



Sparse identification of nonlinear dynamics for stochastic delay differential equations

Dimitri Breda ^a, Dajana Conte ^b, Raffaele D'Ambrosio ^c, Ida Santaniello ^b,
Muhammad Tanveer ^a

^a CDLab - Computational Dynamics Laboratory Department of Mathematics, Computer Science and Physics - University of Udine, via delle Scienze 206, Udine, 33100, Italy

^b Department of Mathematics, University of Salerno, Via Giovanni Paolo II 132, Fisciano (SA), 84084, Italy

^c Department of Information Engineering and Computer Science and Mathematics, University of L'Aquila, Vetoio, Loc. Coppito, L'Aquila, 67100, Italy

ARTICLE INFO

2020 MSC:

60H10

34K50

68T05

93E12

Keywords:

Sparse identification

Stochastic differential equations

Delay differential equations

Drift/diffusion estimation

ABSTRACT

A general framework for recovering drift and diffusion dynamics from sampled trajectories is presented for the first time for stochastic delay differential equations. The core relies on the well-established SINDy algorithm for the sparse identification of nonlinear dynamics. The proposed methodology combines recently proposed high-order estimates of drift and covariance for dealing with stochastic problems with augmented libraries to handle delayed arguments. Three different strategies are discussed in view of exploiting only realistically available data. A thorough comparative numerical investigation is performed on different models, which helps guiding the choice of effective and possibly outperforming schemes.

1. Introduction

Stochastic differential equations (SDEs) provide a natural framework for modeling systems subject to random fluctuations. Among them, stochastic ordinary differential equations (SODEs) describe the evolution of systems whose future state depends on their current state and random perturbations, typically modeled by Brownian motion. In contrast, stochastic delay differential equations (SDDEs) incorporate memory effects by allowing the system's dynamics to depend not only on the present but also on past states. The analysis of such systems have been widely explored in the literature, both from the theoretical and numerical perspectives, as well as from a data-driven point of view [1–5]. provide an excellent review of these and relevant topics.

In the last years, there has been a growing interest in the identification of stochastic dynamical systems from observed time-series data. In this context, Sparse Identification of Nonlinear Dynamics (SINDy) has emerged as a powerful data-driven algorithm for discovering governing equations from trajectories [6,7]. The principal aim of SINDy is to identify parsimonious and interpretable models from data, although its performance may decline when only limited measurements are available, due to its sensitivity to data sparsity and noise. Several works have applied SINDy to SODEs [8,9] and deterministic delay differential equations (DDEs) [10–13], while its extension to SDDEs has not yet been fully investigated, owing to its technical and computational difficulties. This work aims to address these challenges by extending the SINDy framework to the SDDEs setting. In this regard, we contrast our

E-mail addresses: dimitri.breda@uniud.it (D. Breda), dajconte@unisa.it (D. Conte), raffaele.dambrosio@univaq.it (R. D'Ambrosio), isantaniello@unisa.it (I. Santaniello), tanveer.muhammad@spes.uniud.it (M. Tanveer).

<https://doi.org/10.1016/j.cam.2025.117247>

Received 30 July 2025

Available online 11 December 2025

0377-0427/© 2025 The Authors.

Published by Elsevier B.V. This is an open access article under the CC BY license

(<http://creativecommons.org/licenses/by/4.0/>).

approach with [14], which is restricted to the Stratonovich formulation and whose reliance on small-delay Taylor approximations limits generalization to arbitrary delays.

We explore and compare several estimations strategies, including Kramers-Moyal (KM) formulas, forward differences (FD), central differences (CD) and trapezoidal-like schemes (TR), all based on Itô-Taylor expansions, for reconstructing drift and diffusion from time-series data and extend these estimation strategies to the framework of SDDEs, relying on an augmented state with delay. The resulting approximations are then used as input for the sparse regression performed by the SINDy algorithm. In this respect, we investigate three practical approaches (A, B1, B2) for recovering drift and diffusion data using the SINDy algorithm combined with the discussed Itô-Taylor-based estimators, under different data availability scenarios. The proposed approaches address the fundamental challenge of obtaining reliable estimates at each time step and are designed to complement the developed estimation methods. These strategies differ in terms of data requirements, statistical robustness, obtained accuracy and computational cost in terms of CPU time.

The paper is organized as follows. In Section 2, we introduce the notation and main concepts related to SODEs and SDDEs, focusing on how to recover the KM formulas and their higher-order extensions for estimating drift and diffusion terms, based on Itô-Taylor expansions. Section 3 provides a concise overview of the original SINDy framework for ordinary differential equations (ODEs) and its extensions first to delay differential equations (DDEs) and second to SODEs. The discussion then moves to SDDEs, highlighting how the method can be adapted to account for both delayed and stochastic components. In Section 4, we turn to the practical use of the estimators introduced in Section 2 depending on the available data. We propose three different strategies for recovering drift and diffusion terms: Approach A, which generates synthetic paths from a single observed trajectory; Approach B1, which computes averaged estimates before sparse regression using multiple trajectories; and Approach B2, which applies sparse regression separately to each trajectory and averages the resulting models.

Their comparative performance is evaluated in Section 5, which presents a computational analysis on three representative SDDDE models. For each model, the three approaches A, B1 and B2 are tested in combination with all four estimation methods KM, FD, CD and TR discussed in Section 2. Synthetic data are generated via Euler-Maruyama simulations. Reconstruction quality is assessed using absolute errors on the identified sparse coefficients and root mean square errors (RMSE) on both reconstructed trajectories and estimated drift and diffusion components. Additionally, we analyze the impact of data availability and CPU time on the overall performance.

2. Theoretical background

In Section 2.1 we recall the basics of SODEs first and SDDDEs then: notation, key ingredients and well-posedness of the relevant initial value problems. In Section 2.2 we briefly review the associated Itô-Taylor expansions, on which Kramers-Moyal and higher-order estimates of drift and diffusion rely. They form the necessary theoretical foundation for extending the sparse identification framework of SINDy to both SODEs and SDDDEs. Useful general references for the underlying theoretical framework are [2–5,15,16].

2.1. Stochastic ordinary and delay differential equations

SODEs describe the evolution in time $t \geq 0$ of a real random variable $X(t) \in \mathbb{R}^n$ through the interplay of deterministic and stochastic forces. In its classical formulation without delays a SODE reads

$$dX(t) = f(X(t))dt + g(X(t))dW(t), \tag{1}$$

where $W(t)$ is a q -dimensional standard Wiener process, the *drift* function $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ governs the deterministic evolution and represents the expected direction of motion at each point, while the *diffusion* function $g : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times q}$ controls the intensity and structure of random fluctuations. We assume the *covariance* matrix $G := gg^T/2 \in \mathbb{R}^{n \times n}$ to be positive definite, thus ensuring that g is non-degenerate and can be uniquely determined from G up to a right orthogonal transformation [17,18]. The latter assumption is not restrictive and indeed many physical systems affected by noise fit into this framework [5,19]. Of course Eq. (1) is just a symbolic notation for

$$X(t) = X(0) + \int_0^t f(X(\sigma))d\sigma + \int_0^t g(X(\sigma))dW(\sigma),$$

where the last integral is intended in the Itô sense as far as this work is concerned. Under the standard global Lipschitz and linear growth conditions

$$\begin{aligned} \|f(x) - f(y)\| + \|g(x) - g(y)\| &\leq L\|x - y\|, \quad x, y \in \mathbb{R}^n, \\ \|f(x)\| + \|g(x)\| &\leq C(1 + \|x\|), \quad x \in \mathbb{R}^n, \end{aligned}$$

for L and C positive constants, the initial value problem for Eq. (1) admits a unique strong solution with continuous sample paths (for a proof see, e.g., [5]). We observe that alternative hypotheses may also be assumed, e.g., one-sided Lipschitz continuous drift and locally Lipschitz continuous diffusion, in order to ensure bounded moments for the solution [20].

Many real-world systems exhibit memory effects, for which the current evolution of the state depends also on past values [21,22]. In the presence of randomness SDDDEs capture this phenomenon by incorporating explicit time delays into the dynamics [1,16,23,24]. Here we consider SDDDEs with a single discrete and constant delay $\tau > 0$ affecting both drift and diffusion:

$$dX(t) = f(X(t), X(t - \tau))dt + g(X(t), X(t - \tau))dW(t). \tag{2}$$

Extension to multiple delays is not conceptually demanding, even though it might involve some technicalities. In Eq. (2) drift and diffusion are described, respectively, by $f : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ and $g : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^{n \times q}$. A related initial value problem requires specifying an initial history through a random variable $\varphi(t)$ defined on $[-\tau, 0]$ by imposing $X(t) = \varphi(t)$ for $t \in [-\tau, 0]$. Unlike standard SODEs, solutions to SDDEs are generally non-Markovian due to their dependence on the past [16]. Nevertheless, the initial value problem for Eq. (2) becomes well-posed under appropriate Lipschitz and linear growth conditions whose general form reads

$$\begin{aligned} \|f(\varphi) - f(\psi)\| + \|g(\varphi) - g(\psi)\| &\leq L\|\varphi - \psi\|_C, \quad \varphi, \psi \in C, \\ \|f(\varphi)\| + \|g(\varphi)\| &\leq C(1 + \|\varphi\|_C), \quad \varphi \in C, \end{aligned}$$

for drift and diffusion given as maps $C \rightarrow \mathbb{R}^n$ acting on the full history defined as a function $[-\tau, 0] \ni s \mapsto X(t)(s) := X(t + s)$ for $X(t) \in C := C([-\tau, 0]; \mathbb{R}^n)$ the classic state space of continuous functions equipped with the uniform norm $\|\varphi\|_C := \max_{s \in [-\tau, 0]} \|\varphi(s)\|$ [1,16].

As anticipated qualitatively, our goal is to estimate the assumed to be unknown drift f and diffusion g from trajectory data given in terms of the m time samples $X(t_1), \dots, X(t_m)$, possibly related to multiple paths. Note once more that one can recover g uniquely once G has been identified under the assumption that the latter is positive definite.

2.2. Estimates of drift and diffusion via Itô-Taylor expansions

Practical estimation methods for f and G are based on Itô-Taylor expansions, which express increments of test functions of a stochastic process in terms of differential operators and multiple stochastic integrals. They provide a systematic way to approximate solutions of SODEs over short time intervals, analogous to Taylor series for solutions of ODEs. Based on such expansions, the recent paper [9] discusses several higher-order extensions of the standard first-order Kramers-Moyal formulas. The latter were originally proposed in [8] as a basis for extending SINDy to SODEs. Here we first recall from [9] the main ingredients, presenting the estimation methods of interest in this work, to be later extended to SDDEs and then compared in the numerical Section 5. The extension to SDDEs is illustrated only at the end relying on an obvious argument. For a complete reference on stochastic-Taylor expansions for SODEs see [3, Chapter 5]; as for SDDEs we refer to [25–27]. We assume throughout that f and g in Eq. (1) (and then in Eq. (2)) are sufficiently smooth with bounded derivatives to ensure the validity of the considered expansions as well as meaningful estimates of related remainders.

For the SODE Eq. (1) let

$$L^0 := \sum_{i=1}^n f_i(x) \frac{\partial}{\partial x_i} + \sum_{i,j=1}^n g_{i,j}(x) \frac{\partial^2}{\partial x_i \partial x_j}, \quad L^j := \sum_{i=1}^n g_{i,j}(x) \frac{\partial}{\partial x_i}, \quad j = 1, \dots, q,$$

be differential operators acting on sufficiently smooth maps $\mathbb{R}^n \rightarrow \mathbb{R}$, where subscripts denote components of vectors and matrices. L^0 captures both the drift and the second-order effects of diffusion (the Itô correction), while L^j represents the first-order effect of the j -th noise component. For a specific test function ϕ the first few terms of the corresponding Itô-Taylor expansion read

$$\begin{aligned} \phi(X(t + \Delta t)) &= \phi(X(t)) + L^0 \phi(X(t)) \Delta t + \sum_{j=1}^n L^j \phi(X(t)) \Delta W_j(t) \\ &\quad + \frac{1}{2} (L^0)^2 \phi(X(t)) \Delta t + \sum_{j=1}^n L^j L^0 \phi(X(t)) I_{j,0} \\ &\quad + \sum_{j=1}^n L^0 L^j \phi(X(t)) I_{0,j} + \sum_{j,k=1}^n L^j L^k \phi(X(t)) I_{j,k} + \dots \end{aligned} \tag{3}$$

where $\Delta W_j(t) := W_j(t + \Delta t) - W_j(t)$ is the increment of the j -th component of the underlying Brownian motion and the terms I^j represent the multiple stochastic integrals

$$\begin{aligned} I_{j,0} &:= \int_t^{t+\Delta t} \int_t^{\sigma_1} dW_j(\sigma_2) d\sigma_1, \\ I_{0,j} &:= \int_t^{t+\Delta t} \int_t^{\sigma_1} d\sigma_2 dW_j(\sigma_1), \\ I_{j,k} &:= \int_t^{t+\Delta t} \int_t^{\sigma_1} dW_j(\sigma_2) dW_k(\sigma_1). \end{aligned}$$

Eq. (3) can be used in weak form by taking the conditional expectation of $\phi(X(t + \Delta t))$ given $X(t)$, or in strong form by working pathways with $\phi(X(t + \Delta t))$ thus retaining stochasticity. For instance, using the weak form with $\phi(x) := x_i$ for some $i = 1, \dots, n$ leads to

$$\mathbb{E}[x_i(t + \Delta t) \mid X(t) = x] = x_i + f_i(x) \Delta t + O(\Delta t^2)$$

from which the well-known Kramers-Moyal formula

$$f_i(x) \approx \mathbb{E} \left[\frac{X_i(t + \Delta t) - X_i(t)}{\Delta t} \mid X(t) = x \right] \tag{4}$$

follows as a component-wise estimate of the drift. Similarly, for the covariance

$$G_{i,j}(x) \approx \mathbb{E} \left[\frac{(X_i(t + \Delta t) - X_i(t))(X_j(t + \Delta t) - X_j(t))}{2\Delta t} \mid X(t) = x \right] \tag{5}$$

follows by choosing $\phi(x) := (x_i - X_i(t))(x_j - X_j(t))$. In the strong (pathways) version Eqs. (4) and (5) become

$$f_i(X(t)) \approx \frac{X_i(t + \Delta t) - X_i(t)}{\Delta t} \tag{6}$$

$$G_{i,j}(X(t)) \approx \frac{(X_i(t + \Delta t) - X_i(t))(X_j(t + \Delta t) - X_j(t))}{2\Delta t}. \tag{7}$$

In either weak or strong form, in the sequel we denote by KM these standard first-order estimates [28–30].

Following [9], improvements rely in general on higher-order finite-differences. In this work we consider only second-order estimates resorting respectively to forward, central and trapezoidal-like differences as illustrated next (only in the strong version for brevity).

- **Forward Differences (FD).** A second-order forward finite-differences scheme follows for both drift and covariance by considering more terms in the Itô-Taylor expansion Eq. (3):

$$f_i(X(t)) \approx \frac{4(X_i(t + \Delta t) - X_i(t)) - (X_i(t + 2\Delta t) - X_i(t))}{2\Delta t}, \tag{8}$$

$$G_{i,j}(X(t)) \approx \frac{1}{4\Delta t} [4(X_i(t + \Delta t) - X_i(t))(X_j(t + \Delta t) - X_j(t)) - (X_i(t + 2\Delta t) - X_i(t))(X_j(t + 2\Delta t) - X_j(t))]. \tag{9}$$

- **Central Differences (CD).** Similarly, second order can be reached in principle by

$$f_i(X(t)) \approx \frac{X_i(t + \Delta t) - X_i(t - \Delta t)}{2\Delta t}, \tag{10}$$

$$G_{i,j}(X(t)) \approx \frac{(X_i(t + \Delta t) - X_i(t - \Delta t))(X_j(t + \Delta t) - X_j(t - \Delta t))}{4\Delta t}. \tag{11}$$

Just “in principle” as the formulas above should lead to Stratonovich SODEs rather than Itô ones [9, Remark 2]. We rather show in the experimental Section 5 that they still work correctly, at least for the models considered therein.

- **Trapezoidal Rule (TR).** Finally, to avoid erroneous convergence of CD the authors of [9] propose the following trapezoidal-like approximations:

$$\frac{f_i(X(t + \Delta t)) + f_i(X(t))}{2} \approx \frac{X_i(t + \Delta t) - X_i(t)}{\Delta t}, \tag{12}$$

$$G_{i,j}(X(t + \Delta t)) + G_{i,j}(X(t)) \approx \frac{1}{2\Delta t} \{ [2(X_i(t + \Delta t) - X_i(t)) - \Delta t(f_i(X(t + \Delta t)) + f_i(X(t)))] \cdot [2(X_j(t + \Delta t) - X_j(t)) - \Delta t(f_j(X(t + \Delta t)) + f_j(X(t)))] \}. \tag{13}$$

All the KM, FD, CD and TR estimates are employed and compared in Section 5. As discussed in Section 4, resorting to their weak or strong form depends on the specific strategy to be adopted in view of the available data.

We close this section by observing that extension to SDDEs follows straightforwardly for the Itô-Taylor expansions just by formally replacing the state variable $X(t) \in \mathbb{R}^n$ of the SODE Eq. (1) with an augmented state variable $Z(t) := (X(t), X(t - \tau)) \in \mathbb{R}^{2n}$ from the SDDE Eq. (2). As far as the related weak estimates of drift and covariance are concerned, we stress that now conditioning must include both current and delayed states for consistent parameter estimation due to the non-Markovian nature of SDDEs. As an instance Eqs. (4) and (5) become

$$f_i(z) \approx \mathbb{E} \left[\frac{X_i(t + \Delta t) - X_i(t)}{\Delta t} \mid X(t) = z_1, X(t - \tau) = z_2 \right],$$

$$G_{i,j}(z) \approx \mathbb{E} \left[\frac{(X_i(t + \Delta t) - X_i(t))(X_j(t + \Delta t) - X_j(t))}{2\Delta t} \mid X(t) = z_1, X(t - \tau) = z_2 \right],$$

by choosing $\phi(z) = z_{1,i}$ in the Itô-Taylor expansion. Above, the block partition of $z \in \mathbb{R}^{2n}$ is intended as $(z_1, z_2) \in \mathbb{R}^n \times \mathbb{R}^n$ with z_1 representing the current time value and z_2 the delayed counterpart.

3. Sparse identification of nonlinear dynamics with SINDy

SINDy is a data-driven framework designed to uncover governing equations directly from observed time-series data. It leverages the assumption that most dynamical systems have an inherently sparse structure: their governing equations typically involve right-hand sides (RHSs) with a limited number of significant terms. This sparsity assumption allows the construction of accurate and interpretable models when the true underlying equations are unknown or when dimensionality reduction is required. We first summarize the original approach for ODEs in Section 3.1. Then we briefly illustrate the main aspects of the recent extensions to DDEs in Section 3.2 and to SODEs in Section 3.3. Finally we discuss the approach followed in this work for tackling SDDEs in Section 3.4.

3.1. SINDy Algorithm for ODEs

For general references see the pioneering work [7] as well as [6,31].

Consider a dynamical system governed by an ODE $x' = f(x)$ with unknown RHS $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$. Suppose to know or measure the state $x(t)$ at equidistant or even irregularly spaced discrete time points t_1, \dots, t_m . SINDy approximates each component $f_i(x), i = 1, \dots, n$, as a sparse linear combination of a predefined set of candidate basis functions $\{\theta_j(x)\}_{j=1}^p$:

$$f_i(x) = \sum_{j=1}^p \theta_j(x) \xi_{j,i}. \tag{14}$$

The choice of basis functions $\theta_j(x)$ significantly affects the effectiveness of the approach (common basis functions include polynomials, trigonometric functions and other nonlinear functions that may reflect the anticipated system behaviour or prior knowledge).

To estimate the coefficients $\xi_{j,i}$, the collected data are organized as

$$\mathbf{X} := \begin{bmatrix} x_1(t_1) & \dots & x_n(t_1) \\ \vdots & \ddots & \vdots \\ x_1(t_m) & \dots & x_n(t_m) \end{bmatrix} \in \mathbb{R}^{m \times n}, \quad \mathbf{X}' := \begin{bmatrix} x'_1(t_1) & \dots & x'_n(t_1) \\ \vdots & \ddots & \vdots \\ x'_1(t_m) & \dots & x'_n(t_m) \end{bmatrix} \in \mathbb{R}^{m \times n},$$

where the derivatives may either be directly measured or estimated numerically [32]. A library matrix

$$\Theta(\mathbf{X}) := \begin{bmatrix} 1 & \mathbf{X} & \mathbf{X}^2 & \dots & \mathbf{X}^d & \dots & \sin(\mathbf{X}) & \cos(\mathbf{X}) & \dots \end{bmatrix} \in \mathbb{R}^{m \times p}.$$

is then constructed by evaluating the candidate basis functions at all time points. Above, \mathbf{X}^d denotes a matrix containing all possible polynomial combinations of degree d in the components of x . The total number p of basis functions depends on the chosen function types and their complexity (e.g., a polynomial library of degree d include $p = \binom{n+d}{d}$ basis functions).

Determining the coefficients $\xi_{j,i}$ involves solving $\mathbf{X}' = \Theta(\mathbf{X})\Xi$ for $\Xi := [\xi_1 \ \xi_2 \ \dots \ \xi_n] \in \mathbb{R}^{p \times n}$ where each column vector $\xi_i := (\xi_{1,i}, \dots, \xi_{p,i}) \in \mathbb{R}^p$ corresponds to the sparse representation Eq. (14). Each ξ_i is determined by independently solving

$$\xi_i = \operatorname{argmin}_{\xi \in \mathbb{R}^p} \left(\left\| \mathbf{X}'_i - \Theta(\mathbf{X})\xi \right\|_2 + \lambda \|\xi\|_1 \right), \tag{15}$$

with sparse regression algorithms (such as STLS [7] or LASSO [33]). Above \mathbf{X}'_i denotes the i -th column of \mathbf{X}' and the hyperparameter $\lambda \geq 0$ controls the balance between fitting accuracy (the first term, in $\|\cdot\|_2$) and sparsity (the second term, in $\|\cdot\|_1$).

3.2. SINDy For DDEs

Recent extensions to DDEs such as $x'(t) = f(x(t), x(t - \tau))$ for a single constant delay $\tau > 0$ have been proposed in [10–13]. The following summary refers mainly to [10,12]. The core of SINDy for DDEs is the construction of an augmented library $\Theta(\mathbf{X}, \mathbf{X}_\tau)$ including delayed time samples $x(t_i - \tau), i = 1 \dots, m$. If these values are not at disposal they can be provided by piecewise linear interpolation of the available data \mathbf{X} . Then the SINDy framework extends by solving $\mathbf{X}' = \Theta(\mathbf{X}, \mathbf{X}_\tau)\Xi(\tau)$ via sparse regression, where now the sparse coefficients $\Xi(\tau)$ depend on the value of τ , assumed to be known in principle. The case of multiple delays can be treated straightforwardly. Different is the realistic situation where also the delays are unknown (possibly both in number and values). The main idea is to minimize the reconstruction error $\tau \mapsto \|\mathbf{X}' - \Theta(\mathbf{X}, \mathbf{X}_\tau)\Xi(\tau)\|_2$ as a function of the unknown delay(s) $\tau = (\tau_1, \dots, \tau_k =: \bar{\tau})$ by using an “external” optimizer (not to be confused with the “internal” minimization of SINDy). In [12] the use of Bayesian tools has largely improved the brute force approach originally followed in [13]. In [10] a further substantial improvement has been reached by employing particle swarm techniques [34–36]. Moreover, in [10] a different approach has been proposed, based on first discretizing the unknown DDE into a finite number of ODEs, to then apply the standard SINDy algorithm, possibly with external optimization. The great advantage of this “pragmatic” approach is that of having to deal only with the maximum delay $\bar{\tau}$, while the presence of possible multiple intermediate delays $(\tau_1, \dots, \tau_{k-1})$ is hidden in the collocation process that reduce the DDE to ODEs [37]. In this case the augmented library includes samples at time-shifted collocation nodes rather delayed samples. Both approaches can be extended by including stochastic terms as explained in Section 3.4.

3.3. SINDy For SODEs

The application of SINDy to SODEs such as Eq. (1) has been first considered in [8] and more recently in [9] (see also the references therein and [38,39]). The main difference from the use on deterministic differential equations, where data \mathbf{X}' are either available or

can be easily recovered from \mathbf{X} , is the need for reasonable data for drift and diffusion. In view of understanding just the principle underlying the extension of SINDy to SODEs, for the moment we can assume that such data are somehow reconstructed or estimated in a non-parametric fashion and are then available as $\mathbf{f}(\mathbf{X}) \in \mathbb{R}^{m \times n}$ for drift and $\mathbf{G}(\mathbf{X}) \in \mathbb{R}^{m \times n^2}$ for diffusion in the form of vectorized covariance. It is then enough to choose two possibly different libraries of candidate functions $\Theta_f \in \mathbb{R}^{m \times p_f}$ and $\Theta_G \in \mathbb{R}^{m \times p_G}$ and solving independently

$$\alpha_i = \operatorname{argmin}_{\alpha \in \mathbb{R}^p} \|\mathbf{f}_i(\mathbf{X}) - \Theta_f(\mathbf{X})\alpha\|_2 + \lambda_f \|\alpha\|_1, \quad i = 1, \dots, n, \tag{16}$$

$$\beta_{i,j} = \operatorname{argmin}_{\beta \in \mathbb{R}^p} \|\mathbf{G}_{i,j}(\mathbf{X}) - \Theta_G(\mathbf{X})\beta\|_2 + \lambda_G \|\beta\|_1, \quad i, j = 1, \dots, n. \tag{17}$$

To note the freedom of choosing also different sparse promoting parameters among drift and diffusion. We demand to Section 4 a detailed discussion on how obtaining $\mathbf{f}(\mathbf{X})$ and $\mathbf{G}(\mathbf{X})$ following three different approaches. First we deal with SINDy for SDDEs as the driving novelty of the present work.

3.4. SINDy For SDDEs

To the best of the authors' knowledge the only work on SINDy for SDDEs is the recent paper [14]. The approach proposed therein heavily relies on assuming a single small delay in order to approximate SDDEs by SODEs via Taylor expansion. As such, the methodology is not suitable for generic delay values. Moreover, the stochastic integral is restricted to the Stratonovich type and an extension to multiple delays seems far from being trivial.

Here instead we pursue a general extension of SINDy to SDDEs by combining the arguments of both Section 3.2 for DDEs and Section 3.3 for SODEs. Indeed, the key observation is that in principle the presence of delay terms asks for an augmented library including delayed samples \mathbf{X}_τ , while the presence of stochastic terms requires suitable strategies to recover data $\mathbf{f}(\mathbf{X})$ for drift and $\mathbf{G}(\mathbf{X})$ for diffusion through covariance. As such the two issues appear basically unrelated and this would lead to extend Eqs. (16) and (17) as

$$\alpha_i = \operatorname{argmin}_{\alpha \in \mathbb{R}^p} \|\mathbf{f}_i(\mathbf{X}, \mathbf{X}_\tau) - \Theta_f(\mathbf{X}, \mathbf{X}_\tau)\alpha\|_2 + \lambda_f \|\alpha\|_1, \quad i = 1, \dots, n,$$

$$\beta_{i,j} = \operatorname{argmin}_{\beta \in \mathbb{R}^p} \|\mathbf{G}_{i,j}(\mathbf{X}) - \Theta_G(\mathbf{X}, \mathbf{X}_\tau)\beta\|_2 + \lambda_G \|\beta\|_1, \quad i, j = 1, \dots, n,$$

together with all the possible variations in view of multiple or unknown delays, as well as the pragmatic reduction of the delay dependence via collocation as mentioned in Section 3.2. Nevertheless, a more trustworthy approach should instead consider the explicit dependence on the delayed samples also for the reconstruction of data for drift and diffusion, hence

$$\alpha_i = \operatorname{argmin}_{\alpha \in \mathbb{R}^p} \|\mathbf{f}_i(\mathbf{X}, \mathbf{X}_\tau) - \Theta_f(\mathbf{X}, \mathbf{X}_\tau)\alpha\|_2 + \lambda_f \|\alpha\|_1, \quad i = 1, \dots, n, \tag{18}$$

$$\beta_{i,j} = \operatorname{argmin}_{\beta \in \mathbb{R}^p} \|\mathbf{G}_{i,j}(\mathbf{X}, \mathbf{X}_\tau) - \Theta_G(\mathbf{X}, \mathbf{X}_\tau)\beta\|_2 + \lambda_G \|\beta\|_1, \quad i, j = 1, \dots, n. \tag{19}$$

Indeed, as discussed at the end of Section 2.2 in view of Itô-Taylor expansions for SDDEs, in general weak forms require conditioning also w.r.t. past values and thus Eqs. (16) and (17) fit better into the framework. The issue becomes evident in the forthcoming Section 4 about one of the possible strategies to recover the necessary data.

4. Estimation of drift and diffusion from data

In this section we deal with obtaining the data $\mathbf{f}(\mathbf{X})$ and $\mathbf{G}(\mathbf{X})$ as just anticipated in Section 3.4. In Section 2.2 general moment-based estimators are described relying on Itô-Taylor expansions. Independently of the specific choice or order, KM, FD, CD and TR methods all face a fundamental challenge: how to get sufficiently-many data at each time instant to ensure reliable estimates of drift and diffusion. The answer depends, among other factors, on whether temporal data are available for a single trajectory alone, or samples from multiple independent realizations are at disposal (realistically from repeated experiments). In the first case the approach we follow, hereafter named “approach A” or simply “A” for brevity, relies on creating more realizations via suitable tools and is described in Section 4.1. In the second case, “approach B” or simply “B” in the sequel, we discuss two different strategies, namely B1 and B2, described respectively in Section 4.2 and in Section 4.3. The key difference between B1 and B2 is whether averages are taken before or after the application of SINDy’s sparse regression. Then in Section 5 we perform a thorough experimental comparison on different models of SDDEs, illustrating different trade-off’s between reconstruction error, computational cost, data requirements and implementation complexity.

First to proceed we underline that most of the adopted techniques we refer to in the sequel are taken from the recent literature but none of them deal with the presence of delays. We thus extend their application by considering when necessary an augmented state $Z(t) := (X(t), X(t - \tau)) \in \mathbb{R}^{2n}$ as already discussed at the end of Section 2.2. Nevertheless, we are aware that the general claims about the properties and effectiveness of such tools that hold in the case without delays are not rigorously proved for SDDEs in this work (as this would lead us out of scope). We reserve to dive into this direction in the next future, relying for the moment on the experimental results of Section 5.

4.1. Approach a: More realizations from conditional expectations

In this section we resort to the methodology discussed recently in [40]. The underlying basis is used also in the first work [8] on SINDy for SODEs by employing KM estimates.

Let δt be the stepsize chosen to represent the Brownian motion W underlying Eq. (2) and assume that samples of a single path $X(t)$ are available with time stepsize $\Delta t = K\delta t$ for some integer $K \geq 1$ (for our simulation we have used $K = 1$ in all the results presented in Section 5). If not, one may think at interpolated samples, also to get delayed samples and hence $Z(t)$. Given the relevant time instants $t_i, i = 1, \dots, m$ and a positive integer M , for each $i = 1, \dots, m - 1$ the proposed methodology first requires the following three steps:

- A(i) initialize M independent sample paths $X^{(k)}(t), k = 1, \dots, M$, of Eq. (2) each starting at time t_i from $Z(t_i)$;
- A(ii) approximate a single step of each path at time $t_i + \Delta t$ by using a numerical time-integrator, e.g., Euler-Maruyama as adapted to Eq. (2) [1], viz.

$$X_{i+1}^{(k)} = X_i^{(k)} + f(X_i^{(k)}, X_{i-\kappa}^{(k)})\Delta t + g(X_i^{(k)}, X_{i-\kappa}^{(k)})(W^{(k)}(t_{i+1}) - W^{(k)}(t_i)),$$

where we implicitly assume that τ is an integer multiple κ of Δt ;

- A(iii) record the increment $\Delta X_i^{(k)} := X_{i+1}^{(k)} - X_i^{(k)}$.

Then estimates of drift and diffusion (again through covariance) are obtained as

$$\mathbf{f}(Z(t_i)) = \frac{1}{M\Delta t} \sum_{k=1}^M \Delta X_i^{(k)},$$

$$\mathbf{G}(Z(t_i)) = \frac{1}{\Delta t} \left(\frac{1}{M} \sum_{k=1}^M \Delta X_i^{(k)} (\Delta X_i^{(k)})^T - \mathbf{f}(Z(t_i))\mathbf{f}(Z(t_i))^T (\Delta t)^2 \right).$$

As $M \rightarrow \infty$ and $\Delta t \rightarrow 0$, these estimates are expected to converge to the true values $f(Z(t_i))$ and $G(Z(t_i))$ (which is indeed the case without delay [40]). In practice one chooses K large enough ($\delta t \ll \Delta t$) to control discretization bias and M large enough to reduce sampling variance. Higher-order schemes (as, e.g., Milstein [41] and stochastic Runge-Kutta [3]) or finite-time corrections [42] can be used when Δt is not negligible.

A final comment is fundamental, also in view of approaches B1 and B2 as described next: the above procedure requires an assumed specification or initial approximation of both f and g , otherwise step A(ii) is not possible. In practice, one supplies these functions via a suitable parametric family, a non-parametric initializer or prior structural information and then refines them iteratively until a desired tolerance [40]. In Section 5 we adopt the true forms of f and g for the sake of avoiding complexities unrelated to the current scope.

4.2. Approach B1: Multiple trajectories with pre-regression averaging

An alternative to approach A can be followed when a large dataset from several independent realizations of the system is available. Consider M independent trajectories of Eq. (2), each sampled at $t_i, i = 1, \dots, m$, with uniform time step Δt . Let $Z^{(k)}(t_i)$ be the augmented state of the k -th trajectory at time t_i and $\Delta X^{(k)}(t_i) = X^{(k)}(t_{i+1}) - X^{(k)}(t_i)$ be the relevant increment of the current time state. To estimate drift and diffusion at a specific point $z \in \mathbb{R}^{2n}$ in the augmented state space we aggregate data points from all trajectories falling within a predefined neighborhood of z [43,44]. Given a tolerance $\epsilon > 0$ let $N_\epsilon(z)$ denote the cardinality of the set $S_\epsilon(z) := \{(k, i) : \|Z^{(k)}(t_i) - z\| \leq \epsilon\}$ of such observations. This effectively performs a kernel-like local averaging, where ϵ controls the trade-off between statistical variance (too few points if ϵ is small) and bias (over-smoothing if ϵ is large). This and similar techniques are widely used in non-parametric drift and diffusion estimation methods for stochastic processes for instance, through kernel density estimation or binning schemes in continuous-time models. The required estimates are finally obtained by averaging the observed increments and quadratic variations, respectively, over all data points in the neighborhood of z , viz.

$$\mathbf{f}(z) = \frac{1}{N_\epsilon(z)\Delta t} \sum_{k=1}^M \sum_{i:(k,i) \in S_\epsilon(z)} \Delta X_{t_i}^{(k)}, \tag{20}$$

$$\mathbf{G}(z) = \frac{1}{N_\epsilon(z)\Delta t} \sum_{k=1}^M \sum_{i:(k,i) \in S_\epsilon(z)} \Delta X_{t_i}^{(k)} (\Delta X_{t_i}^{(k)})^T. \tag{21}$$

The estimates presented as such are based on KM in an ensemble setting. Generalization to FD, CD and TR (presented in Section 2.2) is straightforward. Convergence to $f(z)$ and $G(z)$ is expected as $M \rightarrow \infty$ (or as the total number of observations $N_\epsilon(z)$ increases) also for SDDs. The accuracy depends on balancing the bias (from finite Δt and neighborhood size ϵ) and variance (from limited amount of data). As a final comment we observe that this approach facilitates a sparse, data-driven reconstruction of the underlying non-Markovian stochastic dynamics of SDDs without requiring explicit knowledge of the functional forms of f and g beforehand. Moreover, we remark once more that B1 is characterized by taking averages before the application of SINDY’s sparse regression, contrary to B2 as discussed next.

4.3. Approach B2: Multiple trajectories with post-regression averaging

A second strategy for leveraging an ensemble of M trajectories consists in applying SINDY’s sparse regression to each trajectory individually and then averaging the identified model parameters. This method, described in the experimental setup of [9], contrasts

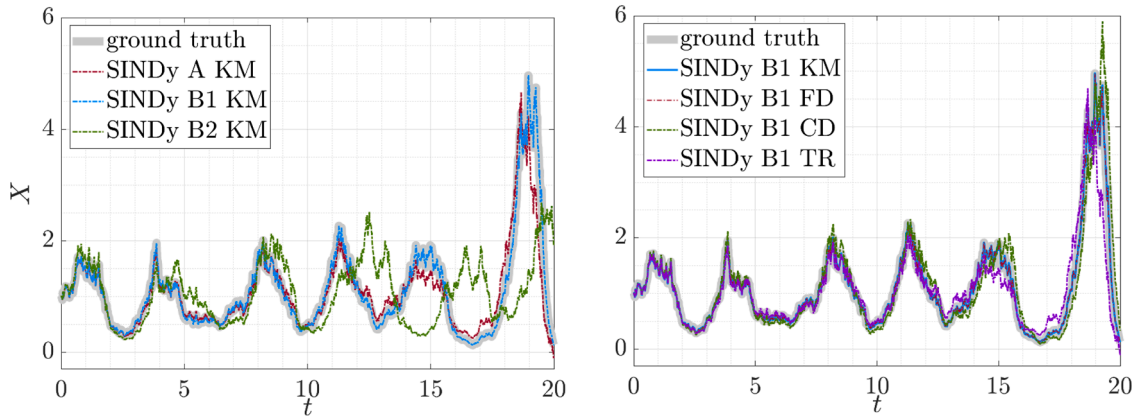


Fig. 1. comparison of trajectories for the logistic SDDE Eq. (22): ground truth and approaches A, B1 and B2 all with KM (left). Ground truth versus estimation methods KM, FD, CD and TR all with approach B1 (right).

with B1 by averaging in the parameter space after regression, rather than in the data space before regression. Briefly, if $\alpha_i^{(k)}$ and $\beta_{i,j}^{(k)}$ are the sparse coefficients identified by SINDy from the k -th trajectory through respectively Eqs. (16) and (17), the final model coefficients are retrieved by averaging across the entire ensemble:

$$\alpha_i = \frac{1}{M} \sum_{k=1}^M \alpha_i^{(k)}, \quad \beta_{i,j} = \frac{1}{M} \sum_{k=1}^M \beta_{i,j}^{(k)}.$$

Practically B2 creates an ensemble of models, with the final model being the average of this ensemble. The underlying assumption is that while each individually identified model may contain variance due to the specific realization of the stochastic process, averaging over many such models yields a result that is expected to converge to the true underlying system dynamics.

5. Numerical experiments and results

In this section we perform a thorough computational analysis by collecting several experiments on the three SDDEs described below. The experiments consist in applying the approaches A, B1 and B2 discussed in Section 4, each of them using all of KM, FD, CD and TR as underlying estimators from Section 2.2. We recall that M denotes both the numbers of paths realized with A through Euler-Maruyama simulations, as well as the cardinality of the ensemble of available trajectories for B1 and B2. A full comparison is first performed for all the three examples by assuming $M = 1000$ and $\epsilon = 10^{-4}$ for B1. Further tests regard limited comparisons to separately assess the role of M and ϵ also w.r.t. CPU time and errors. As for the latter, we consider both the absolute errors $|\bar{\xi}_i - \xi_i|$ on the i -th sparse coefficient of the chosen SINDy library, $i = 1, \dots, p$, as well as the RMSE on relevant trajectories (they regard both trajectories for the state $X(t)$ as well as for separated drift and diffusion dynamics reconstruction). Finally, data are artificially provided by numerical simulation via Euler-Maruyama and we use STLS [7] for sparse regression because of its robustness compared to other methods. All tests are performed on a Windows 11 system (CPU 5 GHz, RAM 16 GB) using MATLAB implementations for SINDy (MATLAB version R2024b). Before discussing the obtained results in Section 5.1 we present the three specific SDDEs that we use as models to test.

As a first model we consider from [45,46] the stochastic version with multiplicative noise

$$dX(t) = \alpha X(t)(1 - X(t - \tau))dt + \sigma X(t)dW(t) \tag{22}$$

of the so-called delay logistic or Hutchinson equation [47]. Eq. (22) models the evolution of a single ($n = 1$) population $X(t)$ at time t with intrinsic growth rate $\alpha = 2$, time delay $\tau = 1$ and noise intensity $\sigma = 0.4$ with underlying scalar ($q = 1$) Brownian motion. Data are obtained by simulating over the time window $[0, 20]$ with time step $\Delta t = 0.01$ via Euler-Maruyama, prescribing the history $X(s) = \cos(s)$, $s \in [-\tau, 0]$. 80% of the samples are used for training and 20% for validation. SINDy is applied with polynomial libraries of degree $d = 2$ for both drift and diffusion and $\lambda = 0.025$ is used. Results of the full comparison are collected in Table 1 and discussed in Section 5.1.

As a second model we consider from [48] the predator-prey interaction with delay and multiplicative noise

$$\begin{cases} dX_1(t) = X_1(t)[\alpha - \beta X_1(t) - \gamma X_2(t - \tau)]dt + \sigma_1 X_1(t)dW_1(t) \\ dX_2(t) = X_2(t)[- \delta + \kappa X_1(t - \tau)]dt + \sigma_2 X_2(t)dW_2(t), \end{cases} \tag{23}$$

where $X_1(t)$ and $X_2(t)$ denote prey and predator populations, respectively, with parameters $\alpha = 1$, $\delta = 0.5$, $\beta = \gamma = \kappa = 0.1$, $\sigma_1 = \sigma_2 = 0.4$, $\tau = 1$ and W_1, W_2 independent scalar Wiener processes (here $n = q = 2$). Again, data are obtained by simulating over the time window $[0, 20]$ with time step $\Delta t = 0.01$ via Euler-Maruyama, prescribing the constant history $X(s) \equiv (5, 2)^T$. 80% of the samples are

Table 1
 Comparison of errors on estimated coefficients, RMSE on trajectories and CPU time for the logistic SDDF Eq. (22) for both drift and diffusion, all estimation methods KM, FD, CD and TR from Section Eq. (2.2) and all data reconstruction strategies A, B1 and B2 from Section 4. Diffusion estimation might remain the same (KM in particular, first part above) or change (all of KM, FD, CD and TR, second part below).

library terms	true coeff.	$ \xi_i - \hat{\xi}_i $											
		KM			TR								
		A	B1	B2	A	B1	B2						
Same diffusion													
drift	$\xi_2 = 2$ $\xi_3 = 2$	$2.13e^{-3}$ $5.07e^{-3}$	$1.29e^{-4}$ $6.68e^{-5}$	$5.76e^{-1}$ $7.47e^{-2}$	$2.64e^{-2}$ $2.53e^{-2}$	$1.28e^{-4}$ $7.14e^{-5}$	$5.54e^{-1}$ $6.79e^{-2}$	$1.11e^{-2}$ $1.08e^{-2}$	$3.87e^{-4}$ $2.08e^{-4}$	$3.36e^{-2}$ $6.57e^{-2}$	$2.13e^{-3}$ $2.10e^0$	$2.13e^{-3}$ $1.17e^{-3}$	$1.02e^{-1}$ $1.60e^{-2}$
diffusion	$\xi_4 = 0.16$	$6.39e^{-3}$	$7.54e^{-9}$	$9.17e^{-3}$	$6.39e^{-3}$	$7.54e^{-9}$	$9.17e^{-3}$	$6.39e^{-3}$	$7.54e^{-9}$	$9.17e^{-3}$	$6.39e^{-3}$	$7.54e^{-9}$	$9.17e^{-3}$
Performance metrics													
CPU [s]		≈ 4	≈ 10	≈ 10	≈ 4	≈ 10	≈ 10	≈ 4	≈ 10	≈ 10	≈ 4	≈ 10	≈ 10
RMSE	full drift diffusion	$2.99e^{-1}$ $1.11e^{-1}$ $9.99e^{-2}$	$1.10e^{-3}$ $1.13e^{-4}$ $4.92e^{-6}$	$8.79e^{-1}$ $3.79e^{-1}$ $2.43e^{-1}$	$3.43e^{-1}$ $2.18e^{-2}$ $9.97e^{-2}$	$4.29e^{-4}$ $9.75e^{-5}$ $2.85e^{-6}$	$9.80e^{-1}$ $4.68e^{-1}$ $2.21e^{-1}$	$3.31e^{-1}$ $9.00e^{-3}$ $9.97e^{-2}$	$1.30e^{-3}$ $3.07e^{-4}$ $1.53e^{-5}$	$1.10e^0$ $1.37e^{-1}$ $1.97e^{-1}$	$3.73e^0$ $5.47e^0$ $9.99e^{-2}$	$1.43e^{-3}$ $1.70e^{-3}$ $1.22e^{-5}$	$1.68e^0$ $1.13e^0$ $2.20e^{-1}$
Different diffusion													
diffusion	$\xi_4 = 0.16$	$6.39e^{-3}$	$7.54e^{-9}$	$8.31e^{-2}$	$7.91e^{-2}$	$1.50e^{-5}$	$7.92e^{-2}$	$8.48e^{-2}$	$7.76e^{-3}$	$9.47e^{-2}$	$1.84e^{-1}$	$2.90e^{-6}$	$6.95e^{-2}$
Performance metrics													
CPU [s]		≈ 4	≈ 10	≈ 10	≈ 4	≈ 10	≈ 10	≈ 4	≈ 10	≈ 10	≈ 4	≈ 10	≈ 10
RMSE	full drift diffusion	$2.99e^{-1}$ $1.11e^{-2}$ $9.99e^{-2}$	$1.10e^{-3}$ $1.13e^{-4}$ $4.92e^{-6}$	$9.43e^{-1}$ $1.15e^0$ $1.57e^0$	$5.25e^{-1}$ $1.14e^{-2}$ $3.08e^{-1}$	$6.00e^{-3}$ $9.06e^{-4}$ $3.89e^{-5}$	$1.08e^0$ $3.59e^{-1}$ $1.05e^0$	$6.49e^{-1}$ $9.00e^{-3}$ $2.69e^{-1}$	$2.99e^{-2}$ $2.32e^{-2}$ $1.54e^{-2}$	$1.89e^0$ $1.56e^{-1}$ $8.97e^{-1}$	$2.62e^0$ $5.47e^0$ $6.21e^{-1}$	$3.00e^{-2}$ $1.10e^{-3}$ $7.64e^{-6}$	$9.83e^{-1}$ $2.90e^{-1}$ $2.41e^{-1}$

Table 2
 Comparison of errors on estimated coefficients, RMSE on trajectories and CPU time for the predator-prey SDDE Eq. (23) for both drift and diffusion, all estimation methods KM, FD, CD and TR from Section Eq. (2.2) and all data reconstruction strategies A, B1 and B2 from Section 4. Diffusion estimation might remain the same (KM in particular, first part above) or change (all of KM, FD, CD and TR, second part below).

library	true	$ \xi_j - \tilde{\xi}_j $												
		KM			FD			CD			TR			
terms	coeff.	A	B1	B2	A	B1	B2	A	B1	B2	A	B1	B2	
drift	$X(t)$	$\xi_2 = 1$	$1.28e^{-1}$	$9.20e^{-5}$	$2.27e^0$	$6.90e^{-2}$	$5.19e^{-4}$	$2.26e^0$	$1.16e^{-2}$	$5.52e^{-4}$	$1.80e^{-1}$	$3.23e^{-1}$	$1.85e^{-5}$	$4.51e^0$
	$Y(t)$	$\xi_3 = 0.5$	$1.42e^{-3}$	$2.86e^{-7}$	$1.13e^0$	$7.14e^{-2}$	$2.32e^{-4}$	$1.55e^0$	$1.78e^{-2}$	$2.56e^{-4}$	$3.60e^{-1}$	$7.86e^{-3}$	$1.38e^{-6}$	$15.58e^0$
	$X^2(t)$	$\xi_6 = 0.1$	$6.16e^{-3}$	$8.38e^{-6}$	$1.65e^{-1}$	$6.59e^{-3}$	$3.98e^{-5}$	$1.73e^{-1}$	$1.64e^{-3}$	$5.09e^{-5}$	$7.88e^{-2}$	$1.68e^{-2}$	$5.06e^{-5}$	$6.33e^{-1}$
	$X(t)Y(t-\tau)$	$\xi_9 = 0.1$	$1.05e^{-2}$	$3.07e^{-6}$	$1.42e^{-1}$	$6.55e^{-3}$	$5.86e^{-5}$	$1.45e^{-1}$	$7.63e^{-4}$	$1.60e^{-5}$	$2.20e^{-2}$	$1.36e^{-1}$	$3.79e^{-5}$	$5.60e^{-1}$
	$Y(t)X(t-\tau)$	$\xi_{11} = 0.1$	$7.35e^{-4}$	$1.84e^{-8}$	$5.07e^{-2}$	$1.08e^{-2}$	$2.35e^{-5}$	$6.35e^{-2}$	$3.07e^{-3}$	$2.77e^{-5}$	$1.15e^{-2}$	$1.30e^{-2}$	$1.45e^{-7}$	$2.46e^{-1}$
Same diffusion														
diffusion	$X^2(t)$	$\xi_6 = 0.16$	$6.36e^{-4}$	$2.34e^{-6}$	$1.10e^{-2}$	$6.36e^{-4}$	$2.34e^{-6}$	$1.10e^{-2}$	$6.36e^{-4}$	$2.34e^{-6}$	$1.10e^{-2}$	$6.36e^{-4}$	$2.34e^{-6}$	$1.10e^{-2}$
	$Y^2(t)$	$\xi_{10} = 0.16$	$1.16e^{-4}$	$1.06e^{-9}$	$3.55e^{-3}$	$1.16e^{-4}$	$1.06e^{-9}$	$3.55e^{-3}$	$1.16e^{-4}$	$1.06e^{-9}$	$3.55e^{-3}$	$1.16e^{-4}$	$1.06e^{-9}$	$3.55e^{-3}$
Performance metrics														
CPU [s]		≈ 5	≈ 30	≈ 30	≈ 5	≈ 30	≈ 30	≈ 5	≈ 30	≈ 30	≈ 5	≈ 30	≈ 30	≈ 30
RMSE	full	$1.07e^{-1}$	$1.00e^{-3}$	$2.89e^0$	$4.84e^{-1}$	$4.80e^{-3}$	$2.13e^0$	$1.50e^{-1}$	$7.20e^{-3}$	$4.30e^0$	$1.70e^0$	$5.37e^{-2}$	$5.10e^0$	
	drift	$3.10e^{-2}$	$8.80e^{-5}$	$2.66e^0$	$1.00e^{-1}$	$5.40e^{-4}$	$8.89e^0$	$2.63e^{-2}$	$7.50e^{-4}$	$1.27e^0$	$1.39e^0$	$4.66e^{-4}$	$2.03e^1$	
	diffusion	$2.06e^{-2}$	$7.57e^{-6}$	$1.59e^{-1}$	$2.28e^{-2}$	$6.62e^{-6}$	$8.87e^{-2}$	$2.28e^{-2}$	$1.11e^{-5}$	$6.93e^{-2}$	$2.60e^{-2}$	$6.68e^{-5}$	$1.41e^{-1}$	
Different diffusion														
diffusion	$X^2(t)$	$\xi_6 = 0.16$	$6.36e^{-4}$	$2.34e^{-6}$	$1.10e^{-2}$	$7.91e^{-2}$	$9.35e^{-6}$	$7.13e^{-2}$	$7.96e^{-2}$	$3.13e^{-3}$	$1.04e^{-1}$	$1.73e^{-3}$	$1.30e^{-8}$	$1.10e^0$
	$Y^2(t)$	$\xi_{10} = 0.16$	$1.16e^{-4}$	$1.06e^{-9}$	$3.55e^{-3}$	$8.06e^{-2}$	$2.32e^{-8}$	$7.76e^{-2}$	$7.97e^{-2}$	$2.65e^{-4}$	$1.05e^{-1}$	$7.61e^{-5}$	$4.49e^{-9}$	$7.24e^{-1}$
Performance metrics														
CPU [s]		≈ 5	≈ 30	≈ 30	≈ 5	≈ 30	≈ 30	≈ 5	≈ 30	≈ 30	≈ 5	≈ 30	≈ 30	≈ 30
RMSE	full	$1.07e^{-1}$	$1.00e^{-3}$	$2.89e^0$	$1.17e^0$	$2.70e^{-3}$	$2.21e^0$	$1.27e^0$	$1.70e^0$	$1.14e^0$	$1.70e^0$	$5.88e^{-2}$	$6.58e^1$	
	drift	$3.10e^{-2}$	$8.80e^{-5}$	$2.66e^0$	$1.00e^{-1}$	$5.76e^{-4}$	$8.71e^0$	$2.63e^{-2}$	$2.34e^{-2}$	$1.16e^0$	$1.39e^0$	$3.90e^{-4}$	$6.53e^1$	
	diffusion	$2.06e^{-2}$	$7.57e^{-6}$	$1.59e^{-1}$	$4.24e^0$	$3.38e^{-4}$	$7.61e^{-1}$	$4.22e^0$	$6.06e^{-2}$	$6.72e^{-1}$	$5.42e^{-2}$	$1.55e^{-6}$	$1.87e^0$	

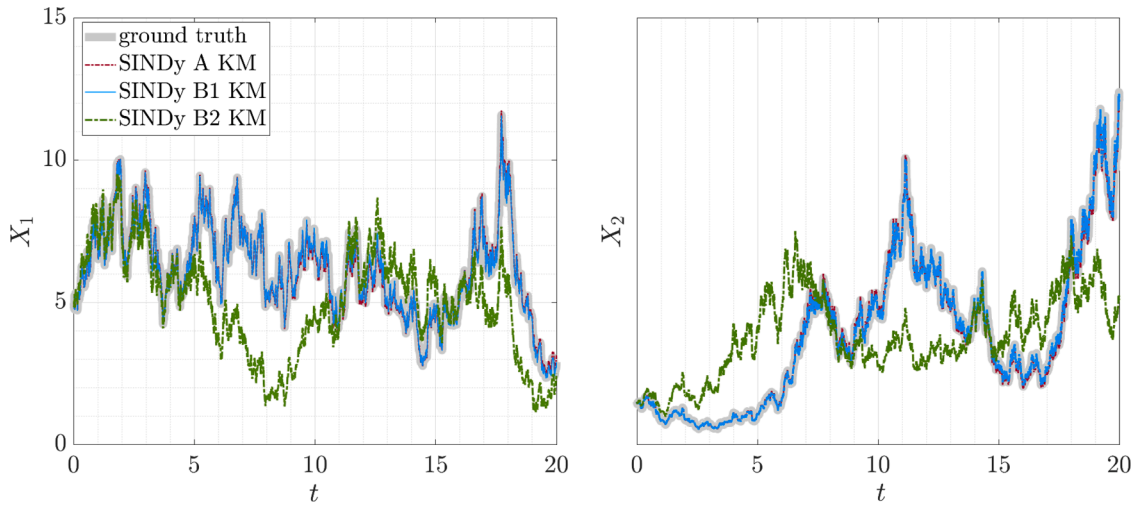


Fig. 2. comparison of trajectories for the predator-prey SDDE Eq. (23): ground truth and approaches A, B1 and B2 all with KM.

used for training and 20% for validation. SINDy is applied with polynomial libraries of degree $d = 2$ for both drift and diffusion and $\lambda = 0.025$ is used. Results of the full comparison are collected in Table 2 and discussed in Section 5.1.

Finally, as a third and last model we consider from [49] the SDDE

$$dX(t) = rX(t)dt + \sigma(X(t), X(t - \tau))X(t)dW(t) \tag{24}$$

modeling the price dynamics of a financial asset with memory, where $X(t)$ denotes the asset (or option) price at time t , $r = 0.05$ is the constant risk-free interest rate, $\tau = 0.002$, the function σ represents the volatility (see below), which is a continuous function of also the history of the stock price and $W(t)$ is a standard Wiener process. Eq. (24) generalizes the classical Black-Scholes framework by incorporating historical dependence into the volatility term, capturing memory effects observed in real markets. Once more, data are obtained by simulating over the time window $[0, 20]$ with time step $\Delta t = 0.01$ via Euler-Maruyama, prescribing the constant history $X(s) \equiv 100$. 80% of the samples are used for training and 20% for validation. For SINDy identification, custom libraries are constructed. The drift is identified using a polynomial library of degree $d = 1$ and the diffusion is identified using a polynomial library of degree $d = 4$. Both libraries are augmented with the logarithmic term $\ln(X(t)/X(t - \tau))$ (see Eq. (25) below) and $\lambda = 0.025$ is used. Results of the full comparison are collected in Table 3 and discussed in Section 5.1.

As proposed in [49], the volatility function σ is based on a continuous-time analogue of the GARCH model, viz.

$$\sigma^2(X(t), X(t - \tau)) = \frac{\gamma V}{\alpha + \gamma} + \frac{\alpha}{\tau(\alpha + \gamma)} \ln^2 \left(\frac{X(t)}{X(t - \tau)} \right), \tag{25}$$

where $V = 0.127$ is the long-run average variance, $\alpha = 0.6$ and $\gamma = 0.4$ are positive model parameters controlling the weighting of past return and reversion strength. This formulation ensures that σ remains strictly positive and reflects volatility clustering effects observed in empirical financial time series. Such memory-aware volatility plays a crucial role in capturing more realistic option price dynamics compared to models assuming constant or purely time-dependent volatility.

5.1. Discussion

The analysis of the results collected in Tables 1–3 reveals that the optimal estimation strategy is highly dependent on the specific model, the performance metric of interest (error on the reconstructed coefficients, RMSE on trajectories or CPU time) and the chosen methodology (A, B1 or B2 with KM, FD, CD or TR). In general the pre-regression averaging strategy of B1 overwhelmingly outperforms the other choices in terms of accuracy for all the three considered models. On the contrary, B2 consistently yields the highest absolute errors for every estimated coefficient and the highest RMSE values for the full system as well as for separated drift and diffusion and given the poor results, it is not considered a viable method. As far as A is concerned, the potential accuracy seems to depend on the prior available knowledge concerning the data via the (one-step) generation of relevant realizations. On the other hand, B1 operates without structural constraints, it frequently demonstrates a superior ability to identify the precise values of individual coefficients, thus highlighting its power in pure parameter and model discovery. Figs. 1 to 3 present a comparison between the ground truth and the SINDy-predicted trajectories for each model. Fig. 1 (left) compares different estimation approaches for fixed KM, while Fig. 1 (right) compares various methods for estimation, holding method B1 constant. Based on the similar results for the other models, the KM approach was used for all subsequent figures (2 and 3).

Figs. 4 through 6 further clarify the comparative performance of the estimation strategies and numerical methods across different scenarios. Fig. 4 presents a comprehensive comparison of the RMSE for the state reconstructions across all methodologies as functions of the estimator (KM, FD, CD, TR) and data reconstruction approaches (A, B1, B2). It confirms the dominant accuracy of

Table 3 comparison of errors on estimated coefficients, RMSE on trajectories and CPU time for the option-pricing SDE Eq. (24) with Eq. (25) for both drift and diffusion, all estimation methods KM, FD, CD and TR from Section Eq. (2.2) and all data reconstruction strategies A, B1 and B2 from Section 4. Diffusion estimation might remain the same (KM in particular, first part above) or change (all of KM, FD, CD and TR, second part below). Recall that $G(Z(t)) = X^2(t) \ln^2(X(t)/X(t - \tau))$.

library	true	$ \xi_i - \hat{\xi}_i $											
		KM			FD			CD			TR		
terms	coeff.	A	B1	B2	A	B1	B2	A	B1	B2	A	B1	B2
drift	$\xi_2 = 0.05$	$1.49e^0$	$4.26e^{-6}$	$1.32e^0$	$1.48e^0$	$3.65e^{-6}$	$1.32e^0$	$3.03e^{-1}$	$6.26e^{-6}$	$1.32e^0$	$3.09e^0$	$7.57e^{-5}$	$1.32e^0$
same diffusion													
diffusion	$\xi_5 = 0.058$	$5.70e^{-1}$	$2.35e^{-5}$	$9.20e^{-1}$	$5.70e^{-1}$	$2.35e^{-5}$	$9.20e^{-1}$	$5.70e^{-1}$	$2.35e^{-5}$	$9.20e^{-1}$	$5.70e^{-1}$	$2.35e^{-5}$	$9.20e^{-1}$
	$\xi_{5,3} = 300$	$8.51e^2$	$1.48e^0$	$1.46e^2$	$8.51e^1$	$1.48e^0$	$1.46e^2$	$8.51e^1$	$1.48e^0$	$1.46e^2$	$8.51e^1$	$1.48e^0$	$1.46e^2$
performance metrics													
CPU [s]		≈ 4	≈ 10	≈ 10	≈ 4	≈ 10	≈ 10	≈ 4	≈ 10	≈ 10	≈ 4	≈ 10	≈ 10
RMSE	full	$3.00e^0$	$6.48e^{-6}$	$3.72e^0$	$1.23e^0$	$9.67e^{-6}$	$4.37e^0$	$1.12e^0$	$6.19e^{-6}$	$4.42e^0$	$7.88e^0$	$6.46e^{-3}$	$2.39e^1$
	drift	$5.35e^{-1}$	$2.85e^{-5}$	$2.50e^0$	$7.72e^{-1}$	$1.05e^{-5}$	$2.32e^0$	$2.97e^{-1}$	$7.83e^{-5}$	$2.48e^0$	$1.07e^0$	$2.14e^{-4}$	$9.68e^1$
	diffusion	$9.00e^0$	$1.17e^{-3}$	$2.67e^0$	$1.13e^0$	$4.66e^{-3}$	$6.69e^{-1}$	$1.13e^0$	$1.53e^{-3}$	$1.29e^0$	$9.00e^0$	$2.11e^{-2}$	$1.54e^0$
different diffusion													
diffusion	$\xi_5 = 0.058$	$5.70e^{-1}$	$2.35e^{-5}$	$9.20e^{-1}$	$3.08e^0$	$7.37e^{-7}$	$2.50e^{-1}$	$7.97e^{-1}$	$3.75e^{-6}$	$5.67e^{-1}$	$6.16e^{-1}$	$1.54e^{-5}$	$4.82e^0$
	$\xi_{5,3} = 300$	$8.51e^2$	$1.48e^0$	$1.46e^2$	$2.93e^2$	$7.59e^{-1}$	$7.46e^1$	$2.94e^2$	$5.43e^{-1}$	$2.46e^2$	$8.85e^2$	$1.40e^0$	$2.90e^2$
performance metrics													
CPU [s]		≈ 4	≈ 10	≈ 10	≈ 4	≈ 10	≈ 10	≈ 4	≈ 10	≈ 10	≈ 4	≈ 10	≈ 10
RMSE	full	$3.00e^0$	$6.48e^{-6}$	$3.72e^0$	$3.67e^1$	$5.15e^{-6}$	$4.24e^1$	$2.63e^0$	$1.03e^{-6}$	$2.91e^0$	$7.76e^0$	$4.40e^{-2}$	$2.76e^1$
	drift	$5.35e^{-1}$	$2.85e^{-5}$	$2.50e^0$	$5.36e^{-1}$	$5.88e^{-4}$	$5.88e^0$	$2.97e^{-1}$	$5.27e^{-5}$	$2.28e^0$	$1.07e^0$	$7.83e^{-6}$	$9.13e^1$
	diffusion	$9.00e^0$	$1.17e^{-3}$	$2.67e^0$	$1.61e^0$	$1.58e^{-5}$	$1.25e^0$	$6.47e^1$	$1.88e^{-4}$	$4.43e^0$	$9.30e^0$	$4.15e^{-3}$	$1.13e^0$

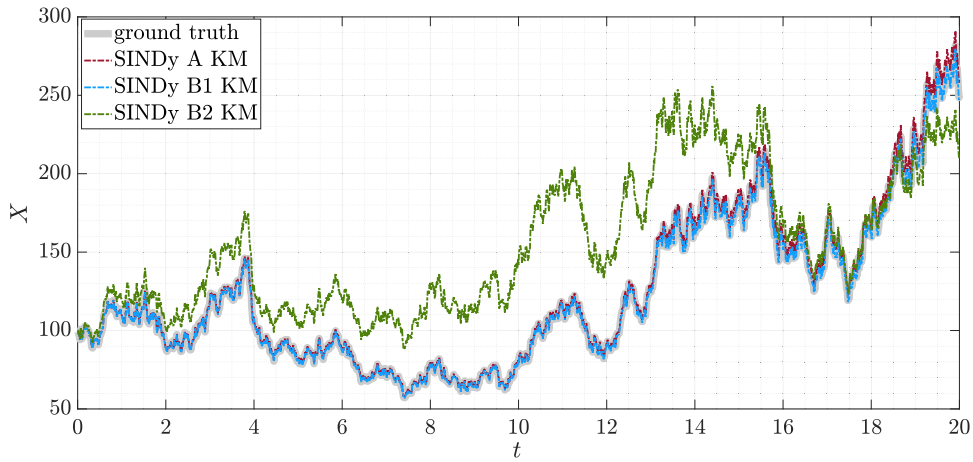


Fig. 3. comparison of trajectories for the option pricing SDDE Eq. (24): ground truth and approaches A, B1 and B2 all with KM.

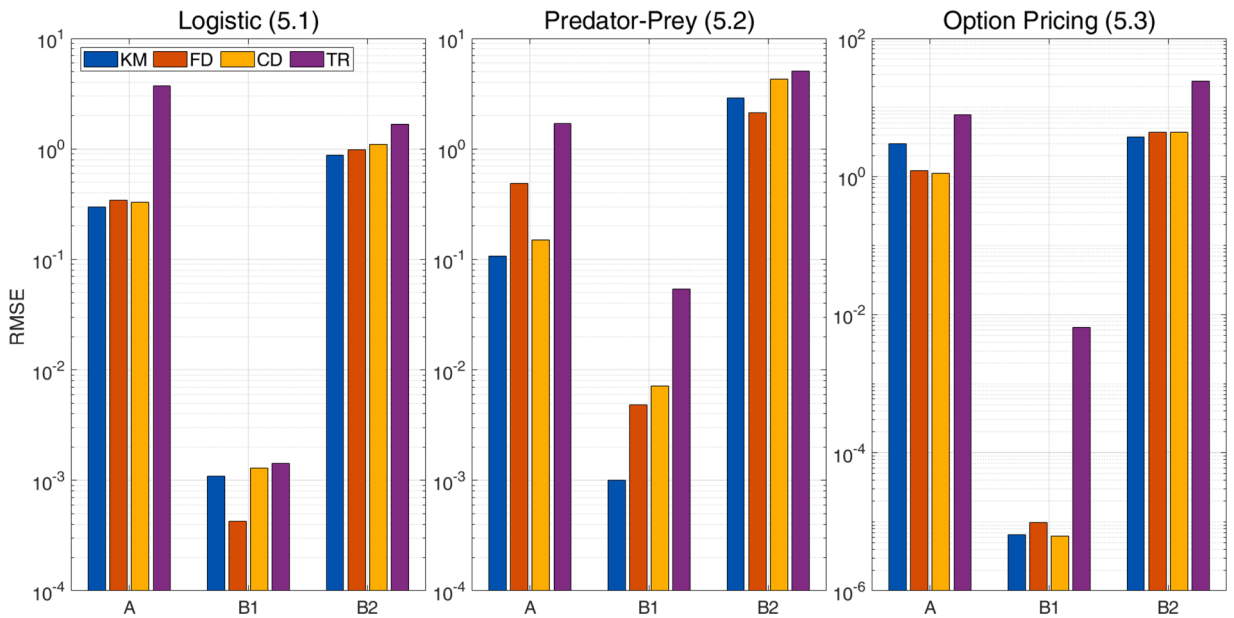


Fig. 4. comparison of RMSE for the SDDEs Eq. (22), Eq. (23) and Eq. (24) with all the approaches A, B1 and B2 against all the estimators KM, FD, CD and TR.

the pre-regression averaging approach B1 regardless of the estimator choice, with notably lower RMSE values reflecting superior reconstruction performance. Conversely, B2 again consistently yields higher RMSE values, reinforcing its unsuitability for reliable system identification. Figs. 5 and 6 in particular explore critical methodological factors influencing the accuracy and computational cost, further guiding practical implementation. Fig. 5 examines the effect of the neighborhood tolerance parameter ϵ in B1 on the absolute errors of estimated coefficients and CPU time. This insight underlines the necessity of balancing statistical variance and bias during neighborhood selection to optimize performance. We anyway observe that decreasing ϵ too much does not lead to significant improvements and better results are obtained for diffusion rather than drift. Fig. 6 evaluates the impact of ensemble size M on estimation accuracy restricted only to the logistic SDDE model Eq. (22) (results are similar for the other models). We observe that, as the size M increases, Approach A does not notably improve the estimation of the diffusion coefficient; however, it does lead to better results for the drift term, but only for substantial increase of M , underlying the data-hungry nature of this approach. Conversely, Approach B2 consistently performs poorly across all values of M . In contrast, Approach B1 consistently outperforms the others in all cases, although its computational cost grows with increasing M (indeed larger values of M are not even considered).

Collectively, these figures validate the numerical superiority of B1 for SDDE model discovery. They provide actionable guidelines for tuning hyperparameters such as neighborhood size and ensemble cardinality, ensuring practical and effective application of the proposed SINDy framework for stochastic delay differential equations. These insights align with and reinforce the overall recommendation of prioritizing B1 in real-world discovery scenarios where model structures are unknown.

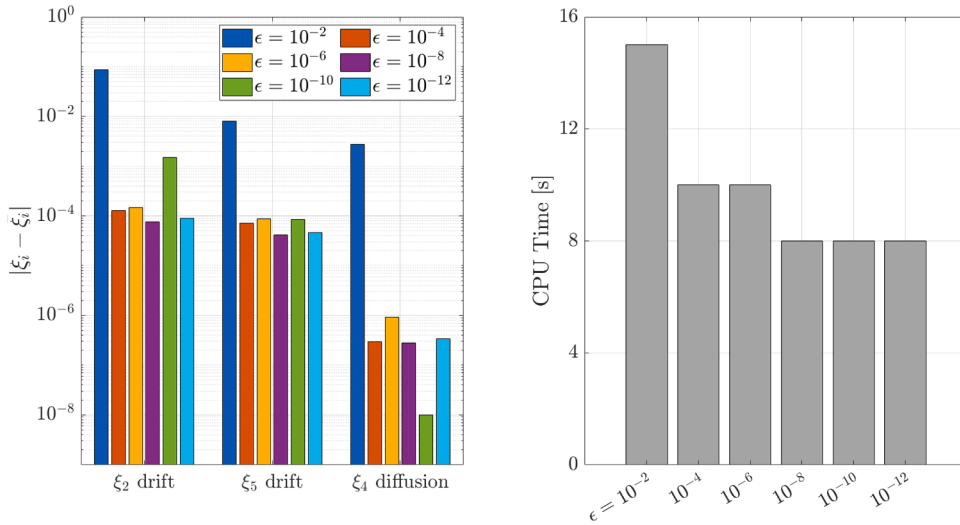


Fig. 5. comparison of the absolute errors $|\xi_i - \bar{\xi}_i|$ (left) and CPU time (right) for varying ϵ for the logistic SDDE Eq. (22) with B1-FD.

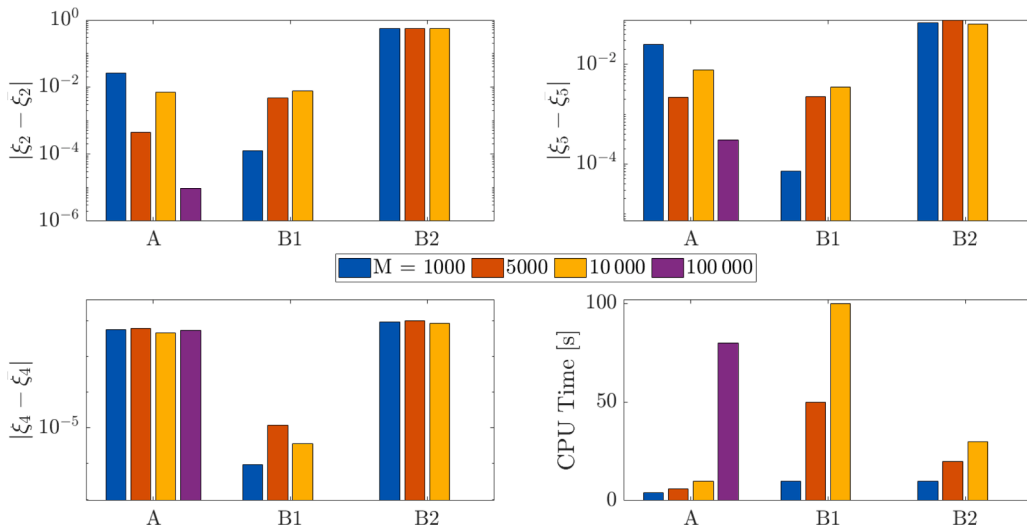


Fig. 6. comparison of the absolute errors $|\xi_i - \bar{\xi}_i|$ for drift (top row) and diffusion (bottom left) for increasing M for the logistic SDDE Eq. (22) using A, B1 and B2 all with FD.

In conclusion, the selection of an optimal strategy depends entirely on the analytical objective. If the underlying model structure is known in advance and the goal is a rapid, computationally inexpensive estimation of the overall system behavior, A is a suitable choice. However, in a more realistic research scenario where the functional forms of the drift and diffusion terms are unknown and must be determined from the data, B1 is the decisively better option. Its ability to perform system identification without prior assumptions makes it a more robust, powerful and scientifically sound method for discovery-oriented analysis. The higher computational cost is justified by its capacity to derive the model directly from observations, minimizing user-introduced bias and leading to more accurate and reliable parameter estimation.

6. Conclusion and future work

This study presents a comprehensive evaluation of sparse identification strategies for SDDEs, focusing on the combined influence of estimator accuracy, trajectory ensemble construction and underlying system complexity. The comparative results underscore the superiority of exploiting the availability of multiple independent trajectories with pre-regression averaging (B1). Across all considered models including biological and financial systems with nonlinear delays this combination consistently yielded the lowest reconstruction errors and the most robust recovery of drift and diffusion dynamics. In contrast, strategies relying on single-path realizations-based reconstruction (A) or post-regression averaging (B2) performed reliably only under constrained or idealized conditions and often failed in more realistic, noise-sensitive scenarios. The analysis further reveals that estimator choice is not merely

a computational convenience but a critical element influencing bias, robustness and interpretability. Together, these findings offer practical guidance for researchers applying sparse learning methods to complex stochastic systems with memory, highlighting the importance of ensemble-based strategies and carefully selected numerical estimators for successful model discovery and scientific insight.

Looking forward, the methodology offers promising opportunities for extension, including the incorporation of more complex and problem-specific candidate libraries to capture richer nonlinear and nonlocal dynamics, application and validation on real-world empirical datasets to address practical challenges such as measurement noise and missing data and the development of hybrid estimation techniques that combine the strengths of statistical and machine learning approaches. Additionally, integrating structure-preserving machine learning frameworks, such as physics-informed neural networks, could further enhance model interpretability and stability by respecting underlying physical constraints. Altogether, these avenues hold the potential to broaden the applicability and deepen the impact of sparse identification methods for SDDEs, facilitating robust, interpretable and computationally feasible data-driven discovery in complex stochastic systems with memory.

Data availability

Data will be made available on request.

Acknowledgements

DB, DC, RDA and MT are members of INdAM research group GNCS; DB is a member of UMI research group “Modellistica socio-epidemiologica”; DC and RDA are members of UMI research group “Matematica per l’intelligenza Artificiale e il machine learning”. The work of DB, DC, RDA and IS was supported by the European Union - NextGenerationEU with the project “MODellistica Numerica e Data-driven per l’Innovazione sostenibile - MONDI” (CUP: G25F21003390007). The work of MT was supported by the Italian Ministry of University and Research (MUR) through a PhD grant PNRR DM351/22.

References

- [1] E. Buckwar, Introduction to the numerical analysis of stochastic delay differential equations, *J. Comput. Appl. Math.* 125 (1–2) (2000) 297–307.
- [2] D.J. Higham, An algorithmic introduction to numerical simulation of stochastic differential equations, *SIAM Rev.* 43 (3) (2001) 525–546.
- [3] P.E. Kloeden, E. Platen, *Stochastic differential equations*, Springer, 1992.
- [4] S.-E.A. Mohammed, Stochastic differential systems with memory: theory, examples and applications, in: *Stochastic Analysis and Related Topics VI: Proceedings of the Sixth Oslo-Silivri Workshop* Geilo, Springer, 1996, pp. 1–77.
- [5] B. Øksendal, B. Øksendal, *Stochastic differential equations*, Springer, 2003.
- [6] S.L. Brunton, J.N. Kutz, *Data-driven Science and Engineering*, Cambridge University Press, 2022.
- [7] S.L. Brunton, J.L. Proctor, J.N. Kutz, Discovering governing equations from data by sparse identification of nonlinear dynamical systems, *Proc. Natl. Acad. Sci.* 113 (15) (2016) 3932–3937.
- [8] L. Boninsegna, F. Nüske, C. Clementi, Sparse learning of stochastic dynamical equations, *J. Chem. Phys.* 148 (24) (2018).
- [9] M. Wanner, I. Mezić, On higher order drift and diffusion estimates for stochastic sindy, *SIAM J. Appl. Dyn. Syst.* 23 (2) (2024) 1504–1539.
- [10] E. Bozzo, D. Breda, M. Tanveer, Sparse identification of time delay systems via pseudospectral collocation, *IFAC-PapersOnLine* 58 (27) (2024) 108–113.
- [11] A. Köpeczi-Bócz, H.S. Tamás, D. Takács, Data-driven delay identification with sindy, in: *International Conference on Nonlinear Dynamics and Applications*, Springer, 2023, pp. 481–491.
- [12] A. Pecile, N. Demo, M. Tezzele, G. Rozza, D. Breda, Data-driven discovery of delay differential equations with discrete delays, *J. Comput. Appl. Math.* 461 (2025) 116439.
- [13] A. Sandoz, V. Ducret, G.A. Gottwald, G. Vilmart, K. Perron, Sindy for delay-differential equations: application to model bacterial zinc response, *Proc. R. Soc. A* 479 (2023) 20220556.
- [14] P. Han, Q. Guo, H. Zhang, L. Wang, Approximation identification for the stochastic time-delayed dynamical system, *Phys. A: Stat. Mech. Appl.* 654 (2024) 130135.
- [15] D. Higham, P. Kloeden, *An introduction to the numerical simulation of stochastic differential equations*, Philadelphia, 2021.
- [16] X. Mao, Robustness of exponential stability of stochastic differential delay equations, *IEEE Trans. Autom. Control* 41 (3) (1996) 442–447.
- [17] N.O. Connell, Interacting diffusions on positive definite matrices, *Probab. Theory Relat. Fields* 180 (3) (2021) 679–726.
- [18] A. Raj, U. Simsekli, A. Rudi, Efficient sampling of stochastic differential equations with positive semi-definite models, *Adv. Neural Inf. Process. Syst.* 36 (2023) 64266–64305.
- [19] C.W. Gardiner, et al, *Handbook of stochastic methods*, 3, Springer, Berlin, Berlin, 1985.
- [20] D.J. Higham, X. Mao, A.M. Stuart, Strong convergence of euler-type methods for nonlinear stochastic differential equations, *SIAM J. Numer. Anal.* 40 (3) (2002) 1041–1063.
- [21] V. Da, The color of environmental noise, *Ecology* 85 (2004) 1146–1152.
- [22] N.G.V. Kampen, *Stochastic processes in physics and chemistry*, 1, Elsevier, 1992.
- [23] S. Guillouez, I. Heuroux, A. Longtin, Small delay approximation of stochastic delay differential equations, *Phys. Rev. E* 59 (4) (1999) 3970.
- [24] T. Tian, K. Burrage, P.M. Burrage, M. Carletti, Stochastic delay differential equations for genetic regulatory networks, *J. Comput. Appl. Math.* 205 (2) (2007) 696–707.
- [25] W. Cao, Z. Zhang, G.E. Karniadakis, Numerical methods for stochastic delay differential equations via the wong-Zakai approximation, *SIAM J. Sci. Comput.* 37 (1) (2015) 295–A318.
- [26] N. Rosli, A. Bahar, S. Yeak, X. Mao, A systematic derivation of stochastic taylor methods for stochastic delay differential equations, *Bull. Malays. Math. Sci. Soc.* 36 (3) (2013).
- [27] N. Rosli, A. Bahar, S.H. Yeak, H.A. Rahman, Stochastic taylor methods for stochastic delay differential equations, 2013, pp. 241–251.
- [28] D. Kleinhans, Estimation of drift and diffusion functions from time series data: a maximum likelihood framework, *Phys. Rev. E Stat. Nonlin. Soft Matter. Phys.* 85 (2) (2012) 26705.
- [29] D. Kleinhans, R. Friedrich, A. Nawroth, J. Peinke, An iterative procedure for the estimation of drift and diffusion coefficients of langevin processes, *Phys. Lett. A* 346 (1–3) (2005) 42–46.
- [30] H. Risken, *Fokker-planck equation*, in: *The Fokker-Planck Equation: Methods of Solution and Applications*, Springer, 1989, pp. 63–95.
- [31] K. Champion, B. Lusch, J.N. Kutz, S.L. Brunton, Data-driven discovery of coordinates and governing equations, *Proc. Natl. Acad. Sci.* 116 (2019) 22445–22451.
- [32] R. Chartrand, Numerical differentiation of noisy, nonsmooth data, *Int. Sch. Res. Not.* 2011 (1) (2011) 164564.

- [33] R. Tibshirani, Regression shrinkage and selection via the lasso, *J. R. Stat. Soc. Ser. B* 58 (1) (1996) 267–288.
- [34] M.R. Bonyadi, Z. Michalewicz, Particle swarm optimization for single objective continuous space problems: a review, *Evol. Comput.* 25 (1) (2017) 1–54.
- [35] J. Kennedy, R. Eberhart, Particle swarm optimization, 4, IEEE, 1995.
- [36] Y. Shi, R. Eberhart, A modified particle swarm optimizer, in: *IEEE International Conference on Evolutionary Computation*, IEEE, 1998, pp. 69–73.
- [37] D. Breda, O. Diekmann, M. Gyllenberg, F. Scarabel, R. Vermiglio, Pseudospectral discretization of nonlinear delay equations: new prospects for numerical bifurcation analysis, *SIAM J. Appl. Dyn. Syst.* 15 (1) (2016) 1–23.
- [38] B. Lenfesty, S. Bhattacharyya, K. Wong-Lin, Uncovering dynamical equations of stochastic decision models using data-driven SINDY algorithm, 2025, pp. 1–19.
- [39] Y. Wang, H. Fang, J. Jin, G. Ma, X. He, X. Dai, Z. Yue, C. Cheng, H.-T. Zhang, D. Pu, Et al. data-driven discovery of stochastic differential equations, *Engineering* 17 (2022) 244–252.
- [40] X. Chen, I. Timofeyev, Non-parametric estimation of stochastic differential equations from stationary time-series, *J. Stat. Phys.* 186 (2022) 1–31.
- [41] G. Mil'shtejn, Approximate integration of stochastic differential equations, *Theory Probab. Appl* 19 (3) (1975) 557–562.
- [42] M. Ragwitz, H. Kantz, Indispensable finite time corrections for fokker-Planck equations from time series data, *Phys. Rev. Lett.* 87 (25) (2001) 254501.
- [43] F.M. Bandi, P.C. Phillips, Fully nonparametric estimation of scalar diffusion models, *Econometrica* 71 (1) (2003) 241–283.
- [44] N. Mohammadi, L.V. Santoro, V.M. Panaretos, Nonparametric estimation for SDE with sparsely sampled paths: an FDA perspective, *Stoch. Process. Their Appl.* 167 (2024) 104239.
- [45] O. Babasola, E.O. Omondi, K. Oshinubi, N.M. Imbusi, Stochastic delay differential equations: a comprehensive approach for understanding biosystems with application to disease modelling, *AppliedMath* 3 (4) (2023) 702–721.
- [46] Z. Wu, H. Huang, L. Wang, Stochastic delay logistic model under regime switching, 2012, Wiley Online Library, 2012.
- [47] G.E. Hutchinson, Circular causal systems in ecology, *Ann. N. Y. Acad. Sci.* 50 (4) (1948) 221–246.
- [48] B. Du, Y. Wang, X. Lian, A stochastic predator-prey model with delays, 2015, pp. 1–16.
- [49] Y. Kazmerchuk, A. Swishchuk, J. Wu, The pricing of options for securities markets with delayed response, *Math. Comput. Simul.* 75 (3–4) (2007) 69–79.