A reconstruct-then-bootstrap test for the sufficiency of diffusion processes

Kathleen Medriano[a] **and Joachim Vandekerckhove**[a]

Draft of December 8, 2025.

Diffusion processes are commonly used in the modeling of time series that exhibit stationarity and Markovianity. However, those two properties do not guarantee that a diffusive process is sufficient for the time series. In this paper, we develop a test for the sufficiency of a diffusion process for an observed time series. To develop the test we capitalize on the Kramers–Moyal (KM) expansion: a Taylor expansion of the integral form of the master equation that describes Markov continuous-time processes. In the idealized case, if the observed data indeed arise from a true diffusion process, then the KM expansion should truncate naturally after the second term. In theory, this means that any higher-order (≥ 3) KM coefficients should be zero. However, in practice, the discrete nature of measurement introduces artificial higher-order KM coefficients, even when the underlying process is truly diffusive. Nonetheless, for genuinely diffusive systems, it is expected that the sampling distribution of a statistic associated with higher-order coefficients will be different than non-diffusive ones. This is a viable avenue for testing the appropriateness of a diffusion model given an observed time series. We take advantage of this and propose a meaningful statistic that could inform whether or not a diffusion model is sufficient for a given time series. We then build a test that involves reconstructing the diffusion equation to generate surrogate paths, yielding a bootstrap distribution against which the observed statistic could be compared. We evaluate the sensitivity and selectivity of the proposed test in Monte Carlo studies.

Fokker-Planck | Kramers-Moyal | Stochastic Time Series | Diffusion

In time series analysis, it is common practice to draw inferences about a postulated continuous-time model on the basis of discretely sampled observations. This introduces a problem since identifying continuous-time models from discrete-time data is subject to the so-called *aliasing* phenomenon: distinct continuous-time processes may look identical when sampled at regular time intervals (Aït-Sahalia, 2002). Nevertheless, it is common for modelers to use a parsimonious diffusion process to model time series data.[1]

For example, in neuroscience the assumption of white noise for single-neuron activity implies a diffusion process (Vellmer & Lindner, 2021). Certain exploratory methods that consider patterns of EEG under anaesthesia (Bahraminasab, Ghasemi, Stefanovska, McClintock, & Friedrich, 2009) and in epileptic dynamics (Prusseit & Lehnertz, 2007) implicitly rely on assuming a diffusion model. In the field of affective dynamics, diffusion processes have been used to model affective states of individuals (Oravecz, Tuerlinckx, & Vandekerckhove, 2009; Oravecz & Vandekerckhove, 2020). They are also used to model heart rate fluctuations and cardio and respiratory interactions (Ghasemi, Sahimi, Peinke, & Tabar, 2006; Ghasemi, Peinke, Reza Rahimi Tabar, & Sahimi, 2006; Smelyanskiy, Luchinsky, Stefanovska, & McClintock, 2005), as well as prices of assets in finance (Detemple & Rindisbacher, 2011) and even turbulence in earth sciences (Tutkun, 2017; Peinke, Tabar, & Wächter, 2019).

A diffusion process is a Markov process with continuous sample paths (Aït-Sahalia, 2002). *Markov process* here means that the current value only depends on the most immediate past value and not on earlier observations. Since the discrete nature of measurement limits one to having only discretely sampled data, there is a need for a testing procedure that can help determine whether the underlying model that gave rise to the data is truly diffusive – this is informative of the kind of variation present in the system of interest. Further, it helps in avoiding model misspecification – specifically, fitting diffusion process to a time series when the variation exhibited in it is not due to a diffusion process.

In the development of our testing procedure, we will be able to capitalize on an important theoretical result known as the *Pawula theorem* (Pawula, 1967). This theorem, which we explain below, describes a particular regularity in data resulting from diffusion processes. Our proposed testing procedure is based on these regularities, which allow us to develop a meaningful statistic for the sufficiency of diffusion for an observed time series. Then we build a test that uses a *reconstruct-then-bootstrap* strategy The comparison of the test statistic to its expected distribution under the assumption of "pure diffusion" allows us to decide whether an observed time series can be sufficiently explained by diffusion.

The outline of the paper is as follows. In Section 1 we introduce the notation and the mathematical prerequisites for building the test. We detail the form of the diffusion processes that are within the scope of the test. We introduce the Pawula theorem and Kramers-Moyal expansion and show how they are necessary for the test construction. In Section 2, we illustrate how the test statistic is derived and further enumerate the steps taken in the test. Section 3 describes a Monte Carlo study we conducted to evaluate the sensitivity and selectivity of the test. Section 4 contains examples of applications of the new test on two datasets – self-reported arousal level time series of participants in a mobile-health study and EEG recordings from pre-surgical epilepsy patients. In Sections 5 and 6 we discuss and conclude.

Notation. Before we formally introduce diffusion processes and the necessary mathematical results from which we ground the construction of the test, we review the notation used throughout

[1] To head off any confusion: The current paper deals with diffusion processes *as models for time series* and is not related to the use of diffusion models for choice response times (Vandekerckhove, Tuerlinckx, & Lee, 2011) in any important way.

[a] University of California, Irvine
[†] To whom correspondence should be addressed. E-mail: joachim@uci.edu.

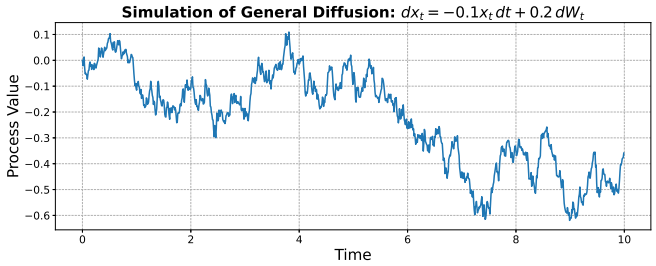

Fig. 1. An example of a path of a diffusion process of the form: $dx_t = -0.1x_t dt + 0.2dW_t$. The characteristic attribute of diffusion processes are the random perturbations that make the path non-smooth.

this paper.

A stochastic process can be viewed as a collection of random variables indexed by a variable such as time. It is therefore convenient to think of these random variables as functions of the indexing variable. In this work, we use time t as the indexing variable and x as the random state variable. The notation $x(t)$ denotes the state x at time t; for brevity, we will sometimes write x_t to mean the same.

The term dx_t represents an infinitesimal increment of the process. Expressions such as $a(x_t)$ or $b(x_t)$ indicate that a and b are functions of the state variable x_t. The differential dt denotes an infinitesimal change in time, while dW_t denotes an increment of a Wiener process, where dW_t follows a Gaussian distribution with mean 0 and standard deviation dt: $dW_t \sim \mathcal{N}(0, dt)$.

The probability density of x at time t is written as $p(x,t)$. Its partial derivative with respect to time is written as $\frac{\partial p(x,t)}{\partial t}$, and $\left(\frac{\partial^n}{\partial x^n}\right)$ denotes the n-th partial derivative with respect to x. Superscripts such as x^n indicate powers, while parentheses such as $x^{(n)}$ indicate the n-th order of a quantity—for example, $K^{(n)}(x_t)$ denotes the n-th conditional moment which is a function of the state variable x_t.

Square brackets, $[\cdot]$, are used to group terms that go together, and finally angle-bracket notation $\langle\cdot\rangle$ denotes the expectation operator.

Diffusion processes. Diffusion processes describe the random evolution of systems over time, typically resulting from the cumulative effect of many small, random perturbations such as small impacts of molecules on pollen particles that then appear to move about erratically (Einstein, 1905; Uhlenbeck & Ornstein, 1930; Wiener, 1923). These processes are characterized by continuous but non-differentiable paths typically depicted as noisy random walks (Boninsegna, Nüske, & Clementi, 2018).

Figure 1 illustrates an example path from a diffusion process with mean-reverting properties. The path exhibits random fluctuations as it approaches the mean forward in time. It is this noisy looking behavior that is a signature of diffusion processes.

Langevin representation. Formally, the diffusion process is best known as a description at the level of individual paths by the stochastic differential equation (SDE):

$$dx_t = a(x_t)\,dt + b(x_t)\,dW_t, \qquad [1]$$

where $a(x_t)\,dt$ is the drift term representing the deterministic component of change and $b(x_t)\,dW_t$ is the diffusion term responsible for the microscopic random perturbations. In the diffusion term, $b(x_t)$ describes the magnitude of random fluctuations. It scales the increment dt of the Wiener process dW_t. Similar to $b(x_t)$, the term $a(x_t)$ may be thought of as a scaling factor of the deterministic change in time, dt. Both $a(x_t)$ and $b(x_t)$ are functions of the state variable x_t. We note that these scaling factors could made time-varying—meaning that they could be made explicit functions of time, t–but in this paper we restrict our discussion to only those that are functions of the endogenous state variable x_t.

This SDE is commonly referred to as the *Langevin equation* and serves as a microscopic description of the diffusion process (Tabar, 2019). Through the SDE, we gain a picture of how the state, x_t, of a stochastic system changes over time due to the change of increments, dx_t, over time.

Fokker-Planck representation. Another way in which diffusion processes can be described is through the evolution of the probability density: $p(x,t)$. Here, $p(x,t)$ corresponds to the likelihood of the system being at state x at time t (Paul & Baschnagel, 1999). Meaning, instead of describing how the state of a stochastic system changes over time, one could instead describe how the stochasticity of the system's state changes over time. For diffusion processes, this is characterized through the Fokker-Planck equation, which shows how $p(x,t)$ changes over time:

$$\frac{\partial p(x,t)}{\partial t} = -\left(\frac{\partial}{\partial x}\right)\left[a(x_t)\,p(x,t)\right] + \left(\frac{\partial^2}{\partial x^2}\right)\left[\frac{1}{2}b^2(x_t)\,p(x,t)\right] \qquad [2]$$

Equations 1 and 2 are equivalent descriptions of the diffusion process – it is possible to derive one from the other (Paul & Baschnagel, 1999). Later, we will see that the test we aim to develop will rely heavily on this equivalence.

Kramers-Moyal expansion and the Pawula Theorem. The Fokker-Planck equation (Eq. 2) tells us how the probability density of a diffusion process evolves over time. Diffusion processes are a subclass of Markov processes. The evolution of the probability density of the broader class of Markov processes can be represented through the so-called *Kramers–Moyal (KM) expansion*. KM expansion is derived with a Taylor expansion of the integral form of the master equation that describes Markov continuous-time processes (Gorjão & Meirinhos, 2019).

The KM expansion is a partial differential equation that is first-order in time and infinite order in the state variable. It is expressed as an infinite series involving derivatives of $p(x,t)$, where each term is weighted by a corresponding coefficient (Tabar, 2019). Formally, the KM expansion takes the form:

$$\frac{\partial p(x,t)}{\partial t} = \sum_{n=1}^{\infty}\left(-\frac{\partial^n}{\partial x^n}\right)\left[D^{(n)}(x_t)\,p(x,t)\right] \qquad [3]$$

where:

- $p(x,t)$ is a probability density function describing the likelihood of state x at time t;
- $D^{(n)}(x_t)$, with $n \in \mathbb{N}$, is the n-th KM coefficient, defined as:

$$D^{(n)}(x_t) = \frac{1}{n!}\lim_{\tau\to 0}\frac{1}{\tau}K^{(n)}(x_t); \qquad [4]$$

  and

- $K^{(n)}(x_t)$ is the n-th conditional moment, defined as:

$$\begin{aligned} K^{(n)}(x_t) &= \left\langle \left[x\left(t+\tau\right) - x\left(t\right)\right]^n \middle| x(t)\right\rangle \\ &= \int \left[x\left(t+\tau\right) - x\left(t\right)\right]^n \\ &\quad \times p\left(x(t+\tau),\, t+\tau \mid x\left(t\right),\, t\right)\, dx\left(t+\tau\right) \end{aligned} \qquad [5]$$

Here, τ corresponds to the sampling interval. These conditional moments, $K^{(n)}(x_t)$, are functions of the conditioning variable $x\,(t)$. Consequently, the KM coefficients, $D^{(n)}(x_t)$, are also functions of $x\,(t)$.

The KM expansion is helpful since it describes how the stochasticity of a Markov process evolves over time. One drawback is that it is an infinite-order PDE and in real world applications it is infeasible to work with infinite sum of terms. Fortunately, the Pawula theorem provides that there are only three possible cases for the KM expansion (Tabar, 2019): Either (1) the KM expansion is truncated after the first term, implying that the process under study is deterministic; or (2) the expansion is truncated after the second term; or (3) the expansion must retain all terms and cannot be truncated.

The truncation at the first term is equivalent to having an ordinary differential equation (ODE) as a description for how a system's state changes over time. In this description, only the elapsed time, dt and some function of the current state, $a(x_t)$, determine the succeeding values in the process. The truncation of the KM expansion at the second term is due to the observation that if any even-order coefficient beyond the second vanishes (e.g., $D^{(4)}(x_t) = 0$), then it is consistent to set all higher-order terms with $n \geq 3$ to zero (Tabar, 2019). Hence, in the case that $D^{(4)}(x_t) = 0$, the expansion reduces to:

$$\frac{\partial p(x,t)}{\partial t} = -\frac{\partial}{\partial x}\left[D^{(1)}(x_t)\,p(x,t)\right] + \frac{\partial^2}{\partial x^2}\left[D^{(2)}(x_t)\,p(x,t)\right] \quad [6]$$

The last case is the only other possibility since premature truncation at any other term will lead to a mathematical contradiction, namely that $p(x,t)$ is not guaranteed to be a valid density.

Equation 6 is precisely the Fokker-Planck characterization of diffusion processes in Equation 2. This equivalence implies that for diffusion processes, there is a relationship between the first two KM coefficients and the drift and diffusion terms in Equation 1. Specifically:

$$\begin{aligned} D^{(1)}(x_t) &= a(x_t) \\ D^{(2)}(x_t) &= \frac{1}{2}b^2(x_t) \end{aligned} \quad [7]$$

Hence, we can rewrite the SDE in terms of the first and second KM coefficients to yield:

$$dx_t = D^{(1)}(x_t)\,dt + \sqrt{2D^{(2)}(x_t)}\,dW_t, \quad [8]$$

This tells us that the first two KM coeficients are sufficient in characterizing diffusion processes and higher-order (≥ 3) KM coefficients are not needed in its microscopic description.

Testing the sufficiency of diffusion. Based on the implications of the Pawula Theorem for the KM expansion, among Markov processes, one could develop a test to assess the sufficiency of diffusion by checking whether for a given time series, any of its higher-order KM coefficients (≥ 3), such as $D^{(4)}(x_t)$, are negligible.

Characterizing the hypothesis space. We build a diffusion sufficiency test based on this observation and borrow logic from the Neyman-Pearson (NP) hypothesis testing framework.[2] A fundamental element of the test construction is the observation that if an observed time series is already Markov and exhibits stochasticity, then there are only two possibilities: either it is diffusive (corresponds to the truncation of the KM expansion at the second term) or it is not (meaning all terms in the KM expansion are needed). Hence, we can partition the hypothesis space into a null, $\mathcal{H}_0$, and alternative hypothesis, $\mathcal{H}_1$, and say that given a time series, ω, it belongs either in:

$$\mathcal{H}_0 :\ \omega \in \Omega_0 = \left\{ \begin{array}{l} \omega \text{ is a path sufficiently described} \\ \text{by a diffusion process specified in Eq. 1} \end{array} \right\},$$

or

$$\mathcal{H}_1 :\ \omega \in \Omega_1 = \left\{ \begin{array}{l} \omega \text{ is a path NOT sufficiently described} \\ \text{by a diffusion process specified in Eq. 1} \end{array} \right\}.$$

The strict binarization of the hypothesis space into two parts is different from the usual binarization of hypothesis that happens in the traditional NP framework. In the traditional NP framework, the null hypothesis (e.g. "no effect") is often reduced to a precise condition such as testing whether a parameter (e.g. a regression coefficient) satisfies $\beta = 0$. However, in practice, the notion of "no effect" may more accurately correspond to a small effect such as $|\beta| < \epsilon$ where $\epsilon \geq 0$, this subtlety is not easily captured by the traditional NP framework (Etz, Haaf, Rouder, & Vandekerckhove, 2018; Etz, Goodman, & Vandekerckhove, 2022).

By contrast, strictly partitioning the hypothesis space into "is it diffusive" and "is it not diffusive" is supported by the implications of the Pawula theorem: if an observed time series is already Markov and exhibits stochasticity, either we have something that is entirely diffusive or not. This binarization is consistent with the notion of an M-closed world in Bayesian model comparison (Bernardo & Smith, 2009), where the set of models under consideration is assumed to include the true data-generating process. Unlike the more common M-open setting, where the truth may lie outside the chosen family of models, the structure imposed by the Pawula theorem ensures that our candidate space ("diffusive" vs. "non-diffusive") is exhaustive and mutually exclusive, and thus the inference problem can legitimately be posed in M-closed terms.

Constructing the test statistic. The next important step is to come up with a test statistic able to arbitrate between the two scenarios. We note that we do not have a good characterization of the entirety of the space that is not diffusive. Therefore, in building a test for the sufficiency of diffusion, we can only capitalize on what we know about diffusive processes: Its higher-order KM coefficients (≥ 3), such as $D^{(4)}(x_t)$, are effectively zero.

A complication that arises is that KM coefficients have a known estimation bias in samples with finite temporal resolution. Even for a purely diffusive process, finite-time discretization introduces a coarse-graining effect: unobserved small-time-scale fluctuations distort the increment distribution, generating spurious higher-order Kramers–Moyal coefficients (Tabar, 2019). The implication is that, even if something were truly diffusive, merely checking whether any of the higher-order KM coefficient is zero (or close to zero) is insufficient.

In addition, there is ambiguity in determining what constitutes "close to zero," which can introduce subjectivity into the decision-making process. Further, since the KM coefficients are functions of the state variable x_t, this evaluation of whether it is "close to zero," must be carried out across the entire domain of x_t, rather than at a few selected points.

For these reasons, we construct the test statistic to be the *distance of the fourth KM coefficient, $D^{(4)}(x_t)$, from zero*. This test statistic measures the degree of departure from expected behavior if the time series were truly generated by a diffusion process.

[2] While we would prefer it, the computational complexity of our test is not conducive to a fully Bayesian implementation.

Borrowing logic from the NP framework, we now need to construct the expected distribution of this test statistic under the null hypothesis. Then we will compare the observed test statistic against its expected distribution. In our case, the null distribution will indicate which values of the test statistic values we should expect if the assumption of diffusion holds. Then the arbitration will simply amount to checking whether the observed statistic falls reasonably within its null distribution.

Method

In this section, we detail the specifics of the test construction. In designing the test, we first construct a test statistic from the observed time series that captures whether data can be reasonably explained by a diffusion process. Then we describe what we expect about the distribution of the test statistic under the null hypothesis, $\mathcal{H}_0$. Lastly, we describe the reconstruct-then-bootstrap approach used in approximating the null distribution of the test statistic and discuss the decision rule for the test.

L^2-norm of $\hat{D}^{(4)}(x_t)$ as κ-statistic. We choose the fourth KM coefficient in building the test statistic. Since KM coefficients are based on conditional moments, the fourth conditional moment captures the heaviness of the tails of the probability density function. Intuitively, heavier tails are associated with greater discontinuities or non-diffusive variation. Additionally, the proof of the Pawula theorem involves establishing an inequality in which all even-order KM coefficients (excluding the second) are bounded above by a multiple of the fourth KM coefficient (Tabar, 2019). Therefore, assurance that $D^{(4)}(x_t)$ is close to zero also indirectly suggests that all higher even-order KM coefficients are negligible.

Analyzing $D^{(4)}(x_t)$ directly can be cumbersome, primarily because it is a function defined over a domain rather than a single scalar value. This functional nature complicates the process of drawing conclusions, particularly when the goal is to make a decision based on the sufficiency of diffusion. To simplify such decision-making, we can reduce the problem to a scalar comparison by evaluating a single summary statistic rather than a function in its entirety.

If we are interested in testing whether diffusion is sufficient, and we know from theoretical considerations that $D^{(4)}(x_t)$ should be identically zero under conditions of sufficient diffusion, then a natural approach is to measure how far $D^{(4)}(x_t)$ deviates from zero. We can take a common distance measure, in this case the L^2-norm. This leads us to consider a scalar test statistic that becomes a meaningful indicator of deviation from the expected behavior. This approach is not uncommon, having previously been used in creating a two-samples test for functional data (Zhang, Peng, & Zhang, 2010). From here on, we call this statistic the κ-statistic:

$$\kappa = \|\hat{D}^{(4)}(x_t)\|_2 \tag{9}$$

Doing this reduces comparison of functions to a scalar test which aids interpretability and a more manageable decision process. Since time series in practice are finite, then a discrete approximation of κ is:

$$\kappa = \|\hat{D}^{(4)}(x_t)\|_2 \approx \left(\sum_{t=1}^{m} |\hat{D}^{(4)}(x_t)|^2 \right)^{1/2} \tag{10}$$

Sampling distribution of κ-statistic under the assumption of diffusion sufficiency. To operationalize the notion of diffusion sufficiency, we have defined a scalar summary statistic—referred to here as the *κ-statistic*—that quantifies the distance of the empirically estimated fourth-order KM coefficient, $\hat{D}^{(4)}(x_t)$, from zero. Under the assumption that the process is truly diffusive, the true value of $D^{(4)}(x_t)$ should be identically zero.

A key insight is that the behavior of *κ-statistic* under a truly diffusive process should be consistent with the stochastic properties of the first two KM coefficients, $\hat{D}^{(1)}(x_t)$ and $\hat{D}^{(2)}(x_t)$, which are assumed to be sufficient for characterizing the dynamics of a diffusion process. In particular, if the underlying process is indeed diffusive, then repeated sampling (e.g., simulations) using only $\hat{D}^{(1)}(x_t)$ and $\hat{D}^{(2)}(x_t)$ to generate surrogate paths should yield a sampling distribution for the *κ-statistic* that is statistically similar to the sampling distribution obtained from the observed data.

That is, we expect consistency between the distribution on the LHS and RHS:

$$\mathcal{D}(\kappa \mid \hat{D}^{(1)}(x_t), \hat{D}^{(2)}(x_t), \tau, \mathcal{H}_0) \approx \mathcal{D}(\kappa \mid D^{(1)}(x_t), D^{(2)}(x_t)) \tag{11}$$

Here $\mathcal{D}(\cdot)$ denotes the sampling distribution and $\mathcal{H}_0$ indicates the null assumption that the process is diffusive. The LHS is the sampling distribution of κ, given estimates of the first two KM coefficients and under the assumptions that (a) the null hypothesis of sufficiency of diffusion holds and (b) there is a finite sampling rate. The RHS is the sampling distribution of κ given that the true data generating process follows the form specified in Equation 1. Significant discrepancies between the two distributions would suggest that higher-order terms (e.g., $D^{(4)}(x_t)$) play a non-negligible role, thereby violating the null assumption of diffusion sufficiency, $\mathcal{H}_0$.

Diffusion sufficiency test procedure. Algorithm 1 outlines the step-by-step procedure for evaluating the sufficiency of a diffusion process using the proposed K-statistic.

To test whether a given path ω belongs to Ω_0—that is, whether it can be sufficiently described by a diffusion process—we proceed as follows. We conduct pretests to check whether ω is stationary and Markov. Given parameters $M > 0$ (the number of surrogate paths) and $0 < q < 1$ (the quantile threshold), we compute estimates of the Kramers–Moyal (KM) coefficients $\hat{D}^{(1)}(x_t)$, $\hat{D}^{(2)}(x_t)$, and $\hat{D}^{(4)}(x_t)$ from the observed path. Using these estimates, we calculate the observed statistic $\kappa_{\text{observed}} = \|\hat{D}^{(4)}(x_t)\|_2$. Next, we reconstruct a diffusion stochastic differential equation (SDE) of the form:

$$dx_t = \hat{D}^{(1)}(x_t)\, dt + \sqrt{2\hat{D}^{(2)}(x_t)}\, dW_t \tag{12}$$

and use it to generate surrogate realizations of the process under the null hypothesis $\mathcal{H}_0$ (that the path is purely diffusive). Specifically, we perform a bootstrap procedure by generating $j = 1, ...M$ surrogate paths ω_j from the reconstructed SDE. For each surrogate ω_j, we estimate its fourth KM coefficient $\hat{D}_j^{(4)}(x_t)$ and compute the corresponding statistic $\kappa_j = \|\hat{D}_j^{(4)}(x_t)\|_2$, thereby obtaining the null distribution $D(\kappa \mid \mathcal{H}_0)$. Finally, we compare the observed statistic κ_{observed} with this null distribution. If κ_{observed} lies below the q-th quantile of $D(\kappa \mid \mathcal{H}_0)$, we conclude that $\omega \in \Omega_0$; otherwise, we conclude that $\omega \in \Omega_1$, indicating that the path cannot be sufficiently described by a diffusion process.

Pretest for stationarity and Markovianity. Our proposed test requires the observed time series to first pass stationarity and Markovianity checks. To this end we used the Augmented-Dickey Fuller (ADF) test for stationarity and the Conditional Mutual Information (CMI) test to check for Markovianity could be used. The ADF test evaluates whether a time series is stationary by checking for the presence of a unit root, with added lag terms to account for

Algorithm 1 Test whether $\omega \in \Omega_0$ (a path sufficiently described by diffusion)

Require: ω is stationary and Markov; $M > 0$ (number of surrogate paths); $0 < q < 1$ (quantile threshold)

Given ω, compute estimates of KM coefficients ($\hat{D}^{(1)}(x_t)$, $\hat{D}^{(2)}(x_t)$, $\hat{D}^{(4)}(x_t)$)

Compute observed statistic κ_{observed}:

$\kappa_{\text{observed}} = \|\hat{D}^{(4)}(x_t)\|_2$

Reconstruct diffusion SDE:

$dx_t = \hat{D}^{(1)}(x_t)\,dt + \sqrt{2\hat{D}^{(2)}(x_t)}\,dW_t$

Bootstrap the null distribution $\mathcal{D}(\kappa|\mathcal{H}_0)$:

for $j = 1$ to M **do**

  Generate surrogate path ω_j using the reconstructed SDE

  Estimate fourth KM coefficient $\hat{D}_j^{(4)}(x_t)$ for ω_j

  Compute κ_j using $\hat{D}_j^{(4)}(x_t)$

Compare κ_{observed} against $\mathcal{D}(\kappa|\mathcal{H}_0)$:

if κ_{observed} < q-th quantile of $\mathcal{D}(\kappa|\mathcal{H}_0)$ **then**

  conclude $\omega \in \Omega_0$

else

  conclude $\omega \in \Omega_1$

autocorrelation (Dickey & Fuller, 1979). The CMI test assesses whether a process follows a Markov property of a given order by measuring how much new information the present adds about the future after considering the past (Papapetrou & Kugiumtzis, 2013).

Estimation of Kramers-Moyal (KM) Coefficients. To estimate the KM coefficients, we employ a non-parametric approach. Specifically, we used the Nadaraya-Watson estimator to compute the KM coefficients from a time series, as described in (Gorjão & Meirinhos, 2019). This kernel-based method estimates conditional moments, $K^{(n)}(x_t)$, and offers several advantages over traditional histogram-based approaches. Unlike histograms, it operates in a continuous (non-binned) space and allows flexibility in kernel selection, leading to more reliable results. Additionally, using kernel convolution avoids the high memory demands associated with sequential array summation. It is also particularly effective for short time series, as it incorporates all data points, possibly assigning greater influence to some based on the kernel.

Recall from Equation 4:

$$\begin{aligned} D^{(n)}(x_t) &= \frac{1}{n!}\lim_{\tau\to 0}\frac{1}{\tau}K_\tau^{(n)}(x_t) \\ &= \frac{1}{n!}\lim_{\tau\to 0}\frac{1}{\tau}\left\langle [x(t+\tau)-x(t)]^n \,\middle|\, x(t)\right\rangle \end{aligned}$$

And so, given a time series of length m, one can estimate the n-th conditional moment in the above equation as:

$$\begin{aligned} &\left\langle [x(t+\tau)-x(t)]^n \,\middle|\, x(t)=x\right\rangle \\ &= \left\langle [x_{j+1}-x_j]^n \,\middle|\, x_j = x\right\rangle \\ &= \frac{\sum_{j=1}^m k\left(\frac{x_j-x}{h}\right)(x_{(j+1)}-x_j)^n}{\sum_{j=1}^m k\left(\frac{x_j-x}{h}\right)} \end{aligned} \qquad [13]$$

where the change of notation in the first line formalizes the fact that the expectation is being taken over a finite sample. In the second line, h is the bandwidth of the kernel $k(u)$, which satisfies the condition $\int u^2 k(u)\,du < \infty$. In general, the kernel k is any smooth function that meets the following criteria: $k(x) \geq 0$, $\int k(x)\,dx = 1$, and $\int xk(x)\,dx = 0$. For this application we used a uniform kernel and the bandwidth was estimated using Silverman's rule of thumb (Silverman, 2018).

Reconstruct the diffusion SDE. To reconstruct the SDE in Equation 12 using the estimated coefficients $\hat{D}^{(1)}(x_t)$ and $\hat{D}^{(2)}(x_t)$, it is necessary to first fit smooth functional models to these quantities. Since the KM coefficients are computed empirically at discrete points—resulting in vector-valued estimates rather than actual functions that could be evaluated—they cannot be directly used in simulation. To address this, we can fit a regression model to the empirical estimates of $\hat{D}^{(1)}(x_t)$ and $\hat{D}^{(2)}(x_t)$ across the observed range of x_t. For example, a spline regression or a Gaussian process model can flexibly map a vector to a function. These fitted models can then be used as the drift and diffusion terms, respectively, in the reconstruction of the SDE to generate surrogate time series.

Bootstrap the null distribution, $\mathcal{D}(\kappa|\mathcal{H}_0)$. We bootstrap the distribution of the κ-statistic under the null assumption of diffusion sufficiency by simulating a large number (e.g., M = 1000 or more) of time series using the reconstructed SDE. We use the same time resolution and duration as the original data in this surrogate series generation. Using the same KM estimation procedure as in §, we estimate the 4th KM coefficients then compute κ-statistic for the surrogate data.

For each simulated time series j, estimate $\hat{D}_j^{(4)}(x_t)$ and compute the corresponding κ-statistic:

$$\kappa_j = \|\hat{D}_j^{(4)}(x_t)\|_2 \approx \left(\sum_{t=1}^n |\hat{D}_j^{(4)}(x_t)|^2\right)^{1/2}$$

forming an empirical distribution of κ-values under the assumption of diffusion sufficiency.

Decision rule. We note that since the κ-statistic is computed using a distance measure, its value is bounded below by zero. The farther the observed κ-statistic from the lower bound, the more pronounced the role of higher-order KM coefficients is.

Hence, we can employ a decision rule where if κ_{observed} is below a certain quantile of the bootstrap distribution then we conclude that the time series is sufficiently described by diffusion, otherwise we reject the null hypothesis and conclude that the process exhibits variation that is not fully explained by diffusion.

The choice of quantile determines the sensitivity and selectivity of the test with higher quantile suggesting we become more permissive in what counts as diffusion.

Monte Carlo Study

We conducted simulation studies to evaluate the performance of the test. In addition, we checked how the choice of quantile affects the test results.

Setup and choice of parameters. We use the following model to simulate paths that are diffusive and that contain additional variation in the form of jumps:

$$dx_t = a(x_t)\,dt + b(x_t)\,dW_t + \xi\,dJ_t(\lambda) \qquad [14]$$

where W_t is a standard Brownian motion, $a(x_t)$ and $b(x_t)$ are some functions of the state, ξ is a random variable indicating the magnitude jumps and $J_t(\lambda)$ is a Poisson process counting the jumps, with rate λ. The jump term provides a source of variability in addition to that due to diffusion.

Table 1. Parameter settings for simulation studies

	Diffusion parameters		Jump parameters	
Case	$a(x_t)$	$b(x_t)$	σ_ξ	λ
OU-1	$-0.1\,x_t$ (linear)	0.1 (constant)	–	–
OU-2	$-0.1\,x_t$ (linear)	0.3 (constant)	–	–
OU-3	$-0.1\,x_t$ (linear)	1 (constant)	–	–
CIR-1	$0.1(0.15 - x_t)$ (linear)	$0.1\sqrt{x_t}$ (square root)	–	–
CIR-2	$0.1(0.55 - x_t)$ (linear)	$0.3\sqrt{x_t}$ (square root)	–	–
CIR-3	$0.1(5.1 - x_t)$ (linear)	$1\sqrt{x_t}$ (square root)	–	–
JD-1	$-0.1\,x_t$ (linear)	0.3 (constant)	0.1	0.1
JD-2	$-0.1\,x_t$ (linear)	0.3 (constant)	0.5	0.1
JD-3	$-0.1\,x_t$ (linear)	0.3 (constant)	1.0	0.1
JD-4	$-0.1\,x_t$ (linear)	0.3 (constant)	0.1	0.5
JD-5	$-0.1\,x_t$ (linear)	0.3 (constant)	0.5	0.5
JD-6	$-0.1\,x_t$ (linear)	0.3 (constant)	1.0	0.5

Table 2. Simulation studies results.

	← more conservative			more liberal →		
Case	q=0.999	q=0.99	q=0.98	q=0.97	q=0.96	q=0.95
OU-1	0.000	0.000	0.010	0.020	0.025	0.025
OU-2	0.000	0.000	0.000	0.005	0.005	0.015
OU-3	0.000	0.020	0.040	0.050	0.060	0.070
CIR-1	0.015	0.045	0.055	0.080	0.090	0.125
CIR-2	0.020	0.090	0.110	0.140	0.155	0.170
CIR-3	0.005	0.035	0.080	0.115	0.135	0.150
JD-1	0.555	0.785	0.825	0.845	0.885	0.890
JD-2	0.560	0.690	0.740	0.760	0.780	0.785
JD-3	0.360	0.495	0.550	0.570	0.575	0.580
JD-4	0.760	0.895	0.930	0.945	0.945	0.945
JD-5	0.645	0.800	0.835	0.840	0.855	0.860
JD-6	0.725	0.905	0.940	0.945	0.955	0.960

Note: Reported results are empirical rejection rates of the test based on 200 Monte Carlo replications for each case and where each time series considered is of length $n = 1000$.

We explore settings where $a(x_t)$ is a linear function of the state and $b(x_t)$ is either a constant function or a square root function of the state. For our purposes, we assume that the magnitude of jumps follows a normal distribution centered at 0 with modifiable variance, $\xi \sim \mathcal{N}(0, \sigma_\xi^2)$. Further, we choose to modify the rate of jumps, λ in $J_t\,(\lambda)$.

Table 1 shows the different cases we considered for the simulation study. In total we have 12 cases, half of which generated time series that are considered purely diffusive, and another half that contain jumps. Among the diffusive cases, we considered either an Ornstein-Uhlenbeck (OU) process or a Cox-Ingersoll-Ross (CIR) process. These two differ in the magnitude of diffusive variation with OU having a constant term and CIR containing a square root function. In the simulation of the time series for CIR we ensured that the Feller condition was met to ensure positivity of the generated series (Albrecher, Mayer, Schoutens, & Tistaert, 2007). In addition to guaranteeing stationarity in the generated series, we chose the mean or attractor parameter in $a(x_t)$ to be of the form $\mu = \frac{v^2}{2s} + 0.1$, where speed s and volume v correspond to the constant multipliers of $a(x_t)$ and $b(x_t)$, respectively.

Since $J_t\,(\lambda)$ in the model (Eq. 14) is a Poisson process, if the rates of jump λ are too small, the simulated processes may not violate the notion of diffusion (because at low but nonzero jump rates, the number of jumps may be 0). Hence, in simulating the six cases that correspond to the alternative hypothesis, we required for the procedure to only generate series with at least one jump. We vary the jump rate, λ, and the variance of the jump sizes, ξ, to test the sensitivity of the test. All models were simulated using the Euler-Maruyama scheme and we chose the simulated time series to be of length $n = 1000$.

Choice of regression model for the reconstruction of the diffusion SDE. To be able to use the estimated first and second KM coefficients for the SDE reconstruction, we fitted spline regression models to the empirical estimates of $\hat{a}(x_t) = \hat{D}^{(1)}(x_t)$ and $\hat{b}(x_t) = \sqrt{2\,\hat{D}^{(2)}(x_t)}$ across the observed range of x_t. Cubic spline basis were used with 10 interior knots to allow flexibility and to be able to capture potential nonlinear relationships. The coefficients for the spline basis functions were then estimated via ordinary least squares regression. These fitted regression models were then used to generate $M = 1000$ surrogate paths used to bootstrap $\mathcal{D}(\kappa|\mathcal{H}_0)$.

Monte Carlo results. Table 2 presents a summary of the simulation results. It reports the empirical rejection rates of the proposed test, computed from 200 replications for each simulation setting. The table also provides the rejection frequencies corresponding to the chosen quantile threshold used in the decision rule.

Overall, the results demonstrate that the test performs well in correctly identifying processes that are genuinely diffusive, as evidenced by low rejection rates under the true diffusion cases. When the quantile is set to 0.999, the test almost never rejects the null hypothesis for purely diffusive processes—indicating that the test recognizes them as sufficiently described by diffusion alone. Conversely, for processes that include additional variation through jumps, the observed κ_{observed} often exceeds the maximum κ_j within the null distribution $\mathcal{D}(\kappa \mid \mathcal{H}_0)$, leading to a rejection of the null as expected.

Furthermore, we observe the anticipated pattern when varying the chosen quantile: lowering the quantile threshold makes the test more liberal, increasing the probability of rejecting the null hypothesis.

Figure 2 illustrates representative examples from the OU-2, CIR-2, and JD-2 cases. These examples highlight the accuracy of the test in distinguishing between diffusive and jump-diffusive time series, confirming the method's reliability across different process types. The presented JD-2 case example had three jump instances and with the chosen quantile threshold of 95%, we see that the test reliably diagnoses the time series as not diffusive.

We report receiver operating characteristic (ROC) curves in Figure 3 summarizing the performance of the proposed test across various quantile thresholds for different null–alternative pairs. Four ROC curves are shown, corresponding to the test's performance on the following pairs: (1) OU-2 vs. JD-5, (2) OU-3 vs. JD-2, (3) OU-3 vs. JD-1, and (4) CIR-1 vs. JD-1. Although ROC curves were generated for all 36 possible null–alternative combinations, these four were selected to illustrate the range from the best to the weakest performance of the test.

When applied to the OU-2 and JD-5 series, the test achieved an area under the curve (AUC) of 0.968, indicating near-perfect discrimination. For the OU-3 and JD-2 pair, the AUC was 0.869; for OU-3 and JD-1, 0.778; and for CIR-1 and JD-1, 0.701, representing the lowest observed performance. The reported AUC values demonstrate strong test performance. An AUC close to 1 signifies

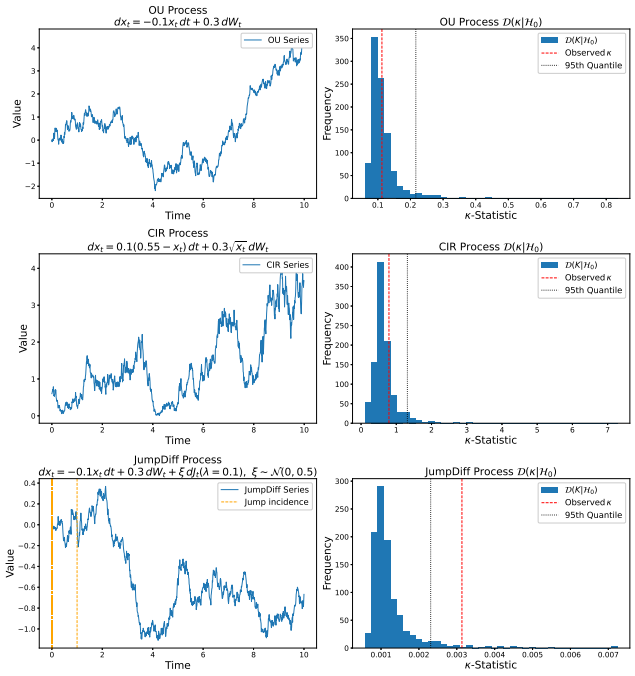

Fig. 2. Example results of the proposed test for three representative cases: OU-2, CIR-2, and JD-2 (see Tab. 1). The panels on the left display the corresponding time series for each case. The OU-2 (top) and CIR-2 (middle) series originate from pure diffusion processes, while the JD-2 series (bottom) includes an additional source of variation in the form of jumps, indicated by the orange dashed lines. The blue histograms in the right panels show the distribution $\mathcal{D}(\kappa \mid \mathcal{H}_0)$ for each example. The observed κ-statistic is marked by the red dashed line, and the 95th quantile of $\mathcal{D}(\kappa \mid \mathcal{H}_0)$ is shown as the black dotted vertical line. With the chosen quantile threshold of $q = 0.95$, all three examples are correctly classified.

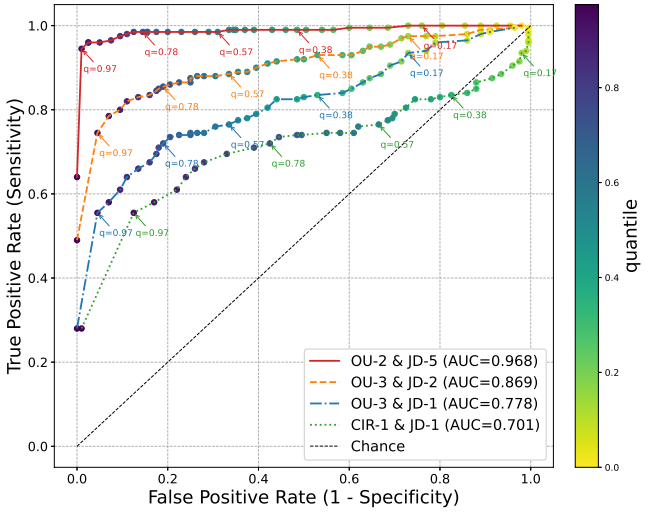

Fig. 3. Receiver operating characteristic (ROC) curves illustrating the performance of the proposed test across various quantile thresholds for different null-alternative process pairs. The figure displays results for four representative cases: (1) OU-2 vs. JD-5, (2) OU-3 vs. JD-2, (3) OU-3 vs. JD-1, and (4) CIR-1 vs. JD-1. These examples were selected from the 36 possible null–alternative combinations to reflect the range from the strongest to the weakest test performance. The corresponding areas under the curve (AUCs) are 0.968, 0.869, 0.778, and 0.701, respectively. Higher AUC values indicate better discriminative ability between diffusive and non-diffusive (jump) processes. Overall, the results demonstrate that the proposed test achieves strong classification performance and maintains a good balance between sensitivity and specificity across a variety of process pairs.

that the method effectively distinguishes between diffusive and non-diffusive (jump) processes, achieving a good balance between sensitivity and specificity.

Empirical application

Mobile health data application. We applied the proposed test to a subset of data from a mobile health (mHealth) intervention study designed to promote psychological well-being among college students (Heshmati et al., 2025). In the study, participants completed brief daily wellness surveys assessing various dimensions of psychological well-being (PWB). Participants received notifications on their phones prompting them to report their well-being levels. During the initial laboratory session, participants were randomly assigned to one of three groups: a control group (Control), a positive practice intervention group (PPI), or a meditation-enhanced positive practice group (PCPI). The study began with a 14-day baseline phase without intervention, during which participants only completed momentary PWB assessments. This was followed by a 15-day intervention phase, during which participants in the control group performed an evening working-memory task, while those in the intervention groups engaged in positive or contemplative practices (e.g., meditation, reflecting on three good things they did during the day). Then followed a 28 day post-intervention phase.

We implemented our new test on two participants' self-reported emotional arousal levels – that is, the intensity of emotions experienced at the time of reporting. One participant (ID 171) belonged to the control group, while the other (ID 139) was part of the PCPI group. The emotional arousal level series of these participants passed our pretests for stationarity and Markovianity in §.

Figure 4 shows the test results for both participants' emotional arousal trajectories over the course of the study. Using a quantile threshold of 95%, the time series for Participant 139 was classified as diffusive, whereas Participant 171's was not, suggesting the presence of additional sources of variation beyond diffusion.

Qualitatively, the two participants showed distinct temporal patterns. Participant 139's emotional arousal reports remained relatively stable between 0.4 and 0.8, with only minor fluctuations. In contrast, Participant 171's trajectory displayed marked shifts between high and low values during the first half of the study, characterized by clusters of consecutive reports near each jump – most notably between weeks 1.5-2 and 4-4.5. In the latter half, Participant 171 reported generally higher arousal levels, interrupted by brief low-value episodes that quickly recover in the subsequent measurement. The pattern exhibited in the first half of the study is consistent with the behavior expected from a jump-diffusion process.

EEG application. We apply our approach to electroencephalogram (EEG) recordings collected from pre-surgical epilepsy patients (Andrzejak et al., 2001). The analysis considers three representative single-channel EEG time series (Fig. 5), each with a duration of 23.6 seconds and preprocessed to remove artifacts. The EEG data were obtained from the epileptogenic zone—regions of the brain capable of precipitating epileptic activity—using a 128-channel amplifier system with an average common reference. Signals were band-pass filtered between 0.53 and 40Hz, digitized with 12-bit precision, and sampled at 173.61Hz (Andrzejak et al., 2001, pp. 2-3). These EEG time series were also subjected to our pretests for stationarity and Markovianity in §.

Using a quantile threshold of 95%, segment F097 was identified as sufficiently diffusive, while segments F086 and F076 were

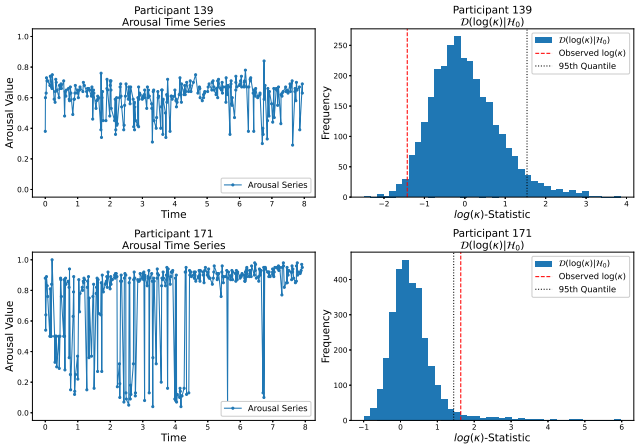

Fig. 4. Test results for two participants' self-reported emotional arousal levels over the course of the study. Using a quantile threshold of 95%, the time series for subject 139 (intervention group) was classified as sufficiently diffusive, while subject 171 (control group) was not, indicating additional sources of variation beyond diffusion. Subject 139's trajectory remained relatively stable between 0.4 and 0.8, whereas subject 171's exhibited pronounced shifts between high and low values in the early phase (weeks 1.-2 and 4-4.5) and higher arousal with brief low-value dips later in the study, consistent with jump-diffusion behavior.

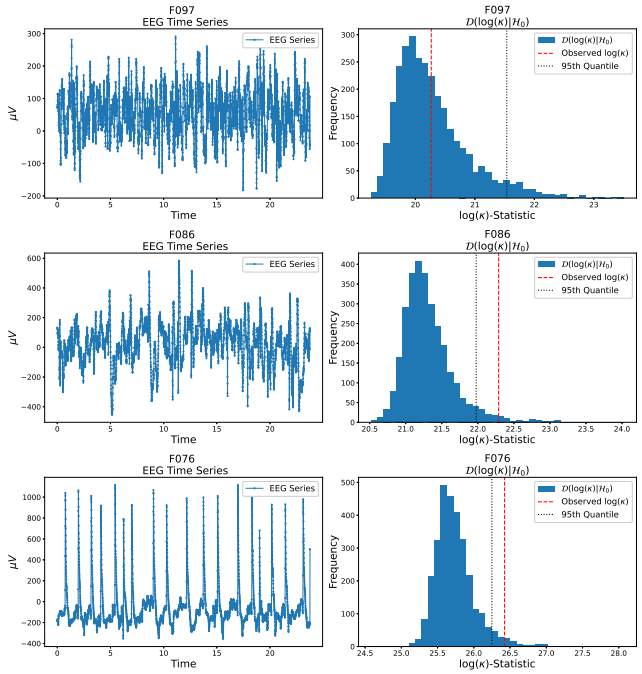

Fig. 5. Test results for three representative EEG segments. Using a quantile threshold of 95%, segment F097 was classified as sufficiently diffusive, whereas segments F086 and F076 were not.

not. This outcome aligns with the qualitative distinctions observed among the three segments. Segment F076 shows clear periodic positive jumps, suggesting variability that deviates from purely diffusive behavior. In contrast, segment F097 closely resembles the canonical diffusion-like time series and is accordingly classified as diffusive by the test. Segment F086 appears largely diffusive but displays intermittent jump-like behavior, leading the test to classify it as not diffusive.

Discussion

The simulation results provide evidence that the proposed test reliably distinguishes between purely diffusive and jump-diffusive processes. Across all simulation settings, the test showed low rejection rates under true diffusion, confirming its ability to avoid false positives. Conversely, when jump components were introduced, the test successfully rejected the null hypothesis. The ROC analyses further support these findings, showing high AUC values across a range of diffusion–jump comparisons. These results indicate that the test achieves a favorable balance between sensitivity and specificity, with performance improving as the frequency of jumps increase.

Applying the test to the mHealth well-being data further demonstrates its value in real-world behavioral settings. Psychological and emotional states, as captured through intensive self-reports, often exhibit both gradual fluctuations and abrupt transitions driven by contextual or internal events. In our application, one participant's emotional arousal trajectory (Participant 139) was characterized as diffusive, consistent with stable affective dynamics over time. In contrast, another participant (Participant 171) showed strong deviations from diffusion, displaying intermittent jumps between emotional states — a pattern that may reflect episodic shifts in mood or responsiveness to intervention demands. These findings align with the simulation outcomes and illustrate how the test can capture meaningful distinctions in temporal dynamics and be used as a preliminary check for the sufficiency of diffusion as a model.

Another application of the test was made to EEG data collected from pre-surgical epilepsy patients. Among three representative EEG time series, we found two that showed clear deviations from purely diffusive behavior. This has implications for the study of epileptic dynamics since earlier work that tried to characterize epileptic dynamics from EEG recordings relied on the assumption of diffusion (Prusseit & Lehnertz, 2007). More recent work examined jump-diffusion processes epilepsy (Anvari, Tabar, Peinke, & Lehnertz, 2016) hinting that additional variation is necessary to effectively capture the the dynamics. Our test results provide further evidence for models that go beyond simple diffusion.

At present, the test considers the diffusion class defined by Equation 1, in which neither $a(x_t)$ nor $b(x_t)$ are explicit functions of time, t – time factors in only indirectly as it affects x_t. This is due to a current limitation in the estimation procedure for the KM coefficients, where instead of estimating $D^{(n)}(x,t)$ we are restricted to only estimating $D^{(n)}(x_t)$. The current estimation procedure relies on kernel-based methods and incorporating potential time dependence, t, would require substantially longer time series to maintain estimation precision. Further methodological advances in the estimation of KM coefficients are therefore necessary to capture possible time-varying effects, $D^{(n)}(x,t)$.

As a consequence, rejection of the null hypothesis under the current test indicates that purely endogenous diffusion—that is, diffusion dynamics independent of time caused only by serial dependence of the variate on its previous value—is not sufficient to explain the observed behavior. Accordingly, users of the test are encouraged to further examine whether such deviation arises from additional sources of variation, such as discontinuous jumps, or whether incorporating time-varying effects would adequately account for the observed dynamics.

Conclusion

We develop a test to assess whether a diffusion process is sufficient to describe an observed time series. The test is based on

the results from Kramers–Moyal (KM) expansion and the Pawula theorem, which says that if an observed time series is already Markov and exhibits stochasticity, then there are only two possibilities: either it is diffusive which corresponds to the truncation of the KM expansion at the second term or it is not, meaning all terms in the KM expansion are needed. We evaluated the performance of the test by checking how it fares when decision threshold are varied and found overall good performance of the test.

Applying the proposed test to real-world data from a mobile health well-being intervention study revealed evidence for the presence of additional sources of variation beyond diffusion. Specifically, one example participant's self-reported emotional arousal trajectory deviated from the null hypothesis of pure diffusion, displaying temporal patterns consistent with jump-like dynamics. In addition, application to EEG data from pre-surgical epileptic patients revealed that epileptic dynamics might be best studies by adding variation beyond diffusion.

Methodological advancement in the estimation of time-varying KM coefficients can aid in future work to extend the test to diffusion processes that are not purely endogenous and that vary over time.

References

Aït-Sahalia, Y. (2002). Telling from discrete data whether the underlying continuous-time model is a diffusion. *The Journal of Finance*, *57*(5), 2075–2112.

Albrecher, H., Mayer, P., Schoutens, W., & Tistaert, J. (2007). The little Heston trap. *Wilmott*(1), 83–92.

Andrzejak, R. G., Lehnertz, K., Mormann, F., Rieke, C., David, P., & Elger, C. E. (2001). Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Phys. Rev. E*, *64*, 061907.

Anvari, M., Tabar, M. R. R., Peinke, J., & Lehnertz, K. (2016). Disentangling the stochastic behavior of complex time series. *Scientific reports*, *6*(1), 35435.

Bahraminasab, A., Ghasemi, F., Stefanovska, A., McClintock, P., & Friedrich, R. (2009). Physics of brain dynamics: Fokker–Planck analysis reveals changes in EEG δ–θ interactions in anæsthesia. *New journal of physics*, *11*(10), 103051.

Bernardo, J. M., & Smith, A. F. (2009). *Bayesian theory* (Vol. 405). Sussex: John Wiley & Sons.

Boninsegna, L., Nüske, F., & Clementi, C. (2018). Sparse learning of stochastic dynamical equations. *The Journal of chemical physics*, *148*(24), 241723.

Detemple, J., & Rindisbacher, M. (2011). Diffusion models of asset prices. *Handbook of Computational Finance*, 35–60.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, *74*(366a), 427–431.

Einstein, A. (1905). Über die von der molekularkinetischen Theorie der wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Annalen der Physik*, *322*(8), 549–560. doi: 10.1002/andp.19053220806

Etz, A., Goodman, S. N., & Vandekerckhove, J. (2022). Statistical inference in behavioral research: Traditional and Bayesian approaches. In L. J. Jussim, J. A. Krosnick, & S. T. Stevens (Eds.), *Research integrity: Best practices for the social and behavioral sciences* (p. 175–202). Oxford University Press.

Etz, A., Haaf, J., Rouder, J., & Vandekerckhove, J. (2018). Bayesian inference and testing any hypothesis you can specify. *Advances in Methods and Practices in Psychological Science*, *1*, 281–295. doi: 10.1177/2515245918773087

Ghasemi, F., Peinke, J., Reza Rahimi Tabar, M., & Sahimi, M. (2006). Statistical properties of the interbeat interval cascade in human hearts. *Int. J. Mod. Phys. C*, *17*(04), 571–580.

Ghasemi, F., Sahimi, M., Peinke, J., & Tabar, M. R. R. (2006). Analysis of non-stationary data for heart-rate fluctuations in terms of drift and diffusion coefficients. *Journal of biological physics*, *32*(2), 117–128.

Gorjão, L. R., & Meirinhos, F. (2019). kramersmoyal: Kramers–Moyal coefficients for stochastic processes. *Journal of Open Source Software*, *4*(44), 1693. doi: 10.21105/joss.01693

Heshmati, S., Muth, C., Li, Y., Roeser, R. W., Smyth, J. M., Vandekerckhove, J., ... Oravecz, Z. (2025). Who benefits from mobile health interventions? a dynamical systems analysis of psychological well-being in early adults. *Applied Psychology: Health and Well-Being*, *17*(3), e70037.

Oravecz, Z., Tuerlinckx, F., & Vandekerckhove, J. (2009). A hierarchical Ornstein-Uhlenbeck model for continuous repeated measurement data. *Psychometrika*, *74*, 395–418.

Oravecz, Z., & Vandekerckhove, J. (2020). A joint process model of consensus and longitudinal dynamics. *Journal of Mathematical Psychology*, *98*, 102386. doi: 10.31234/osf.io/xyghj

Papapetrou, M., & Kugiumtzis, D. (2013). Markov chain order estimation with conditional mutual information. *Physica A: Statistical Mechanics and its Applications*, *392*(7), 1593–1601.

Paul, W., & Baschnagel, J. (1999). *Stochastic processes*. Berlin: Springer.

Pawula, R. (1967). A modified version of Price's theorem. *IEEE Transactions on Information Theory*, *13*(2), 285–288.

Peinke, J., Tabar, M. R., & Wächter, M. (2019). The Fokker–Planck approach to complex spatiotemporal disordered systems. *Annual Review of Condensed Matter Physics*, *10*(1), 107–132.

Prusseit, J., & Lehnertz, K. (2007). Stochastic qualifiers of epileptic brain dynamics. *Phys. Rev. Lett.*, *98*(13), 138103.

Silverman, B. W. (2018). *Density estimation for statistics and data analysis*. New York: Routledge.

Smelyanskiy, V., Luchinsky, D., Stefanovska, A., & McClintock, P. (2005). Inference of a nonlinear stochastic model of the cardiorespiratory interaction. *Phys. Rev. Lett.*, *94*, 098101.

Tabar, R. (2019). *Analysis and data-based reconstruction of complex nonlinear dynamical systems*. Cham: Springer.

Tutkun, M. (2017). Markovian properties of velocity increments in boundary layer turbulence. *Physica D: Nonlinear Phenomena*, *351*, 53–61.

Uhlenbeck, G. E., & Ornstein, L. S. (1930). On the theory of the Brownian motion. *Physical Review*, *36*(5), 823–841.

Vandekerckhove, J., Tuerlinckx, F., & Lee, M. D. (2011). Hierarchical diffusion models for two-choice response times. *Psychological Methods*, *16*, 44–62. doi: 10.1037/a0021765

Vellmer, S., & Lindner, B. (2021). Fokker–Planck approach to neural networks and to decision problems: A unique method for stochastic models in computational and cognitive neuroscience. *Eur. Phys. J.: Spec. Top.*, *230*(14), 2929–2949.

Wiener, N. (1923). Differential-space. *Journal of Mathematics and Physics*, *2*(1–4), 131–174. doi: 10.1002/sapm192321131

Zhang, C., Peng, H., & Zhang, J.-T. (2010). Two samples tests for functional data. *Communications in Statistics—Theory and Methods*, *39*(4), 559–578.