PAPER • OPEN ACCESS

Data-driven reconstruction of stochastic dynamical equations based on statistical moments

To cite this article: Farnik Nikakhtar *et al* 2023 *New J. Phys.* **25** 083025

View the article online for updates and enhancements.

New Journal of Physics

The open access journal at the forefront of physics

PAPER

Data-driven reconstruction of stochastic dynamical equations based on statistical moments

Farnik Nikakhtar[1,*] ⓘ, Laya Parkavousi[2], Muhammad Sahimi[3] ⓘ, M Reza Rahimi Tabar[4,5], Ulrike Feudel[5,*] and Klaus Lehnertz[6,7,8,*] ⓘ

1   Department of Physics, Yale University, New Haven, CT 06511, United States of America
2   Max Planck Institute for Dynamics and Self-Organization (MPI-DS), 37077 Göttingen, Germany
3   Mork Family Department of Chemical Engineering & Materials Science, University of Southern California, Los Angeles, CA 90089-1211, United States of America
4   Department of Physics, Sharif University of Technology, Tehran 11155-9161, Iran
5   Theoretical Physics/Complex Systems, ICBM, University of Oldenburg, 26129 Oldenburg, Germany
6   Department of Epileptology, University of Bonn Medical Centre, Venusberg Campus 1, 53127 Bonn, Germany
7   Helmholtz-Institute for Radiation and Nuclear Physics, University of Bonn, Nussallee 14–16, 53115 Bonn, Germany
8   Interdisciplinary Center for Complex Systems, University of Bonn, Brühler Straße 7, 53175 Bonn, Germany
*   Authors to whom any correspondence should be addressed.

E-mail: farnik.nikakhtar@yale.edu, ulrike.feudel@uni-oldenburg.de and klaus.lehnertz@ukbonn.de

Abstract

Stochastic processes are encountered in many contexts, ranging from generation sizes of bacterial colonies and service times in a queueing system to displacements of Brownian particles and frequency fluctuations in an electrical power grid. If such processes are Markov, then their probability distribution is governed by the Kramers–Moyal (KM) equation, a partial differential equation that involves an infinite number of coefficients, which depend on the state variable. The KM coefficients must be evaluated based on measured time series for a data-driven reconstruction of the governing equations for the stochastic dynamics. We present an accurate method of computing the KM coefficients, which relies on computing the coefficients' conditional moments based on the statistical moments of the time series. The method's advantages over state-of-the-art approaches are demonstrated by investigating prototypical one-dimensional stochastic processes with well-known properties.

1. Introduction

The analysis of time series reflecting the dynamics of complex systems, ranging from physics via engineering to medicine and the neurosciences, needs to be based on assessing the interactions and strengths of random forces operating within the systems. Modeling of such systems and time series has been pursued for decades [1–5]. Time series with a finite Markov–Einstein scale can be modeled by a data-driven reconstruction of the governing equations, leading to stochastic differential equations (SDEs). These consist of deterministic and stochastic terms, with the latter corresponding to the fluctuating behavior of the measured time series. As described below, the parameters of the SDE, known as Kramers–Moyal (KM) coefficients, are estimated directly from the measured data [3–5].

Not only does the modeling aim to understand the physics of the system and to extract as much information from the data as possible, but it also aims to make probabilistic predictions, namely to address the question: given the state of a system at time t, what is the probability of finding it in a particular state at time $t + \tau$? If the state of a system at time t depends statistically only on its state at the previous time step, but not on earlier ones, then the stochastic process represents a Markov process. A given time series may be considered as Markov above the Markov–Einstein scale, which can be estimated directly from observations via some statistical tests [3–5].

It is known that the probability distributions—both marginal and conditional—of Markov processes are governed by a first-order partial differential equation in time and an infinite series of derivatives with respect to the state variable x. The equation is known as the KM equation [1, 2] and is given by

$$\frac{\partial}{\partial t} p(x,t|x_0,t_0) = \sum_{l=0}^{\infty} \left(-\frac{\partial}{\partial x}\right)^l \left\{ D^{(l)}(x,t) p(x,t|x_0,t_0) \right\}, \tag{1}$$

subject to the initial condition $p(x,t_0|x_0,t_0) = \delta(x-x_0)$. Here, $D^{(l)}(x,t)$ are the conditional moments of the probability density functions of the transition rates, which—if all known—can identify the probability distribution of the transition rates. They are the aforementioned KM coefficients and are given by

$$D^{(l)}(x,t) = \lim_{\tau \to 0} \frac{1}{l!} \frac{K^{(l)}(x,t,\tau)}{\tau}, \tag{2}$$

with $K^{(l)}(x,t,\tau)$ defined as

$$\begin{aligned}
K^{(l)}(x,t,\tau) &= \int dx' (x'-x)^l p(x',t+\tau|x,t) \\
&= \left\langle \left. [x(t+\tau) - x(t)]^l \right|_{x(t)=x} \right\rangle,
\end{aligned} \tag{3}$$

and with $\langle \cdots \rangle$ denoting averaging over the conditional distribution. State-of-the-art procedures for estimating the conditional moments $K^{(l)}(x,t,\tau)$ and, henceforth, the KM coefficients involve histogram-based [5] and kernel-based regression [6–8] as well as the maximum likelihood method [9–11], with the first two methods estimating the conditional probability distribution function (PDF) $p(x',t+\tau|x,t)$ for given x and x' in order to calculate the average in equation (3). Other methods are the unbiased estimation method [12] and the statistical inference approach [13]; see also [14] for double kernel-based regression for unevenly sampled time series.

In general, estimating the conditional PDF using the histogram-based regression with a fixed binning, given the time series with finite number of data points, will lead to a sparse matrix (for $p(x',t+\tau|x,t)$) and, therefore, will not provide reliable results for KM coefficients. Furthermore, the kernel-based regression requires an *a priori* selection of the kernel and its band-width to compute the conditional moments, which represents the method's main obstacle for yielding accurate results. In addition to such limitations, due to the sparsity of the matrix of conditional PDFs, such procedures (histogram/kernel) are applicable in one and rarely in two dimensions [3, 15, 16]. On the other hand, similar to other optimization problems, maximum likelihood estimates can be sensitive to the choice of the starting values and are computationally expensive. To overcome such limitations, we introduce in this paper a new approach, the moments method, for computing the KM coefficients using the short-time correlation functions and statistical moments of time series.

2. Estimation of KM conditional moments using statistical moments

If $y_l(t) = [x(t+\tau) - x(t)]^l$ represents the lth increments of $x(t)$, and assuming that the conditional moments in equation (3) can be approximated by an l'-order polynomial (see the discussion below for the choice of l'), one can write

$$\begin{aligned}
K^{(l)}(x,t,\tau) &= \left\langle y_l(t,\tau) \Big| x(t) = x \right\rangle \\
&= \phi_0 + \phi_1 x + \phi_2 x^2 + \phi_3 x^3 + \cdots + \phi_{l'} x^{l'},
\end{aligned} \tag{4}$$

where we omitted the dependence of ϕ_i on t and τ to simplify notation. We find that the coefficients $\phi_0, \phi_1, \ldots, \phi_{l'}$ are the solution of a system of linear equations (see appendix A)

$$\begin{pmatrix} \langle y_l \rangle \\ \langle x y_l \rangle \\ \vdots \\ \langle x^{l'} y_l \rangle \end{pmatrix} = \begin{pmatrix} 1 & \langle x \rangle & \langle x^2 \rangle & \cdots & \langle x^{l'} \rangle \\ \langle x \rangle & \langle x^2 \rangle & \langle x^3 \rangle & \cdots & \langle x^{l'+1} \rangle \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \langle x^{l'} \rangle & \langle x^{l'+1} \rangle & \langle x^{l'+2} \rangle & \cdots & \langle x^{2l'} \rangle \end{pmatrix} \begin{pmatrix} \phi_0 \\ \phi_1 \\ \vdots \\ \phi_{l'} \end{pmatrix}, \tag{5}$$

where $\langle x^k \rangle$ is the statistical moment of $x(t)$ of order k. The left-hand side of equation (5) involves terms such as $\langle x^i (x(t+\tau) - x(t))^l \rangle$, which require the estimation of short-time correlation functions of $\langle x^i x(t+\tau)^j \rangle$ and of statistical moments of $\langle x^{i+j} \rangle$. Therefore, to compute the KM coefficients, it is necessary to have knowledge about the short-time correlation functions and statistical moments of a time series. By solving the

matrix equation for $\phi_0, \phi_1, \ldots,$ and $\phi_{l'}$, the conditional moments and the KM coefficients of order l are determined by [cf equation (2)]

$$
\begin{aligned}
D^{(l)}(x,t) &= \lim_{\tau \to 0} \frac{1}{l!} \frac{1}{\tau} \langle [x(t+\tau) - x(t)]^l | x(t) = x \rangle \\
&= \lim_{\tau \to 0} \frac{1}{l!} \frac{1}{\tau} \left[\phi_0 + \phi_1 x + \cdots + \phi_{l'} x^{l'} \right].
\end{aligned}
\tag{6}
$$

The ϕ coefficients are a function of the time lag τ. These coefficients are estimated ($\tilde{\phi}$) for different time lags $\tau = (1, 2, 3, \ldots) \, \mathrm{d}t$ and a linear regression is performed for the first four time lags. The slopes of the linear regressions are the ϕ coefficient. Here, $\mathrm{d}t$ is the sampling interval of the time series $x(t)$. For a non-stationary process, the moments matrix on the right hand side and the vector on the left hand side of equation (5) will be time-dependent, implying that the coefficients ϕ_i can also depend explicitly on time (see appendix B, also for a validation of our method from an analytical perspective).

For a given time series, the order of the polynomial l' is selected based on resolving of $x^{2l'} p(x)$, especially at the tails of $p(x)$. By 'resolving', we mean that for large values of x, given a sufficient number of data points, estimates of $x^{2l'} p(x)$ will be reliable. Polynomials of higher orders l' require a larger amount of data points in order to safely estimate the statistical moments $\langle x^{2l'} \rangle$. For a given number of data points n, increasing l' beyond some upper bound l'_c will no longer resolve $x^{2l'} p(x)$ and, therefore, one should stop the expansion in powers of x at order l'_c. In appendix C, we determine numerically the upper order of the moments, l'_c, for the three examples presented in the following.

The proposed method is also applicable to any non-polynomial KM coefficients, such as fractional or sinusoidal functions. Examples of non-polynomial expressions in equation (3) include the host-immune-tumor model (also known as dynamical model of cancer growth) and the Kuramoto–Sakaguchi phase oscillator model with higher-order interactions [17], to name but a few.

Before we demonstrate advantages of the proposed method over state-of-the-art approaches, we briefly discuss the truncation of the KM equation (1) to rate the number of KM coefficients that should be derived in case of an unknown process. The KM equation offers three possibilities: (i) truncating the expansion at $l = 1$ represents a deterministic process; (ii) truncating at $l = 2$ results in the Fokker–Planck equation for diffusion processes; and (iii) including all terms up to $l \to \infty$. Checking the condition $D^{(l)}(x,t) = 0$ for $l \geqslant 3$ in a given time series is not a straightforward task as it requires knowledge of the KM conditional moments $K^{(l)}(x,t,\tau)$ in powers of the sampling time τ. Lehnertz *et al* have shown that for a process to be classified as diffusive, the condition $K^{(4)}(x,t,\tau) \simeq 3(K^{(2)}(x,t,\tau))^2$ must be satisfied [18]. If this holds, then the first two KM coefficients are sufficient to describe the measured stochasticity of the time series. Otherwise, the dynamics can be approximated using a jump-diffusion process which requires knowledge of the KM coefficients of orders 1, 2, 4, and 6 [5, 24].
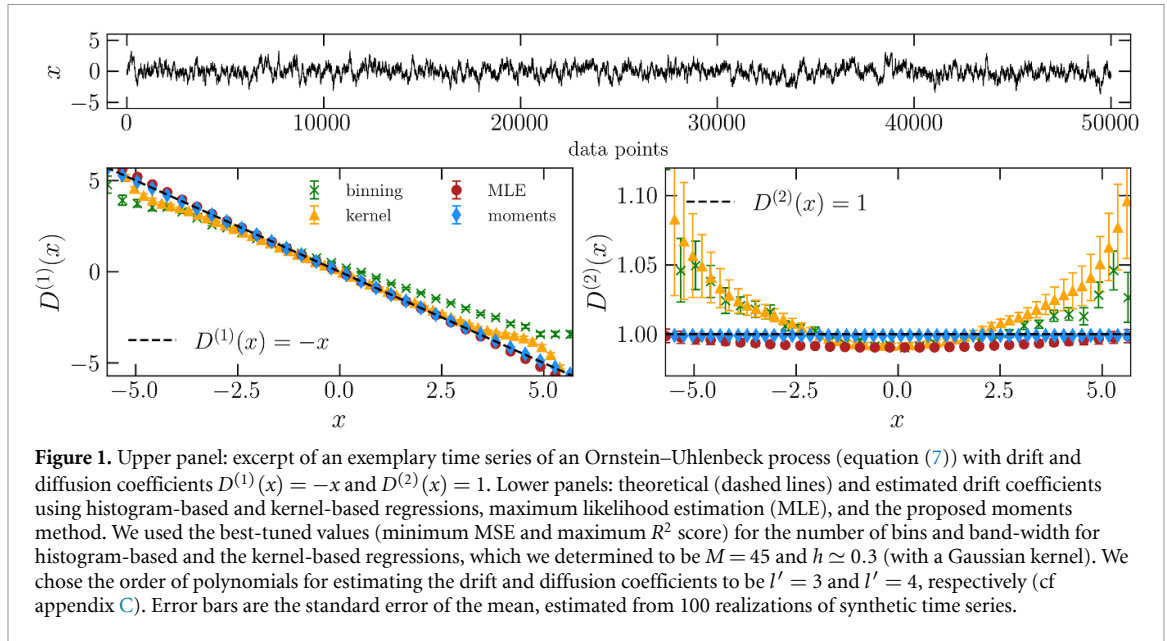
In what follows, we reconstruct one-dimensional stochastic dynamical equations from three prototypical stochastic processes with well-known properties, of which two examples represent diffusion processes and the third one is a jump-diffusion process.

We generated synthetic time series by a numerical simulation of the corresponding dynamical equations using the adaptive integration method for Itô SDEs [19, 20]. Then, we estimated the KM coefficients $D^{(l)}(x,t)$ of the time series using four methods, namely histogram-based and kernel-based regressions, maximum likelihood estimation, and the moments methods proposed in this paper. For each example, the standard error of the mean of KM coefficients was computed based on 100 realizations of the applied noises and with $n = 10^6$ data points. The mean-squared error (MSE) of the estimated as well as the coefficient of determination (R^2 score) between estimated and actual values for the drift and diffusion coefficients (the first two KM coefficients) and the jump parameters, i.e. amplitude and jump-rate, are reported in table 1.

Example 1: Diffusion processes: Ornstein–Uhlenbeck process. The Pawula theorem [21] states that the KM expansion can be truncated at $l = 2$ if $D^{(4)}(x,t)$ vanishes. The KM equation then reduces to the Fokker–Planck equation, with the state variable $x(t)$ describing the diffusion process. In that case, $x(t)$ will be governed by the Langevin equation which (using the Itô description) has the following form:

$$
\mathrm{d}x = D^{(1)}(x,t) \, \mathrm{d}t + \sqrt{2D^{(2)}(x,t)} \, \mathrm{d}W(t) ,
\tag{7}
$$

where $D^{(1)}(x,t)$ and $D^{(2)}(x,t)$ are the drift and diffusion coefficients, respectively. $\{W(t), t \geqslant 0\}$ is a scalar Wiener process.

We consider the Langevin equation (7) with $D^{(1)}(x) = -x$, $D^{(2)}(x) = 1$, and integrate it with $\mathrm{d}t = 0.01$ to generate synthetic time series. Our findings for $D^{(1)}(x)$ and $D^{(2)}(x)$ are shown in figure 1. For the linear process considered, the MSEs of $D^{(1)}(x)$ and $D^{(2)}(x)$ estimated with the moments method are smaller (with

Figure 1. Upper panel: excerpt of an exemplary time series of an Ornstein–Uhlenbeck process (equation (7)) with drift and diffusion coefficients $D^{(1)}(x) = -x$ and $D^{(2)}(x) = 1$. Lower panels: theoretical (dashed lines) and estimated drift coefficients using histogram-based and kernel-based regressions, maximum likelihood estimation (MLE), and the proposed moments method. We used the best-tuned values (minimum MSE and maximum R^2 score) for the number of bins and band-width for histogram-based and the kernel-based regressions, which we determined to be $M = 45$ and $h \simeq 0.3$ (with a Gaussian kernel). We chose the order of polynomials for estimating the drift and diffusion coefficients to be $l' = 3$ and $l' = 4$, respectively (cf appendix C). Error bars are the standard error of the mean, estimated from 100 realizations of synthetic time series.

R^2 closer to unity) than the respective values estimated with the other methods (table 1). For an analytical solution of equation (5) for the coefficients ϕ_i and the KM coefficients, as well as for the condition to stop increasing the order of moments, see appendices B and C.

Example 2: Diffusion processes: the noisy genetic model. As a second example from the family of diffusion processes, we consider a nonlinear multiplicative process, namely, the noisy genetic model in the context of population genetics [22], which exhibits a noise-induced transition from unimodal to bimodal PDFs with increasing noise intensity. The noisy genetic model is described as a one-variable diffusive SDE [23] (see appendix D)

$$\mathrm{d}x = [\alpha - x + \beta_0 x(1-x)]\,\mathrm{d}t + \sigma x(1-x)\,\mathrm{d}W(t)\,, \tag{8}$$

where $x \in [0,1]$, and $\alpha \in (0,1)$ stands for the mutation rate. Here, β_0 denotes the gene selection factor and, independent of α and β_0, the deterministic part of the dynamical system (8) has only one equilibrium point in the corresponding interval, which is stable. Moreover, the intensity σ is a real constant.

There is a critical σ_c^2 such that for $\sigma^2 < \sigma_c^2$, the stationary solution of the corresponding Fokker–Planck equation, $p_s(x)$, has only one extremum, namely a maximum. When $\sigma^2 > \sigma_c^2$, the stationary probability distribution has three extrema, with x_2 being a minimum and x_1 and x_3 each being a maximum. Therefore, when $\sigma^2 < \sigma_c^2$, $p_s(x)$ is unimodal at $\sigma^2 = \sigma_c^2$, the probability distribution becomes flat, followed by a transition to bimodal distribution functions for more intense noise.

We integrate equation (8) with $\mathrm{d}t = 0.01$ to generate synthetic time series for $\sigma_1^2 < \sigma_c^2$, $\sigma_3^2 > \sigma_c^2$, and $\sigma_2^2 \approx \sigma_c^2$. Our results for the drift and diffusion coefficients are summarized in figure 2. We find that for all values of σ^2, the MSEs corresponding to the proposed moments method are comparably smaller (with R^2 closer to unity), as presented in table 1.

Example 3: Jump-diffusion process with a bistable potential. It has become evident that a non-vanishing $D^{(4)}(x,t)$ is a signature of jump discontinuities in a time series [18, 24]. In this case, one needs the KM coefficients of at least order six, $D^{(6)}(x,t)$, and in some other cases, up to order eight [5] to estimate jump amplitude and rate, and the KM equation will contain all the terms up to $l \rightarrow \infty$ [24–34]. For non-vanishing $D^{(4)}(x,t)$, one models a time series as a jump-diffusion process, which is given by a (Itô) dynamical stochastic equation [5, 24]

$$\mathrm{d}x(t) = D^{(1)}(x,t)\,\mathrm{d}t + \sqrt{D(x,t)}\,\mathrm{d}W(t) + \xi\,\mathrm{d}J(t)\,, \tag{9}$$

where $\{W(t), t \geqslant 0\}$ is a scalar Wiener process, $D^{(1)}(x,t)$ is the drift coefficient, $D(x,t)$ is the diffusion function, and $J(t)$ is a Poisson jump process. The jump rate is $\lambda(x)$ that can be state-dependent with size ξ, which we assume to have some symmetric distribution with finite even-order statistical moments $\langle \xi^{2j} \rangle$ for $j \geqslant 1$. For a jump-diffusion process (9), all functions and parameters can be determined based on measured
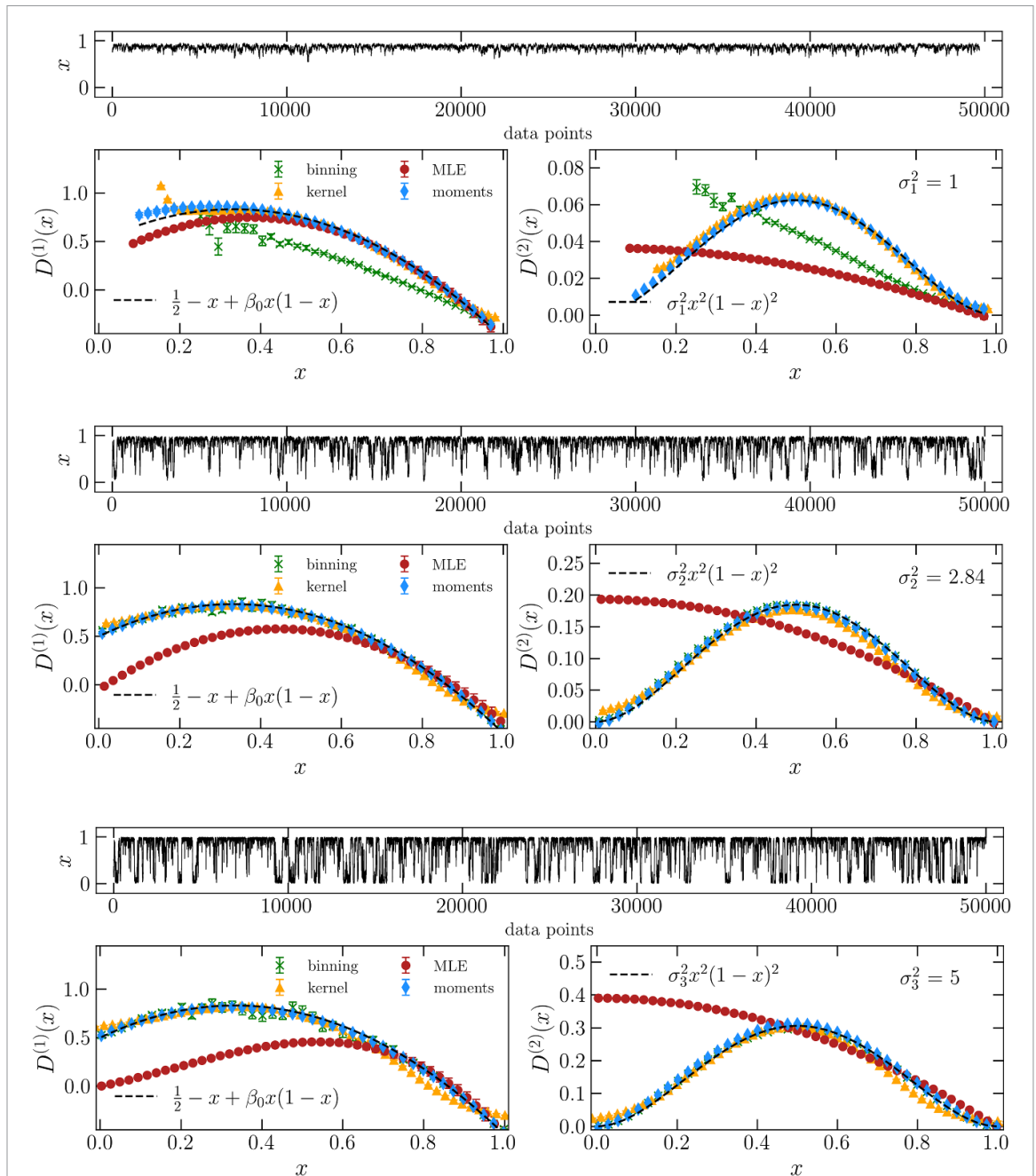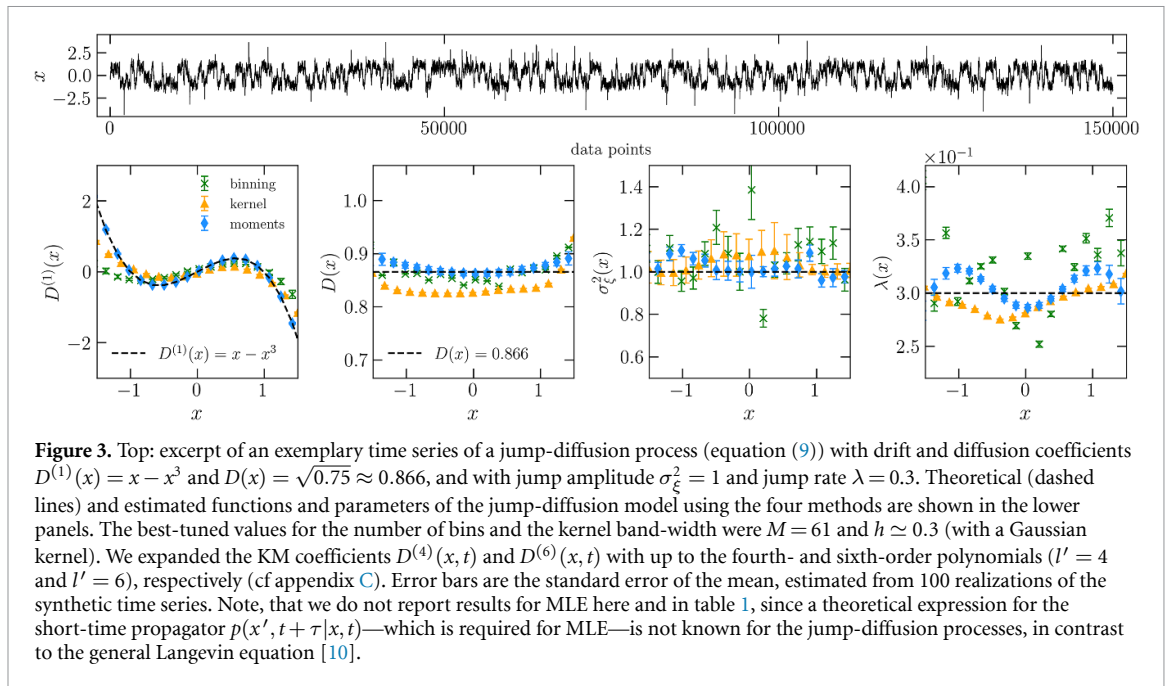
Figure 2. Top to bottom: excerpts of exemplary time series of the noisy genetic model (equation (8)) as well as drift and diffusion coefficients $D^{(1)}(x) = \frac{1}{2} - x + \beta_0 x(1-x)$ (with $\beta_0 = 3$) and $D^{(2)}(x) = \sigma^2 x^2(1-x)^2$ for $\sigma^2 < 2.84$ (top), $\sigma^2 \approx 2.84$ (middle), and $\sigma^2 > 2.84$ (bottom). Theoretical (dashed lines) and the estimated drift and diffusion coefficients using the four methods. The best-tuned values for the number of bins and for the kernel band-width were $M = 41$ and $h \simeq 0.3$ (with a Gaussian kernel). The order of polynomials for estimating the drift and diffusion coefficients were $l' = 3$ and $l' = 4$, respectively (cf appendix C). Error bars are the standard error of the mean (SEM) estimated from 100 realizations of the synthetic time series.

time series by estimating the KM conditional moments [5, 24] (see appendix E). For simplicity, we assume that the jump size ξ has a Gaussian PDF with variance σ_ξ^2. In that case, $\langle \xi^{2k+1} \rangle = 0$ and $\langle \xi^{2k} \rangle = \frac{2n!}{2^k k!} \langle \xi^2 \rangle^k$.

We consider a dynamical system with a cubic nonlinearity ($D^{(1)}(x,t) = x - x^3$) with additive white noise, $D(x,t) = \sqrt{0.75}$, jump amplitude $\sigma_\xi^2 = 1$, and jump rate $\lambda = 0.3$. The deterministic part of the dynamics has three fixed points at 0, one unstable fixed point, and ± 1 as stable ones. In figure 3, we show the numerical values for both the reconstructed drift and diffusion coefficients, as well as jump amplitude and jump rate. We find that the MSEs corresponding to the proposed moments method are comparably smaller, see table 1.

Table 1. Mean squared errors (MSEs) of estimated and R^2 scores between estimated and actual values of the drift and diffusion coefficients, as well as jump rate and jump amplitude for the examples considered. Results are given for the histogram-based and the kernel-based regressions, maximum likelihood estimation (MLE), and the proposed moments method. Since no theoretical expression for the short-time propagator of $p(x', t + \tau | x, t)$ is known for the jump-diffusion processes, MLE could not be applied. MSEs and R^2 scores for the Ornstein–Uhlenbeck process were estimated in the interval $x \in [-2, 2]$, for the noisy genetic model in the interval $x \in [0.1, 1]$, and for the jump-diffusion process in the interval $x \in [-1.5, 1.5]$.

Example	Drift coefficient MSE/R^2	Diffusion coefficient MSE/R^2	Jump rate MSE/R^2	Jump amplitude MSE/R^2	Method
Ornstein–Uhlenbeck process	0.2168/0.7133	0.0001/−17.2624	N/A	N/A	Histogram-based
Ornstein–Uhlenbeck process	0.0106/0.9915	0.0001/−0.0623	N/A	N/A	Kernel-based
Ornstein–Uhlenbeck process	0.0001/0.9999	0.0001/−9.2731	N/A	N/A	MLE
Ornstein–Uhlenbeck process	0.0001/0.9999	0.0001/0.9999	N/A	N/A	Moments
Noisy genetic model ($\sigma < \sigma_c$)	0.0567/0.4069	0.0003/0.3845	N/A	N/A	Histogram-based
Noisy genetic model ($\sigma < \sigma_c$)	0.0041/0.9708	0.0001/0.9739	N/A	N/A	Kernel-based
Noisy genetic model ($\sigma < \sigma_c$)	0.0055/0.9474	0.0005/−3.3394	N/A	N/A	MLE
Noisy genetic model ($\sigma < \sigma_c$)	0.0011/0.9921	0.0001/0.9865	N/A	N/A	Moments
Noisy genetic model ($\sigma \simeq \sigma_c$)	0.0007/0.9949	0.0001/0.9931	N/A	N/A	Histogram-based
Noisy genetic model ($\sigma \simeq \sigma_c$)	0.0029/0.9784	0.0001/0.9712	N/A	N/A	Kernel-based
Noisy genetic model ($\sigma \simeq \sigma_c$)	0.0684/−0.0221	0.0058/−0.6734	N/A	N/A	MLE
Noisy genetic model ($\sigma \simeq \sigma_c$)	0.0001/0.9999	0.0001/0.9972	N/A	N/A	Moments
Noisy genetic model ($\sigma > \sigma_c$)	0.0015/0.9891	0.0001/0.9962	N/A	N/A	Histogram-based
Noisy genetic model ($\sigma > \sigma_c$)	0.0091/0.9429	0.0003/0.9739	N/A	N/A	Kernel-based
Noisy genetic model ($\sigma > \sigma_c$)	0.1322/−1.3872	0.0263/−0.8569	N/A	N/A	MLE
Noisy genetic model ($\sigma > \sigma_c$)	0.0003/0.9977	0.0001/0.9959	N/A	N/A	Moments
Jump-diffusion process	0.1684/−2.3667	0.0003/−0.0035	0.0199/−0.2823	0.0013/−0.3256	Histogram-based
Jump-diffusion process	0.0782/0.4234	0.0013/−1.0006	0.0032/−1.1983	0.0002/−0.3163	Kernel-based
Jump-diffusion process	–/–	–/–	–/–	–/–	MLE
Jump-diffusion process	0.0001/0.9999	0.0001/−0.5863	0.0057/−0.3062	0.0002/−0.3062	Moments

Figure 3. Top: excerpt of an exemplary time series of a jump-diffusion process (equation (9)) with drift and diffusion coefficients $D^{(1)}(x) = x - x^3$ and $D(x) = \sqrt{0.75} \approx 0.866$, and with jump amplitude $\sigma_\xi^2 = 1$ and jump rate $\lambda = 0.3$. Theoretical (dashed lines) and estimated functions and parameters of the jump-diffusion model using the four methods are shown in the lower panels. The best-tuned values for the number of bins and the kernel band-width were $M = 61$ and $h \simeq 0.3$ (with a Gaussian kernel). We expanded the KM coefficients $D^{(4)}(x,t)$ and $D^{(6)}(x,t)$ with up to the fourth- and sixth-order polynomials ($l' = 4$ and $l' = 6$), respectively (cf appendix C). Error bars are the standard error of the mean, estimated from 100 realizations of the synthetic time series. Note, that we do not report results for MLE here and in table 1, since a theoretical expression for the short-time propagator $p(x', t + \tau | x, t)$—which is required for MLE—is not known for the jump-diffusion processes, in contrast to the general Langevin equation [10].

3. Conclusions

We introduced a new method to estimate conditional averages of increments $(x(t + \tau) - x(t))^l$, with $l = 0, 1, \ldots$, that are needed for a data-driven reconstruction of stochastic dynamical equations from experimental time series. The method consists of determining a conditional average in terms of statistical moments of the measured time series. It replaces the estimation of the sparse conditional probability distributions with solving a system of linear equations that has polynomial time complexity. Using three paradigmatic linear and nonlinear stochastic processes, we demonstrated that MSEs (as well as coefficients of determination of KM coefficients) estimated with the proposed method are smaller (larger) than the respective values estimated with current state-of-the-art methods.

While the examples presented in this work focus on stationary moments, we acknowledge the potential extension of our theoretical approaches to non-stationary processes (as discussed in appendix B). We note that our method has not been implemented for non-stationary data at this stage. However, we are actively working on proving the method's applicability to time-dependent Fokker–Planck equations and expanding its scope to non-stationary time series. We intend to provide further updates and findings in the near future. As a final remark, we note that the proposed method is easily extendable to multivariate time series, which will enable one to characterize pairwise, as well as higher-order interactions in the dynamics [17]. Our moments method provides a way to rapidly and reliably estimate functions and parameters needed for a reconstruction of dynamic equations of complex systems.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

Appendix A. Proof of relation (5)

For two random variables $y(t)$ and $x(t)$ with joint probability density function $p(y, x)$, the conditional expectation of $y(t)$, given that $x(t) = x$, is given by

$$\langle y(t) | x(t) = x \rangle = \int y p(y|x) \mathrm{d}y \tag{A.1}$$

where $p(y|x)$ is the conditional probability, which is obtained via $p(x,y)$ and $p(x)$ using Bayes' theorem. In particular, by setting $y_l(t)$ as $(x(t+\tau) - x(t))^l$, the conditional KM moments will be given by

$$K^{(l)}(x, t, \tau) = \left\langle \left[x(t+\tau) - x(t) \right]^l \middle| x(t) = x \right\rangle, \tag{A.2}$$

in accordance with equation (3) in the main text [1]. Let us assume that the conditional moments $K^{(l)}(x,t,\tau)$ have a functional form $A_l(x)$, whose parameters depend implicitly on l and the time-lag τ and time t. By definition of conditional expectation and using Bayes' theorem, we write

$$\int y p(y|x)\,\mathrm{d}y = \int y \frac{p(y,x)}{p(x)}\,\mathrm{d}y = A_l(x) \tag{A.3}$$

for a given function $A_l(x)$, implying that

$$\int y p(y,x)\,\mathrm{d}y = p(x) \cdot A_l(x). \tag{A.4}$$

By integrating both sides of equation (A.4) over x, one finds,

$$\int \mathrm{d}x \int y p(y,x)\,\mathrm{d}y = \int \mathrm{d}y\, y p(y) = \int \mathrm{d}x\, p(x) \cdot A_l(x). \tag{A.5}$$

$A_l(x)$ can have any functional form [17]. An extreme case for the time series is white noise, for which $K^{(l)}(x,t,\tau)$ will be a polynomial of order l.

Let us consider $A_l(x)$ as a polynomial function of x of order $l' \geqslant l$. Therefore, we must determine $l' + 1$ free parameters, and we have

$$\langle y(t)|x(t)=x \rangle = \phi_0 + \phi_1 x + \phi_2 x^2 + \phi_3 x^3 + \cdots + \phi_{l'} x^{l'}, \tag{A.6}$$

where we can choose $y_l \equiv y_l(t,\tau) = (x(t+\tau) - x(t))^l$ for a given l. Using equation (A.4), one finds

$$\int y_l p(x,y_l)\,\mathrm{d}y = \phi_0 p(x) + \phi_1 x p(x) + \cdots + \phi_{l'} x^{l'} p(x). \tag{A.7}$$

To determine the unknown coefficients ϕ_i, $i = 0, \ldots, l'$ in equation (A.7), we need $l' + 1$ independent relations. For the first of our systems of equations, we integrate both sides of the above equation over variable x to obtain

$$\langle y_l \rangle = \phi_0 + \phi_1 \langle x \rangle + \cdots + \phi_l' \langle x^{l'} \rangle. \tag{A.8}$$

Multiplying equation (A.7) by $x, x^2, \ldots, x^{l'}$ and integrating over x, we obtain a system of linear equations

$$\begin{pmatrix} \langle y_l \rangle \\ \langle x y_l \rangle \\ \vdots \\ \langle x^{l'} y_l \rangle \end{pmatrix} = \begin{pmatrix} 1 & \langle x \rangle & \langle x^2 \rangle & \cdots & \langle x^{l'} \rangle \\ \langle x \rangle & \langle x^2 \rangle & \langle x^3 \rangle & \cdots & \langle x^{l'+1} \rangle \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \langle x^{l'} \rangle & \langle x^{l'+1} \rangle & \langle x^{l'+2} \rangle & \cdots & \langle x^{2l'} \rangle \end{pmatrix} \begin{pmatrix} \phi_0 \\ \phi_1 \\ \phi_2 \\ \vdots \\ \phi_{l'} \end{pmatrix} \tag{A.9}$$

which proves the relation (5) in the main text.

Appendix B. Analytical solutions

Example 1: Stationary process, Ornstein–Uhlenbeck process.

As we mentioned in the main text in the first example, the Ornstein–Uhlenbeck (OU) process can be written as a Langevin equation of the form

$$\mathrm{d}x(t) = -\theta x(t)\,\mathrm{d}t + \mu\,\mathrm{d}W(t)\,, \tag{B.1}$$

with drift and diffusion coefficients given by $D^{(1)}(x,t) = -\theta x$, and $\sqrt{2D^{(2)}(x,t)} = \mu$. Here μ and θ are positive constants. For $\theta > 0$, the process (B.1) has a stationary probability distribution [5]. The mean of the process can be set to zero and the covariance of $x(t)$ and $x(s)$ in the stationary state can be written as [5]

$$\mathrm{cov}(x(t),x(s)) = \frac{\mu^2}{2\theta} \exp\left(-\theta|t-s|\right). \tag{B.2}$$

The statistical moments of the OU process would be as follows

$$\langle x^k \rangle = \begin{cases} (k-1)!! \left(\frac{\mu^2}{2\theta}\right)^{k/2} & k \text{ even} \\ 0 & k \text{ odd} \end{cases}.$$ (B.3)

The double factorial is the product of all the integers from 1 up to k that have the same parity as k.

To demonstrate our moments method, let us assume that the conditional KM moments are a cubic function of x, so by having the analytical form of the covariance and statistical moments, the matrix elements of the system of the linear equations (equation (A.9)) can be determined as

$$\begin{pmatrix} \langle y \rangle \\ \langle xy \rangle \\ \langle x^2 y \rangle \\ \langle x^3 y \rangle \end{pmatrix} = \begin{pmatrix} 1 & 0 & \frac{\mu^2}{2\theta} & 0 \\ 0 & \frac{\mu^2}{2\theta} & 0 & 3(\frac{\mu^2}{2\theta})^2 \\ \frac{\mu^2}{2\theta} & 0 & 3(\frac{\mu^2}{2\theta})^2 & 0 \\ 0 & 3(\frac{\mu^2}{2\theta})^2 & 0 & 15(\frac{\mu^2}{2\theta})^3 \end{pmatrix} \begin{pmatrix} \phi_0 \\ \phi_1 \\ \phi_2 \\ \phi_3 \end{pmatrix}.$$ (B.4)

For the first order KM moment (to obtain the drift coefficient), we set $y \equiv y_1(t,\tau) = x(t+\tau) - x(t)$, so the vector in the lhs of equation (B.4) would be obtained as

$$\begin{pmatrix} 0 \\ \frac{\mu^2}{2\theta}(e^{-\theta\tau} - 1) \\ 0 \\ 3(\frac{\mu^2}{2\theta})^2(e^{-\theta\tau} - 1) \end{pmatrix}.$$ (B.5)

The solution for the parameters vector in this system of linear equations is

$$\begin{pmatrix} \phi_0 \\ \phi_1 \\ \phi_2 \\ \phi_3 \end{pmatrix} = \begin{pmatrix} 0 \\ e^{-\theta\tau} - 1 \\ 0 \\ 0 \end{pmatrix},$$ (B.6)

which means $K^{(1)} = (e^{-\theta\tau} - 1)x(t)$, and the drift coefficient is

$$\begin{aligned} D^{(1)}(x,t) &= \lim_{\tau \to 0} \frac{K^{(1)}(x,t)}{\tau} \\ &= \lim_{\tau \to 0} \frac{1}{\tau}(1 - \theta\tau - 1 + \mathcal{O}(\tau^2))x(t) = -\theta x(t). \end{aligned}$$ (B.7)

For the second KM moment (to obtain the diffusion coefficient), we set $y \equiv y_2(t,\tau) = (x(t+\tau) - x(t))^2$, so the vector in the lhs of equation (B.4) and the solution to the linear system would be as follows

$$\begin{pmatrix} \frac{\mu^2}{\theta}(1 - e^{-\theta\tau}) \\ 0 \\ \frac{\mu^2}{2\theta}(4 - 3e^{-\theta\tau} + 2e^{-2\theta\tau}) \\ 0 \end{pmatrix},$$ (B.8)

and

$$\begin{pmatrix} \phi_0 \\ \phi_1 \\ \phi_2 \\ \phi_3 \end{pmatrix} = \begin{pmatrix} \frac{\mu^2}{\theta}(1 - e^{-\theta\tau}) \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$ (B.9)

Note that the moments matrix in the rhs of equation (B.4) is not a function of y and does not change. Therefore, the diffusion coefficient is

$$D^{(2)}(x,t) = \frac{1}{2} \lim_{\tau \to 0} \frac{K^{(2)}(x,t)}{\tau}$$
$$= \lim_{\tau \to 0} \frac{1}{2\tau} \frac{\mu^2}{\theta} (1 - 1 + \theta\tau + \mathcal{O}(\tau^2))) = \frac{\mu^2}{2}. \tag{B.10}$$

Example 2: Non-stationary process, Wiener process.

As an example of a non-stationary process, we consider the Wiener process whose derivative is a white noise. It is known that the Wiener process has a vanishing drift and a constant diffusion coefficient which equals to $1/2$ [5]. The covariance of a Wiener process is

$$\mathrm{cov}(x(t), x(s)) = \min(t, s), \tag{B.11}$$

where $\min(t,s)$ is the minimum of two chosen times (t, s). With $x(0) = 0$ and $t_0 = 0$ the statistical moments of Wiener process are given by $\langle x^{2k+1} \rangle = 0$ and

$$\langle x^{2k} \rangle = \frac{\Gamma(k + 1/2)}{\Gamma(1/2)} 2^k t^k \qquad k = 1, 2, \ldots,$$

where $\Gamma(x)$ is the Euler gamma function. Here $\langle \cdots \rangle$ denotes ensemble averaging.

For the moments matrix we have

$$\begin{pmatrix} 1 & \langle x \rangle & \langle x^2 \rangle & \langle x^3 \rangle \\ \langle x \rangle & \langle x^2 \rangle & \langle x^3 \rangle & \langle x^4 \rangle \\ \langle x^2 \rangle & \langle x^3 \rangle & \langle x^4 \rangle & \langle x^5 \rangle \\ \langle x^3 \rangle & \langle x^4 \rangle & \langle x^5 \rangle & \langle x^6 \rangle \end{pmatrix} = \begin{pmatrix} 1 & 0 & t & 0 \\ 0 & t & 0 & 3t^2 \\ t & 0 & 3t^2 & 0 \\ 0 & 3t^2 & 0 & 15t^3 \end{pmatrix}. \tag{B.12}$$

To find the drift coefficient, we have to solve the following linear set of equations with $y \equiv y_1(t, \tau) = (x(t + \tau) - x(t))$:

$$\begin{pmatrix} \langle y \rangle \\ \langle xy \rangle \\ \langle x^2 y \rangle \\ \langle x^3 y \rangle \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & t & 0 \\ 0 & t & 0 & 3t^2 \\ t & 0 & 3t^2 & 0 \\ 0 & 3t^2 & 0 & 15t^3 \end{pmatrix} \begin{pmatrix} \phi_0 \\ \phi_1 \\ \phi_2 \\ \phi_3 \end{pmatrix}, \tag{B.13}$$

This gives us $(\phi_0, \phi_1, \phi_2, \phi_3) = (0, 0, 0, 0)$, and then $D^{(1)}(x,t)$ will be

$$D^{(1)}(x,t) = \lim_{\tau \to 0} \frac{K^{(1)}(x,t)}{\tau} = 0. \tag{B.14}$$

Similarly, to find the diffusion coefficient one has to solve the following set of linear equations with $y \equiv y_2(t, \tau) = (x(t + \tau) - x(t))^2$:

$$\begin{pmatrix} \langle y \rangle \\ \langle xy \rangle \\ \langle x^2 y \rangle \\ \langle x^3 y \rangle \end{pmatrix} = \begin{pmatrix} \tau \\ 0 \\ t\tau \\ 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & t & 0 \\ 0 & t & 0 & 3t^2 \\ t & 0 & 3t^2 & 0 \\ 0 & 3t^2 & 0 & 15t^3 \end{pmatrix} \begin{pmatrix} \phi_0 \\ \phi_1 \\ \phi_2 \\ \phi_3 \end{pmatrix}, \tag{B.15}$$

This gives $(\phi_0, \phi_1, \phi_2, \phi_3) = (\tau, 0, 0, 0)$. To compute higher-order covariances in the lhs of equation (B.15), we have used Wick's theorem [5]

$$\langle x(i)^3 x(j) \rangle = 3 \langle x(i)^2 \rangle \langle x(i) x(j) \rangle \tag{B.16}$$

$$\langle x(i)^2 x(j)^2 \rangle = \langle x(i) x(i) \rangle \langle x(j) x(j) \rangle + 2 \langle x(i) x(j) \rangle^2. \tag{B.17}$$

Finally, the diffusion coefficient is given by

$$D^{(2)}(x,t) = \frac{1}{2} \lim_{\tau \to 0} \frac{K^{(2)}(x,t)}{\tau} = \frac{1}{2}. \tag{B.18}$$

Appendix C. Stop-condition for the order of polynomials in the expansion of the KM coefficients

As shown in relation (A.9), to estimate the KM coefficients with a polynomial of order l' one needs to estimate statistical moments of x up to the order $2l'$. To study how such moments can be reliably estimated from given time series (each with length $n\,\mathrm{d}t$), one should ensure that the tails of $x^l\,p(x)$ (for $l = 1, 2, \ldots, 2l'$) are well resolved. In what follows, we show how one can check the reliability of the estimation of the tail of the PDF for the three examples considered in the main text, where the drift and diffusion coefficients are polynomials of orders 1, 2, and 3, respectively. For these examples, we need to estimate the statistical moments at least up to the order of six. In figures C1–C3, we plot $x^2\,p(x)$, ..., $x^{10}\,p(x)$, estimated from time series of the three examples and for different $n\,\mathrm{d}t$. For given $n\,\mathrm{d}t \in \{10, 100, 1000, 10\,000\}$, we estimate the error bars from 10 realizations of the generated time series for each example. As shown in figure C1, for example 1 and with increasing $n\,\mathrm{d}t$ the tails of PDF of $p(x)$ are well-resolved, rare events are detected, and $x^{2k}\,p(x)$ possesses smaller error bars (in the size of symbols), particularly in the tails. Similar discussions are valid for the other examples. We note that in the main text, the results have been reported from ensemble averaging over 100 realizations of time series for each example.

In addition, in figures C4 and C5, we plot the statistical moments $\langle x^{2k} \rangle$, with $k = 1, 2, \ldots, 5$ (from 10 realizations) for different $n\,\mathrm{d}t$. Considering example 1, for $n\,\mathrm{d}t > 1000$ the statistical moment $\langle x^8 \rangle$ and those of higher orders are not approaching constant values up to $n\,\mathrm{d}t = 10\,000$. Therefore by considering the error bars one can choose $l' = 6$.

For a given stationary time series, we can analyze the statistical moments $\langle x^{2k} \rangle$ with varying numbers of data points n. In this case, the error bars are calculated based on a number (e.g. 100) of randomly selected data chunks, each consisting of n data points from the time series.

In summary, using the saturating behavior of $\langle x^{2k} \rangle$ for different k with increasing $n\,\mathrm{d}t$, one can decide in a data-driven way that the KM coefficients can be approximated with a polynomial of order l'_c. In figure C6, we plot the errors of estimations of $\langle x^{2k} \rangle$, with $k = 1, 2, \ldots, 5$ (from 10 realizations) for different $n\,\mathrm{d}t$ for the three examples. The errors decrease with $n\,\mathrm{d}t$ as $1/(n\,\mathrm{d}t)^\gamma$, with $\gamma \simeq 0.5$.

Finally, we note that in cases where we have access to ground truth, we can provide specific details about the minimum order of the polynomial regression. To determine the appropriate minimum order, it is important to utilize a distance measure that allows us to select the minimum order without sacrificing essential information required for accurate estimation of the KM coefficients. By reducing the polynomial order, e.g. from 4 to 3, and subsequently to 2 and 1, we can assess the impact on the accuracy of coefficient estimation. The choice of the distance measure depends on the specific context and analysis requirements. The ultimate objective is to identify the minimum polynomial order that preserves the necessary information for accurate estimation of the KM coefficients, while avoiding unnecessary complexity.
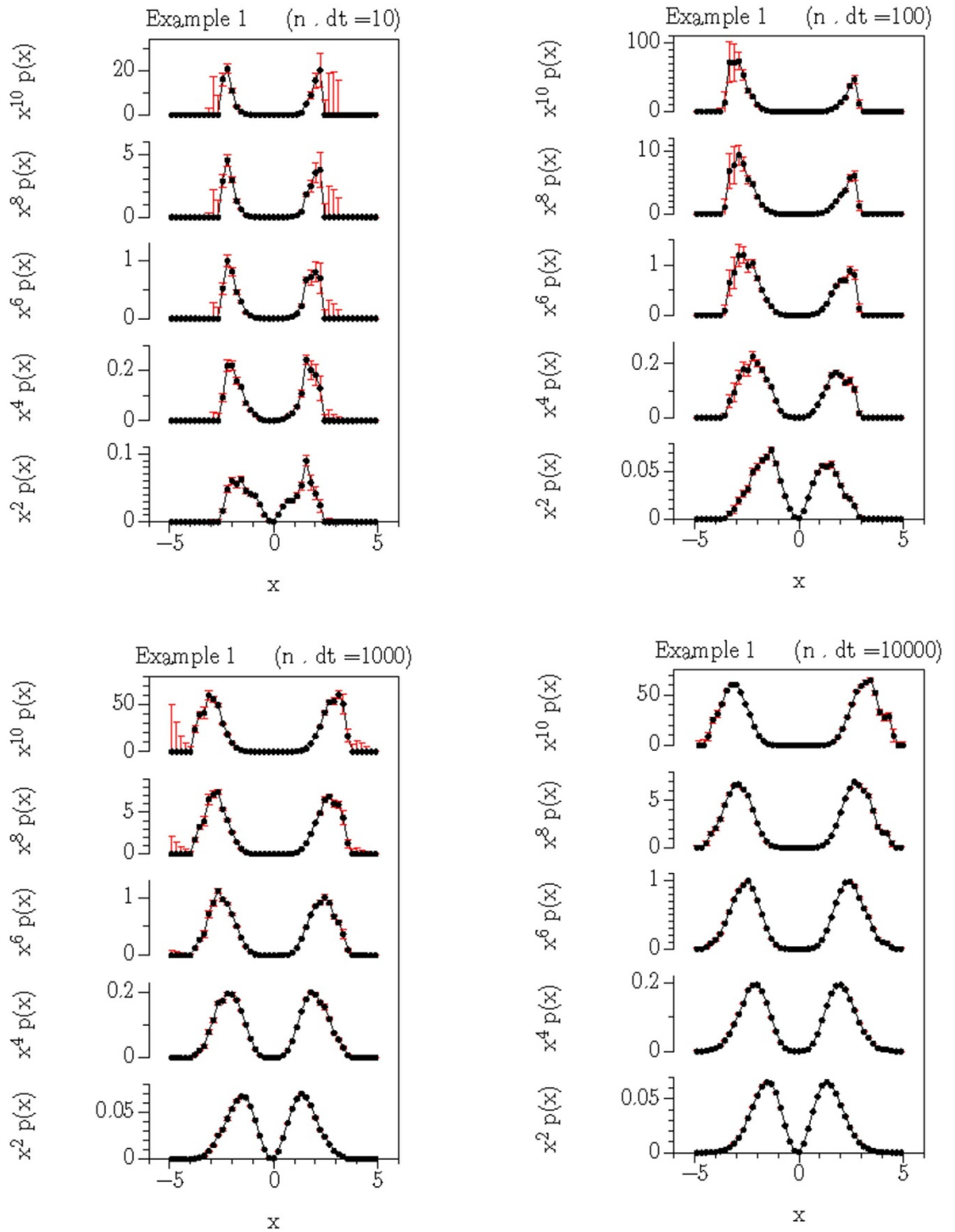
Figure C1. Plots of $x^2 p(x), \ldots, x^{10} p(x)$, estimated from time series of example 1 for different $n\,\mathrm{d}\,t$. We estimate the error bars from 10 realizations of the generated time series.
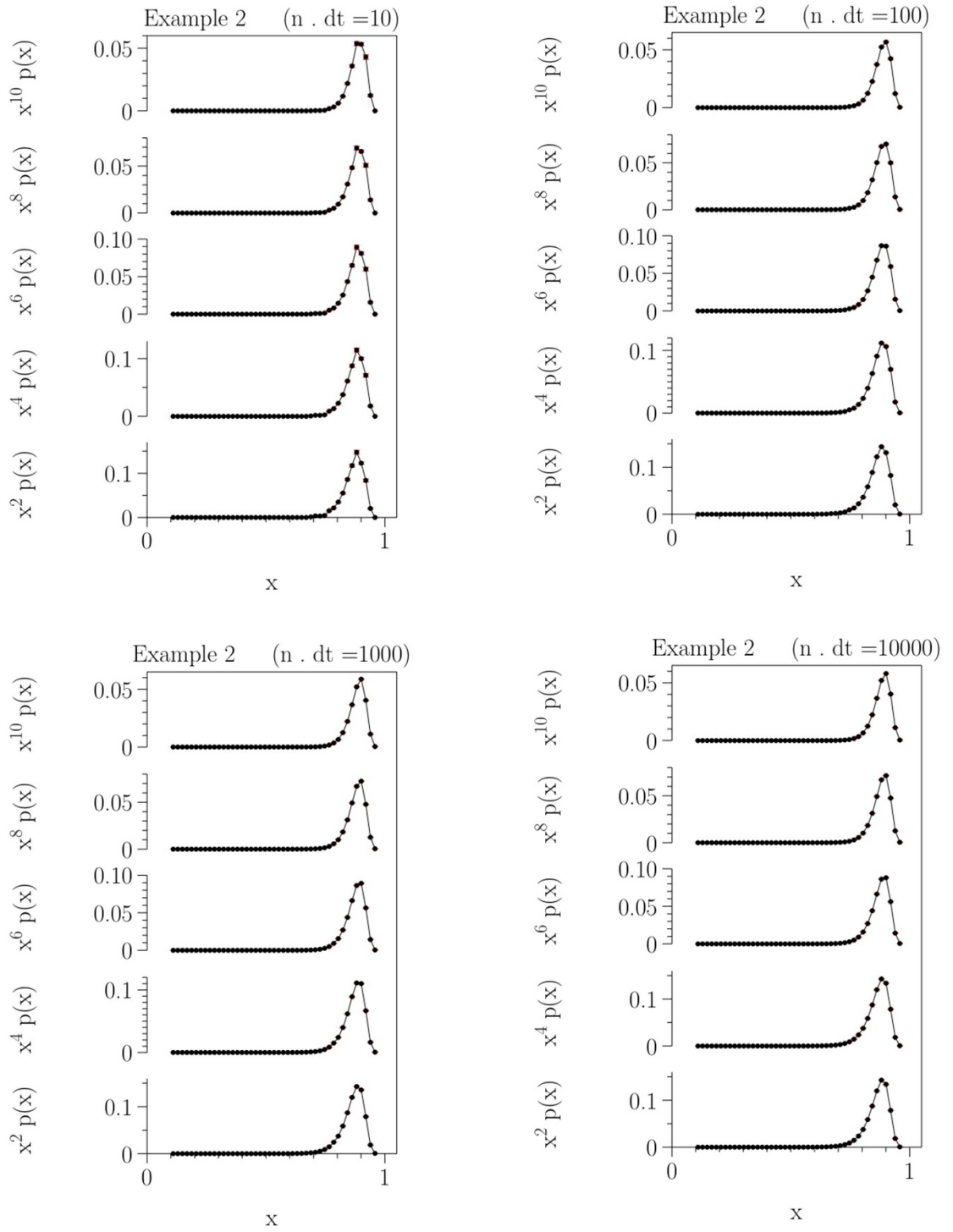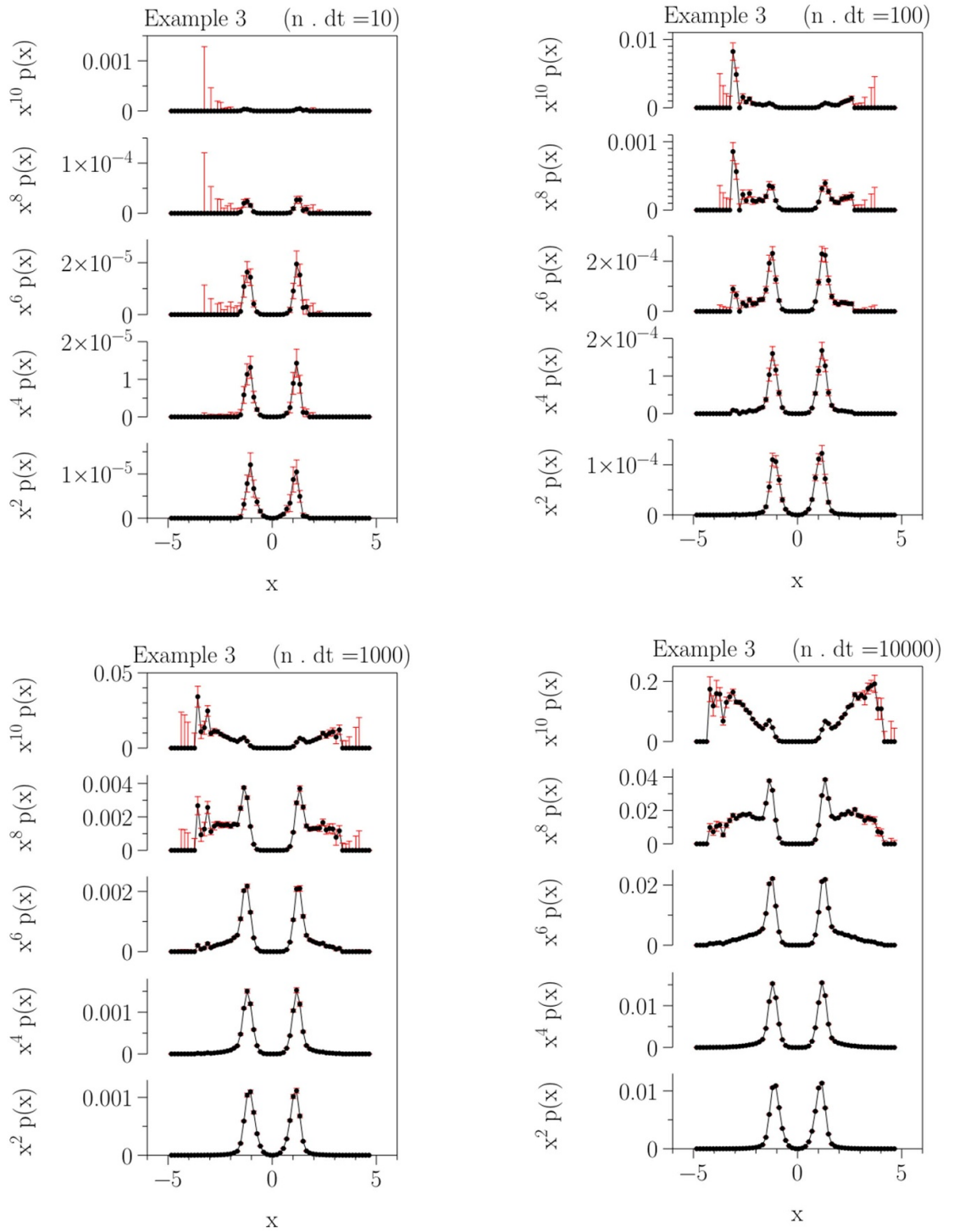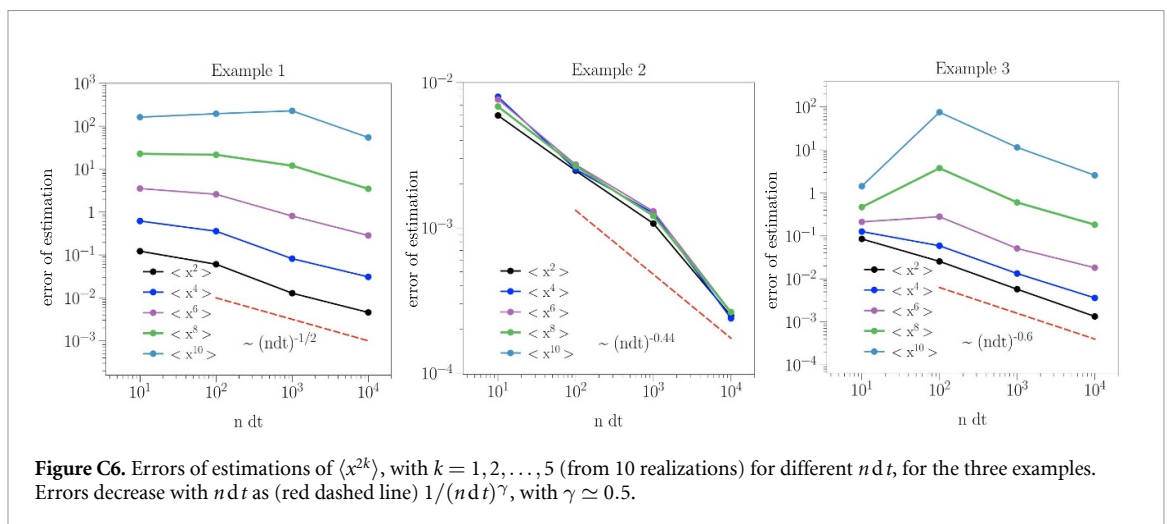
Figure C2. Same as figure C1 but for example 2.
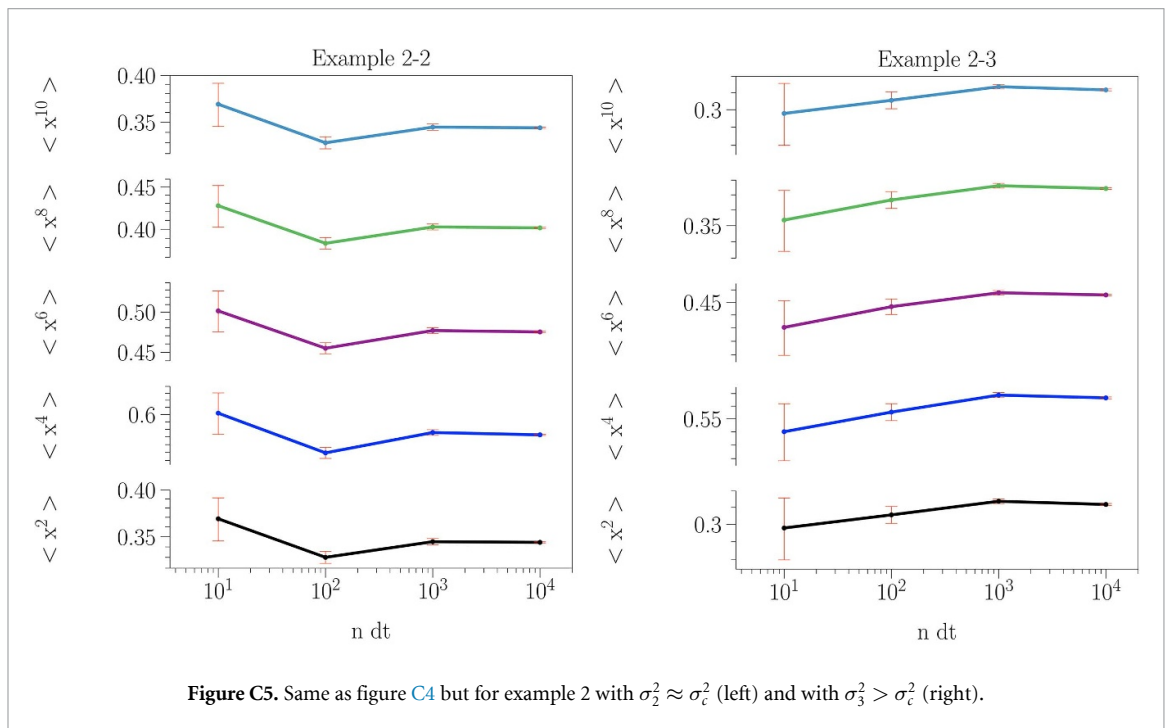
Figure C3. Same as figure C1 but for example 3.

Figure C4. Dependence of statistical moments $\langle x^{2k} \rangle$, with $k = 1, 2, \ldots, 5$ on $n\,d\,t$ from 10 realizations for various $n\,d\,t$ for the three examples of the main text. The middle panel is for example 2 with $\sigma_1^2 < \sigma_c^2$.

Figure C5. Same as figure C4 but for example 2 with $\sigma_2^2 \approx \sigma_c^2$ (left) and with $\sigma_3^2 > \sigma_c^2$ (right).

Figure C6. Errors of estimations of $\langle x^{2k} \rangle$, with $k = 1, 2, \ldots, 5$ (from 10 realizations) for different $n\,d\,t$, for the three examples. Errors decrease with $n\,d\,t$ as (red dashed line) $1/(n\,d\,t)^{\gamma}$, with $\gamma \simeq 0.5$.

Appendix D. The noisy genetic model

We consider a haploid group x as our data and suppose that each haploid may have genes A or B. The simple genotype selection model can be described as a one-variable deterministic differential equation [23]

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \alpha - x + \beta x(1-x), \tag{D.1}$$

where x is the ratio between the number of genes A and the total number of genes (A+B), so that $x \in [0, 1]$, and $\alpha \in (0, 1)$ stands for the mutation rate of gene B. Here, β denotes the gene selection factor, where $\beta = 1$ means that the selection is completely propitious to gene A haploid, whereas $\beta = -1$ implies that the selection is completely propitious to gene B haploid. Independent of α and β, the dynamical system (D.1) has only one equilibrium point in the corresponding interval given by

$$x_s = \frac{\beta - 1 + \sqrt{(\beta - 1)^2 + 4\alpha\beta}}{2\beta}, \tag{D.2}$$

which is stable. Now, consider the system (D.1) coupled to a noisy environment, with β fluctuating. We assume that $\beta(t) = \beta_0 + \sigma\eta$, where η is a white noise with unit intensity and σ is a real constant. The Langevin equation (noisy genetic model) is given by

$$\mathrm{d}x = [\alpha - x + \beta_0 x(1-x)]\,\mathrm{d}t + \sigma x(1-x)\,\mathrm{d}W(t) , \tag{D.3}$$

where $\mathrm{d}W(t) = \eta(t)\,\mathrm{d}t$ and $W(t)$ is a Wiener process. The noisy genetic model has two stationary points, that can be obtained by solving the corresponding Fokker–Planck (FP) equation, one of which coincides with the stationary point determined from the deterministic component (the drift coefficient) of the dynamics. The stationary solution of the FP equation of the Langevin dynamics (D.3) for $\alpha = 1/2$ is

$$p_s(x) = \frac{C}{x(1-x)} \exp\left[-\frac{1}{\sigma^2}\left\{\frac{1}{2x(1-x)} + \beta_0 \ln\left(\frac{1-x}{x}\right)\right\}\right] , \tag{D.4}$$

where C is a normalization constant.

There is a critical σ_c^2 in equation (D.4) such that for $\sigma^2 < \sigma_c^2$, the stationary solution of the FP equation, $p_s(x)$, is unimodal, i.e. it has only one extremum, namely a maximum. At $\sigma^2 = \sigma_c^2$, the probability distribution becomes flat, followed by a transition to a bimodal PDF for more intense noise. In fact, the model expressed by equation (D.3) exhibits a noise-induced transition with changing σ. The critical σ_c^2 depends on β_0. In order to have a non-trivial drift (non-linear) in (D.3), we consider the case $\beta_0 \neq 0$, say $\beta_0 = 3$. The critical noise intensity in this case is $\sigma_c^2 \simeq 2.84$. For $\sigma^2 < \sigma_c^2$ (we choose $\sigma^2 = 1$), $x \approx 0.885$ is the only extremum, which is a maximum. For $\sigma^2 > \sigma_c^2$ (we choose $\sigma^2 = 5$), the extrema are located at $x_1 \approx 0.023$, $x_2 \approx 0.435$, and $x_3 \approx 0.982$, with x_1 and x_3 each being a maximum and x_2 is the minimum.

Appendix E. Functions and parameters of the jump-diffusion process

For the jump-diffusion process, which is described by a dynamical stochastic equation (9) (in the main text), all functions and parameters can be determined based on measured time series by estimating the KM conditional moments [5, 24]:

$$K^{(1)}(x,t) = D^{(1)}(x,t)\tau$$
$$K^{(2)}(x,t) = [D(x,t) + \lambda(x,t)\langle\xi^2\rangle]\tau$$
$$K^{(j)}(x,t) = \lambda(x,t)\langle\xi^j\rangle\tau, \ \text{for } j > 2, \tag{E.1}$$

where the KM conditional moments are $K^{(l)}(x,t,\tau) = \langle [x(t+\tau) - x(t)]^l|_{x(t)=x}\rangle$, as given in equation (3) in the main text.

Using the last relation in equation (E.1) with $j = 4$ and $j = 6$, we first estimate the state-dependent (Gaussian) jump amplitude $\sigma_\xi^2(x)$ and jump rate $\lambda(x)$ via

$$\sigma_\xi^2(x,t) = \frac{M^{(6)}(x,t)}{5M^{(4)}(x,t)}, \qquad \lambda(x,t) = \frac{M^{(4)}(x,t)}{3\sigma_\xi^4(x,t)}, \tag{E.2}$$

where $M^{(l)}(x,t) = \lim_{\tau\to 0} \frac{K^{(l)}(x,t,\tau)}{\tau}$. We remind that $D^{(l)}(x,t)$ is related to $M^{(l)}(x,t)$ as $D^{(l)}(x,t) = M^{(l)}(x,t)/l!$.

Once the jump components $\sigma_\xi^2(x,t)$ and $\lambda(x,t)$ are identified, the second moment $M^{(2)}(x,t) = D(x,t) + \lambda(x,t)\langle\xi^2\rangle$ identifies the diffusion function $D(x,t)$, while the first moment yields the estimate for the drift function $D^{(1)}(x,t)$; see equation (E.1). In practice, we can also define the averaged jump amplitude and jump rate as $\overline{\sigma_\xi^2} = \frac{1}{N_r}\sum_{k=1}^{N_r}\sigma_\xi^2(x_k)$ and $\overline{\lambda} = \frac{1}{N_r}\sum_{k=1}^{N_r}\lambda(x_k)$, where N_r is the number of data points in some range of the state variable over which the coefficients are being evaluated. For stationary processes, we estimate the diffusion coefficient $D(x)$ from the relation $M^{(2)}(x) \simeq D(x) + \overline{\sigma_\xi^2}\,\overline{\lambda}$ [5]. We note that the jump rate per unit of time will be $\lambda_{\text{unit time}}(x) = \lambda(x)\,dt$.

ORCID iDs

Farnik Nikakhtar ⬤ https://orcid.org/0000-0002-3641-4366
Muhammad Sahimi ⬤ https://orcid.org/0000-0002-8009-542X
Klaus Lehnertz ⬤ https://orcid.org/0000-0002-5529-8559

References

[1] Risken H 1996 *The Fokker–Planck Equation* 2nd edn (Springer)
[2] Gardiner C W 1983 *Handbook of Stochastic Methods* (Springer)
[3] Friedrich R, Peinke J, Sahimi M and Tabar M R R 2011 Approaching complexity by stochastic methods: from biological systems to turbulence *Phys. Rep.* **506** 87
[4] Peinke J, Tabar M R R and Wächter M 2019 The Fokker–Planck approach to complex spatiotemporal disordered systems *Annu. Rev. Condens. Matter Phys.* **10** 107–32
[5] Tabar M R R 2019 *Analysis and Data-Based Reconstruction of Complex Nonlinear Dynamical Systems: Using the Methods of Stochastic Processes* (Springer) (https://doi.org/10.1007/978-3-030-18472-8)
[6] Nadaraya E A 1964 On estimating regression, theory of probability and its applications *Theory Probab. Appl.* **9** 141
[7] Watson G S 1964 Smooth regression analysis *Sankhya: Indian J. Stat.* A **26** 359–72 (available at: http://www.jstor.org/stable/25049340)
[8] Lamouroux D and Lehnertz K 2009 Kernel-based regression of drift and diffusion coefficients of stochastic processes *Phys. Lett.* A **373** 3507
[9] Kleinhans D and Friedrich R 2007 Maximum likelihood estimation of drift and diffusion functions *Phys. Lett.* A **368** 194
[10] Kleinhans D 2012 Estimation of drift and diffusion functions from time series data: a maximum likelihood framework *Phys. Rev.* E **85** 026705
[11] García C A, Otero A, Félix P, Presedo J and Márquez D G 2017 Nonparametric estimation of stochastic differential equations with sparse Gaussian processes *Phys. Rev.* E **96** 022104
[12] Hindriks R, Bijma F, van Dijk B W, van der Werf Y D, van Someren E J W and van der Vaart A W 2011 Dynamics underlying spontaneous human alpha oscillations: a data-driven approach *Neuroimage* **57** 440
[13] Kassel J A and Kantz H 2022 Statistical inference of one-dimensional persistent nonlinear time series and application to predictions *Phys. Rev. Res.* **4** 013206
[14] Davis W and Buffett B 2022 Estimation of drift and diffusion functions from unevenly sampled time-series data *Phys. Rev.* E **106** 014140
[15] Gorjão L R, Heysel J, Lehnertz K and Tabar M R R 2019 Analysis and data-driven reconstruction of bivariate jump-diffusion processes *Phys. Rev.* E **100** 062127
[16] Aslim E, Rings T, Zabawa L and Lehnertz K 2021 Enhancing the accuracy of a data-driven reconstruction of bivariate jump-diffusion models with corrections for higher orders of the sampling interval *J. Stat. Mech.* 033406
[17] Tabar M R R, Nikakhtar F, Parkavosi L, Akhshi A, Feudel U and Lehnertz K submitted
[18] Lehnertz K, Zabawa L and Tabar M R R 2018 Characterizing abrupt transitions in stochastic dynamics *New J. Phys.* **20** 113043
[19] Rackauckas C and Nie Q 2017 Adaptive methods for stochastic differential equations via natural embeddings and rejection sampling with memory *Discrete Contin. Dyn. Syst.* B **22** 2731
[20] Rößler A 2010 Runge-Kutta methods for the strong approximation of solutions of stochastic differential equations *SIAM J. Numer. Anal.* **48** 922
[21] Pawula R F 1967 Approximation of the linear Boltzmann equation by the Fokker–Planck equation *Phys. Rev.* **162** 186
[22] Wio H S and Lindenberg K 2003 Noise induced phenomena: a sampler *AIP Conf. Proc.* **658** 1
[23] Arnold L, Horsthemke W and Lefever R 1978 White and coloured external noise and transition phenomena in nonlinear systems *Z. Phys.* B **29** 367
[24] Anvari M, Tabar M R R, Peinke J and Lehnertz K 2016 Disentangling the stochastic behavior of complex time series *Sci. Rep.* **6** 35435
[25] Giraudo M T and Sacerdote L 1997 Jump-diffusion processes as models for neuronal activity *Biosystems* **40** 75
[26] Sirovich R, Sacerdote L and Villa A E P 2014 Cooperative behavior in a jump diffusion model for a simple network of spiking neurons *Math. Biosci. Eng.* **11** 385
[27] Daly E and Porporato A 2006 Probabilistic dynamics of some jump-diffusion systems *Phys. Rev.* E **73** 026108
[28] Cont R and Tankov P 2004 *Financial Modelling With Jump Processes* (Chapman and Hall)
[29] Stanton R 1997 A nonparametric model of term structure dynamics and the market price of interest rate risk *J. Financ.* **52** 1973–2002
[30] Johannes M 2004 The statistical and economic role of jumps in continuous-time interest rate models *J. Financ.* **59** 227
[31] Bandi F M and Nguyen T 2003 On the functional estimation of jump-diffusion models *J. Econom.* **116** 293
[32] Madan D and Seneta E 1990 The variance gamma (V.G.) model for share market returns *J. Bus.* **63** 511
[33] Aït-Sahalia Y and Jacod J 2009 Testing for jumps in a discretely observed process *Ann. Stat.* **37** 184
[34] Lee S and Mykland P A 2008 Jumps in financial markets: a new nonparametric test and jump clustering *Rev. Financ. Stud.* **21** 2535