Journal of Physics:
Complexity

Assessing the suitability of the Langevin equation for analyzing measured data through downsampling

Journal of Physics: Complexity

OPEN ACCESS

RECEIVED
3 July 2024

REVISED
24 February 2025

ACCEPTED FOR PUBLICATION
28 February 2025

PUBLISHED
17 March 2025

PAPER

Assessing the suitability of the Langevin equation for analyzing measured data through downsampling

Pyei Phyo Lin[1,4], **Matthias Wächter**[1,*], **Joachim Peinke**[1] and **M Reza Rahimi Tabar**[1,2,3]

[1] ForWind, Institute of Physics, University of Oldenburg, Oldenburg, Germany
[2] Theoretical Physics/Complex Systems, ICBM, University of Oldenburg, Oldenburg, Germany
[3] Department of Physics, Sharif University of Technology, Tehran 11155-9161, Iran
[4] Institute for Fluid Dynamics and Ship Theory, Hamburg University of Technology, Hamburg, Germany
[*] Author to whom any correspondence should be addressed.

E-mail: matthias.waechter@uol.de, pyei.phyo.lin@uol.de, peinke@uol.de and tabar@uol.de

Keywords: stochastic processes, Langevin equation, jump–diffusion process, generalized Langevin equation

Abstract

The measured time series from complex systems are renowned for their complex stochastic behavior, characterized by random fluctuations stemming from external influences and nonlinear interactions. These fluctuations take diverse forms, ranging from continuous trajectories reminiscent of Brownian motion to noncontinuous trajectories featuring jump events. The Langevin equation is a versatile framework for modeling stochastic systems, effectively describing the complex behavior of measured data that exhibit continuous stochastic variability and adhere to Markov properties. However, the traditional modeling framework of the Langevin equation falls short when it comes to capturing the presence of abrupt changes, particularly jumps, in trajectories that exhibit non-continuity. Such non-continuous changes pose a significant challenge for general processes and have profound implications for risk management. Moreover, the discrete nature of observed physical phenomena, measured with a finite sample rate, adds another layer of complexity. In such cases, data points often appear as a series of discontinuous jumps, even when the underlying trajectory is continuous. In this study, we present an analytical framework that goes beyond the limitations of the Langevin equation. Our approach effectively distinguishes between diffusive or Brownian-type trajectories and non-diffusive trajectories such as those with jumps. By introducing downsampling techniques, where we artificially lower the sample rate, we derive a set of measures and criteria to analyze the data and differentiate between diffusive and non-diffusive behaviors. To further demonstrate its versatility and practical applicability, we have applied our proposed method to real-world data in various scientific fields, such as trapped particles in optical tweezers, market price, neuroscience, turbulence and renewable energy. For real-world data that lack Markov properties, we estimate the functions and parameters using the generalized Langevin equation, which incorporates a memory kernel to account for non-Markovian dynamics.

1. Introduction

In a complex world, empirical data ranging from natural sciences (e.g. physical, climatological, biological and ecological systems) to social sciences (e.g. infectious diseases spread, finance and economics) can be characterized by the stochastic dynamics. A critical assumption for such data is that they belong to the class of Markov processes, where the future state depends only on the present state and not on the past history. The classical method to approach such systems is to approximate the data with the Langevin equation or diffusion process [1–3]. The Langevin equation is a stochastic differential equation which is driven by Gaussian white noise generating Markov and continuous (Brownian type) stochastic process [4, 5]. Recent studies have highlighted the presence of jumps, which are discontinuous or non-Brownian events, in complex systems. These jumps can have significant implications for risk management and have received considerable attention in recent years [6–10]. Jumps introduce higher uncertainties in the stochastic features

of the underlying system and contribute to the non-Gaussian characteristics of the increment statistics in short-time scales [11].

In the presence of discontinuous events, the traditional modeling framework of the Langevin equation, which generates only continuous trajectories, is inadequate for capturing these phenomena [5]. To address this limitation, an effective approximation is to introduce an additional noise term known as jump-noise in the Langevin equation with distributed jumps in time and amplitude. This gives rise to jump–diffusion processes [5, 7, 11–13]. In this framework, the jump noise represents the discontinuous paths within the diffusion process.

The nature of experimental data is such that it is often measured with limited sampling time or spatial resolution. As a result, the recorded trajectory may appear discontinuous, despite the underlying process being continuous. Furthermore, certain real-world datasets, such as those related to stock market indices, glaciers, and sea levels, exhibit distinct trajectories.

In real-world datasets, it is unlikely that they strictly follow Markov properties that Langevin or jump–diffusion dynamics are assumed to be Markov processes. Normally they are rarely Markovian and may exhibit finite Markov–Einstein time scales t_{M}. This time scale represents the minimum interval during which the data can be effectively treated as a Markov process. For such data, we should consider alternative modelling approaches, such as fractional Klein–Kramers equations [14], fractional diffusion processes [15, 16] and Lévy-driven Langevin dynamics [5].

One standard approach to studying such non-Markov is to use the generalized Langevin equation (GLE) [17, 18], which exhibits memory kernel originated from non-Markovian properties. The GLE is derived through perturbation theory, linear response theory, and projector operator formalism [19–22].

Here, we present an analytical framework that extends the Langevin equation to model both continuous diffusive and non-diffusive trajectories, including those with jumps, while maintaining the Markov property. By applying downsampling techniques, the method differentiates between these behaviors. The approach is demonstrated on real-world data from fields such as trapped particles in optical tweezers, market price analysis, neuroscience, turbulence and renewable energy. For data lacking Markov properties, the GLE is used, incorporating a memory kernel to account for non-Markovian dynamics.

This paper is organized as follows. Firstly, we present the criteria to distinguish the diffusive and non-diffusive (jumpy) nature of the simulated stochastic processes with different integration time steps. Then, we discuss the applicability of these criteria to analyze the simulated data with a finite sampling time by means of downsampling and their behavior at different downsampling time scales. Finally, we test these criteria on the real-world experimental data. All derivations and proofs are presented in the appendices.

2. Diffusive and non-diffusive behaviors

A given Markov time series typically consists of three independent time scales: the sampling interval τ_{s}, the correlation timescale T_{C}, and the average timescale between jumps T_{J} (if present). The timescales T_{C} and T_{J} can be estimated using Kramers–Moyal (KM) coefficients of order one, four and six, see below [11]. For time series $x(t)$, KM coefficients of order n, $K^{(n)}(x,t)$ are given by

$$K^{(n)}(x,t) = \lim_{\tau_{\mathrm{s}}\to 0} \frac{M^{(n)}(x,t,\tau_{\mathrm{s}})}{\tau_{\mathrm{s}}} \tag{1}$$

where $M^{(n)}(x,t,\tau_{\mathrm{s}})$ are KM conditional moments, defined as

$$M^{(n)}(x,t,\tau_{\mathrm{s}}) = \int \mathrm{d}x' \,(x'-x)^n \, p(x', t+\tau_{\mathrm{s}}|x,t) \ . \tag{2}$$

Here, τ_{s} represents the sampling time interval of the measured or simulated time series.

The Nadaraya–Watson estimator [23, 24], which is a kernel estimator, can be used to estimate the conditional moments such that

$$M^{(n)}(x,t,\tau_{\mathrm{s}}) = \frac{\sum_i k\left(\frac{x_{i\tau_{\mathrm{s}}}-x}{h}\right)\left(x_{(i+1)\tau_{\mathrm{s}}}-x_{i\tau_{\mathrm{s}}}\right)^n}{\sum_i k\left(\frac{x_{i\tau_{\mathrm{s}}}-x}{h}\right)}. \tag{3}$$

Here, we use a Gaussian kernel $k(\cdot)$, defined as $k(x) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)$, where h is the bandwidth or smoothing parameter. The kernel-based method enables the real-time calculation of conditional moments h [25, 26]. For our analyses, we have selected a kernel bandwidth of $h = 0.1$.

In definition (1), we omitted the $n!$ term in the denominator to simplify the expressions that allowing us to distinguish between diffusive and non-diffusive processes.

For a diffusive process that generates a continuous trajectory like Brownian motion [5], the dynamics of a state variable $x(t)$ can be described by a Langevin equation that includes both deterministic and stochastic contributions, given by [27]:

$$\mathrm{d}x = D^{(1)}(x,t)\,\mathrm{d}t + \sqrt{D^{(2)}(x,t)}\,\mathrm{d}W_t\,, \tag{4}$$

where $D^{(1)}(x,t)$ (drift) and $D^{(2)}(x,t)$ (diffusion coefficient) are given functions of the state variable x and time t. They are expressed in terms of the first and second KM coefficients as $D^{(1)}(x,t) = K^{(1)}(x,t)$ and $D^{(2)}(x,t) = K^{(2)}(x,t)$. For a diffusive process, all higher-order KM coefficients are zero, i.e. $K^{(n)}(x,t) = 0$ for $n \geqslant 3$, as demonstrated by Pawula [28]. In equation (4), the increment of Wiener process is defined by $\mathrm{d}W_t = \eta(t)\mathrm{d}t$ where $\eta(t)$ represents a zero-mean Gaussian white noise with unit intensity, satisfying $\langle \eta(t)\eta(t')\rangle = \delta(t-t')$.

In practice, verifying the condition $K^{(n)}(x,t) = 0$ for $n \geqslant 3$ is not straightforward, as it requires knowledge of the expansions of KM conditional moments $M^{(n)}(x,t,\tau_s)$ in terms of the sampling time τ_s. By employing the expansion of KM conditional moments, it has been demonstrated that the necessary condition for a process to belong to the diffusive category is $M^{(4)}(x,t,\tau_s) \simeq 3(M^{(2)}(x,t,\tau_s))^2$ [29]. In an alternative representation, we define $\Theta(x,t,\tau_s)$ as:

$$\Theta(x,t,\tau_s) = 3\frac{\left(M^{(2)}(x,t,\tau_s)\right)^2}{M^{(4)}(x,t,\tau_s)}\,. \tag{5}$$

If $\Theta(x,t,\tau_s) = 1$, it indicates that the data fulfill the necessary condition (criterion) for belonging to the diffusion processes. This criterion has been examined for various linear and nonlinear drift and diffusion coefficients, and we have observed $\Theta(x,t,\tau_s) = 1$; for more details, refer to appendix appendix A.

A diffusive process has a continuous sample path [5]. However, in reality, some processes exhibit sudden changes that resemble discontinuous sample paths. These discontinuities can be effectively modeled as jump processes. A straightforward approach involves adding an additional compound Poisson jump term to the Langevin equation, resulting in a jump–diffusion process. This approach can be expressed as [11]:

$$\mathrm{d}x = D^{(1)}(x,t)\,\mathrm{d}t + \sqrt{D^{(2)}(x,t)}\,\mathrm{d}W_t + \xi\,\mathrm{d}J_t\,. \tag{6}$$

In this model, we consider the noise $\xi \sim \mathcal{N}(0,\sigma_\xi^2)$, which represents the size of jumps and follows a normal distribution with variance σ_ξ^2. The jump amplitude is also defined as σ_ξ^2 and can be state- and time-dependent. Additionally, the Poisson distributed jump process $J_t \sim P(\lambda t)$ is present, which is a zero-one jump process with a jump rate of $\lambda(x,t)$. Indeed, the final term in equation (6) simulates abrupt changes (jumps) in the trajectories of $x(t)$ with varying amplitudes. In the jump–diffusion dynamical equation (6), we assume that jump events are rare and can be modeled via a Poisson process.

Two unknown functions, jump rate $\lambda(x,t)$ and jump amplitude, σ_ξ^2 can be found using the relations [5, 11],

$$\sigma_\xi^2(x,t) = \frac{K^{(6)}(x,t)}{5K^{(4)}(x,t)}, \qquad \lambda(x,t) = \frac{K^{(4)}(x,t)}{3\sigma_\xi^4(x,t)}\,. \tag{7}$$

For jump–diffusion processes described by equation (6), the expansion of KM conditional moments can be performed, allowing for the definition of a function $Q(x,t,\tau_s)$ (referred to as the Q-criterion, [29]). This function is given by (see appendix A.1),

$$Q(x,t,\tau_s) = \frac{M^{(6)}(x,t,\tau_s)}{5M^{(4)}(x,t,\tau_s)}\,. \tag{8}$$

The function $Q(x,t,\tau_s)$ serves as a necessary criterion for identifying the presence of jumps in a given time series, assuming that the underlying processes are governed by jump–diffusion dynamics.

For small values of $\tau_s(\ll 1)$ in the functions $\Theta(x,t,\tau_s)$ and $Q(x,t,\tau_s)$, which are obtained from second, fourth, and sixth-order KM conditional moments, we observe distinct behavior for three types of time series (see appendix A for the derivation), as

$$\Theta(x,t,\tau_s) = \begin{cases} 1, & \text{diffusive} \\ \dfrac{\left(D^{(2)}(x,t)+\lambda(x,t)\sigma_\xi^2(x,t)\right)^2}{\lambda(x,t)\sigma_\xi^4(x,t)}\tau_s, & \text{jumpy} \\ 1, & \text{white noise at } x=0, \end{cases} \tag{9}$$

and

$$Q(x,t,\tau_s) = \begin{cases} D^{(2)}(x,t)\,\tau_s, & \text{diffusive,} \\ \sigma_\xi^2(x,t), & \text{jumpy,} \\ \sigma_\eta^2, & \text{white noise at, } x=0\,. \end{cases} \tag{10}$$

Here, the behaviors of $\Theta(x,t,\tau_s)$ and $Q(x,t,\tau_s)$ for various processes, including diffusive processes (Langevin dynamics), jump–diffusion processes and pure white noise are given. It is worth noting that the requirement for $x=0$ is needed only in the case of Gaussian white noise. For all other situations, the aforementioned relations hold for any value of x. Also σ_η^2 is the variance of white noise.

The function $\Theta(x,t,\tau_s)$ exhibits a linear behavior with respect to τ_s for jump–diffusion processes, while it remains unity for diffusive and white noise trajectories (at $x=0$). Simultaneously, the function $Q(x,t,\tau_s)$ has a constant value of $\sigma_\xi^2(x,t)$ for small τ_s in the case of jump–diffusion processes. However, for diffusive processes, it displays a linear relationship with τ_s, specifically given by $D^{(2)}(x,t)\tau_s$.

In appendix A.3, we performed the numerical integration of the corresponding Langevin and jump–diffusion dynamics for various values of τ_s. This analysis allowed us to examine the behaviors of $\Theta(x,t,\tau_s)$ and $Q(x,t,\tau_s)$ as functions of τ_s for diffusive processes, jump–diffusion processes, and pure uncorrelated (white) noise. Using the Θ- and Q-criteria, we can effectively identify whether a stochastic time series generated from specific dynamical equations originates from a diffusive (continuous) process or a jumpy (discontinuous) process [29].

3. Limitations in the applicability of the $\Theta(x,t,\tau_s)$ and $Q(x,t,\tau_s)$ criteria

The results presented above for $\Theta(x,t,\tau_s)$ and $Q(x,t,\tau_s)$ are derived from the expansion of KM coefficients for Langevin and jump–diffusion dynamical equations with varying time steps, τ_s. To make them applicable to real-world time series, especially those displaying unique trajectories, their expressions require modification.

In the following, we demonstrate that the obtained behaviors of $Q(x,t,\tau)$ hold true for time scales τ_s that are shorter than both the correlation timescale T_C of the time series and the average timescale between jumps T_J. However, when dealing with a real-world time series sampled with τ_s, it is important to note that we typically lack direct access to the ground truth values of T_C and especially T_J. Consequently, assessing the behaviors of $\Theta(x,t,\tau_s)$ and $Q(x,t,\tau)$ can potentially lead to misleading conclusions. Therefore, in this study, our aim is to investigate this matter more deeply to gain a thorough understanding of the implications and limitations associated with using these criteria in real-world scenarios.

Based on a given dynamical equation, it is possible to generate a set of stochastic processes with different time scales T_C and T_J in order to check the Θ- and Q-criteria for the corresponding time series for different τ_s. In order to verify these criteria using synthetic data, we conducted numerical integration of a Langevin and jump–diffusion dynamical equations with specific parameters.

Example 1: the drift coefficient is given by $D^{(1)}(x,t) = -\gamma x$, where $\gamma = 1\ \mathrm{s}^{-1}$. The diffusion coefficient is denoted as $D^{(2)}(x,t) = D$, with $D = 1\ \mathrm{s}^{-1}$. It is also known as Ornstein-Uhlenbeck (OU) process.

Example 2: additionally, we considered an additional jump term (OU jump–diffusion process), characterized by a jump rate of $\lambda = 100\ \mathrm{s}^{-1}$ and a variance of $\sigma_\xi^2 = 0.01$.

In the cases involving pure noisy data, we employed normal Gaussian white noise with a variance of $\sigma_\eta^2 = 1$. The chosen parameters in the Langevin and jump–diffusion dynamics yield two timescales. The correlation timescale is given by $T_C = \frac{1}{\gamma} = 1$ s, and the average timescale between jumps is denoted as $T_J = \frac{1}{\lambda} = 0.01$ s.

From the simulated data with different time step τ_s-values, we estimate $\Theta(x,\tau=\tau_s)$ and $Q(x,\tau=\tau_s)$ as shown in figure 1. Theoretical predictions as described in equations (9) and (10) are also plotted with the solid lines. As shown in figure 1, for jump–diffusion process, equation (10) is valid only for $\tau_s < \frac{1}{\lambda} = T_J$. For small τ_s as shown in figure 1, all those estimates are rather close to the preset values and emphasize the accuracy of the predictions of Θ and Q-criteria for four types of the synthetic data. The subplot (ii) in figure 1 exhibits a notable departure from the expected theoretical prediction of the KM conditional moments' expansion for the jump–diffusion process, specifically observed in the case of $\tau_s > T_J$. In fact, for $\tau_s > T_J$, OU jump-diffusion process behaves like OU (diffusion) process with the apparant diffusion coefficient $\widetilde{D}$ which has the relation $\widetilde{D}\,\tau_s = D\,\tau_s + \sigma_\xi^2$. The proof can be seen in appendix A.3. To conclude this section, we find that the behaviors reported in equations (9) and (10) are valid only for $\tau_s < T_J$ and T_C. It is essential to consider these constraints when applying the criteria to real-world time series data.

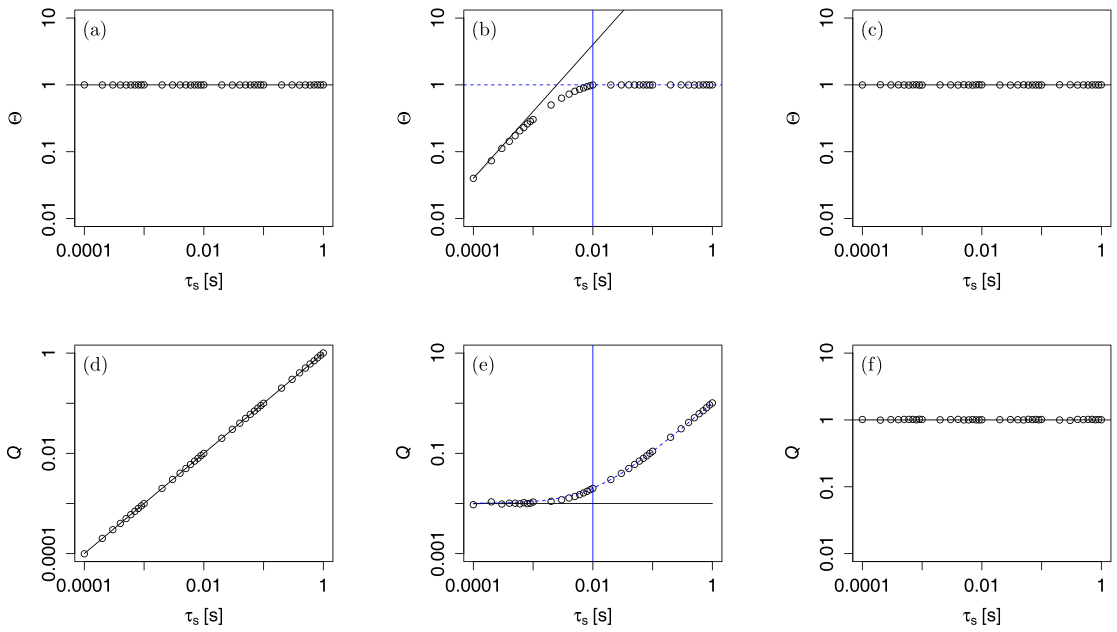

Figure 1. $\Theta = \Theta(x = 0, \tau = \tau_s)$ and $Q = Q(x = 0, \tau = \tau_s)$ are calculated for OU (diffusion) process as described in example 1, (a) and (d), OU jump–diffusion process as in example 2, (b) and (e), and standard normal Gaussian white noise, (c) and (f), are evaluated for each τ_s and plotted in double logarithmic scale. For the evaluation, OU processes are generated for different τ_s. The theoretical predictions according to equations (9) and (10) are presented with black solid lines. The blue vertical line is the time scale $T_J = \frac{1}{\lambda}$. The blue dashed curve $Q = D\,\tau_s + \sigma_\xi^2$ and line $\Theta = 1$ show the behavior after $\tau_s > T_J$. Standard error of the mean are shown with shadows, which is smaller than the symbol size in these cases.

4. Downsampling of the data

To analyze data that is originally sampled with a finite time step τ_s and data with unique trajectories, the only practical approach to study diffusive and non-diffusive behaviors is downsampling. When downsampling data, we encounter situations where the sampling interval τ_{ds} (which belongs to the set $\in [\tau_s, 2\tau_s, \ldots]$) can be larger than both the timescales T_J and T_C, or it may lie between these two timescales.

Initially, we focus on diffusive processes that are governed by Langevin equations. In appendix B, we provide a proof that downsampling time series generated from a Langevin equation does not alter the behavior of $\Theta(x, t, \tau_{ds})$. This is a key result of our study. As a consequence, diffusive processes maintain their diffusive nature after downsampling. The main concept is that the Wiener process for the downsampled time series remains a Wiener process, with the only difference being the change from τ_s to τ_{ds}.

However, downsampling behaves differently for processes that exhibit jump discontinuities. Let us consider the case where time scales have the property $T_J < T_C$. To this end, we note that on average, there are $\lambda\tau_{ds}$ jump events within each downsampled step of τ_{ds}. Consequently, within each step, we observe multiple jumps with different amplitudes. The average jump amplitude for the downsampled time interval τ_{ds} is given by (a proof is provided in appendix B):

$$\sigma^2_{\xi,\mathrm{eff}} = \sum_{i=1}^{\lambda\tau_{ds}} \sigma_\xi^2 = \sigma_\xi^2\,\lambda\tau_{ds}, \tag{11}$$

where λ represents the jump rate of the original time series (before downsampling).

From equation (11), we can observe that $\sigma^2_{\xi,\mathrm{eff}}$ now exhibits a linear dependence on τ_{ds}. As a result, it becomes challenging to distinguish the jump discontinuity process from the diffusive process using the Q-criterion presented in equation (10), which we have shown was applicable for small τ_s.

As shown in appendix B, similar to jump events, the Gaussian noise η in the diffusion part of the jump–diffusion process also accumulates over each time step within the downsampled interval τ_{ds}. This results in an effective variance given by $\sigma^2_{\eta,\mathrm{eff}} = \sum_{i=1}^{N_\tau} \sigma_\eta^2 = N_\tau = \frac{\tau_{ds}}{\tau_s}$, where $\sigma_\eta^2 = 1$ represents the unit variance of η, and τ_s is the sampling time step of the original data. As a result, the effective diffusion coefficient $D^{(2)}_{\mathrm{eff}}(x,t)$ satisfies the relation $D^{(2)}_{\mathrm{eff}}(x,t)\tau_{ds} = D^{(2)}\tau_s N_\tau$, which implies that the effective diffusion coefficient remains the same as the original diffusion coefficient ($D^{(2)}_{\mathrm{eff}}(x,t) = D^{(2)}(x,t)$). A detailed proof can

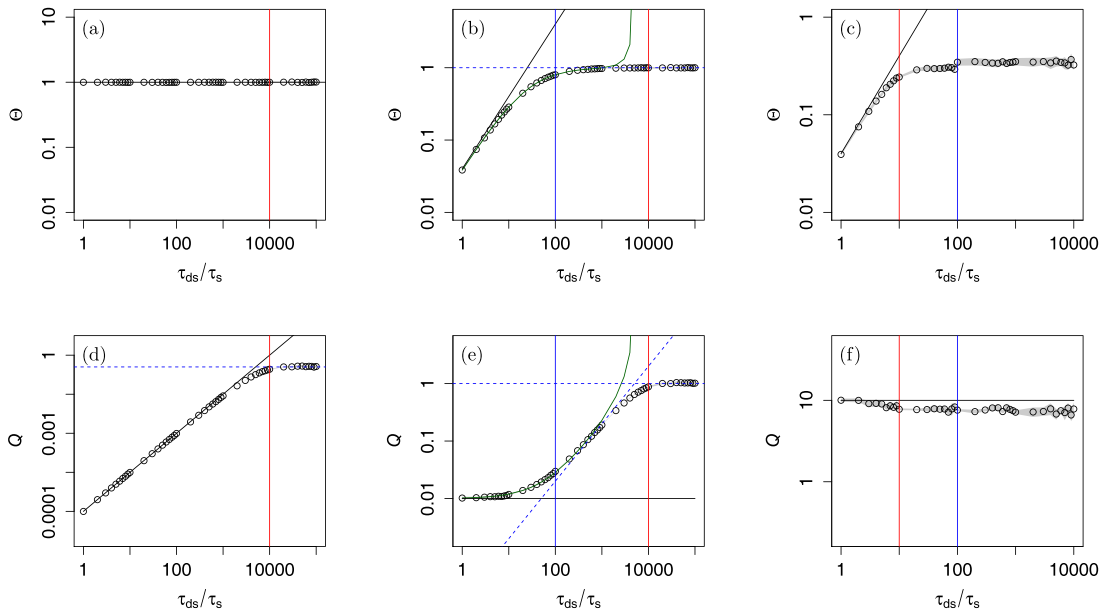

Figure 2. $\Theta = \Theta(x = 0, \tau = \tau_s)$ and $Q = Q(x = 0, \tau = \tau_s)$ are calculated for the downsampled OU (diffusion) process as described in example 1, (a) and (d), downsampled OU jump–diffusion process for $T_J < T_C$ as in example 2, (b) and (e), and $T_J > T_C$ as in example 3, (c) and (f), are evaluated for each downsampled period τ_s and plotted in double logarithmic scale. The theoretical predictions according to equations (9) and (10) are presented with black solid lines, and equation (12) in (e), $\Theta = 1$ in (b) and $Q = \sigma_x^2$ in (d) and (e) with blue dashed lines. The blue vertical line is the timescale $T_J = \frac{1}{\lambda}$ and the red vertical line $T_C = \frac{1}{\gamma}$. Standard error of the mean are shown with shadows, which is smaller than the symbol size in these cases. The green solid lines in (b) and (e) the approximation up to $\mathcal{O}(\tau^4)$ of conditional moments while calculating Θ and Q in equations (9) and (10) as described in appendix A.1.

be found in the appendix B. Thus, the effective function Q for the jump–diffusion in the range of $T_J < \tau_{ds} < T_C$ can be written as

$$Q_{\text{eff}} = D^{(2)}\, \tau_{ds} + \lambda \sigma_\xi^2\, \tau_{ds}\,. \tag{12}$$

Therefore, in equation (10), the condition for jumpy time series changes from a constant σ_ξ^2 to a linear relation with τ_{ds}, given by $(D^{(2)} + \lambda\sigma_\xi^2)\tau_{ds}$. If $\tau_{ds} > T_C$, the downsampled data will resemble white noise. In this case, the Q-criterion will have a value equal to the variance of the resulting time series. We also note that in the case of $T_C < T_J$, equations (9) and (10) are valid for $\tau_{ds} < T_C$.

In order to check the aforementioned expressions, i.e. equations (9), (10) and (12) numerically, we generate OU process and OU jump–diffusion process with $\gamma = 1\ \text{s}^{-1}$, $D = 1\ \text{s}^{-1}$, $\lambda = 100\ \text{s}^{-1}$ and $\sigma_\xi^2 = 0.01$ as described in examples 1 and 2. In example 2, $T_J < T_C$.

Examples 3: for the case of $T_J > T_C$, we generate additional OU jump–diffusion process with $\gamma = 1000\ \text{s}^{-1}$, $D = 1000\ \text{s}^{-1}$, $\lambda = 100\ \text{s}^{-1}$ and $\sigma_\xi^2 = 10$.

All the samples are simulated with the integrating time step $\tau_s = 10^{-4}$ s. The processes are realized with different time-lags τ_{ds} using downsampling, and then the Θ- and Q-criterion are calculated. To estimate the uncertainties, we computed $\Theta = \Theta(x = 0, \tau_{ds})$ and $Q = Q(x = 0, \tau_{ds})$ across multiple realizations of the process, and subsequently calculated their mean values and standard deviations. The results are plotted in figure 2. The blue vertical line is the timescale $T_J = \frac{1}{\lambda}$ and the red vertical line $T_C = \frac{1}{\gamma}$.

For diffusion process, example 1, we can observe the same diffusive behavior for $\tau_{ds} < T_C = \frac{1}{\gamma}$ as described in equations (9) and (10) shown by black solid lines in figures 2(a) and (d). For downsampled time steps larger than correlation length, $\tau_{ds} > T_C$, the results of the criteria shows the behavior of uncorrelated Gaussian white noise in which $\Theta = 1$ and $Q = \sigma_x^2$ where $\sigma_x^2 = 0.5$ of OU process x, shown by blue horizontal dashed in figure 2(d) for this case.

In the first case of jump–diffusion process, example 2, where $T_J < T_C$ shown in figures 2(b) and (e), we can see that the approximations in equations (9) and (10) are still valid for $\tau_{ds} \ll T_J$. (If we consider calculating Θ and Q in equations (9) and (10) up to $\mathcal{O}(\tau^4)$ of conditional moments as described in appendix A.1, we can have a better approximation shown in solid green line). In the range of $T_J < \tau_{ds} < T_C$, the effective value of Q becomes the sum of effective diffusion and jump terms described in equation (12), and Θ also becomes unity after a few time steps in this range which apparently show the diffusive behavior.

These are shown with the blue dashed lines. After $\tau_{ds} > T_C$, the downsampled process becomes uncorrelated with Gaussian distribution, where $\Theta = 1$ and Q become the variance of the process x, $Q = \sigma_x^2 = 1$. The variance of OU jump-diffusion process can also be calculated analytically with equation (C.12) derived in appendix C.

In the second case of jump–diffusion process, example 3, where $T_J > T_C$ shown in figures 2(c) and (f), we can see that the approximations in equations (9) and (10) are still valid for $\tau_{ds} \ll T_C$ (shown by black solid lines). For $\tau_{ds} > T_J$, Q and Θ become constant.

Aa a remark, to assess the robustness of our results against finite-order time steps Δt, we apply the Itô–Taylor expansion method introduced by [30]. This approach reduces errors in estimating KM coefficients for finite Δt. Our analysis shows only a few percent deviations in KM coefficient estimates as the expansion order increases from first to third order for finite-order time corrections.

5. Statistical tests

Before proceeding with real-world time series, we present two statistical tests and evaluate the confidence levels in the estimations of the drift term and the overall stochastic behavior across the entire dynamics for all examples (1–3). Here, we provide two statistical tests to evaluate the accuracy of the deterministic and stochastic components of dynamics.

We elaborate on how we estimated the accuracy of our method at various stages of the estimation procedure, with data with length $N\Delta t$ [31]. Specifically:

- We calculated the coefficient of determination (R^2) to evaluate the correspondence between the true drift functions $D^{(1)}(x,t)$ and the estimated drift functions $\widetilde{D^{(1)}}(x,t)$. This comparison was based on simulated $x(t)$ with various integration times $T = N\Delta t$.
- We considered the entire dynamics, including both the deterministic drift term $D^{(1)}(x,(t))$ and the stochastic components, such as the diffusion terms in the Langevin equation and the diffusion and jump terms in jump–diffusion processes. To evaluate the similarity, we compared the cumulative distribution functions (CDFs) of time series generated from the true functions and parameters with the CDFs of time series generated using the estimated functions and parameters. To achieve this, we employed a two-sample Kolmogorov–Smirnov (KS) test. Also, we present the KS statistics along with the corresponding significance levels.

5.1. Comparing true and estimated drift terms: coefficient of determination

In cases where we have access to ground truth in examples (1–3), we can assess the similarity between two drift functions – $D^{(1)}(x,t)$, which is based on known (preset) coupling strengths and parameters and $\widetilde{D^{(1)}}(x,t)$, which we estimate using our method – by calculating the coefficient of determination (R^2-score). In case of a perfect match, $R^2 = 1$.

To assess the quality of these estimations, we examined the time dependency of $D^{(1)}(x,t)$ and of $\widetilde{D^{(1)}}(x,t)$ in all examples by substituting the simulated $x(t)$. This process yielded two sets of time series, each representing the values of the two drifts at each time point t. Since the precision of the estimated drift function depends on integration time T, R^2 values will also be influenced by T.

In figure 3, we present plots of R^2 as a function of integration time T. For $T \geqslant 10\,\mathrm{s}$ or $N \geqslant 10\frac{T_C}{\Delta t}$ in example 1, $T \geqslant 50\,\mathrm{s}$ or $N \geqslant 50\frac{T_C}{\Delta t}$ in example 2, and $T \geqslant 0.5\,\mathrm{s}$ or $N \geqslant 500\frac{T_C}{\Delta t}$ in example 3, $R^2 > 0.9$. Note that $R^2 = 1$ indicates a perfect match. In example 1 and 2, $T_C = 1\,\mathrm{s}$ and in example 3, $T_C = 0.001\,\mathrm{s}$. Therefore, we need at least $N \geqslant 10\frac{T_C}{\Delta t}$ data points and much more data points may be necessary if the correlation timescale is T_C is relatively short.

5.2. Comparing true and reconstructed dynamics: KS test

After estimating the functions and parameters for both the deterministic drift $D^{(1)}(x,t)$ and the stochastic components in examples (1–3), we can generate time series of the dynamics and compare them with time series created using the preset functions and parameters. To do so, we computed the CDFs for each of these time series ($\mathcal{F}_e$ and $\mathcal{F}_p$, corresponding to estimated and preset CDFs, respectively), and applied the non-parametric two-sample KS test to assess any differences between them.

The maximum 'distance' between the two CDFs is defined as:

$$D_{KS} = \sup_x \left| \mathcal{F}_e(x) - \mathcal{F}_p(x) \right|,$$

with D_{KS} approaching zero for identical CDFs. For each example, we report the D_{KS} values and their corresponding significance level p_{KS}.

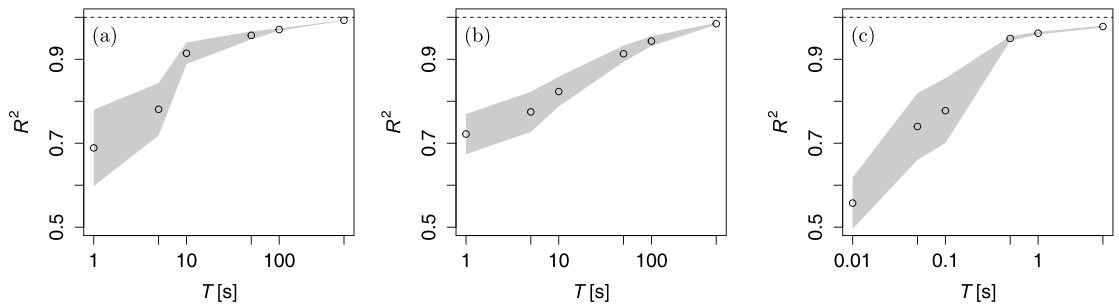

Figure 3. The coefficient of determination (R^2-score) for all examples (1-3). As shown, for $T \geqslant 10$ s or $N \geqslant 10\frac{T_C}{\Delta t}$ in example 1, $T \geqslant 50$ s or $N \geqslant 50\frac{T_C}{\Delta t}$ in example 2, and $T \geqslant 0.5$ s or $N \geqslant 500\frac{T_C}{\Delta t}$ in example 3, $R^2 > 0.9$. Note that $R^2 = 1$ indicates a perfect match. Gray shading represents standard errors of the mean.

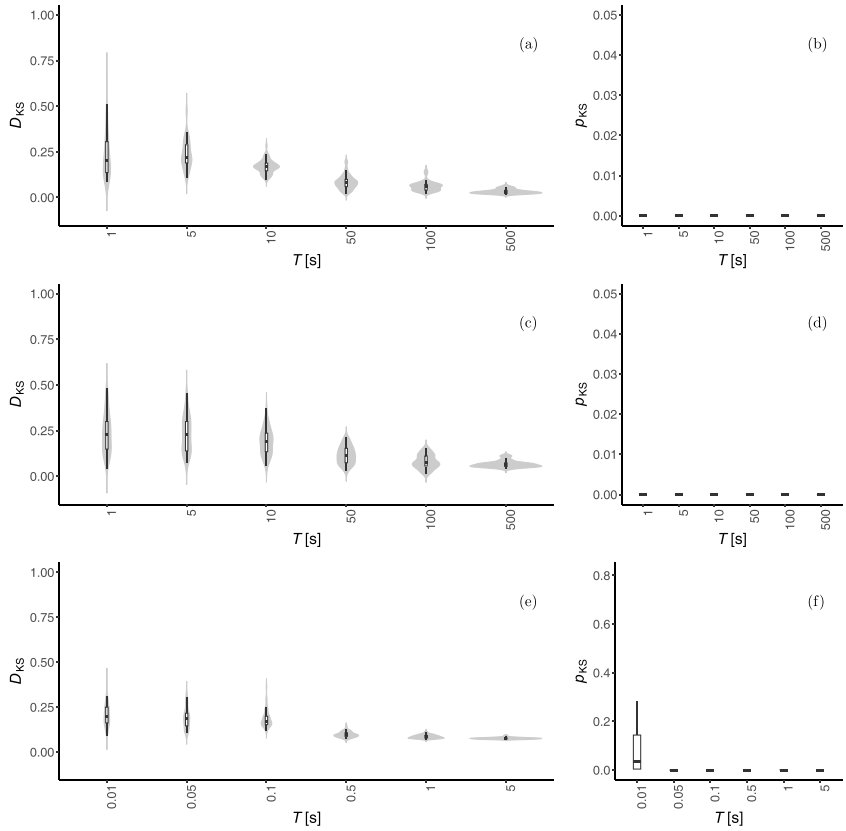

Figure 4. KS statistic of a comparison between reset and reconstructed dynamics of examples (1–3). Distributions of D_{KS} are presented in (a), (c) and (e), with shaded gray areas and p_{KS}-values in (b), (d) and (f), over 50 realizations of applied noises. Whiskers (boxes in black lines) indicate the first and the third quartile of the distributions, and the central lines inside the boxes mark the median values.

In figure 4, we present the dependence of the KS statistic D_{KS} on integration time T, comparing true and reconstructed dynamics for all examples. It is clear that D_{KS} decreases as T increases.

5.3. Proposed guideline to analyze the real-world datasets

From the KM coefficients of Gaussian white-noise-driven Langevin equations, which describe the standard diffusion process, we have shown that the explicit criterion $\Theta(\tau) = 1$. A similar relation has been found for the jump–diffusion process, derived under the assumption of compound Poisson jumps with normally distributed jump sizes, as an example. Thus, the main results of our criteria are given in equations (9) and (10) and are shown in figure 1 and 2.

In real-world datasets, it is unlikely that they strictly follow Langevin or jump–diffusion dynamics, particularly since they are rarely Markovian and may exhibit finite Markov–Einstein time scales.

Based on the our main results and also from the results of the statistical tests performed, we propose the following steps to analyze the real-world datasets:

- Estimate the correlation timescale $T_C = \frac{1}{\gamma}$ using the approximation $D^{(1)}(x) = -\gamma x$ and the average timescale between jumps $T_J = \frac{1}{\lambda}$ using equation (7) near the point $x \simeq 0$,
- Check if there is the number of data points is at least $N \geqslant 10\frac{T_C}{\tau_s}$ where τ_s is the sampling time of the dataset. For small T_C, more number of data points may be necessary as shown in section 5.1,
- Estimate the conditional moments and calculate Θ and Q as in equations (5) and (8)
- Check if the results follow equations (9) and (10),
- Consider other approaches if the results clearly deviate from equations (9) and (10), such as GLE for non-Markovian dataset.

The proposed guideline is applied to the real-word dataset in sections 6 and 7.

6. Real-world examples

To evaluate our criteria on real-world datasets, we analyze five distinct examples:

- The spatial positions of a polystyrene dielectric bead (diameter: 1 μm, Bangs Laboratories Inc. USA) trapped in optical tweezers [32].
- Log-returns of the DAX stock market index [33].
- Intracranial electroencephalographic (iEEG) data reflecting epileptic brain dynamics [11].
- Velocity measurements of low-temperature helium turbulence in a free jet, with Reynolds number $Re = 757\,000$ [34].
- Solar clear sky index, defined as the ratio of measured solar irradiance to its theoretical prediction under clear-sky conditions at specific latitudes and longitudes [35, 36].

All datasets with sampling time τ_s are one-dimensional quantities. Each dataset is normalized by subtracting its mean and dividing by its standard deviation, resulting in dimensionless data with zero mean and a standard deviation of one. The Θ- and Q-criteria are computed at $x = 0$ using the normalized data. These evaluations are performed after downsampling the original dataset with a downsampling time step τ_{ds}.

For each time series sampled with timescale τ_s, we apply the methods described in the appendix A to estimate the drift, diffusion, jump rate, and jump amplitude. This involves analyzing the KM coefficients of order 1, 4, and 6 to derive the values of T_C and T_J for the respective data sets. $T_C = \frac{1}{\gamma}$ is calculated using the approximation $D^{(1)}(x) = -\gamma x$ and $T_J = \frac{1}{\lambda}$ using equation (7) near the point $x \simeq 0$ where Θ and Q are evaluated.

The data are categorized into three groups. In the first category, we analyze measurements of the spatial positions of a dielectric bead trapped in optical tweezers. The $\Theta = \Theta(x = 0, \tau_{ds})$ and $Q = Q(x = 0, \tau_{ds})$ values, plotted against τ_{ds} for a single dataset with a duration of 3s and a sampling rate of 22kHz, are shown in figures 5(a) and (b).

The results from optical tweezers indicate diffusive behavior. The correlation timescale $T_C = \frac{1}{\gamma}$, estimated from the drift function $D^{(1)}(x) = -\gamma x$, shows the following trends: for $\tau_{ds} < T_C$, Q exhibits a linear relationship, and $\Theta = 1$. Beyond the correlation timescale ($\tau_{ds} > T_C$), $Q = \sigma_x^2 = 1$ and $\Theta = 1$, reflecting uncorrelated Gaussian white noise behavior.

Interestingly, [37] examined the same setup with molecular motors added to the bead. Their results revealed a jump component with the motors, providing a contrast to this analysis, where no jumps are observed without motors.

In the second category, we analyze the log-returns of the DAX stock market index and iEEG data for brain dynamics. The results for $\Theta = \Theta(x = 0, \tau_{ds})$ and $Q = Q(x = 0, \tau_{ds})$ versus τ_{ds} for the DAX data are plotted in figures 5(c) and (d), while those for the epileptic brain data are shown in figures 5(e) and (f).

For both datasets, Θ deviates from unity, indicating that they do not satisfy the necessary conditions for a purely diffusive process. The correlation time $T_C = \frac{1}{\gamma}$ was evaluated from the drift function, approximately $D^{(1)}(x) = -\gamma x$, while the jump timescale $T_J = \frac{1}{\lambda}$ was estimated using equation (7). For small τ_{ds}, Q remains constant, but Θ shows deviations from unity and non-constant behavior with respect to τ_{ds}. This suggests that a jump–diffusion process is a better fit for modeling these dynamics.

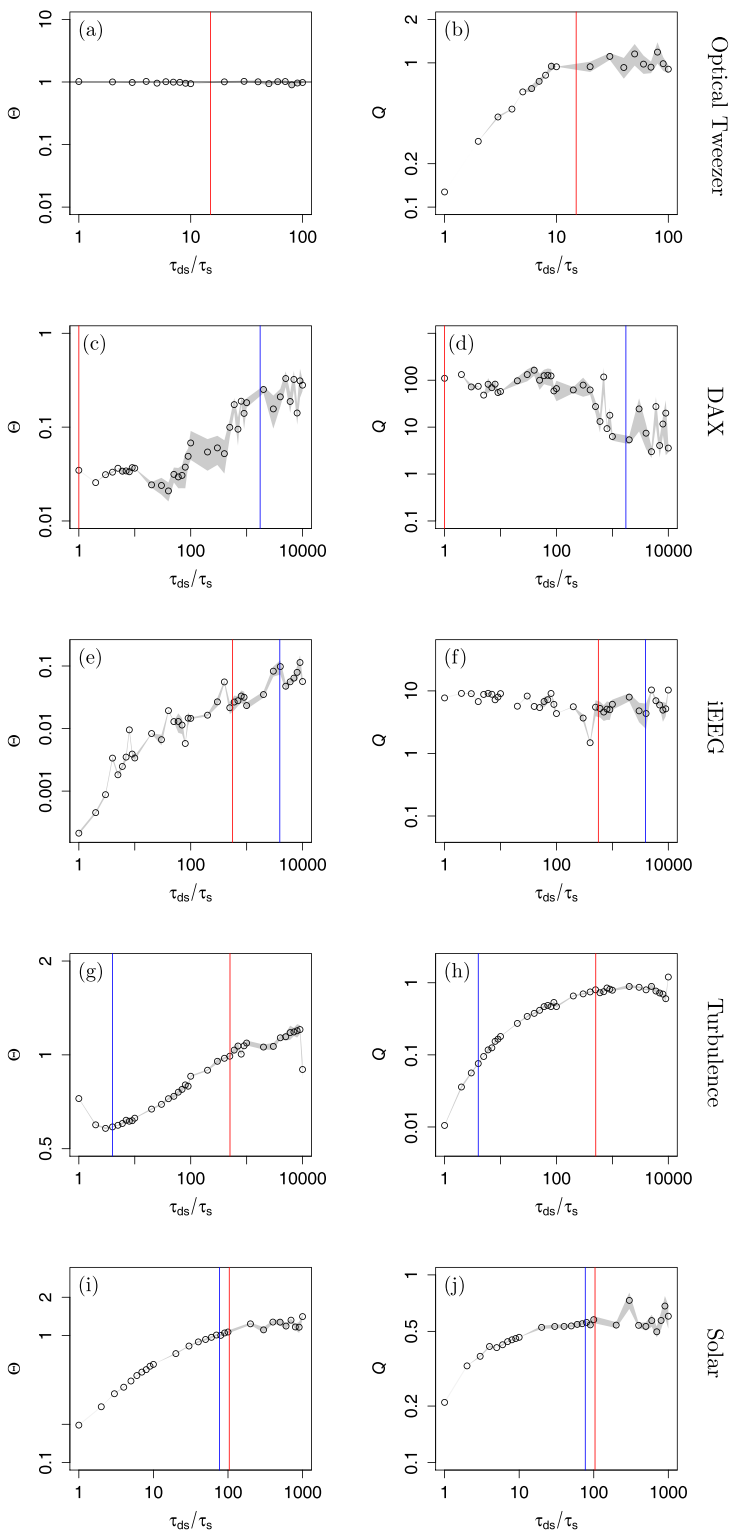

Figure 5. Θ- and Q-criteria for downsampled displacement data of a dielectric bead in optical tweezers (a), (b), log-return of DAX data (c), (d), iEEG brain data (e), (f), turbulence data (g), (h), and solar clear sky index data (i), (j), are plotted in double logarithmic scale. Vertical blue lines indicate the jump timescale $T_J = \frac{1}{\lambda}$, while vertical red lines denote the correlation timescale $T_C = \frac{1}{\gamma}$. Gray shading represents standard errors of the mean, some of which are smaller than the symbol size.

The previous work [11] also examined jump–diffusion behavior in the epileptic brain dynamics. However, it has not provided conclusive evidence to support the use of Langevin or jump–diffusion dynamics. Our new results provide stronger validation for the Langevin and jump–diffusion frameworks while exploring their application in broader contexts. It particularly focuses on real-world data, where only downsampling is feasible for effective analysis.

Focusing on the DAX data, the analysis captures the intrinsic characteristics of financial time series, which often exhibit heavy tails and jumps due to market shocks dynamics. The observed non-diffusive behavior aligns with well-established findings in financial modeling, where jump–diffusion processes, such as those in Merton's or Kou's models, are commonly employed to describe log-returns. These jumps reflect abrupt market movements caused by economic or geopolitical events, further reinforcing the applicability of the Θ- and Q-criteria to detect such discontinuities [38, 39].

In the third category, we look into the velocity of free jet low temperature helium turbulence and the solar clear sky index. The results of $\Theta = \Theta(x = 0, \tau_{\rm ds})$ and $Q = Q(x = 0, \tau_{\rm ds})$ against $\tau_{\rm ds}$ for the turbulence data are plotted in figures 5(g) and (h). The results of $\Theta = \Theta(x = 0, \tau_{\rm ds})$ and $Q = Q(x = 0, \tau_{\rm ds})$ against $\tau_{\rm ds}$ for the solar clear sky index are plotted in figures 5(i) and (j). As Θ is deviated from unity, they do not fulfill the necessary conditions for the diffusion process. Similarly to previous data analysis, we evaluated again the correlation time $T_{\rm C} = \frac{1}{\gamma}$ where the drift function is defined by $D^{(1)}(x) = -\gamma x$. We also estimated $T_{\rm J} = \frac{1}{\lambda}$ by equation (7). In these two cases, $T_{\rm J} < T_{\rm C}$. For small $\tau_{\rm ds}$, both Q and Θ show the nonlinear behavior on τ. Thus, one could only conclude that they are not diffusive from the current study. Turbulence is known to have much more complex nature, and it is also not Markovian in time but is in scale [2, 40]. For the third category, we can conclude that these time series do not belong to the class of diffusion processes.

As a remark, our key assumption for these criteria is that the underlying stochastic process is Markovian. To further investigate, we analyze the time series of turbulence and the solar clear sky index, estimating their Markov–Einstein time scale $t_{\rm M}$, which defines the minimum interval during which the data can be treated as Markovian. Verifying Markovianity typically requires statistical tests [5]; here, we apply the Wilcoxon test [41]. For the solar clear sky index sampled at 1 Hz, $t_M \approx 20$ s, while for turbulence data with Reynolds number $Re = 757\,000$, $t_{\rm M} \approx 104$ data points. Thus, these datasets necessitate non-Markovian modeling. One suitable approach is the GLE [42], incorporating a memory kernel to replace the standard Langevin equation. In the following section, we present results for these two non-Markovian datasets using the GLE framework.

7. Generalized Langevin equation

Langevin and jump–diffusion equations with white noise imply Markovian dynamics, which can be confirmed for real-world time series through statistical tests like those in [5]. The Markov–Einstein time scale $t_{\rm M}$ identifies the time threshold above which data can be treated as Markovian. This is critical for systems such as turbulence or biological processes [2]. For time scales below $t_{\rm M}$, non-Markovian models, like the GLE, are necessary to account for memory effects.

To study real-world time series with a finite Markov–Einstein scale, the memory kernel can be estimated using the GLE [17, 18], which exhibits non-Markovian behavior. Derived through perturbation theory, linear response theory, and projector operator formalism [19–22], the GLE is commonly used to describe tracer particles, requiring access to forces acting on the particles and their velocity at each moment in time. The memory function is related to the velocity autocorrelation function, determining both the dissipative and random forces as per the generalized fluctuation-dissipation theorem [22, 43, 44].

The GLE can also be applied beyond tracer particles, such as modeling collective variables like polymer end-to-end distances [18]. Examples include molecular dynamics [45–49], climate modeling [50–55], biophysics [56], and finance [53, 54, 57, 58].

For estimating the memory kernel from time series, we adapt the method proposed in [59]. The GLE, with its memory kernel $\mathcal{K}$, is given by:

$$\dot{X}_t = D^{(1)}(X_t) + \sum_{k=0}^{K} \mathcal{K}_k \Delta X_{t,k} + \sqrt{D^{(2)}(X_t)}\,\eta_t. \tag{13}$$

Here, $D^{(1)}$ and $D^{(2)}$ are the drift and diffusion functions, respectively. η_t represents white noise, i.e. $\langle \eta_t \eta_{t'} \rangle \propto \delta(t - t')$ and the increment $\Delta X_{t,k} := X_t - X_{t-k\Delta t}$ with a finite time step Δt.

Maximum likelihood estimation and Bayesian estimation are commonly used for parameter estimation in GLE. While Bayesian estimation is computationally intensive for large datasets, we employ an efficient Bayesian method with a piecewise constant approximation for drift and diffusion, which reduces computational costs [59, 60]. In figure 6, we present plots for the estimations of $D^{(1)}$ and $D^{(2)}$, as well as the memory kernel $\mathcal{K}_s$, for the turbulence and solar clear sky index time series. Long-range memory is evident from this analysis, supporting the non-Markovian properties of these time series.

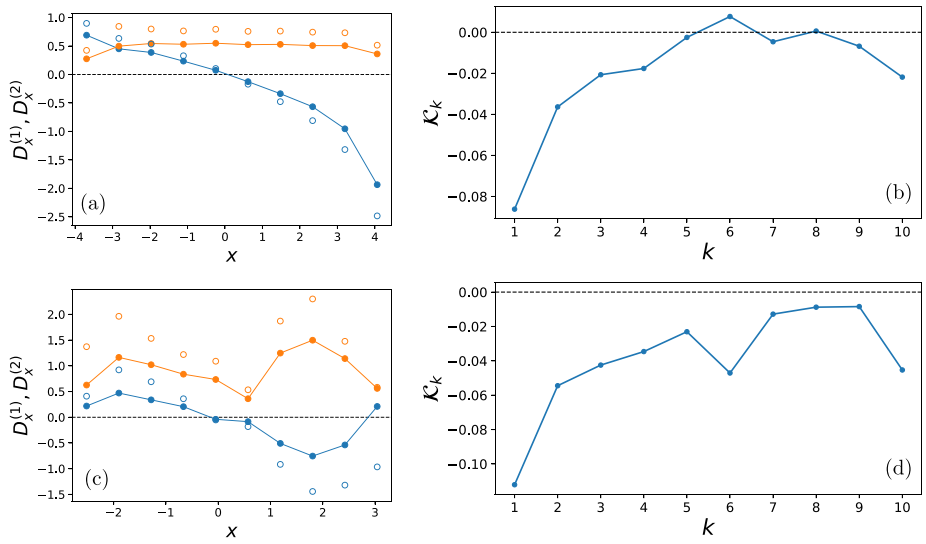

Figure 6. Estimated drift coefficients $D^{(1)}$ (blue) and diffusion coefficients $D^{(2)}$ (orange) for turbulence (a) and solar clear sky index (c). Open circles show the estimated coefficients of standard Langevin equation, equation (4), and used as initial guess for the GLE, equation (13). Solid circles shows the results of GLE where initial guess is adjusted slightly. Memory kernel $\mathcal{K}_k$ of GLE is estimated for kernel length $K = 10$ for turbulence (b) and solar clear sky index (d). For detailed explanation of the methods and algorithms, the readers are referred to [59]. Long-range memory is evident from this analysis, supporting the non-Markovian properties of these time series.

8. Conclusion

In many models describing the dynamics of natural measurements, it is common to assume a standard Gaussian white noise-driven Langevin equation to account for observed variability. However, it is not immediately clear whether this assumption holds true. Evaluating the validity of a Langevin equation to describe the data presents a significant challenge, especially when dealing with finite data points or processes that exhibit unique trajectories [26, 34, 36, 61].

In this study, we have developed a set of criteria to assess the suitability of modeling the data using a Langevin equation. By recognizing the limitations of Langevin dynamics, alternative modeling approaches become relevant. These may include jump–diffusion stochastic dynamics [11], GLEs [43, 62], fractional Klein–Kramers equations [14], fractional diffusion processes [15, 16] and Lévy-driven Langevin dynamics [5].

By focusing on higher-order terms in KM coefficient estimations for high-frequency data, we attempt to perform model selection. Alternative methods for model selection include Akaike information criterion (AIC) and Bayesian information criterion (BIC), which are statistical metrics used to assess the quality of a model relative to other models by balancing goodness-of-fit and model complexity [63–65].

It is well-established that systems exhibiting higher levels of stochasticity explore a broader portion of the phase space, with noisy time series providing richer information about the underlying dynamics. Consequently, increased noise intensity can be advantageous for our method, potentially leading to improved results compared to AIC or BIC [31]. Intuitively, data with greater stochastic components-such as diffusion and jump processes-are more likely to fully explore the phase space. For long time series with strong stochasticity, we utilized both time-averaged and ensemble-averaged KM coefficients, assuming ergodicity to enhance the reliability of our analysis. Ergodicity can be verified using the method outlined in [66]. Additionally, it is possible that other models, such as diffusing diffusivity [67], might not be distinguishable from jump–diffusion processes by our method.

Revealing novel aspects of the data that cannot be captured by white-noise driven Langevin equations can illuminate the presence of abrupt changes in time series. These changes often arise from complex physical phenomena that extend beyond the scope of Langevin equations. Consequently, this exploration uncovers new phenomena, enhances our understanding of measured variability in diverse datasets, and opens up avenues for further research.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://cloud.uol.de/s/CQ2sSbJKNgNnDs6.

Acknowledgment

We thank Hanieh Hatami for her help with the computations in GLE.

Appendix A. Expansion of conditional moments at finite time step

The conditional probability distribution of the process $x = x(t)$ which satisfies the KM differential equation can be written as [5, 29],

$$\frac{\partial p\left(x,t|x',t'\right)}{\partial t} = \mathcal{L}_{\mathrm{KM}}\, p\left(x,t|x',t'\right) \tag{A.1}$$

with initial condition $p(x,t|x',t) = \delta(x-x')$ and the KM operator $\mathcal{L}_{\mathrm{KM}}$ is given by,

$$\mathcal{L}_{\mathrm{KM}} = \sum_{n=1}^{\infty} \frac{1}{n!} \left(-\frac{\partial}{\partial x}\right)^n K^{(n)}\left(x,t\right) \quad . \tag{A.2}$$

The formal solution of (A.1) reads

$$p\left(x,t+\tau|x',t\right) = \exp\left\{\tau\ \mathcal{L}_{\mathrm{KM}}\right\}\ \delta\left(x-x'\right) \quad . \tag{A.3}$$

The nth-order conditional moments $M^{(n)}(x,t,\tau)$ with finite τ can be written as,

$$\begin{aligned} M^{(n)}\left(x_i,t,\tau\right) &= \int_{-\infty}^{\infty} \left(x-x_i\right)^n \exp\left\{\tau\ \mathcal{L}_{\mathrm{KM}}\right\}\ \delta\left(x-x_i\right) \mathrm{d}x \\ &= \exp\left\{\tau\ \mathcal{L}_{\mathrm{KM}}^{\dagger}\right\} \left(x-x_i\right)^n |_{x=x_i} \end{aligned} \tag{A.4}$$

where $\mathcal{L}_{\mathrm{KM}}^{\dagger}$ is the adjoint operator of $\mathcal{L}_{\mathrm{KM}}$ and is given by,

$$\mathcal{L}_{\mathrm{KM}}^{\dagger} = \sum_{n=1}^{\infty} \frac{1}{n!}\, K^{(n)}\left(x,t\right) \left(\frac{\partial}{\partial x}\right)^n , \tag{A.5}$$

where $K^{(n)}(x,t) = \lim_{\tau\to 0} \frac{1}{\tau} M^{(n)}(x,t,\tau)$ are the nth-order KM coefficients and $M^{(n)}(x,t,\tau)$ are given by equation (A.4). Next, the explicit cases of the diffusion and jump–diffusion processes are discussed.

A.1. Conditional moments of diffusion and jump–diffusion processes

Diffusion process, in general, can be described by Langevin equation in equation (4). We can evaluate its conditional moments $M_{\mathrm{d}}^{(n)}(x,\tau)$ using equation (A.4). For diffusion process or Langevin equation, the adjoint KM operator $\mathcal{L}_{\mathrm{KM}}^{\dagger}$ becomes the adjoint Fokker–Planck (FP) operator $\mathcal{L}_{\mathrm{FP}}^{\dagger}$. For better readability, let the drift be $D^{(1)}(x,t) = K^{(1)}(x,t) = a$ and the diffusion $D^{(2)}(x,t) = K^{(2)}(x,t) = b^2$ and the adjoint FP-operator becomes

$$\mathcal{L}_{\mathrm{FP}}^{\dagger} = a\,\frac{\partial}{\partial x} + \frac{1}{2}\,b^2\,\frac{\partial^2}{\partial x^2}\,. \tag{A.6}$$

In this formulation, a and b can still be the function of x and t. The second-, fourth- and sixth-order conditional moments up to first non-vanishing term can be derived from equations (A.4) and (A.6) as follow [29]:

$$M_{\mathrm{d}}^{(2)}(x,\tau) = b^2\tau + \mathcal{O}\left(\tau^2\right),$$

$$M_{\mathrm{d}}^{(4)}(x,\tau) = 3b^4\tau^2 + \mathcal{O}\left(\tau^3\right),$$

$$M_{\mathrm{d}}^{(6)}(x,\tau) = 15b^6\tau^3 + \mathcal{O}\left(\tau^4\right). \tag{A.7}$$

In order to distinguish from the jump–diffusion process which will be discussed next, the subscript 'd' is used here for the diffusion process.
OU process is considered here to be more concrete such that

$$\mathrm{d}x = -\gamma x\,\mathrm{d}t + \sqrt{D}\,\mathrm{d}W_t\,. \tag{A.8}$$

where γ and D are positive real constants. Expansion of $M_{\mathrm{d}}^{(n)}(x,t,\tau)$ of OU process for small τ up to $\mathcal{O}(\tau^3)$ (up to $\mathcal{O}(\tau^4)$ for sixth-order conditional moment) reads

$$M_{\mathrm{d}}^{(2)}(x,\tau) = D\tau + \left(\gamma^2x^2 - \gamma D\right)\tau^2 + \mathcal{O}\left(\tau^3\right),$$

$$M_{\mathrm{d}}^{(4)}(x,\tau) = 3D^2\tau^2 + \mathcal{O}\left(\tau^3\right),$$

$$M_{\mathrm{d}}^{(6)}(x,\tau) = 15D^3\tau^3 + \mathcal{O}\left(\tau^4\right). \tag{A.9}$$

Jump-diffusion process, in general, is described in equation (6). The KM-coefficients of such jump–diffusion are

$$K^{(1)}(x,t) = D^{(1)}(x,t)$$

$$K^{(2)}(x,t) = \left[D^{(2)}(x,t) + \langle\xi^2\rangle\lambda(x)\right]$$

$$K^{(2n)}(x,t) = \langle\xi^{2n}\rangle\lambda(x), \quad \text{for } 2n > 2. \tag{A.10}$$

Therefore, the adjoint KM-operator $\mathcal{L}_{\mathrm{KM}}^{\dagger}$ for jump–diffusion process becomes

$$\mathcal{L}_{\mathrm{KM}}^{\dagger} = \underbrace{D^{(1)}(x,t)}_{\mathbf{A}}\frac{\partial}{\partial x} + \underbrace{\frac{\left[D^{(2)}(x,t) + \langle\xi^2\rangle\lambda(x)\right]}{2!}}_{\mathbf{B}}\frac{\partial^2}{\partial x^2}$$
$$+ \underbrace{\frac{\langle\xi^4\rangle\lambda(x)}{4!}}_{\mathbf{C}}\frac{\partial^4}{\partial x^4} + \underbrace{\frac{\langle\xi^6\rangle\lambda(x)}{6!}}_{\mathbf{D}}\frac{\partial^6}{\partial x^6}$$
$$+ \underbrace{\frac{\langle\xi^8\rangle\lambda(x)}{8!}}_{\mathbf{E}}\frac{\partial^8}{\partial x^8} + \cdots\quad. \tag{A.11}$$

Again for better readability, we use the substitution of the coefficients with $\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{D},\mathbf{E},\cdots$, which can still be the function of x and t in this formulation. With it, we can derive the conditional moments of the jump–diffusion equation equation (6) for second-, fourth and sixth-orders of the time interval τ using equations (A.4) and (A.11) such that

$$M_{\mathrm{j}}^{(2)}(x,\tau) = 2\mathbf{B}\tau + \mathcal{O}\left(\tau^2\right),$$

$$M_{\mathrm{j}}^{(4)}(x,\tau) = 4!\mathbf{C}\tau + \mathcal{O}\left(\tau^2\right),$$

$$M_{\mathrm{j}}^{(6)}(x,\tau) = 6!\mathbf{D}\tau + \mathcal{O}\left(\tau^2\right), \tag{A.12}$$

where the subscript 'j' denotes the jump–diffusion process [29]. Here, they are derived up to first non-vanishing term.

OU jump-diffusion process with constant jump rate and jump amplitude is now considered. It is a linear jump–diffusion process, and one finds,

$$\mathrm{d}x = -\gamma x\,\mathrm{d}t + \sqrt{D}\,\mathrm{d}W_t + \xi\;\mathrm{d}J_t\,. \tag{A.13}$$

Adding the jump term in OU process creates the discontinuities in the trajectory. For finite time interval τ, the conditional moments $M_{\mathrm{j}}^{(n)}(x,\tau)$ of an OU jump–diffusion process can be determined up to $\mathcal{O}(\tau^4)$ using equations (A.4) and (A.11) as follows:

$$\begin{aligned}
M_{\mathrm{j}}^{(2)}(x,\tau) &= (D+\lambda\sigma_\xi^2)\tau + \left(\gamma^2x^2 - \gamma(D+\lambda\sigma_\xi^2)\right)\tau^2 \\
&\quad + \left(\frac{1}{3}\gamma^2(D+\lambda\sigma_\xi^2) - \frac{1}{2}\gamma^3x^2\right)\tau^3 + \mathcal{O}(\tau^4), \\
M_{\mathrm{j}}^{(4)}(x,\tau) &= 3\lambda\sigma_\xi^4\tau + 3\left((D+\lambda\sigma_\xi^2)^2 - 2\gamma\lambda\sigma_\xi^4\right)\tau^2 \\
&\quad + (6x^2\gamma^2(D+\lambda\sigma_\xi^2) + 8\gamma^2\lambda\sigma_\xi^4 \\
&\quad - 6\gamma(D+\lambda\sigma_\xi^2)^2)\tau^3 + \mathcal{O}(\tau^4), \\
M_{\mathrm{j}}^{(6)}(x,\tau) &= 15\lambda\sigma_\xi^6\tau + 45\left(\lambda\sigma_\xi^4(D+\lambda\sigma_\xi^2) - \gamma\lambda\sigma_\xi^6\right)\tau^2 \\
&\quad + (45\gamma^2\lambda\sigma_\xi^4x^2 + 15(D+\lambda\sigma_\xi^2)^3 \\
&\quad + 540\gamma^2\lambda\sigma_\xi^6 - 810\gamma\lambda\sigma_\xi^4(D+\lambda\sigma_\xi^2))\tau^3 + \mathcal{O}(\tau^4).
\end{aligned} \tag{A.14}$$

A.2. Conditional moments of gaussian white noise

For a Gaussian distributed white noise with zero mean and variance σ_η^2, $x=\eta\sim N(0,\sigma_\eta^2)$, the nth order conditional moments can be calculated as

$$\begin{aligned}
M_{\mathrm{g}}^{(n)}(x,\tau) &= \left\langle (x(t+\tau)-x(t))^n|_{x(t)=x}\right\rangle, \\
&= \int_{-\infty}^{\infty}(x'-x)^n\cdot p(x',t'|x,t)\cdot\mathrm{d}x', \\
&= \int_{-\infty}^{\infty}(x'-x)^n\cdot p(x',t')\cdot\mathrm{d}x'.
\end{aligned} \tag{A.15}$$

For $x=0$ and $p(x',t') = p(x') = \frac{1}{\sqrt{2\pi\,\sigma_\eta^2}}\exp\left(-\frac{x'^2}{2\sigma_\eta^2}\right)$, we can use Wick's theorem [68, 69] to solve integral in equation (A.15). The $(2n+1)^{\mathrm{th}}$ and the $2n^{\mathrm{th}}$ conditional moments of Gaussian white noise become

$$\begin{aligned}
M_{\mathrm{g}}^{(2n+1)}(x=0,\tau) &= 0, \\
M_{\mathrm{g}}^{(2n)}(x=0,\tau) &= \frac{(2n)!}{2^n n!}\sigma_\eta^{2n}\,,
\end{aligned} \tag{A.16}$$

where σ_η^2 is its variance.

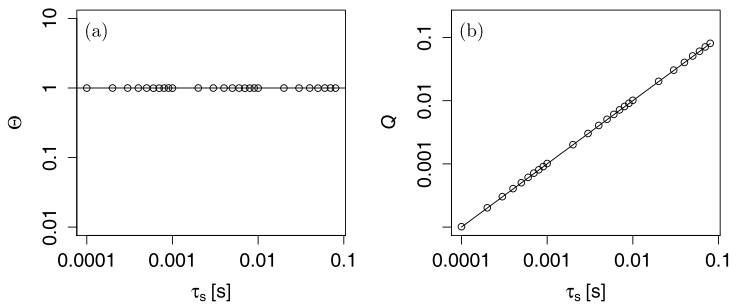

Figure A1. Θ- and Q-criterion of diffusion process at $x = 0$ with $D^{(1)}(x) = -x^3$ and $D^{(2)}(x) = 1 + x^2$, generated for different integration time steps τ_s, are plotted in double logarithmic scale. Theoretical predictions according to equations (9) and (10) are presented with the black solid lines. Standard errors are smaller than the symbol size.

A.3. Q-criterion of jump–diffusion process at time scale $\tau > T_J$

As a first criterion, the Q-criterion is introduced in [29] to distinguish whether given synthetic data simulated with different integration time step τ_s are diffusive or jumpy. Here, this criterion is also applied to the uncorrelated Gaussian white noise.

For a small τ, we obtain the function $Q(x, \tau)$ as described in equation (10) considering up to the first non-vanishing order term [29] which is derived from the fourth- and sixth-order conditional moments.

For diffusive process, case (i), the function Q has a linear relationship with τ while it is constant in first order approximation in τ for jump–diffusion process, case (ii). Additionally, the function Q is also constant for the Gaussian white noise, case (iii) which can be derived by direct computations of conditional moments.

From the simulated data, we can determine $Q = Q(x = 0, \tau = \tau_s)$ for different τ_s. For jump–diffusion process theoretical prediction equation (10) is valid for $\tau_s < \frac{1}{\lambda} = T_J$, which is the average time scale between the jumps. If the integration time step $\tau_s > T_J$, the jump–diffusion process behaves like diffusion process with the apparent diffusion coefficient $\widetilde{D^{(2)}}$ which has the relation, $\widetilde{D^{(2)}}\tau_s = D^{(2)}\tau_s + \sigma_\xi^2$.

To prove this relation, the jump–diffusion equation can be written using Euler's method as

$$x(t+\tau_s) - x(t) = D^{(1)}(x)\ \tau_s + \sqrt{D^{(2)}(x)}\ \Delta W_t + \xi\ \Delta J_t. \tag{A.17}$$

For $\tau_s > \frac{1}{\lambda}$, $\Delta J_t = 1$ due to zero-one jump law, the equation (A.17) becomes

$$x(t+\tau_s) - x(t) = D^{(1)}(x)\ \tau_s + \sqrt{D^{(2)}(x)\ \tau_s}\ \eta + \xi\,. \tag{A.18}$$

$\sqrt{D^{(2)}(x)\ \tau_s}\ \eta + \xi$ is the sum of two Gaussian random variables which is equivalent to $\sqrt{\widetilde{D^{(2)}}(x)\ \tau_s}\ \widetilde{\eta} = \sqrt{D^{(2)}(x)\ \tau_s + \sigma_\xi^2}\ \widetilde{\eta}$. Here, $\widetilde{\eta} \sim N(0,1)$ is the standard normal Gaussian white noise.

Therefore, the jump–diffusion equation for $\lambda > \frac{1}{\tau_s}$ becomes

$$x(t+\tau_s) - x(t) = D^{(1)}(x)\ \tau_s + \sqrt{\widetilde{D^{(2)}}(x)\ \tau_s}\ \widetilde{\eta}, \tag{A.19}$$

respectively,

$$\mathrm{d}x = D^{(1)}(x)\ \mathrm{d}t + \sqrt{\widetilde{D^{(2)}}(x)}\ \mathrm{d}\widetilde{W}_t\,, \tag{A.20}$$

where $\widetilde{D^{(2)}}$ is the apparent diffusion coefficient and $\widetilde{W}_t$ is the Wiener process.

A.4. Θ- and Q-criterion for general diffusion and jump–diffusion processes

Now, the Θ- and Q- criterion are determined for both diffusion and jump–diffusion processes with non-linear drift and multiplicative diffusion term. We numerically integrate equation (4) with $D^{(1)}(x) = -x^3$ and $D^{(2)}(x) = 1 + x^2$, and equation (6) for the case with additional jump term with $\lambda = 100\ \mathrm{s}^{-1}$ and $\sigma_\xi^2 = 1$ for different integration time step τ_s. The result of the non-linear diffusion process is shown in figure A1 and jump–diffusion process in figure A2.

We can observe the similar behavior in the case with linear drift and additive diffusion noise.

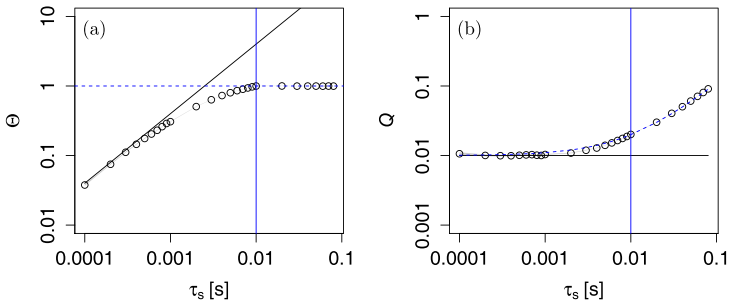

Figure A2. Θ- and Q-criterion of jump–diffusion process at $x = 0$ with $D^{(1)}(x) = -x^3$, $D^{(2)}(x) = 1 + x^2$, $\lambda = 100\ \mathrm{s}^{-1}$ and $\sigma_\xi^2 = 1$, generated for different integration time steps τ_s, are plotted in double logarithmic scale. The vertical blue lines show the jump time scale $T_\mathrm{J} = \frac{1}{\lambda}$. Theoretical predictions according to equations (9) and (10) are presented with the black solid lines. The blue dashed curve $Q = D^{(2)}(0)\tau_\mathrm{s} + \sigma_\xi^2$ and line $\Theta = 1$ show the behavior after $\tau_\mathrm{s} > T_\mathrm{J} = \frac{1}{\lambda}$. Standard errors are smaller than the symbol size.

Appendix B. Consequences of downsampling of the data

In real world, empirical data are given for a fixed sampling time τ_s due to experimental constraints. To handle this effect, we have to downsample the numerical data in which a new time step τ_ds is defined for downsampling time step. For the estimation of the KM conditional moments, it is essential to see in which relation τ_ds is with the correlation time $T_\mathrm{C} = \frac{1}{\gamma}$ and with the average time scale between the jumps $T_\mathrm{J} = \frac{1}{\lambda}$.

Influence of downsampling on the diffusion coefficient is first examined. In order to study the downsampled approximation of the diffusion term, it is integrated over a finite downsampled time τ_ds. In the Itô sense, the integral is generally interpreted as [2, 7]

$$\int_t^{t+\tau_\mathrm{ds}} g(x(s), s)\ \mathrm{d}W_s = g(x(t), t) \int_t^{t+\tau_\mathrm{ds}} \mathrm{d}W_s$$

$$= g(x(t), t) \sum_{k=1}^{N_\tau} (W_k - W_{k-1}). \tag{B.1}$$

Therefore, the integral of diffusion part in (jump–)diffusion process becomes

$$\int_t^{t+\tau_\mathrm{ds}} \sqrt{D^{(2)}(x)}\ \mathrm{d}W_s = \sqrt{D^{(2)}(x)} \sum_{k=1}^{N_\tau} \Delta W_k$$

$$= \sqrt{D^{(2)}(x)\ \tau_\mathrm{s}} \sum_{k=1}^{N_\tau} \eta_k\,, \tag{B.2}$$

where $\eta_k \sim N(0, \sigma_\eta^2 = 1)$ is the standard normal distributed random variable. To estimate the summation $\sum_{k=1}^{N_\tau} \eta_k$, $N_\tau = \frac{\tau_\mathrm{ds}}{\tau_\mathrm{s}}$ is the number of time steps in the downsampled time step τ_ds. Let us define the summation as a new effective noise η_eff for diffusion term as,

$$\eta_\mathrm{eff} = \sum_{k=1}^{N_\tau} \eta_k \tag{B.3}$$

where variance of η_eff will be $\sigma_{\eta,\mathrm{eff}}^2 = N_\tau \sigma_\eta^2 = N_\tau = \frac{\tau_\mathrm{ds}}{\tau_\mathrm{s}}$, (as $\sigma_\eta^2 = 1$), and η_eff tends to a Gaussian random variable. Then, the integral becomes

$$\int_t^{t+\tau_\mathrm{ds}} \sqrt{D^{(2)}(x)}\ \mathrm{d}W_s \simeq \sqrt{D^{(2)}(x)\ \tau_\mathrm{s}}\ \eta_\mathrm{eff} = \sqrt{D^{(2)}(x)\ \tau_\mathrm{ds}}\ \eta \tag{B.4}$$

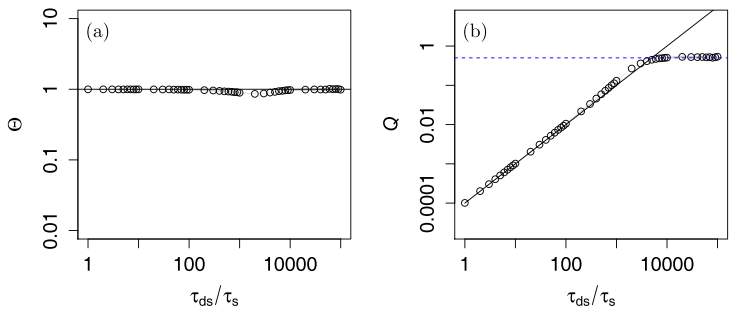

Figure B1. Θ- and Q-criterion of downsampled diffusion process at $x = 0$ with $D^{(1)}(x) = -x^3$ and $D^{(2)}(x) = 1 + x^2$ are plotted in double logarithmic scale. Theoretical predictions according to equations (9) and (10) are presented with the black solid lines. The blue dashed line in (a) shows $Q = \sigma_x^2$. Standard errors are smaller than the symbol size.

where $\eta \sim N(0,1)$ is the standard normal distributed random variable. Therefore, we can conclude that diffusion coefficient remains unchanged with downsampling and summation of the noises approaches to a Gaussian white noise.

Influence of downsampling on the jump amplitude is then examined. The jump term $\xi\ \mathrm{d}J_t$ will provide the following stochastic integral

$$\int_t^{t+\tau_{ds}} \xi\ \mathrm{d}J_s\ . \tag{B.5}$$

To deal with this integral, we use the integral relation [5, 7]

$$\int_t^{t+\tau_{ds}} h\left(x(s), s\right)\ \mathrm{d}J_s \overset{\mathrm{ms}}{=} \sum_{k=1}^{J_t} h\left(x\left(T_k^-\right), T_k^-\right)\ , \tag{B.6}$$

where $\overset{\mathrm{ms}}{=}$ is Itô mean square equals. The T_k^- denotes the limit from the left to jump time T_k. Therefore, the jump integral becomes

$$\int_t^{t+\tau_{ds}} \xi\ \mathrm{d}J_s = \sum_{k=1}^{J_t} \xi_k\ , \tag{B.7}$$

where $\xi_k \sim N(0, \sigma_\xi^2)$ is the Gaussian distributed random variable with zero mean and variance σ_ξ^2. To estimate the summation $\sum_{k=1}^{J_t} \xi_k$, we note that the average time step of jumps in original time series is $\frac{1}{\lambda}$, then the average number of jumps becomes $J_t \simeq \lambda\tau_{ds}$ in the downsampled time step τ_{ds}. Let us define the summation as a new noise ξ_{eff} for jump amplitude as,

$$\xi_{\mathrm{eff}} = \sum_{k=1}^{J_t} \xi_k\ , \tag{B.8}$$

where variance of ξ_{eff} will be $\sigma_{\xi,\mathrm{eff}}^2 \simeq \lambda\tau_{ds}\sigma_\xi^2$ and ξ_{eff} tends to a Gaussian random variable. Then, we find,

$$\int_t^{t+\tau_{ds}} \xi\ \mathrm{d}J_s = \xi_{\mathrm{eff}} \simeq \sqrt{\lambda\tau_{ds}}\ \xi\ , \tag{B.9}$$

where $\xi \sim N(0, \sigma_\xi^2)$ is Gaussian distributed with variance of jump amplitude σ_ξ^2. Therefore, if we downsample with $\tau_{ds} \gg \frac{1}{\lambda}$, $J_t \gg 1$, then zero-one jump law is no longer fulfilled. However, it can be realized as a jump event with the effective jump amplitude

$$\sigma_{\xi,\mathrm{eff}}^2 = \lambda\sigma_\xi^2\tau_{ds}\ . \tag{B.10}$$

In addition, the Θ- and Q-criterion are determined for both downsampled diffusion and jump–diffusion processes with non-linear drift and multiplicative diffusion terms. We numerically integrate equation (4) with $D^{(1)}(x) = -x^3$ and $D^{(2)}(x) = 1 + x^2$, and equation (6) for the case with additional jump term with $\lambda = 100\ \mathrm{s}^{-1}$ and $\sigma_\xi^2 = 1$ with a fixed sampling time $\tau_s = 10^{-4}$ s. The results are shown in figure B1 for diffusion process, and in figure B2 for jump–diffusion process.

We can observe the similar behavior in the case with linear drift and additive diffusion noise.

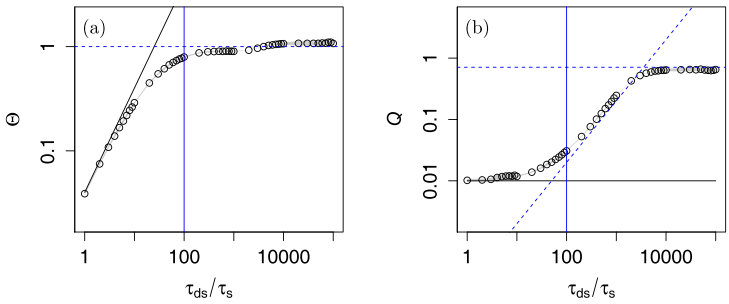

Figure B2. Θ- and Q-criterion of downsampled OU jump–diffusion process at $x = 0$ with $D^{(1)}(x) = -x^3$, $D^{(2)}(x) = 1 + x^2$, $\lambda = 100\ \mathrm{s}^{-1}$ and $\sigma_\xi^2 = 1$ are plotted in double logarithmic scale. The vertical blue lines show the jump time scale $T_\mathrm{J} = \frac{1}{\lambda}$. Theoretical predictions according to equations (9) and (10) are presented with the black solid lines. The blue dashed lines show $Q = (D^{(2)}(0) + \lambda\sigma_\xi^2)\tau_{\mathrm{ds}}$ for $\tau_{\mathrm{ds}} > T_\mathrm{J}$ and $Q = \sigma_x^2$ (horizontal line), and $\Theta = 1$. Standard errors are smaller than the symbol size.

Appendix C. OU process in the presence of jumps

In the presence of compound Poisson jump process $Z_t = \xi J_t$ with constant jump amplitude σ_ξ^2 and jump rate λ, OU jump-diffusion process $x = x(t)$ can be written as equation (A.13).

By introducing the substituting variable $y = xe^{\gamma t}$, applying the Itô's product rule,

$$\mathrm{d}(f \cdot g) = f \cdot \mathrm{d}g + g \cdot \mathrm{d}f + \mathrm{d}f \cdot \mathrm{d}g, \tag{C.1}$$

and using the following properties,

$$(\mathrm{d}t)^i\ (\mathrm{d}W_t)^j = 0, \text{for } i,j \geqslant 1, \tag{C.2}$$

and

$$(\mathrm{d}t)^i\ (\mathrm{d}J_t)^j = 0, \text{for } i,j \geqslant 1, \tag{C.3}$$

we can solve equation (A.13). The solution in $\mathrm{d}t$ precision is

$$x = x_0 \mathrm{e}^{-\gamma t} + \sqrt{D}\int_0^t \mathrm{e}^{-\gamma(t-t')}\mathrm{d}W_{t'} + \int_0^t \mathrm{e}^{-\gamma(t-t')}\mathrm{d}Z_{t'}, \tag{C.4}$$

where $x_0 = x(0)$.

The mean of the OU jump-diffusion process can be calculated as

$$\langle x \rangle = x_0 \mathrm{e}^{-\gamma t}. \tag{C.5}$$

and $\langle x \rangle = 0$ for sufficiently long simulation time $t \to \infty$.

The covariance of OU jump-diffusion process can be calculated using the following properties,

$$(\mathrm{d}W_t)^i\ (\mathrm{d}J_t)^j = 0, \text{for } i,j \geqslant 1, \tag{C.6}$$

$$\mathrm{Cov}(\mathrm{d}W_s, \mathrm{d}W_t) = \langle \mathrm{d}W_s\, \mathrm{d}W_t \rangle = \delta(s-t)\ \mathrm{d}s\, \mathrm{d}t, \tag{C.7}$$

and

$$\mathrm{Cov}(\mathrm{d}Z_s, \mathrm{d}Z_t) = \langle \mathrm{d}Z_s\, \mathrm{d}Z_t \rangle = \lambda\sigma_\xi^2\ \delta(s-t)\ \mathrm{d}s\, \mathrm{d}t, \tag{C.8}$$

such that

$$C_{xx}(t,\tau) = \mathrm{Cov}(x(t)x(t+\tau)) = \frac{D + \lambda\sigma_\xi^2}{2\gamma}\left[\mathrm{e}^{-\gamma\tau} - \mathrm{e}^{-\gamma(\tau+2t)}\right]. \tag{C.9}$$

We can calculate the variance of the OU jump-diffusion process from $C_{xx}(t, \tau = 0)$ which is

$$\mathrm{Var}(x) = \langle x^2 - \langle x \rangle^2 \rangle = \frac{D + \lambda\sigma_\xi^2}{2\gamma}\ \mathrm{e}^{-2t}. \tag{C.10}$$

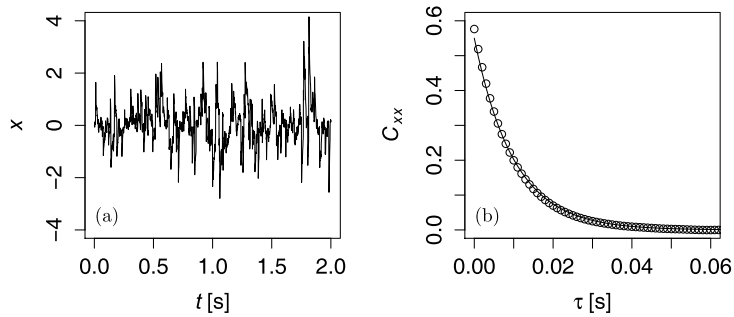

Figure C1. Timeseries of OU jump–diffusion processes, (a), and its covariance function $C_{xx}(\tau)$, (b). The evaluated $C_{xx}(\tau)$ is plotted in open circles and its theoretical result from equation (C.11) are plotted with solid line.

For sufficiently long simulation time $t \to \infty$, the covariance becomes

$$C_{xx}(\tau) = \frac{D + \lambda\sigma_\xi^2}{2\gamma} \, \mathrm{e}^{-\gamma\tau} \,, \tag{C.11}$$

and the variance

$$\mathrm{Var}(x) = \frac{D + \lambda\sigma_\xi^2}{2\gamma} \,. \tag{C.12}$$

The detailed discussions of the properties of stochastic processes that are used in this derivation can be seen in [4, 5, 7, 70]. In order to verify it, a synthetic time series of the OU jump–diffusion process is generated for $\Delta t = 10^{-3}$ s with $\gamma = 100\ \mathrm{s}^{-1}$, $D = 10\ \mathrm{s}^{-1}$ and for the additional jump terms with $\lambda = 100\ \mathrm{s}^{-1}$ and $\sigma_\xi^2 = 1$. In figure C1 the timeseries of the OU jump–diffusion processes is shown. Its covariance function $C_{xx}(\tau)$ is also evaluated and plotted together with the theoretical result from equation (C.11). Here, we can observe that the correlation time T_C depends solely on γ but not on the jump rate λ.

ORCID iDs

Pyei Phyo Lin https://orcid.org/0000-0003-0731-235X
Matthias Wächter https://orcid.org/0000-0001-9212-1122
Joachim Peinke https://orcid.org/0000-0002-0775-7423
M Reza Rahimi Tabar https://orcid.org/0000-0003-1786-975X

References

[1] Chandrasekhar S 1943 *Rev. Mod. Phys.* **15** 1–89
[2] Friedrich R, Peinke J, Sahimi M and Reza R T M 2011 *Phys. Rep.* **506** 87–162
[3] Peinke J, Tabar M R R and Wächter M 2019 *Annu. Rev. Condens. Matter Phys.* **10** 107–32
[4] Gardiner C 1985 Handbook of stochastic methods for physics, chemistry and the natural sciences *Proc. in Life Sciences* (Springer)
[5] Tabar M R R 2019 *Analysis and Data-Based Reconstruction of Complex Nonlinear Dynamical Systems: Using the Methods of Stochastic Processes* (Springer)
[6] Ait-Sahalia Y 2004 *J. Financ. Econ.* **74** 487–528
[7] Hanson F 2007 Applied stochastic processes and control for jump diffusions: modeling, analysis and computation *Advances in Design and Control* (Society for Industrial and Applied Mathematics)
[8] Dumitru A-M and Urga G 2012 *J. Bus. Econ. Stat.* **30** 242–55
[9] Li Q, Li Y, Nolte I, Nolte (Lechner) S and Yu S 2021 *SSRN Electr. J.* (https://doi.org/10.2139/ssrn.3943203)
[10] Wang B and Zheng X 2022 *J. Econ.* **230** 483–509
[11] Anvari M, Tabar M R R, Peinke J and Lehnertz K 2016 *Sci. Rep.* **6** 35435
[12] Lin P P, Peinke I, Hagenmuller P, Wächter M, Tabar M R R and Peinke J 2022 *Cryosphere* **16** 4811–22
[13] Lin P P, Wächter M, Tabar M R R and Peinke J 2023 *PRX Energy* **2** 033009
[14] Dieterich P, Klages R, Preuss R and Schwab A 2008 *Proc. Natl Acad. Sci.* **105** 459
[15] Kassel J A and Kantz H 2022 *Phys. Rev. Res.* **4** 013206
[16] Kassel J A, Walter B and Kantz H 2023 arXiv:2304.025364
[17] Mitterwallner B G, Schreiber C, Daldrop J O, Rädler J O and Netz R 2020 *Phys. Rev.* E **101** 032408
[18] Attard P 2012 *Non-Equilibrium Thermodynamics and Statistical Mechanics: Foundations and Applications* (Oxford University Press)
[19] Zwanzig R 1961 *Phys. Rev.* **124** 983–92
[20] Zwanzig R 1964 *J. Res. Natl Bureau Stand.* B **68B** 143–5
[21] Mori H 1965 *Prog. Theor. Phys.* **33** 423–55
[22] Kob W, Roldán-Vargas S and Berthier L 2011 *Nat. Phys.* **8** 164–7

[23] Nadaraya E A 1964 *Theor. Probab. Appl.* **9** 141–2
[24] Watson G S 1964 *Sankhyā: Indian J. Stat.* A **26** 359–72
[25] Lamouroux D and Lehnertz K 2009 *Phys. Lett.* A **373** 3507–12
[26] Rahvar S *et al* 2024 *Chaos Solitons Fractals* **185** 115069
[27] Risken H 1984 *The Fokker-Planck equation: Methods of Solution and Applications* (Springer)
[28] Pawula R F 1967 *Phys. Rev.* **162** 186–8
[29] Lehnertz K, Zabawa L and Tabar M R R 2018 *New J. Phys.* **20** 113043
[30] Stanton R 1997 *J. Finance* **52** 1973–2002
[31] Tabar M R R, Nikakhtar F, Parkavousi L, Akhshi A, Feudel U and Lehnertz K 2024 *Phys. Rev.* X **14** 011050
[32] Mousavi S M, Reihani S N S, Anvari G, Anvari M, Alinezhad H G and Tabar M R R 2017 *Sci. Rep.* **7** 4832
[33] Ghashghaie S, Breymann W, Peinke J, Talkner P and Dodge Y 1996 *Nature* **381** 767–70
[34] Manshour P, Rahimi Tabar M R and Peinke J 2015 *J. Stat. Mech.* 08031
[35] Anvari M, Lohmann G, Wächter M, Milan P, Lorenz E, Heinemann D, Tabar M R R and Peinke J 2016 *New J. Phys.* **063027** 063027
[36] Madanchi A, Absalan M, Lohmann G, Anvari M and Tabar M R R 2017 *Sol. Energy* **144** 1–9
[37] Osunbayo O, Miles C E, Doval F, Reddy B J N, Keener J P and Vershinin M D 2019 *Soft Matter* **15** 1847–52
[38] Merton R C 1976 *J. Financ. Econ.* **3** 125–44
[39] Kou S G 2002 *Manage. Sci.* **48** 1086–101
[40] Renner C, Peinke J and Friedrich R 2001 *J. Fluid Mech.* **433** 383–409
[41] Manshour P 2024 Stochasticity repository (https://github.com/manshour/stochasticity)
[42] Tepper L, Dalton B and Netz R R 2024 *J. Chem. Theory Comput.* **20** 3061–8
[43] Lei H, Baker N A and Li X 2016 *Proc. Natl Acad. Sci.* **113** 14183
[44] Evans D J and Morriss G P 2008 *Statistical Mechanics of Nonequilibrium Liquids* 2nd edn (Cambridge University Press) (https://doi.org/10.1017/CBO9780511535307)
[45] Horenko I, Hartmann C, Schütte C and Noe F 2007 *Phys. Rev.* E **76** 016706
[46] Satija R and Makarov D E 2019 *J. Phys. Chem.* B **123** 802
[47] Jung G, Hanke M and Schmid F 2018 *Soft Matter* **14** 9368
[48] Grogan F, Lei H, Li X and Baker N A 2020 *J. Comput. Phys.* **418** 109633
[49] Klippenstein V, Tripathy M, Jung G, Schmid F and van der Vegt N F A 2021 *J. Phys. Chem.* B **125** 4931
[50] Wouters J and Lucarini V 2013 *J. Stat. Phys.* **151** 850
[51] Boers N, Chekroun M D, Liu H, Kondrashov D, Rousseau D-D, Svensson A, Bigler M and Ghil M 2017 *Earth Syst. Dyn.* **8** 1171–90
[52] Kondrashov D, Chekroun M D and Ghil M 2015 *Physica* D **297** 33–55
[53] Russo A, Duran-Olivencia M A, Kevrekidis I G and Kalliadasis S 2019 arXiv:1903.09562
[54] Hassanibesheli F, Boers N and Kurths J 2020 *New J. Phys.* **22** 073053
[55] Watkins N W, Chapman S C, Chechkin A, Ford I, Klages R and Stainforth D A 2021 *On Generalized Langevin Dynamics and the Modeling of Global Mean Temperature (Unifying Themes in Complex Systems)* ed X D Braha, M A M de Aguiar, C Gershenson, A J Morales, L Kaufman, E N Naumova, A A Minai and Y Bar-Yam (Springer) pp 433–41
[56] Ferretti F, Chardès V, Mora T, Walczak A M and Giardina I 2020 *Phys. Rev.* X **10** 031018
[57] Takahashi M 1996 *Financ. Eng. Jpn. Mark.* **3** 87
[58] Schmitt D T and Schulz M 2006 *Phys. Rev.* E **73** 056204
[59] Willers C and Kamps O 2024 *J. Comput. Phys.* **497** 112626
[60] Kleinhans D 2012 *Phys. Rev.* E **85** 026705
[61] Movahed M S, Ghasemi F, Rahvar S and Tabar M R R 2011 *Phys. Rev.* E **84** 021103
[62] Kou S C and Xie X S 2004 *Phys. Rev. Lett.* **93** 180603
[63] Brewer M J, Butler A, Cooksley S L and Freckleton R 2016 *Methods Ecol. Evol.* **7** 679–92
[64] Bulteel K, Wilderjans T F, Tuerlinckx F and Ceulemans E 2013 *Behav. Res. Methods* **45** 782–91
[65] Chakrabarti A and Ghosh J K 2011 Aic, Bic and recent advances in model selection *Philosophy of Statistics* (*Handbook of the Philosophy of Science*) vol 7, ed P S Bandyopadhyay and M R Forster (North-Holland) pp 583–605
[66] Schwarzl M, Godec A and Metzler R 2017 *Sci. Rep.* **7** 3878
[67] Chubynsky M V and Slater G W 2014 *Phys. Rev. Lett.* **113** 098302
[68] Isserlis L 1916 *Biometrika* **11** 185–90
[69] Wick G C 1950 *Phys. Rev.* **80** 268
[70] Lin P P 2023 Characterization of jump-diffusion stochastic dynamics: analysis and applications on real world data *PhD Thesis* University of Oldenburg (available at: https://nbn-resolving.org/urn:nbn:de:gbv:715-oops-58821)