friedrich, S. Siegert, J. Peinke, St. Luck, M. Siefert, M. Lindemann, J. Raethjen, G. Deuschl, and G. Pfister, Phys. Lett. A **271**, 217 (2000).
[12] J. Gradisek, S. Siegert, R. Friedrich, and I. Grabec, Phys. Rev. E **62**, 3146 (2000).
[13] R. Friedrich, J. Peinke, M. Sahimi, and T. M. R. Rahimi, Phys. Rep. **506**, 87 (2011).
[14] 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, New York, 2009), pp. 3574–3597.
[15] R. Friedrich, C. Renner, M. Siefert, and J. Peinke, Phys. Rev. Lett. **89**, 149401 (2002).
[16] J. Gottschall and J. Peinke, New J. Phys. **10**, 083034 (2008).
[17] G. A. Pavliotis and A. M. Stuart, J. Stat. Phys **127**, 741 (2007).
[18] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis* (Cambridge University Press, Cambridge, England, 1997).
[19] H. D. I. Abrabanel, R. Brown, J. J. Sidorowich, and L. S. Tsimiring, Rev. Mod. Phys. **65**, 1331 (1993).
[20] E. J. Kostelich and T. Schreiber, Phys. Rev. E **48**, 1752 (1993).
[21] J. P. M. Heald and J. Stark, Phys. Rev. Lett. **84**, 2366 (2000).
[22] P. G. Lind, M. Haase, F. Böttcher, J. Peinke, D. Kleinhans, and R. Friedrich, Phys. Rev. E **81**, 041125 (2010).
[23] B. Lehle, J. Stat. Phys **152**, 1145 (2013).