Non-Markovian reduced models to unravel transitions in non-equilibrium systems

View the article online for updates and enhancements.

Non-Markovian reduced models to unravel transitions in non-equilibrium systems

Mickaël D Chekroun[1,2,*] ⓘ**, Honghu Liu**[3,*] ⓘ
and James C McWilliams[4] ⓘ

[1] Department of Atmospheric and Oceanic Sciences, University of California, Los Angeles, CA 90095-1565, United States of America
[2] Department of Earth and Planetary Sciences, Weizmann Institute of Science, Rehovot 76100, Israel
[3] Department of Mathematics, Virginia Tech, Blacksburg, VA 24061, United States of America
[4] Department of Atmospheric and Oceanic Sciences and Institute of Geophysics and Planetary Physics, University of California, Los Angeles, CA 90095-1565, United States of America

E-mail: mchekroun@atmos.ucla.edu and hhliu@vt.edu

CrossMark

Abstract

This work proposes a general framework for analyzing noise-driven transitions in spatially extended non-equilibrium systems and explaining the emergence of coherent patterns beyond the instability onset. The framework relies on stochastic parameterization formulas to reduce the complexity of the original equations while preserving the essential dynamical effects of unresolved scales. The approach is flexible and operates for both Gaussian noise and non-Gaussian noise with jumps. Our stochastic parameterization formulas offer two key advantages. First, they can approximate stochastic invariant manifolds when these manifolds exist. Second, even when such manifolds break down, our formulas can be adapted through a simple optimization of its constitutive parameters. This allows us to handle scenarios with weak time-scale separation where the system has undergone multiple transitions, resulting in large-amplitude solutions not captured by invariant manifolds or other time-scale separation methods. The optimized stochastic parameterizations capture then

* Authors to whom any correspondence should be addressed.

1

how small-scale noise impacts larger scales through the system's nonlinear interactions. This effect is achieved by the very fabric of our parameterizations incorporating non-Markovian (memory-dependent) coefficients into the reduced equation. These coefficients account for the noise's past influence, not just its current value, using a finite memory length that is selected for optimal performance. The specific memory function, which determines how this past influence is weighted, depends on both the strength of the noise and how it interacts with the system's nonlinearities. Remarkably, training our theory-guided reduced models on a single noise path effectively learns the optimal memory length for out-of-sample predictions. This approach retains indeed good accuracy in predicting noise-induced transitions, including rare events, when tested against a large ensemble of different noise paths. This success stems from our hybrid approach, which combines analytical understanding with data-driven learning. This combination avoids a key limitation of purely data-driven methods: their struggle to generalize to unseen scenarios, also known as the 'extrapolation problem.'

Contents

1. Introduction

Non-equilibrium systems are irreversible systems characterized by a continuous flow of energy that is driven by external forces or internal fluctuations. There are many different types of non-equilibrium systems, and they can be found in a wide variety of fields, including physics, chemistry, biology, and engineering. Non-equilibrium systems can exhibit complex behavior, including self-organization, pattern formation, and chaos. These complex behaviors arise from the interplay of nonlinear dynamics and statistical fluctuations. The emergence of coherent, complex patterns is ubiquitous in many spatially extended non-equilibrium systems; see e.g. [7, 12, 60, 98, 125, 126, 129, 136, 152, 178]. Mechanisms to explain the emergence of these patterns include the development of instability saturated by nonlinear effects whose calculations can be conducted typically near the onset of linear instability; see e.g. [58, 120]. Historically, hydrodynamic systems have been commonly used as prototypes to study instabilities out of equilibrium in spatially extended systems, from both an experimental and a theoretical point of view [59, 60].

To describe how physical instability develops in such systems with infinitely many degrees of freedom, we often focus on the amplitudes' temporal evolution of specific normal modes. The latter correspond typically to those that are mildly unstable and that are only slightly damped in linear theory. When the number of these nearly marginal modes is finite, their amplitudes are governed by ordinary differential equations (ODEs) in which the growth rates of the linear theory have been renormalized by nonlinear terms [58, 60, 98, 120]. Intuitively, the reason for this reduction is a simple separation of time scales. Modes that have just crossed the imaginary axis have a small real part and are evolving slowly on long time scales, all the other fast modes rapidly adapting themselves to these slow modes.

However, for non-equilibrium systems away from the onset of linear instability, such as when the Reynolds number for a fluid flow increases far beyond a laminar regime, the emergence of coherent patterns does not fit within this instability/nonlinear saturation theory as the reduction principle of the fast modes onto the slow ones breaks down [58] and calls for new reduction techniques to explain emergence. Furthermore, in the presence of random fluctuations, reduced equations under the form of deterministic ODEs are inherently incapable to capture phenomena like noise-induced transitions. Noise can have unexpected outcomes such as altering the sequence of transitions, possibly suppressing instabilities [44], the onset of turbulence [21] or, to the opposite, exciting transitions [104, 152]. Examples in which noise-induced transitions have an important role include the generation of convective rolls [1, 81], electroconvection in nematic liquid crystals [104, 144], certain oceanic flows [158, 162], and climate phenomena [69, 123, 148, 154].

This work aims to gain deeper insights into the efficient derivation of reduced models for such non-equilibrium systems subject to random fluctuations [60, 152]. In that respect, we seek reduced models able to capture the emergence of noise-driven spatiotemporal patterns and predict their transitions triggered by the subtle coupling between noise and the nonlinear dynamics. Our overall goal is to develop a unified framework for identifying the key variables and their interactions (parameterization) in complex systems undergoing noise-driven transitions. We specifically focus on systems that have experienced multiple branching points (bifurcations), leading to high-multiplicity regimes with numerous co-existing metastable states whose amplitude is relatively large (order one or more) resulting from a combination of strong noise, inherent nonlinear dynamics, or both. The purpose is thus to address limitations of existing approaches like amplitude equations and center manifold techniques in dealing with such regimes.

To address these limitations, the framework of optimal parameterizing manifolds (OPMs) introduced in [37, 39, 41] for forced-dissipative systems offers a powerful solution. This framework allows for the efficient derivation of parameterizations away from the onset of instability, achieved through continuous deformations of near-onset parameterizations that are optimized by data-driven minimization. Typically, the loss function involves a natural discrepancy metric measuring the defect of parameterization of the unresolved variables by the resolved ones. The parameterizations that are optimized are derived from the governing equations by means of backward–forward (BF) systems providing useful approximations of the unstable growth saturated by the nonlinear terms, even when the spectral gap between the stable and unstable modes is small [39, 41, 43]. These approximations, once optimized, do not suffer indeed the restrictions of invariant manifolds (spectral gap condition [41, equation (2.18)]) allowing to handle situations with e.g. weak time-scale separation between the resolved and unresolved modes.

In this article, we carry over this variational approach to the case of spatially extended non-equilibrium systems within the framework of stochastic partial differential equations (SPDEs). In particular, we extend the parameterization formulas obtained in [40, 42, 43] for SPDEs driven by multiplicative noise (parameter noise [81, 152]), to SPDEs driven, beyond this onset, by more realistic, spatio-temporal noise either of Gaussian or of non-Gaussian nature with jumps. For this class of SPDEs, our framework allows for dealing with the important case of cutoff scales larger than the scales forced stochastically.

For this forcing scenario, the stochastic parameterizations derived in this work, give rise to reduced systems taking the form of stochastic differential equations (SDEs) with non-Markovian, path-dependent coefficients depending on the noise's past; see e.g. section 5.2 below. We mention that many time series records from real-world datasets, are known to exhibit long-term memory. This is the case of long-range memory features assumed to represent the internal variability of the climate on time scales from years to centuries [75, 127, 150, 155]. There, the surface temperature is considered as a superposition of internal variability (the response to stochastic forcing) and a forced signal which is the linear response to external forcing. The background variability may be modeled using a stochastic process with memory, or a different process that incorporates non-Gaussianity if this is considered more appropriate [151]. As will become apparent at the end of this paper, our reduction approach is adaptable to PDEs under the influence of stochastic forcing with long-range dependence; see section 9.

In parallel and since Hasselmann's seminal work [95], several climate models have been incorporating stochastic forcing to represent unresolved processes [13, 77, 114, 117, 121, 134]. Our approach shows that reduced models derived from these complex systems are expected to exhibit finite-range memory effects depending on the past history of the stochastic forcing, even if the latter is white in time. These exogenous memory effects, stemming from the forcing

noise's history, are distinct from endogenous memory effects encountered in the reduction of nonlinear, unforced systems [48, 86, 117]. The latter, predicted by the Mori–Zwanzig (MZ) theory, are functionals of the past of the resolved state variables. Endogenous memory effects often appear when the validity of the conditional expectation in the MZ expansion, breaks down [37, 38, 117, 161]. The exogenous memory effects dealt with in this work, have a different origin, arising as soon as the cutoff scale is larger than the scales forced stochastically.

It is worth mentioning that reduced models with non-Markovian coefficients responsible for such exogenous memory effects, have been encountered in the reduction of stochastic systems driven by white noise, albeit near a change of stability. These are obtained from reduction approaches benefiting from timescale separation such as stochastic normal forms [5, 52, 130], stochastic invariant manifold (SIM) approximations [40, 42, 43, 73], or multiscale methods [14, 16, 17, 106]. Our reduction approach is not limited to timescale separation.

In that respect, our reduction framework is tested against a stochastic Allen–Cahn equation (sACE), a powerful tool for modeling non-equilibrium phase transitions in various scientific fields [22, 60, 98, 164]. The latter model is set in a parameter regime with weak timescale separation and far from the instability onset, in which the system exhibits multiple coexisting metastable states connected to the basic state through rare or typical stochastic transition paths.

We show that our resulting OPM reduced systems, when trained over a single path, are not only able to retain a remarkable predictive power to emulate ensemble statistics (correlations, power spectra, etc) obtained as average over a large ensemble of noise realizations, but also anticipate what are the system's typical and rare metastable states and their statistics of occurrence. The OPM reduced system's ability to reproduce accurately such transitions is rooted in its very structure. Its coefficients are nonlinear functionals of the aforementioned non-Markovian terms (see equation (85) below) allowing for an accurate representation of the genuine nonlinear interactions between noise and nonlinear terms in the original sACE, which drive fluctuations in the large-mode amplitudes.

We emphasize, that because the cutoff scale, defining the retained large-scale dynamics, is chosen here to be larger than the forcing scale, any deterministic reduction (e.g. Galerkin, invariant manifold, etc) would filter out the underlying noise's effects. This leaves them blind to the subtle fluctuations driving the system's behavior and leading eventually to stochastic transitions. In contrast, our stochastic parameterization approach sees right through this filter. It tracks how the noise, acting in the 'unseen' part of the system, interacts with the resolved modes through our reduction method's non-Markovian coefficients.

To demonstrate the broad applicability of our reduction approach, we apply it in section 6 to another significant class of spatially extended non-equilibrium systems: jump-driven SPDEs. The unforced, nonlinear dynamics is here chosen to exhibit an S-shaped solution curve. Many systems sharing this attribute are indicative of multistability, tipping points, and hysteresis. Such behaviors are observed in various fields, including combustion theory [8, 76], plasma physics [88, 157], ecology [126], neuroscience [102], climate science [83, 131], and oceanography [165, 172]. By understanding the complex interactions between noise and non-linearity into these phenomena, we can gain insights into critical transitions and tipping points, which are increasingly relevant to addressing global challenges like climate change [19, 20, 113, 117, 173].

Jump processes offer a powerful tool for modeling complex systems with non-smooth dynamics. They offer valuable insights into multistable behaviors [156, 179] and have been applied to various real-world phenomena, such as paleoclimate events [68, 151], chaotic transport [159], atmospheric dynamics [46, 47, 105, 137, 160, 166], and cloud physics [33, 100].

SPDEs driven by jump processes are gaining increasing attention in applications [66]. For instance, jump processes can replace (non-smooth) 'if-then' conditions in such models, enabling more efficient simulations [47]. Nevertheless, efficient reduction techniques to disentangle the jump interactions with other smoother nonlinear components of the model, are still under development [115, 174, 176]. Our stochastic parameterization framework provides a promising solution in this direction, as demonstrated in section 6. By capturing the intricate interplay between jump noise and nonlinear dynamics within the reduced equations, our approach can significantly simplify the analysis of such complex systems.

2. Invariance equation and approximations

2.1. Spatially extended non-equilibrium systems

This article is concerned with the efficient reduction of spatially extended non-equilibrium systems. To do so, we work within the framework of Stochastic Equations in Infinite Dimensions [64] and its Ergodic Theory [63]. Formally, these equations take the following form

$$\mathrm{d}u = (Au + G(u))\,\mathrm{d}t + \mathrm{d}\boldsymbol{\eta}_t, \tag{1}$$

in which $\boldsymbol{\eta}_t$ is a stochastic process, either composed of Brownian motions or jump processes and whose exact structure is specified below. This formalism provides a convenient way to analyze stochastic partial differential equations (SPDEs) by means of semigroup theory [96, 135], in which the unknown u evolves typically in a Hilbert space H.

The operator A represents a linear differential operator while G is a nonlinear operator that accounts for the nonlinear terms. Both of these operators may involve loss of spatial regularity when applied to a function u in H. To have a consistent existence theory of solutions and their stochastic invariant manifolds for SPDEs requires to take into account such loss of regularity effects [63, 64].

General assumptions encountered in applications are made on the linear operator A following [96]. More specifically, we assume

$$A = -\mathcal{L} + \mathcal{B}, \tag{2}$$

where $\mathcal{L}$ is sectorial with domain $D(\mathcal{L}) \subset H$ which is compactly and densely embedded in H. We assume also that $-\mathcal{L}$ is stable, while $\mathcal{B}$ is a low-order perturbation of $\mathcal{L}$, i.e. $\mathcal{B}$ is a bounded linear operator such that $\mathcal{B} : D(\mathcal{L}^\alpha) \to H$ for some α in $[0, 1)$. We refer to [96, section 1.4] for an intuitive presentation of fractional power of an operator and characterization of the so-called operator domain $D(\mathcal{L}^\alpha)$. In practice, the choice of α should match the loss of regularity effects caused by the nonlinear terms so that

$$G\colon D(\mathcal{L}^\alpha) \to H,$$

is a well-defined C^p-smooth mapping that satisfies $G(0) = 0, DG(0) = 0$ (tangency condition), and

$$G(u) = G_k(u, \cdots, u) + O\left(\|u\|_\alpha^{k+1}\right), \tag{3}$$

with $p > k \geqslant 2$, and G_k denoting the leading-order operator (on $D(\mathcal{L}^\alpha)$) in the Taylor expansion of G, near the origin. A broad class of spatially extended stochastic equations from physics can be recast into this framework (see [40, 42, 64, 116, 119] for examples), as well as time-delay systems subject to stochastic disturbances [32, 33, 36].

Throughout this article, we assume that the driving noise in equation (1) is an H-valued stochastic process that takes the form

$$\boldsymbol{\eta}_t(\omega) = \sum_{j=1}^{N} \sigma_j \eta_t^j(\omega) \boldsymbol{e}_j, \; t \in \mathbb{R}, \; \sigma_j \geqslant 0, \; \omega \in \Omega, \tag{4}$$

where the $\boldsymbol{e}_j$ are eigenmodes of A, the η_t^j denote either a finite family of mutually independent jump processes, or mutually independent Brownian motions, W_t^j, over their relevant probability space $(\Omega, \mathbb{P}, \mathcal{F})$ endowed with its canonical probability measure $\mathbb{P}$ and filtration $\mathcal{F}$; see e.g. [5, appendix A.3].

Within this framework, the reduction of spatially extended non-equilibrium systems is organized in terms of resolved and unresolved spatial scales. The resolved scales are typically spanned by large wavenumbers and the unresolved by smaller ones. In that respect, the eigenmodes of the operator A plays a central role throughout this paper to rank the spatial scales. Typically, we assume that the state space $H_{\mathfrak{c}}$ of resolved variables is spanned by the following modes

$$H_{\mathfrak{c}} = \operatorname{span}\{\boldsymbol{e}_1, \cdots, \boldsymbol{e}_{m_c}\}, \tag{5}$$

where the $\boldsymbol{e}_j$ correspond to large-scale modes up to a cutoff-scale associated with some index m_c (see remark 2.1). The subspace of unresolved modes is then the orthogonal complement of $H_{\mathfrak{c}}$ in H, namely

$$H_{\mathfrak{c}} \oplus H_{\mathfrak{f}} = H. \tag{6}$$

To these subspaces, we associate their respective canonical projectors denoted by

$$\Pi_{\mathfrak{c}} : H \to H_{\mathfrak{c}}, \text{and } \Pi_{\mathfrak{f}} : H \to H_{\mathfrak{f}}. \tag{7}$$

Throughout section 3 below and the remaining of this section, we focus on the case of driving Brownian motions; the case of driving jump processes is dealt with in section 6.

Our goal is to provide a general approach to derive reduced models that preserve the essential features of the large-scale dynamics without resolving the small scales.

We present hereafter and in section 3 below, the formulas of the underlying small-scale parameterizations in the Gaussian noise case. The formalism is flexible and easily adaptable to the case of non-Gaussian noises with jumps such as discussed in sections 6 and 7, below.

Remark 2.1. Within our working assumptions, the spectrum $\sigma(A)$ consists only of isolated eigenvalues with finite multiplicities. This combined with the sectorial property of A implies that there are at most finitely many eigenvalues with a given real part. The sectorial property of A also implies that the real part of the spectrum, $\operatorname{Re}(\sigma(A))$, is bounded above (see also [70, theorem II.4.18]). These two properties of $\operatorname{Re}(\sigma(A))$ allow us in turn to label elements in $\sigma(A)$ according to the lexicographical order which we adopt throughout this article:

$$\sigma(A) = \{\lambda_n \mid n \in \mathbb{N}^*\}, \tag{8}$$

such that for any $1 \leqslant n < n'$ we have either

$$\operatorname{Re}(\lambda_n) > \operatorname{Re}(\lambda_{n'}), \tag{9}$$

or

$$\operatorname{Re}(\lambda_n) = \operatorname{Re}(\lambda_{n'}), \text{and } \operatorname{Im}(\lambda_n) \geqslant \operatorname{Im}(\lambda_{n'}). \tag{10}$$

This way, we can rely on a simple labeling of the eigenvalues/eigenmodes by positive integers to organize the resolved and unresolved scales and the corresponding parameterization formulas derived hereafter. In practice, when the spatial domain is 2D or 3D, it is usually physically more intuitive to label the eigenelements by wave vectors. The parameterization formulas presented below can be easily recast within this convention; see e.g. [31].

2.2. Invariance equation and BF systems

As mentioned in Introduction, this work extends the parameterization formulas in [39, 41] (designed for constant or time-dependent forcing) to the stochastic setting. In both deterministic and stochastic contexts, deriving the relevant parameterizations and reduced systems relies on solving BF systems arising in the theory of invariant manifolds. Similar to [39, 41], our BF framework allows us to overcome limitations of traditional invariant manifolds, particularly the spectral gap condition ([41, equation. (2.18)]). This restrictive condition, which requires large gaps in the spectrum of the operator A, is bypassed through data-driven optimization (detailed in section 3.3) of the backward integration time over which the BF systems are integrated.

Before diving into the parameterization formulas retained to address these limitations (section 2.3), we recall the basics of reducing an SPDE to a SIM. In particular, this detour enables us to better appreciate the emergence of BF systems as a natural framework for parameterizations. For that purpose, adapting the approach from [42], we transform the SPDE reduction problem into the more tractable problem of reducing a PDE with random coefficients. This simplification makes the problem significantly more amenable to analysis than its original SPDE form.

To do so, consider the stationary solution $z(t, \omega)$ to the Langevin equation

$$\mathrm{d}z = Az\,\mathrm{d}t + \mathrm{d}\boldsymbol{W}_t, \tag{11}$$

(Ornstein-Uhlenbeck (OU) process), where $\boldsymbol{W}_t$ is defined in equation (4) with the corresponding mutually independent Brownian motions, W_t^j, in place of the η_t^j, i.e.:

$$\boldsymbol{W}_t(\omega) = \sum_{j=1}^{N} \sigma_j W_t^j(\omega)\boldsymbol{e}_j,\ t \in \mathbb{R},\ \omega \in \Omega. \tag{12}$$

Then, for each noise's path ω, the change of variable

$$v = u - z(t, \omega), \tag{13}$$

transforms the SPDE,

$$\mathrm{d}u = (Au + G(u))\,\mathrm{d}t + \mathrm{d}\boldsymbol{W}_t,$$

into the following PDE with random coefficients in the v-variable:

$$\frac{\mathrm{d}v}{\mathrm{d}t} = Av + G(v + z(t, \omega)). \tag{14}$$

Under standard conditions involving the spectral gap between the resolved and unresolved modes [10], the path-dependent PDE (14) admits a SIM, and the underlying stochastic SIM mapping, $\phi(t, X)$ $(X \in H_{\mathfrak{c}})$, satisfies the stochastic invariance equation:

$$\partial_t \phi + \mathcal{L}_A[\phi](X) = -D\phi(t, X, \omega)(\Pi_{\mathfrak{c}}G(X + \phi + z(t, \omega))) + \Pi_{\mathfrak{f}}G(X + \phi + z(t, \omega)), \tag{15}$$

where $\mathcal{L}_A$ denotes the operator acting on differentiable mappings ψ from $H_{\mathfrak{c}}$ into $H_{\mathfrak{f}}$, as follows:

$$\mathcal{L}_A [\psi] (X) = D\psi (X) A_{\mathfrak{c}} X - A_{\mathfrak{f}} \psi (X), \ X \in H_{\mathfrak{c}}, \tag{16}$$

with $A_{\mathfrak{c}} = \Pi_{\mathfrak{c}} A$ and $A_{\mathfrak{f}} = \Pi_{\mathfrak{f}} A$.

To simplify, assume that $H_{\mathfrak{c}}$ and $H_{\mathfrak{f}}$ are chosen such that $\Pi_{\mathfrak{f}} W_t = 0$. Then in particular $z_{\mathfrak{f}} = \Pi_{\mathfrak{f}} z = 0$. Consider now the Lyapunov–Perron integral

$$\mathfrak{J} (t, X, \omega) = \int_{-\infty}^{t} e^{(t-s)A_{\mathfrak{f}}} \Pi_{\mathfrak{f}} G_k \left(e^{(s-t)A_{\mathfrak{c}}} X + z_{\mathfrak{c}} (s, \omega) \right) ds, \tag{17}$$

where $z_{\mathfrak{c}} = \Pi_{\mathfrak{c}} z$. In what follows, we denote by $\langle \cdot, \cdot \rangle$ the Hermitian inner product on H defined as $\langle f, g \rangle = \int f(x) \overline{g}(x) dx$, while G_k denotes the leading-order term of order k in the Taylor expansion of $G(X)$ around $X = 0$, and $\boldsymbol{e}_n^*$ denotes the nth eigenmode of the adjoint operator of A adopting the same labelling convention as for A; see remark 2.1. To the order k we associate the set K of indices $\{1, \cdots, k\}$, and denote by K_z any subset of indices made of (possibly empty) of disjoint elements of K.

We have then the following result.

Theorem 2.1. *Under the previous assumptions, assume furthermore that the following non-resonance condition holds, for any $(j_1, \cdots, j_k)$ in $(1, \cdots, m_c)^k$, $n \geqslant m_c + 1$, and any subset K_z such as defined above:*

$$\left(G_{j_1 \cdots j_k}^n \neq 0 \ and \ \Pi_{q \in K_z} \sigma_q \neq 0 \right) \Longrightarrow \left(\mathrm{Re} \left(\lambda_n - \sum_{p \in K \setminus K_z} \lambda_{j_p} \right) < 0 \right), \tag{18}$$

with $G_{j_1 \cdots j_k}^n = \langle G_k(\boldsymbol{e}_{j_1}, \cdots, \boldsymbol{e}_{j_k}), \boldsymbol{e}_n^* \rangle$.

Then, the Lyapunov–Perron integral $\mathfrak{J}$ (equation (17)) is well-defined almost surely, and is a solution to the following homological equation with random coefficients:

$$\partial_t \psi + \mathcal{L}_A [\psi] (X) = \Pi_{\mathfrak{f}} G_k (X + z_{\mathfrak{c}} (t, \omega)). \tag{19}$$

Moreover, we observe that the integral $\mathfrak{J}$ is (formally) obtained as the limit, when τ goes to infinity, of the q-solution to the following BF auxiliary system:

$$\frac{dp}{dt} = A_{\mathfrak{c}} p, s \in [t - \tau, t], \tag{20a}$$

$$\frac{dq}{dt} = A_{\mathfrak{f}} q + \Pi_{\mathfrak{f}} G_k (p + z_{\mathfrak{c}} (s, \omega)), \quad s \in [t - \tau, t], \tag{20b}$$

$$with \ p (s)|_{s=t} = X \in H_{\mathfrak{c}}, and \ q (s)|_{s=t-\tau} = 0. \tag{20c}$$

That is

$$\lim_{\tau \to \infty} \| \mathfrak{J} (t, X, \omega) - q_\tau (t, X, \omega) \|_{H_{\mathfrak{f}}} = 0, \tag{21}$$

where $q_\tau(t, X, \omega)$ denotes the solution to equation (20b) at time $s = t$, when initialized with $q = 0$ at $s = t - \tau$.

For a proof of this Theorem, see appendix A. This theorem extends to the stochastic context theorem III.1 of [39]. To simplify the notations, we omit below the ω-dependence in certain notations unless specified otherwise.

Theorem 2.1 teaches us that solving the BF system (20a)–(20c) gives the solution to the homological equation (17) which is an approximation of the full invariance equation equation (15). In the deterministic setting, solutions to the homological equation (17) are

known to provide actual approximations of the invariant manifolds with rigorous estimates; see [41, theorem 1] and [39, theorem III.1]. Such results extend to the case of SPDEs driven by multiplicative (parameter) noise; see [40, theorem 2.1] and [42, theorem 6.1 and corollary 7.1]. It is not the scope of this article to deal with rigorous error estimates regarding the approximation problem of stochastic invariant manifolds for SPDEs driven by additive noise, but the rationale stays the same: solutions to the homological equation (17) or equivalently to the BF system (20*a*)–(20*c*) provide actual approximations of the underlying stochastic invariant manifolds, in the additive noise case as well.

Going back to the SPDE variable *u*, the BF system (20*a*)–(20*c*) becomes

$$d\widehat{p} = A_{\mathfrak{c}}\widehat{p}\,ds + \Pi_{\mathfrak{c}}\,d\boldsymbol{W}_s, \qquad s \in [t-\tau, t], \tag{22a}$$

$$d\widehat{q} = (A_{\mathfrak{f}}\widehat{q} + \Pi_{\mathfrak{f}}G_k(\widehat{p}))\,ds, \qquad s \in [t-\tau, t], \tag{22b}$$

$$\text{with } \widehat{p}(s)|_{s=t} = X \in H_{\mathfrak{c}}, \text{ and } \widehat{q}(s)|_{s=t-\tau} = 0. \tag{22c}$$

Note that in this BF system, only the low-mode variable $\widehat{p}$ is forced stochastically, and thus from what precedes, $\lim_{\tau\to\infty}\widehat{q}_\tau(t,X)$, provides a legitimate approximation of the SIM at time *t* when in the original SPDE, only the low modes are forced stochastically, i.e. $N = m_c$ in equation (12). In the next section, we consider the general case when the low and high modes are stochastically forced.

2.3. Approximations of fully coupled BF systems

BF systems have been actually proposed in the literature for the construction of SIMs [61], through a different route than what presented above, i.e. without exploiting the invariance equation. The idea pursued in [61] is to envision SIMs as a fixed point of an integral form of the following fully coupled BF system

$$dp = [A_{\mathfrak{c}}p + \Pi_{\mathfrak{c}}G(p+q)]\,ds + \Pi_{\mathfrak{c}}d\boldsymbol{W}_s, \ \ s \in I_{t,\tau}, \tag{23a}$$

$$dq = [A_{\mathfrak{f}}q + \Pi_{\mathfrak{f}}G(p+q)]\,ds + \Pi_{\mathfrak{f}}d\boldsymbol{W}_s, \ \ s \in I_{t,\tau}, \tag{23b}$$

$$\text{with } p(s)|_{s=t} = X \in H_{\mathfrak{c}}, \text{ and } q(s)|_{s=t-\tau} = 0. \tag{23c}$$

where $I_{t,\tau} = [t-\tau, t]$. Note that in equations (23*a*) and (23*b*), we do not assume here the noise term $\boldsymbol{W}_s$ to be a *finite* sum of mutually independent Brownian motions. In the case of an infinite sum one can thus covers the case of space-time white noise due to [103].

In any case, it is known that this type of *fully coupled nonlinear problems* involving backward stochastic equations does not have always solutions in general and we refer to [61, proposition 3.1] for conditions on *A* and *G* ensuring existence of solutions and thus SIM. This (nonlinear) BF approach to SIM is also subject to a spectral gap condition that requires gap to be large enough as in [10].

Denoting by $q_\tau(t,X)$ the solution to equation (23*b*) at $s = t$, proposition 3.4 of [61] ensures then that $\lim_{\tau\to\infty}q_\tau(t,X)$ exists in $L^2(\Omega)$ and that this limit gives the sought SIM, i.e.

$$\Psi^{\text{sim}}(X,t) = \lim_{\tau\to\infty}q_\tau(t,X), \quad X \in H_{\mathfrak{c}}. \tag{24}$$

The SIM is thus obtained, almost surely, as the asymptotic graph of the mapping $X \mapsto q_\tau(t,X)$, when τ is sent to infinite.

Instead of relying on an integral form of equations (23*a*) and (23*b*), one can address though the existence and construction of a SIM via a more direct approach, exploiting iteration

schemes built directly from equations (23*a*) and (23*b*). It is not the purpose of this article to analyze the convergence of such iterative schemes (subject also to a spectral gap condition as in [61]) but rather to illustrate how informative such schemes can be in designing stochastic parameterizations in practice.

To solve equations (23*a*) and (23*b*), given an initial guess, $(p^{(0)}, q^{(0)})$, we propose the following iterative scheme:

$$\mathrm{d}p^{(\ell)} = \left[A_{\mathfrak{c}}p^{(\ell)} + \Pi_{\mathfrak{c}}G\left(p^{(\ell-1)} + q^{(\ell-1)}\right)\right]\mathrm{d}s + \Pi_{\mathfrak{c}}\mathrm{d}\boldsymbol{W}_s, \tag{25a}$$

$$\mathrm{d}q^{(\ell)} = \left[A_{\mathfrak{f}}q^{(\ell)} + \Pi_{\mathfrak{f}}G\left(p^{(\ell-1)} + q^{(\ell-1)}\right)\right]\mathrm{d}s + \Pi_{\mathfrak{f}}\mathrm{d}\boldsymbol{W}_s, \tag{25b}$$

where $\ell \geqslant 1$ and with $p^{(\ell)}(s)|_{s=t} = X$ in $H_{\mathfrak{c}}$ and $q^{(\ell)}(s)|_{s=t-\tau} = 0$. Here again, the first equation (equation (25*a*)) is integrated backward over $[t-\tau, t]$, followed by a forward integration of the second equation (equation (25*b*)) over the same interval. To simplify the notations, we will often omit to point out the interval $[t-\tau, t]$ in the BF systems below.

To help interpret the parameterization produced by such an iterative scheme, we restrict momentarily ourselves to the case of a nonlinearity G that is quadratic (denoted by B) and to noise terms that are scaled by a parameter ϵ. This case covers the important case of the Kardar–Parisi–Zhang equation [51]. The inclusion of this scaling factor ϵ allows us to group the terms constituting the stochastic parameterization according to powers of ϵ providing in particular useful small-noise expansions; see equation (30) below.

Under this framework, if we start with $(p^{(0)}, q^{(0)}) \equiv (0,0)$, we get then

$$p^{(1)}(s) = \mathrm{e}^{(s-t)A_{\mathfrak{c}}}X - \epsilon \int_s^t \mathrm{e}^{(s-s')A_{\mathfrak{c}}}\Pi_{\mathfrak{c}}\,\mathrm{d}\boldsymbol{W}_{s'}, \tag{26a}$$

$$q^{(1)}(s) = \epsilon \int_{t-\tau}^s \mathrm{e}^{(s-s')A_{\mathfrak{f}}}\Pi_{\mathfrak{f}}\,\mathrm{d}\boldsymbol{W}_{s'}, \tag{26b}$$

which leads to

$$p^{(2)}(s) = \mathrm{e}^{(s-t)A_{\mathfrak{c}}}X - \int_s^t \mathrm{e}^{(s-s')A_{\mathfrak{c}}}\Pi_{\mathfrak{c}}B\left(p^{(1)}(s') + q^{(1)}(s')\right)\mathrm{d}s' - \epsilon \int_s^t \mathrm{e}^{(s-s')A_{\mathfrak{c}}}\Pi_{\mathfrak{c}}\,\mathrm{d}\boldsymbol{W}_{s'},$$

and

$$q^{(2)}(s) = \int_{t-\tau}^s \mathrm{e}^{(s-s')A_{\mathfrak{f}}}\Pi_{\mathfrak{f}}B\left(p^{(1)}(s') + q^{(1)}(s')\right)\mathrm{d}s' + \epsilon \int_{t-\tau}^s \mathrm{e}^{(s-s')A_{\mathfrak{f}}}\Pi_{\mathfrak{f}}\,\mathrm{d}\boldsymbol{W}_{s'}. \tag{27}$$

To further make explicit $q^{(2)}(s)$ that provides the stochastic parameterization we are seeking (after two iterations), we introduce

$$f_i(s) = -\int_s^t \mathrm{e}^{(s-s')\lambda_i}\,\mathrm{d}W_{s'}^i(\omega), \quad i = 1, \ldots, m_c, \tag{28a}$$

$$g_n(s) = \int_{t-\tau}^s \mathrm{e}^{(s-s')\lambda_n}\,\mathrm{d}W_{s'}^n(\omega), \quad n \geqslant m_c + 1. \tag{28b}$$

We can then rewrite $(p^{(1)}, q^{(1)})$ given in equation (26*a*) as follows:

$$p^{(1)}(s) = \sum_{i=1}^{m_c} \left(\mathrm{e}^{(s-t)\lambda_i}X_i + \epsilon f_i(s)\right)\boldsymbol{e}_i, \tag{29a}$$

$$q^{(1)}(s) = \epsilon \sum_{n \geqslant m_c+1} g_n(s) \boldsymbol{e}_n. \tag{29b}$$

By introducing additionally $B_{ij}^n = \langle B(\boldsymbol{e}_i, \boldsymbol{e}_j), \boldsymbol{e}_n^* \rangle$, we get for any $n \geqslant m_c + 1$ the following expansions

$$\Pi_n B\left(p^{(1)}(s), p^{(1)}(s)\right) = \epsilon^2 \sum_{i,j=1}^{m_c} B_{ij}^n f_i(s) f_j(s) + \epsilon \sum_{i,j=1}^{m_c} B_{ij}^n f_i(s) e^{(s-t)\lambda_j} X_j$$

$$+ \epsilon \sum_{i,j=1}^{m_c} B_{ij}^n f_j(s) e^{(s-t)\lambda_i} X_i + \sum_{i,j=1}^{m_c} B_{ij}^n e^{(s-t)(\lambda_i+\lambda_j)} X_i X_j,$$

$$\Pi_n B\left(p^{(1)}(s), q^{(1)}(s)\right) = \epsilon^2 \sum_{i=1}^{m_c} \sum_{n'=m_c+1}^{\infty} B_{i,n'}^n f_i(s) g_{n'}(s) + \epsilon \sum_{i=1}^{m_c} \sum_{n'=m_c+1}^{\infty} B_{in'}^n g_{n'}(s) e^{(s-t)\lambda_i} X_i,$$

and

$$\Pi_n B\left(q^{(1)}(s), q^{(1)}(s)\right) = \epsilon^2 \sum_{n',n''\geqslant m_c+1} B_{n'n''}^n g_{n'}(s) g_{n''}(s).$$

Let us denote by Π_n the projector onto the mode $\boldsymbol{e}_n$. Using the above identities in (27), and setting $s = t$, we get for $n \geqslant m_c + 1$,

$$\Pi_n q^{(2)}(t) = \epsilon^2 a_n(\tau) + \epsilon \sum_{i=1}^{m_c} b_{in}(\tau) X_i + \sum_{i,j=1}^{m_c} c_{ij}^n(\tau) B_{ij}^n X_i X_j + \epsilon \int_{t-\tau}^{t} e^{(t-s')\lambda_n} dW_{s'}^n(\omega), \tag{30}$$

where

$$a_n(\tau) = \underbrace{\int_{t-\tau}^{t} e^{(t-s')\lambda_n} \sum_{i,j=1}^{m_c} B_{ij}^n f_i(s') f_j(s') \, ds'}_{(\alpha)}$$

$$+ \underbrace{\int_{t-\tau}^{t} e^{(t-s')\lambda_n} \left(\sum_{i=1}^{m_c} \sum_{n'=m_c+1}^{\infty} (B_{in'}^n + B_{n'i}^n) f_i(s') g_{n'}(s') \right) ds'}_{(\beta)}$$

$$+ \underbrace{\int_{t-\tau}^{t} e^{(t-s')\lambda_n} \left(\sum_{n',n''\geqslant m_c+1} B_{n'n''}^n g_{n'}(s') g_{n''}(s') \right) ds'}_{(\gamma)}, \tag{31}$$

$$b_{in}(\tau) = \int_{t-\tau}^{t} e^{(t-s')(\lambda_n-\lambda_i)} \sum_{j=1}^{m_c} (B_{ij}^n + B_{ji}^n) f_j(s') \, ds'$$

$$+ \int_{t-\tau}^{t} e^{(t-s')(\lambda_n-\lambda_i)} \sum_{n'=m_c+1}^{\infty} (B_{in'}^n + B_{n'i}^n) g_{n'}(s') \, ds', \tag{32}$$

and

$$c_{ij}^n(\tau) = \int_{t-\tau}^{t} e^{(t-s')(\lambda_n-(\lambda_i+\lambda_j))} \, ds' = \begin{cases} \frac{1-\exp(-\delta_{ij}^n \tau)}{\delta_{ij}^n}, & \text{if } \delta_{ij}^n \neq 0, \\ \tau, & \text{otherwise,} \end{cases}$$

with $\delta_{ij}^n = \lambda_i + \lambda_j - \lambda_n$.

In the classical approximation theory of SIM, one is interested in conditions ensuring convergence of the integrals involved in the random coefficients a_n and b_{in} as $\tau \to \infty$, since $q^{(2)}(t)$ (and its higher-order analogues $q^{(\ell)}(t)$ with $\ell > 2$) aims to approximate the SIM defined in equation (24). In our approach to stochastic parameterization, we do not restrict ourselves to such a limiting case but rather seek optimal backward integration time τ that minimizes a parameterization defect as explained below in section 3.3.

However, computing this parameterization defect involves the computation of the random coefficients in the course of time (see equation (53) below). The challenge is that the structure of a_n and b_{in} involves repeated stochastic convolutions in time, and as such one wants to avoid a direct computation by quadrature. We propose below an alternative and efficient way to compute such random coefficients, for a simpler, more brutal approximation than $q^{(2)}$. As shown in sections 5 and 6, this other class of stochastic parameterizations turns out to be highly performant for the important case where the stochastically forced scales are exclusively part of the neglected scales.

This approximation consists of setting $q^{(1)} = 0$ in equation (27), namely to deal with the stochastic parameterization

$$\Phi(t, X, \omega) = \int_{t-\tau}^t e^{(t-s)A_{\mathfrak{f}}} \Pi_{\mathfrak{f}} G\left(e^{(s-t)A_{\mathfrak{c}}} X\right) ds + \epsilon \int_{t-\tau}^t e^{(t-s)A_{\mathfrak{f}}} \Pi_{\mathfrak{f}} dW_s(\omega), \tag{33}$$

going back to the original model (1) with its nonlinearity G.

Note also that this parameterization, $\Phi(t, X, \omega)$, is exactly the solution $q(s, X, \omega)$ at $s = t$ of the following BF system

$$dp = A_{\mathfrak{c}} p \, ds, \qquad\qquad s \in I_{t,\tau}, \tag{34a}$$

$$dq = (A_{\mathfrak{f}} q + \Pi_{\mathfrak{f}} G(p)) \, ds + \epsilon \Pi_{\mathfrak{f}} dW_s, \quad s \in I_{t,\tau}, \tag{34b}$$

$$\text{with } p(s)|_{s=t} = X \in H_{\mathfrak{c}}, \text{and } q(s)|_{s=t-\tau} = 0, \tag{34c}$$

with $I_{t,\tau} = [t - \tau, t]$.

Section 3.2 details an efficient computation of the single stochastic convolution in equation (33) using auxiliary ODEs with random coefficients. It is worth noting that previous SPDE reduction approaches have, in certain circumstances or under specific assumptions (e.g. $\Pi_{\mathfrak{c}} B(q, q) = 0$ as seen in [17, equation (2.2)] and [121, equation (2.10)]), deliberately opted to avoid directly computing these convolutions via quadrature [145, 146, 170].

The stochastic convolution in equation (33) accounts for finite-range memory effects stemming from the stochastic forcing to the original equation. Such terms have been encountered in the approximation of low-mode amplitudes, albeit in their asymptotic format when $\tau \to \infty$ see e.g. [17, 18]. As shown in our examples below (sections 5 and 6), stochastic convolution terms become crucial for efficient reduction when the cutoff scale, defining the retained large-scale dynamics, is larger than the forcing scale. There, we show furthermore that the finite-range memory content (measured by τ) is a key factor for a skillful reduction when optimized properly. Also, as exemplified in applications, optimizing the nonlinear terms in equation (33) may turn to be of utmost importance to reproduce accurately the average motion of the SPDE dynamics; see section 5.5 below.

The efficient computation of repeated convolutions involved in the coefficients a_n and b_{in} is however more challenging and will be communicated elsewhere. This is not only a technical aspect though, as these repeated convolutions in a_n given by equation (31) characterize important triad interactions reflecting how the noise interacts through the nonlinear terms

into three groups: the low–low interactions in (α), the low–high interactions in (β) and high–high interactions in (γ). By using the simpler parameterization Φ defined in equation (33) we do not keep these interactions at the parameterization level, however as Φ approximates the high-mode amplitude it still allows us to account for triad interactions into the corresponding reduced models; see e.g. equation (85) below.

3. Non-Markovian parameterizations: formulas and optimization

The previous discussion leads us to consider, for each $n \geqslant m_c + 1$, the following scale-aware BF systems

$$\mathrm{d}p = A_{\mathfrak{c}}p\,\mathrm{d}s, \qquad\qquad s \in [t - \tau, t], \qquad\qquad (35a)$$

$$\mathrm{d}q = (\lambda_n q + \Pi_n G(p))\,\mathrm{d}s + \sigma_n \,\mathrm{d}W_s^n, \quad s \in [t - \tau, t], \qquad\qquad (35b)$$

$$\text{with } p(s)|_{s=t} = X \in H_{\mathfrak{c}}, \text{and } q(s)|_{s=t-\tau} = Y \in \mathbb{R}, \qquad\qquad (35c)$$

where Π_n denotes the projector onto the mode $\boldsymbol{e}_n$. Here, τ is a free parameter to be adjusted, no longer condemned to approach ∞ as in the approximation theory of SIM discussed above. Note that compared to equation (34b), we break down the forcing mode by mode for each high mode to allow for adjusting the free backward parameter, τ, per scale to parameterize (scale-awareness). This strategy allows for a greater degree of freedom to calibrate useful parameterizations, as will be apparent in applications dealt with in sections 5 and 6.

Also, compared to equation (34b), the initial condition for the forward integration in equation (35b) is a scale-aware parameter Y in equation (35c). It is aimed at resolving the time-mean of the nth high-mode amplitude, $\langle u(t), \boldsymbol{e}_n^* \rangle$ with u solving equation (1). In many applications, it is enough to set $Y = 0$ though.

In the following, we make explicit the stochastic parameterization obtained by integration of the BF system (35a)–(35c) for SPDEs with cubic nonlinear terms, for which the stochastically forced scales are part of the neglected scales.

3.1. Stochastic parameterizations for systems with cubic interactions

We consider

$$\mathrm{d}u = (Au + G_2(u, u) + G_3(u, u, u))\,\mathrm{d}t + \mathrm{d}\boldsymbol{W}_t. \qquad\qquad (36)$$

Here, A is a partial differential operator as defined in equation (2) for a suitable ambient Hilbert space H; see also [43, chapter 2]. The operators G_2 and G_3 are quadratic and cubic, respectively. Here again, we assume the eigenmodes of A and its adjoint A^* to form a bi-orthonormal basis of H; namely, the set of eigenmodes $\{\boldsymbol{e}_j\}$ of A and $\{\boldsymbol{e}_j^*\}$ of A^* each forms a Hilbert basis of H, and they satisfy the bi-orthogonality condition $\langle \boldsymbol{e}_j, \boldsymbol{e}_k^* \rangle_H = \delta_{j,k}$, where $\delta_{j,k}$ denotes the Kronecker delta function. The noise term $\boldsymbol{W}_t$ takes the form defined in equation (12). Stochastic equations such as equation (36) arise in many areas of physics such as in the description of phase separation [22, 26, 62], pattern formations [60, 152] or transitions in geophysical fluid models [40, 80, 123, 140, 153, 158, 162].

We assume that equation (36) is well-posed in the sense of possessing for any u_0 in H a unique mild solution, i.e. there exists a stopping time $T > 0$ such that the integral equation,

$$u(t) = \mathrm{e}^{At}u_0 + \int_0^t \mathrm{e}^{A(t-s)}(G_2(u(s)) + G_3(u(s)))\,\mathrm{d}s + \int_0^t \mathrm{e}^{A(t-s)}\,\mathrm{d}\boldsymbol{W}_s, \ t \in [0, T], \qquad (37)$$

possesses, almost surely, a unique solution in $C([0,T],H)$ (continuous path in the Hilbert space H) for all stopping time $t \leqslant T$.

Here the stochastic integral can be represented as a series of one-dimensional Itô integrals

$$\int_0^t \mathrm{e}^{A(t-s)} \, \mathrm{d}\boldsymbol{W}_s = \sum_k \sigma_k \int_0^t \mathrm{e}^{\lambda_k(t-s)} \, \mathrm{d}W_s^k \boldsymbol{e}_k, \tag{38}$$

where the W_s^k are the independent Brownian motions in (12). It is known that equation (36) admits a unique mild solution in H with the right conditions on the coefficients of G_2 and G_3 (confining potential) [64]; see also [92, proposition 3.4]. We refer to [63, 64] for background on the stochastic convolution term $\int_0^t \mathrm{e}^{(t-s)A} \, \mathrm{d}\boldsymbol{W}_s$.

Throughout this section we assume that only a subset of unresolved modes in the decomposition (6) of H are stochastically forced according to equation (12), namely that the following condition holds:

(**H**) $\sigma_j = 0$ for $1 \leqslant j \leqslant m_c$ in (12), and that there exists at least one index k in $(m_c + 1, N)$ such that $\sigma_k \neq 0$.

In this case, the BF system (35a)–(35c) becomes

$$\mathrm{d}p = A_{\mathfrak{c}} p \, \mathrm{d}s, \qquad\qquad s \in [t - \tau, t], \tag{39a}$$

$$\mathrm{d}q_n = (\lambda_n q_n + \Pi_n G_2(p) + \Pi_n G_3(p)) \, \mathrm{d}s + \sigma_n \, \mathrm{d}W_s^n, \quad s \in [t - \tau, t], \tag{39b}$$

$$\text{with } p(s)|_{s=t} = X \in H_{\mathfrak{c}}, \text{ and } q_n(s)|_{s=t-\tau} = Y. \tag{39c}$$

The solution q_n to equation (39b) at $s = t$ provides a parameterization $\Phi_n(\tau, X, t)$ aimed at approximating the nth high-mode amplitude, $u_n(t) = \Pi_n u(t) = \langle u(t), \boldsymbol{e}_n \rangle$, of the SPDE solution u when X is equal to the low-mode amplitude $u_{\mathfrak{c}}(t) = \Pi_{\mathfrak{c}} u(t)$. This BF system can be solved analytically. Its solution provides the stochastic parameterization, Φ_n, expressed, for $n \geqslant m_c + 1$, as the following stochastic nonlinear mapping of X in $H_{\mathfrak{c}}$:

$$\begin{aligned}
\Phi_n(\tau, X, t; \omega) &= q_n(t, X, \omega) \\
&= \mathrm{e}^{\lambda_n \tau} Y + \sigma_n \left(W_t^n(\omega) - \mathrm{e}^{\tau \lambda_n} W_{t-\tau}^n(\omega) \right) + Z_\tau^n(t; \omega) \\
&\quad + \sum_{i,j=1}^{m_c} \left(D_{ij}^n(\tau) B_{ij}^n X_i X_j \right) + \sum_{i,j,k=1}^{m_c} \left(E_{ijk}^n(\tau) C_{ijk}^n X_i X_j X_k \right),
\end{aligned} \tag{40}$$

with $X = \sum_{\ell=1}^{m_c} X_\ell \boldsymbol{e}_\ell$.

The stochastic term Z_τ^n is given by

$$Z_\tau^n(t; \omega) = \sigma_n \lambda_n \mathrm{e}^{\lambda_n t} \int_{t-\tau}^t \mathrm{e}^{-\lambda_n s} W_s^n(\omega) \, \mathrm{d}s, \tag{41}$$

and as such, is dependent on the noise's path and the 'past' of the Wiener process forcing the nth scale: it conveys exogenous memory effects, i.e. it is non-Markovian in the sense of [94]. The coefficients B_{ij}^n and C_{ijk}^n in equation (40) are given respectively by

$$B_{ij}^n = \langle G_2(\boldsymbol{e}_i, \boldsymbol{e}_j), \boldsymbol{e}_n^* \rangle, \ C_{ijk}^n = \langle G_3(\boldsymbol{e}_i, \boldsymbol{e}_j, \boldsymbol{e}_k), \boldsymbol{e}_n^* \rangle, \tag{42a}$$

while the coefficients $D_{ij}^n(\tau)$ and $E_{ijk}^n(\tau)$ are given by

$$D_{ij}^n(\tau) = \begin{cases} \dfrac{1 - \exp\left(-\delta_{ij}^n \tau\right)}{\delta_{ij}^n}, & \text{if } \delta_{ij}^n \neq 0, \\ \tau, & \text{otherwise,} \end{cases} \tag{43}$$

and

$$E_{ijk}^n(\tau) = \begin{cases} \dfrac{1-\exp\left(-\delta_{ijk}^n \tau\right)}{\delta_{ijk}^n}, & \text{if } \delta_{ijk}^n \neq 0, \\ \tau, & \text{otherwise}, \end{cases} \tag{44}$$

with $\delta_{ij}^n = \lambda_i + \lambda_j - \lambda_n$ and $\delta_{ijk}^n = \lambda_i + \lambda_j + \lambda_k - \lambda_n$. From equations (43) and (44), one observes that the parameter τ (depending on n) allows, in principle, for balancing the small denominators due to small spectral gaps, i.e. when the δ_{ij}^n or the δ_{ijk}^n are small. This attribute is shared with the parameterization formulas obtained by the BF approach in the deterministic context; see [39, remark III.1]. It plays an important role in the ability of our parameterizations to handle physically relevant situations with a weak time-scale separation; see section 5.5 below.

Remark 3.1. In connection with the stochastic convolution $\xi_{t,\tau} = \int_{t-\tau}^t e^{(t-s)A_{\mathfrak{f}}} \Pi_{\mathfrak{f}} \, dW_s$ in equation (33), we note that the stochastic terms in equation (40), involving the processes W_t^n, $W_{t-\tau}^n$, and Z_τ^n, can be rewritten as a scalar stochastic integral, more precisely:

$$\int_{t-\tau}^t e^{\lambda_n(t-s)} \, dW_t^n = W_t^n - e^{\tau\lambda_n} W_{t-\tau}^n + \sigma_n^{-1} Z_\tau^n(t). \tag{45}$$

Note that the RHS in equation (45) is simply the projection of $\xi_{t,\tau}$ onto mode $\boldsymbol{e}_n$ (equation (38)).

Equation (45) is a direct consequence of Itô's formula applied to the product $e^{-\lambda s} W_s^n$ followed by integration over $[t-\tau, t]$. Indeed, we have (dropping the n-dependence),

$$\begin{aligned} d\left(e^{-\lambda s} W_s\right) &= -\lambda e^{-\lambda s} W_s \, ds + e^{-\lambda s} \, dW_s + d\left(e^{-\lambda s}\right) dW_s \\ &= -\lambda e^{-\lambda s} W_s \, ds + e^{-\lambda s} \, dW_s, \end{aligned}$$

by treating $ds \, dW_s$ as zero due to Itô calculus [132].

By introducing now $\delta W = W_t - e^{\lambda\tau} W_{t-\tau}$ and denoting by J the stochastic convolution $\int_{t-\tau}^t e^{\lambda(t-s)} \, dW_s$, we arrive due to equation (41) at

$$e^{-\lambda t} \delta W + \lambda e^{-\lambda t} Z_\tau = e^{-\lambda t} J, \tag{46}$$

which gives the relation $J = \delta W + \lambda Z_\tau$, where we have set $\sigma_n = 1$ to simplify.

In the expression of Φ_n given by equation (40), we opt for the expression involving the Riemann–Stieltjes integral Z_τ^n for practical purpose. The latter can be indeed readily generalized to the case of jump noise (equations (101)–(102) below) without relying on stochastic calculus involving a jump measure (section 7).

In what follows we drop the dependence of Φ_n on ω and sometimes on t, unless specified. The parameterization of the neglected scales writes then as

$$\Phi_{\boldsymbol{\tau}}(X, t) = \sum_{n \geqslant m_c + 1} \Phi_n(\tau_n, X, t) \boldsymbol{e}_n, \tag{47}$$

with $\boldsymbol{\tau} = (\tau_n)_{n \geqslant m_c + 1}$, and Φ_n defined in equation (40).

We observe that letting τ approach infinity in equation (47) recovers equation (33) when $G = G_2 + G_3$. This implies that the manifold constructed from equation (47) can be viewed as a homotopic deformation of the SIM approximation defined in equation (33), controlled by the parameters τ_n.

This connection motivates utilizing the variational framework presented in [39, 41] to identify the 'optimal' stochastic manifold within the family defined by equation (47). In

section 3.3, we demonstrate that a simple data-driven minimization of a least-squares parameterization error, applied to a single training path, yields trained parameterizations (in terms of τ) with remarkable predictive power. For instance, these parameterizations can be used to infer ensemble statistics of SPDEs for a large set of unseen paths during training.

However, unlike the deterministic case, the stochastic framework necessitates addressing the efficient simulation of the coefficients involved in equation (47) for optimization purposes. This challenge is tackled in the following subsection.

3.2. Non-Markovian path-dependent coefficients: Efficient simulation

Our stochastic parameterization given by (47) contains random coefficients $Z_\tau^n(t;\omega)$, each of them involving an integral of the history of the Brownian motion making them non-Markovian in the sense of [94], i.e. depending on the noise path history. We present below an efficient mean to compute these random coefficients by solving auxiliary ODEs with path-dependent coefficients, avoiding this way the cumbersome computation of integrals over noise paths that would need to be updated at each time t. This approach is used for the following purpose:

(i) To find the optimal vector τ^* made of optimal BF integration times and form in turn the stochastic optimal parameterization Φ_{τ^*}, and

(ii) To efficiently simulate the corresponding optimal reduced system built from Φ_{τ^*} along with its path-dependent coefficients.

According to (41), the computation of Z_τ^n boils down to the computation of the random integral

$$\mathrm{e}^{\lambda_n t} \int_{-\tau+t}^{t} \mathrm{e}^{-\lambda_n s} W_s^n(\omega)\, \mathrm{d}s,$$

which is of the following form:

$$I_\tau^j(t,\omega;\kappa) = \mathrm{e}^{\kappa t} \int_{-\tau+t}^{t} \mathrm{e}^{-\kappa s} W_s^j(\omega)\, \mathrm{d}s, \tag{48}$$

where $\tau > 0$, and $j \geqslant m_c + 1$ due to assumption (**H**). We only need to consider the case that $\kappa < 0$, since we consider here only cases in which the unstable modes are included within the space $H_{\mathfrak{c}}$ of resolved scales.

By taking time derivative on both sides of (48), we obtain that $I_\tau^j(t,\omega;\kappa)$ satisfies the following ODE with path-dependent coefficients also called a random differential equation (RDE):

$$\frac{\mathrm{d}I}{\mathrm{d}t} = \kappa I + W_t^j(\omega) - \mathrm{e}^{\kappa \tau} W_{t-\tau}^j(\omega). \tag{49}$$

Since $\kappa < 0$, the linear part κI in (49) brings a stable contribution to (49). As a result, $I_\tau^j(t,\omega;\kappa)$ can be computed by integrating (49) forward in time starting at $t = 0$ with initial datum given by:

$$I(0,\omega) = I_\tau^j(0,\omega;\kappa), \tag{50}$$

where $I_\tau^j(0,\omega;\kappa)$ is computed using (48).

Now, to determine $Z_\tau^n(t;\omega)$ in (41) it is sufficient to observe that

$$Z_\tau^n(t;\omega) = \sigma_n \lambda_n I_\tau^n(t,\omega;\lambda_n). \tag{51}$$

The random coefficient $Z_\tau^n(t;\omega)$ is thus computed by using (51) after $I_\tau^n(t,\omega;\lambda_n)$ is computed by solving, forward in time, the corresponding RDE of the form (49) with initial datum

$$I(0,\omega) = I_\tau^n(0,\omega;\lambda_n) = \int_{-\tau}^{0} e^{-\lambda_n s} W_s^n(\omega) \, ds. \tag{52}$$

This way, we compute only the integral (52) and then propagate it through equation (49) to evaluate $Z_\tau^n(t;\omega)$, instead of computing for each t the integral in the definition (41). The resulting procedure is thus much more efficient to handle numerically than a direct computation of $Z_\tau^n(t;\omega)$ based on the integral formulation (41) which would involve a careful and computationally more expensive quadrature at each time-step. Note that a similar treatment has been adopted in [43, chapter 5.3] for the computation of the time-dependent coefficients arising in stochastic parameterizations of SPDEs driven by linear multiplicative noise, and in [39, section III.C.2] for the case of non-autonomous forcing.

3.3. Data-informed optimization and training path

Thanks to section 3.2, we are now in position to propose an efficient variational approach to optimize the stochastic parameterizations of section 3.1 in view of handling parameter regimes located away from the instability onset and thus a greater wealth of possible stochastic transitions.

Given a solution path $u(t)$ to equation (36) available over an interval I_T of length T, we aim for determining the optimal parameterization Φ_n (given by (40)) that minimizes—in the τ-variable—the following *parameterization defect*

$$\mathcal{Q}_n(\tau,T) = \overline{\left| u_n(t) - \Phi_n(\tau, u_{\mathfrak{c}}(t), t) \right|^2}, \tag{53}$$

for each $n \geqslant m_c + 1$. Here $\overline{(\cdot)}$ denotes the time-mean over I_T while $u_n(t)$ and $u_{\mathfrak{c}}(t)$ denote the projections of $u(t)$ onto the high-mode e_n and the reduced state space $H_{\mathfrak{c}}$, respectively.

Figure 1 illustrates a schematic of the stochastic OPM obtained through this procedure. It conveys a useful geometric interpretation: the parameterization $\Phi_{\boldsymbol{\tau}}$ (equation (47)) defines a stochastic manifold $\mathfrak{M}_{\boldsymbol{\tau}}(t)$, constructed as the graph of $\Phi_{\boldsymbol{\tau}}(X,t)$ for X in $H_{\mathfrak{c}}$, that satisfies for almost every noise path:

$$\overline{\mathrm{dist}\left(u(t), \mathfrak{M}_{\boldsymbol{\tau}}(t)\right)^2} \leqslant \sum_{n \geqslant m_c + 1} \mathcal{Q}_n(\tau_n, T),$$

where $\mathrm{dist}(u(t), \mathfrak{M}_{\boldsymbol{\tau}}(t))$ denotes the distance of $u(t)$ to the manifold $\mathfrak{M}_{\boldsymbol{\tau}}(t)$. In practical applications, a finite number of modes are resolved and thus parameterized. Therefore, minimizing each $\mathcal{Q}_n(\tau_n, T)$ with respect to τ_n naturally enforces the closeness of $u(t)$ to the manifold $\mathfrak{M}_{\boldsymbol{\tau}}(t)$ in a least-squares sense.

Our primary objective is not to tailor the parameterizations Φ_n to each individual solution path. Instead, we aim to optimize them on a single 'training path'. Subsequently, this optimized parameterization, Φ_n^* is employed to predict dynamical behaviors for any other noise realization, or at least statistically across an ensemble of such realizations.

To achieve this, we utilize the optimized τ_n^* values to construct the stochastic OPM as defined in equation (47). The optimized non-Markovian coefficients within this parameterization are designed to efficiently capture the time-dependent interactions between nonlinear terms and noise, as well as the corresponding transfer of energy between forced and unforced scales. This is particularly relevant for scenarios where stochastic transitions are driven by small-scale forcing.
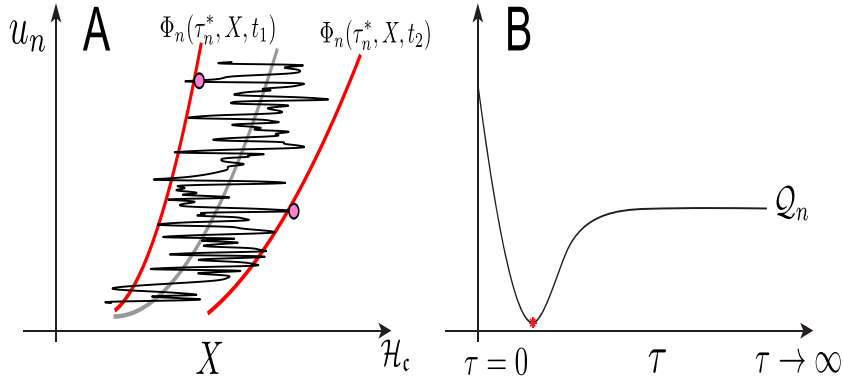
Figure 1. Panel (A): This panel illustrates a training solution path (black curve) for the stochastic optimal parameterization method. This path is shown to be transverse to the underlying stochastic manifolds (red curves) as occurring typically for the reduction of stochastic systems using such objects (see [40, figure 2]) for a concrete simple example). The optimal parameterization (denoted by $\Phi_n(\tau_n^*, X, t)$) aims to approximate the actual state (pink dots) of the system's solution (represented by the black curve) at any given time (here labeled as t_1 and t_2). A key strength of this parameterization is its ability to capture the solution's average behavior by its expected value, $\mathbb{E}(\Phi_n)$ (gray curve). Panel (B): This panel schematically shows how the 'goodness-of-fit' (parameterization defect, $\mathcal{Q}_n$, defined by equation (53)) changes with the parameter τ. The red asterisk marks the value τ_n^* that minimizes this defect.

Sections 5 and 6 below illustrate how this single-path training strategy can be used efficiently to predict from the corresponding reduced systems the statistical behavior as time and/or the noise's path is varied. The next section delves into the justification of this variational approach, and the theoretical characterization of the notion of stochastic OPM by relying on ergodic arguments. The reader interested in applications can jump directly to sections 5 and 6.

4. Non-Markovian optimal parameterizations and invariant measures

The ergodic and mixing properties of dissipative partial differential equations (PDEs) driven by stochastic external forces have been a subject of intense research over the past two decades. It is rather well understood in the case when all deterministic modes are forced; see e.g. [87, 110]. The situation in which only finitely many modes are forced such as considered in this study is much more subtle to handle in order to prove unique ergodicity. The works [91, 92] provide answers in such a degenerate situation based on a theory of hypoellipticity for nonlinear SPDEs. In parallel, the work [93] generalizing the asymptotic coupling method introduced in [90], allows for covering a broad class of nonlinear stochastic evolution equations including stochastic delay equations.

Building upon these results, we proceed under the assumption of a unique ergodic invariant measure, denoted by μ. Given a reduced state space $H_{\mathfrak{c}}$, we demonstrate in this section the following:

(i) The path-dependence of the non-Markovian optimal reduced model arises from the random measure denoted by ρ_ω that is obtained through the disintegration of μ over the underlying probability space Ω.

(ii) The non-Markovian optimal reduced model provides an averaged system in $H_{\mathfrak{c}}$ that is still path-dependent. For a given noise path ω, it provides the reduced system in $H_{\mathfrak{c}}$ that averages out the unresolved variables with respect to the disintegration of ρ_ω over $H_{\mathfrak{c}}$. A detailed explanation is provided in theorem 4.2.

4.1. Theoretical insights

Adopting the framework of random dynamical systems (RDSs) [5], recall that an RDS, $S(t,\omega)$, is said to be *white noise*, if the associated 'past' and 'future' σ-algebras $\mathcal{F}^-$ (see (56)) and $\mathcal{F}^+$ are independent. Given a white noise RDS, $S(t,\omega)$, the relations between random invariant measures for S and invariant measures for the associated Markov semigroup P_t is well known [56]. We recall below these relationships and enrich the discussion with lemma 4.1 below from [72].

Assume that equation (1) generates a white noise RDS, $S(t,\omega)$, associated with a Markov semigroup P_t having μ as an invariant measure, then the following limit taken in the weak* topology:

$$\rho_\omega = \lim_{t\to\infty} S(t,\theta_{-t}\omega)\mu, \tag{54}$$

exists $\mathbb{P}$-a.s. and provides a random probability measure that is S-invariant in the sense that

$$S(t,\omega)\rho_\omega = \rho_{\theta_t\omega}, \quad \mathbb{P}\text{-a.s.}, \ t \geqslant 0. \tag{55}$$

A random probability measure ρ_ω that satisfies (55) is called below a (random) statistical equilibrium. Recall that θ_t here denotes the standard ergodic transformation on the set of Wiener path Ω defined through the helix identity [5, definition 2.3.6], $W_s(\theta_t\omega) = W_{t+s}(\omega) - W_t(\omega)$, ω in Ω.

A statistical equilibrium such as ρ_ω defined above is furthermore Markovian [54], in the sense it is measurable with respect to the past σ-algebra

$$\mathcal{F}^- = \sigma\{\omega \mapsto S(\tau,\theta_{-t}\omega) \,:\, 0 \leqslant \tau \leqslant t\}; \tag{56}$$

see [56, proposition 4.2]; see also [5, theorems 1.7.2 and 2.3.45].

Note that the limit (54) exists in the sense that for every bounded measurable function $f : H \to \mathbb{R}$, the real-valued stochastic process

$$(t,\omega) \mapsto \int_H f(x)\,\mathrm{d}\left(S(t,\theta_{-t}\omega)\mu\right) = \int_H f(S(t,\theta_{-t}\omega)x)\,\mathrm{d}\mu, \tag{57}$$

is a bounded martingale, and therefore converges $\mathbb{P}$-a.s by the Doob's first martingale convergence theorem; see e.g. [132, theorem C.5 p. 302]. This implies, in particular, the $\mathbb{P}$-almost sure convergence of the measure-valued random variable $\omega \mapsto S(t,\theta_{-t}\omega)\mu$ in the topology of weak convergence of $Pr_\Omega(H)$, which in turn implies the convergence in the narrow topology of $Pr_\Omega(H)$; see [55, chapter 3].

Reciprocally, given a Markovian S-invariant random measure ρ_ω of an RDS for which the σ-algebras $\mathcal{F}^-$ and $\mathcal{F}^+$ are independent, the probability measure, μ, defined by

$$\mu = \mathbb{E}(\rho_\bullet) = \int_\Omega \rho_\omega \,\mathrm{d}\mathbb{P}(\omega), \tag{58}$$

is an invariant measure of the Markov semigroup P_t; see [49, theorem 1.10.1].

In case of uniqueness of μ (and thus ergodicity) satisfying $P_t\mu = \mu$, then this one-to-one correspondence property implies that any other random probability measure ν_ω different from the random measure ρ_ω obtained from (54), is non-Markovian which means in particular that $\int_\Omega \nu_\omega \,\mathrm{d}\mathbb{P} \neq \mu$ otherwise we would have $\nu_\omega = \rho_\omega$, $\mathbb{P}$-a.s.

In case a random attractor $\mathcal{A}(\omega)$ exists, since the latter is a forward invariant compact random set, the Markovian random measure ρ_ω must be supported by $\mathcal{A}(\omega)$; see [55, theorem 6.17] (see also [57, proposition 4.5]). We conclude thus, in the case of a unique ergodic measure for P_t, to the existence of *a unique* Markov S-invariant random measure ρ_ω supported by $\mathcal{A}$, and that all other S-invariant random measures (also necessarily supported by $\mathcal{A}$) are non-Markovian.

RDS's where the future and past σ-algebras are independent, are generated for many stochastic systems in practice. This is the case for instance of a broad class of SDEs; see [5, section 2.3]. The problem of generation of white noise RDSs in infinite dimension is much less clarified and it is beyond the scope of this article to address this question.

Instead, we point out below that there exists other ways to associate to an RDS which is not necessarily of white-noise type, a meaningful random probability measure ρ_ω that still for instance satisfies (58). This is indeed the case if the RDS satisfies some weak mixing property with respect to a (non-random) probability μ on H as the following lemma from [72], shows:

Lemma 4.1. *Assume there exists μ in $Pr(H)$ satisfying the following weak mixing property*

$$\frac{1}{t}\int_0^t \mathrm{d}s \int_H \mathbb{E}f(S(s,\theta_{-s}\bullet)u)\,\mathrm{d}\mu \underset{t\to\infty}{\longrightarrow} \int_H f\,\mathrm{d}\mu, \tag{59}$$

for each bounded, continuous $f:H\to\mathbb{R}$.

Then there exists an $\mathcal{F}_0$-measurable random probability measure ρ_ω given by

$$\rho_{\theta_t\omega} = \frac{1}{t}\int_0^t S(s,\theta_{-s}\omega)\mu\,\mathrm{d}s, \tag{60}$$

that satisfies

$$\mathbb{E}(\rho_\bullet) = \int_\Omega \rho_\omega\,\mathrm{d}\mathbb{P}(\omega) = \mu, \tag{61}$$

and is S-invariant in the sense of (55), i.e. that is a statistical equilibrium.

For a proof of lemma 4.1 see that of [72, lemma 2.5].

Thus, either built from an invariant measure of the Markov semigroup, in the case of a white-noise RDS, or from a weakly mixing measure in the sense of (59), a statistical equilibrium satisfying (58) can be naturally associated with equation (36) as long as the appropriate assumptions are satisfied. At this level of generality, we do not enter into addressing the important question dealing with sufficient conditions on the linear part A and and nonlinear terms that ensure for an SPDE like equation (36), the generation of an RDS that is either of white-noise type or weakly mixing in the sense of (59); see [72] for examples in the latter case.

We also introduce the following notion of pullback parameterization defect.

Definition 4.1. Given a parameterization $\Phi:\Omega\times H_\mathfrak{c}\to H_\mathfrak{f}$ that is measurable (in the measure-theoretic sense), and a mild solution u of equation (36), the pullback parameterization defect associated with Φ over the interval $[0,T]$ and for a noise-path ω, is defined as:

$$\mathcal{Q}_T(\omega,\Phi) = \frac{1}{T}\int_0^T \|u_\mathfrak{f}(t,\theta_{-t}\omega) - \Phi(\omega,u_\mathfrak{c}(t,\theta_{-t}\omega))\|^2\,\mathrm{d}t. \tag{62}$$

Given the statistical equilibrium ρ_ω, we denote by $\mathfrak{m}_\omega$ the push-forward of ρ_ω by the projector $\Pi_\mathfrak{c}$ onto $H_\mathfrak{c}$, namely

$$\mathfrak{m}_\omega(B) = \rho_\omega\left(\Pi_\mathfrak{c}^{-1}(B)\right),\ B\in\mathcal{B}(H_\mathfrak{c}), \tag{63}$$

where $\mathcal{B}(H_{\mathfrak{c}})$ denotes the family of Borel sets of $H_{\mathfrak{c}}$; i.e. the family of sets that can be formed from open sets (for the topology on $H_{\mathfrak{c}}$ induced by the norm $\|\cdot\|_{H_{\mathfrak{c}}}$) through the operations of countable union, countable intersection, and relative complement.

The random measure $\mathfrak{m}_{\omega}$ allows us to consider the following functional space $\mathcal{X} = L_{\mathfrak{m}}^2(\Omega \times H_{\mathfrak{c}}; H_{\mathfrak{f}})$ of parameterizations, defined as

$$\mathcal{X} = \left\{ \Phi : \Omega \times H_{\mathfrak{c}} \to H_{\mathfrak{f}} \text{ measurable} \ \Big| \ \int_{H_{\mathfrak{c}}} \|\Phi(\omega, X)\|^2 \, \mathrm{d}\mathfrak{m}_{\omega}(X) < \infty \ a.s. \right\}, \tag{64}$$

namely the Hilbert space constituted by $H_{\mathfrak{f}}$-valued random functions of the resolved variables X in $H_{\mathfrak{c}}$, that are square-integrable with respect to $\mathfrak{m}_{\omega}$, almost surely.

We are now in position to formulate the main result of this section, namely the following generalization to the stochastic context of theorem 4 from [41].

Theorem 4.1. *Assume that one of the following properties hold:*

(i) *The RDS generated by equation (36) is of white-noise type and the Markov semigroup associated with equation (1) admits a (unique) ergodic invariant measure μ.*
(ii) *The RDS generated by equation (36) is weakly mixing in the sense of (59).*

Let us denote by ρ_{ω} the weak*-limit of μ defined in (54) for case (i), and by ρ_{ω} the statistical equilibrium ensured by lemma 4.1, in case (ii). We denote by ρ_{ω}^X the disintegration of ρ_{ω} on the small-scale subspace $H_{\mathfrak{f}}$, conditioned on the coarse-scale variable X.

Assume that the small-scale variable Y has a finite energy in the sense that

$$\int \int \|Y\|^2 \, \mathrm{d}\rho_{\omega} \, \mathrm{d}\mathbb{P} < \infty. \tag{65}$$

Then the minimization problem

$$\min_{\Phi \in \mathcal{X}} \int_{\Omega} \int_{(X,Y) \in H_{\mathfrak{c}} \times H_{\mathfrak{f}}} \|Y - \Phi(\omega, X)\|^2 \, \mathrm{d}\rho_{\omega}(X,Y) \, \mathrm{d}\mathbb{P}(\omega), \tag{66}$$

possesses a unique solution whose argmin is given by

$$\Phi^*(\omega, X) = \int_{H_{\mathfrak{f}}} Y \, \mathrm{d}\rho_{\omega}^X(Y), \ \ X \in H_{\mathfrak{c}}, \ \ \mathbb{P}\text{-a.s.} \tag{67}$$

Furthermore, if the RDS possesses a pullback attractor $\mathcal{A}(\omega)$ in H, and ρ_{ω} is pullback mixing in the sense that for all u in H,

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T f(\omega, S(t, \theta_{-t}\omega)u) \, \mathrm{d}t = \int_{\mathcal{A}(\omega)} f(\omega, v) \, \mathrm{d}\rho_{\omega}(v), \ f(\omega, \cdot) \in C_b(H), \ \mathbb{P}\text{-a.s.}, \tag{68}$$

then $\mathbb{P}$-almost surely

$$\lim_{T \to \infty} \mathcal{Q}_T(\omega, \Phi^*) \leqslant \lim_{T \to \infty} \mathcal{Q}_T(\omega, \Phi), \ \ \Phi \in \mathcal{X}. \tag{69}$$

Proof. The proof follows the same lines of [41, theorem 4]. It consists of replacing:

(i) The probability measure μ therein by the probability measure ρ on $\Omega \times H$ that is naturally associated with the family of random measures ρ_{ω} and whose marginal on Ω is given by $\mathbb{P}$; see [56, proposition 3.6],
(ii) The function space $L_{\mu}^2(H_{\mathfrak{c}} \times H_{\mathfrak{f}}; H_{\mathfrak{f}})$ therein by the function space $\mathcal{E} = L_{\rho}^2(\Omega \times H_{\mathfrak{c}} \times H_{\mathfrak{f}}; H_{\mathfrak{f}})$.

By applying to the ambient Hilbert space $\mathcal{E}$, the standard projection theorem onto closed convex sets [23, theorem 5.2], one defines (given $\Pi_{\mathfrak{c}}$) the conditional expectation $\mathbb{E}_\rho[g|\Pi_{\mathfrak{c}}]$ of g as the unique function in $\mathcal{E}$ that satisfies the inequality

$$\mathbb{E}_\rho\left[\|g - \mathbb{E}_\rho[g|\Pi_{\mathfrak{c}}]\|^2\right] \leqslant \mathbb{E}_\rho\left[\|g - \Psi\|^2\right], \text{ for all } \Psi \in \mathcal{E}. \tag{70}$$

Now by applying the general disintegration theorem of probability measures, applied to ρ (see [41, Eq. (3.18)]), we obtain the following explicit representation of the random conditional expectation

$$\mathbb{E}_\rho[g|\Pi_{\mathfrak{c}}](\omega, X) = \int_{H_{\mathfrak{f}}} g(\omega, X, Y)\, d\rho_{\omega, X}(Y), \tag{71}$$

with $\rho_{\omega,X}$ denoting the disintegrated measure of ρ over $\Omega \times H_{\mathfrak{c}}$. By noting that this disintegrated measure is the same as the disintegration ρ_ω^X (over $H_{\mathfrak{c}}$) of the random measure ρ_ω, we conclude by taking $g(\omega, X, Y) = Y$ as for the proof of [41, Theorem 4] that Φ^* given by (67) solves the minimization problem (66).

The inequality (69) results then from the pullback mixing property (68) and the definition (62) of the pullback parameterization defect. $\qquad\square$

4.2. Non-Markovian optimal reduced model, and conditional expectation

The mathematical framework of section 4.1 allows us to provide a useful interpretation of non-Markovian optimal reduced model for SPDE of type equation (36). To simplify the presentation, we restrict ourselves here to the case $G_3 = 0$.

We denote then by $G(v)$ the vector field in H, formed by gathering the linear and nonlinear parts of equation (36) in this case, namely

$$G(v) = Av + G_2(v, v), \quad v \in H. \tag{72}$$

The theorem formulated below characterizes the relationships between the non-Markovian optimal reduced model and the random conditional expectation associated with the projector $\Pi_{\mathfrak{c}}$ onto the reduced state space $H_{\mathfrak{c}}$; i.e. the resolved modes. Its proof is almost identical to that of [41, theorem 5] with only slight amendment and the details are thus omitted.

Theorem 4.2. *Consider the SPDE of type equation (36) with $G_3 = 0$. Let ρ_ω be a (random) statistical equilibrium satisfying either (54) (in case (i) of theorem 4.1) or ensured by lemma 4.1 (in case (ii) of theorem 4.1).*

Then, under the conditions of theorem 4.1, the random conditional expectation associated with $\Pi_{\mathfrak{c}}$,

$$\mathbb{E}_{\rho_{\theta_t\omega}}[G|\Pi_{\mathfrak{c}}](X) = \int_{Y \in H_{\mathfrak{f}}} G(X+Y)\, d\rho_{\theta_t\omega}^X(Y), \; X \in H_{\mathfrak{c}},$$

satisfies

$$\mathbb{E}_{\rho_{\theta_t\omega}}[G|\Pi_{\mathfrak{c}}](X) = A_{\mathfrak{c}}X + \Pi_{\mathfrak{c}}G_2(X,X) + 2\Pi_{\mathfrak{c}}G_2(X, \Phi^*(\theta_t\omega, X)) + \langle G_2(Y,Y)\rangle_{\rho_{\theta_t\omega}^X}, \tag{73}$$

with

$$\langle G_2(Y,Y)\rangle_{\rho_{\theta_t\omega}^X} = \int_{Y \in H_{\mathfrak{f}}} \Pi_{\mathfrak{c}}G_2(Y,Y)\, d\rho_{\theta_t\omega}^X(Y), \tag{74}$$

for which ρ_ω^X denotes the disintegration of ρ_ω over $H_{\mathfrak{c}}$.

Then, given the reduced state space $H_{\mathfrak{c}}$ and the statistical equilibrium ρ_ω, if $\langle G_2(Y,Y)\rangle_{\rho_{\theta_t\omega}^X} = 0$, the **non-Markovian optimal reduced model** *of equation (36) with noise acting into the 'orthogonal direction' of $H_{\mathfrak{c}}$ (assumption (**H**) in section 3.1), is given by:*

$$\mathrm{d}X = \mathbb{E}_{\rho_{\theta_t\omega}}\left[G|\Pi_{\mathfrak{c}}\right](X)\,\mathrm{d}t. \tag{75}$$

Hereafter, we simply refer to equation (75) as the (random) conditional expectation.

4.3. Practical implications

Thus, theorem 4.2 teaches us that an approximation of the (actual) non-Markovian optimal parameterization, Φ^* given by (67), provides *in fine* an approximation of the conditional expectation involved in equation (75).

The non-Markovian optimal parameterization involves, from its definition, averaging with respect to the unknown probability measure ρ_ω. As such, designing a practical approximation scheme with rigorous error estimates remains challenging. Near the instability onset, the non-Markovian optimal parameterization simplifies to the stochastic invariant manifold. In this regime, probabilistic error estimates have been established for a wide range of SPDEs, including those relevant to fluid dynamics [40]. However, obtaining such guarantees becomes significantly more difficult for scenarios away from the instability onset, even for low-dimensional SDEs [34].

This is where the data-informed optimization approach from section 3.3 offers a practical solution within a relevant class of parameterizations. Variational inequality (69) provides a key insight: the parameterization defect serves as a good measure of a parameterization's quality. Lower defect values indicate a better parameterization, one that is expected to yield a reduced model closer to the (theoretical) non-Markovian optimal reduced model defined by equation (75).

Even with a very good approximation of the actual non-Markovian optimal parameterization, Φ^*, a key question remains: under what conditions does the theoretical conditional expectation (equation (75)) provide a sufficient system's closure on its own? We address this question in sections 5 and 6, through concrete examples. Analyzing two universal models for non-equilibrium phase transitions, we demonstrate that the non-Markovian reduced models constructed using the formulas from section 3 become particularly relevant for understanding and predicting the underlying stochastic transitions when the cutoff scale exceeds the scales forced stochastically (sections 5.3–5.5 and 6.4).

To deal with these non-equilibrium systems, we consider the class $\mathcal{P}$ of parameterizations for approximating the true non-Markovian optimal parameterization, Φ^*. This class $\mathcal{P}$ consists of continuous deformations away from the instability onset point of parameterizations that are valid near this onset (as discussed in section 3.1).

Our numerical results demonstrate that this class $\mathcal{P}$ effectively models and approximates the true non-Markovian terms in the optimal reduced model (equation (75)) across various scenarios. After learning an optimal parameterization within class $\mathcal{P}$ using a single noise realization, the resulting approximation of the optimal reduced model typically becomes an ODE system with coefficients dependent on the specific noise path. These path-dependent coefficients encode the interactions between noise and nonlinearities, as exemplified by the reduced models equations (85) and (104) below, with their ability to predict noise-induced transitions within the reduced state spaces.

5. Predicting stochastic transitions with Non-Markovian reduced models

5.1. Noise-induced transitions in a stochastic Allen–Cahn model

We consider now the following sACE [2, 22, 29]

$$du = \left(\partial_x^2 u + u - u^3 \right) dt + dW_t, \quad 0 < x < L, \ t > 0, \tag{76}$$

with homogeneous Dirichlet boundary conditions. The ambient Hilbert space is $H = L^2(0, L)$ endowed with its natural inner product denoted by $\langle \cdot, \cdot \rangle$.

This equation and variants have a long history. The deterministic part of the equation provides the gradient flow of the Ginzburg–Landau free energy functional [85]

$$\mathcal{E}(u) = \int_0^L \left(\frac{|\partial_x u(x)|^2}{2} + V(u(x)) \right) dx, \tag{77}$$

with the potential V given by the standard double-well function $V(u) = (u^2 - 1)^2/4$. Indeed, one can readily check that $-D\mathcal{E}(u) = \partial_x^2 u + u - u^3$, where $D\mathcal{E}$ denotes the Fréchet derivative of $\mathcal{E}$ at u. The sACE provides a general framework for studying pattern formation and interface dynamics in systems ranging from liquid crystals and ferromagnets to tumor growth and cell membranes [98, 164].

Its universality makes it relevant across diverse fields, contributing to a unified understanding of these complex phenomena. By incorporating the spatio-temporal noise term, dW_t, into the deterministic Allen–Cahn equation, the sACE accounts for inherent fluctuation and uncertainty present in real systems [22]. This allows for a more realistic description of transitions between phases, providing valuable insights beyond what pure deterministic models can offer. Near critical points, where phases coexist, the sACE reveals the delicate balance between deterministic driving forces and stochastic fluctuations. Studying these transitions helps us understand critical phenomena in diverse systems, from superfluid transitions to magnetization reversal in magnets [22, 98].

Phase transitions for the sACE and its variants have indeed been analyzed both theoretically utilizing the large deviation principle [78] and also numerically through different rare event algorithms. For instance, in [11] the expectation of the transition time between two metastable states is rigorously derived for a class of 1D parabolic SPDEs with bistable potential that include sACE as a special case. The phase diagram and various transition paths have also been computed by either an adaptive multilevel splitting algorithm [147] or a minimum-action method [177], where the latter deals with a nongradient flow generalization of the sACE that includes a constant force and also a nonlocal term.

Despite its importance in numerous physical applications, the reduction problem for the accurate reproduction of transition paths in the sACE has received limited attention, with existing works focusing primarily on scenarios near instability onset and low noise intensity (e.g. [15, 18, 27, 142, 170]).

In this study, we consider the sACE placed away from instability onset, after several bifurcations have taken place where multiple steady states coexist; see figure 2. It corresponds to the parameter regime

$$L = 3.9\pi, \ q = 4, \ N = 8, \tag{78}$$

Table 1. Energy distribution of the steady states across the first few eigenmodes.

	e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8
$\phi_1^{\pm}$	94.69%	0	4.79%	0	0.46%	0	0.05%	0
$\phi_2^{a/b}$	0	99.16%	0	0	0	0.83%	0	0
$\phi_3^{a/b}$	0	0	99.93%	0	0	0	0	0

in which the parameters q and N indicate the modes that are forced stochastically in equation (76), according to

$$W_t(x,\omega) = \sum_{j=q+1}^{N} \sigma_j W_t^j(\omega) e_j(x), \, t \in \mathbb{R}, \tag{79}$$

where we set $\sigma_j = \sigma = 0.2$ for $q+1 \leqslant j \leqslant N$. Note that as in section 3.1, e_j, denote the eigenmodes of the operator $Au = \partial_x^2 u + u$ (with homogeneous Dirichlet boundary conditions), given here by

$$e_j(x) = \sqrt{\frac{2}{L}} \sin\left(\frac{j\pi x}{L}\right), \quad j \in \mathbb{N}, \tag{80}$$

with corresponding eigenvalues $\lambda_j = 1 - j^2\pi^2/L^2$. We refer to appendix B for the numerical details regarding the simulation of equation (76).

More specifically, the domain size L is chosen so that when $\sigma = 0$ the trivial steady state has experienced three successive bifurcations leading thus to a total of seven steady states with two stable nodes (ϕ_1^+ and ϕ_1^-) and five saddle points (0, ϕ_2^a, ϕ_2^b, ϕ_3^a, and ϕ_3^b). These steady states are characterized here by their zeros: the saddle states are of sign-change type while the stable ones are of constant sign; see [6, 29]. Note also that the same conclusion holds for the case with either Neumann boundary conditions or periodic boundary conditions; see e.g. [11, figure 2]. In our notation, a and b denote the string made of sign changes over $(0,L)$, with leftmost symbol in a to be +. For instance, for ϕ_3^a (resp. ϕ_3^b), $a = \{+,-,+\}$ (resp. $b = \{-,+,-\}$), while $a = \{+,-\}$ (resp. $b = \{-,+\}$) for ϕ_2^a (resp. ϕ_2^b).

By analyzing the energy distribution of these steady states carried out by the eigenmodes (see table 1), one observes that $\phi_1^{\pm}$ are nearly collinear to e_1, while $\phi_2^{a/b}$ are nearly collinear to e_2, and $\phi_3^{a/b}$ to e_3.

A sketch of the bifurcation diagram as L is varied (for $\sigma = 0$) is shown in figure 2(A); see [6] for detailed analysis. The vertical dashed line marks the domain size $L = 3.9\pi$ considered in our numerical simulations. For this domain size, the system exhibits three unstable modes whose interplay with the nonlinear terms yields the deterministic flow structure depicted in figure 2(B). There, the deterministic global attractor is made of the seven steady states mentioned above (black filled/empty circles) connected by heteroclinic orbits shown by solid/-dashed curves with arrows. To this attractor, we superimposed a solution path to the stochastic model (76) emanating from $u_0 = 0$ (light grey 'rough' curve). We adopt the Arnol'd's convention for this representation: black filled circles indicate the stable steady states while the empty ones indicate the saddle/unstable steady states.

These steady states organize the stochastic dynamics when the noise is turned on: a solution path to equation (76) transits typically from one metastable states to another; the latter corresponding to profiles in the physical space that are random perturbations of an underlying steady state's profile; see insets of figure 4 below. For the parameter setting (78), we observe though, over a large ensemble of noise paths, that the SPDE's dynamics exhibit transition paths
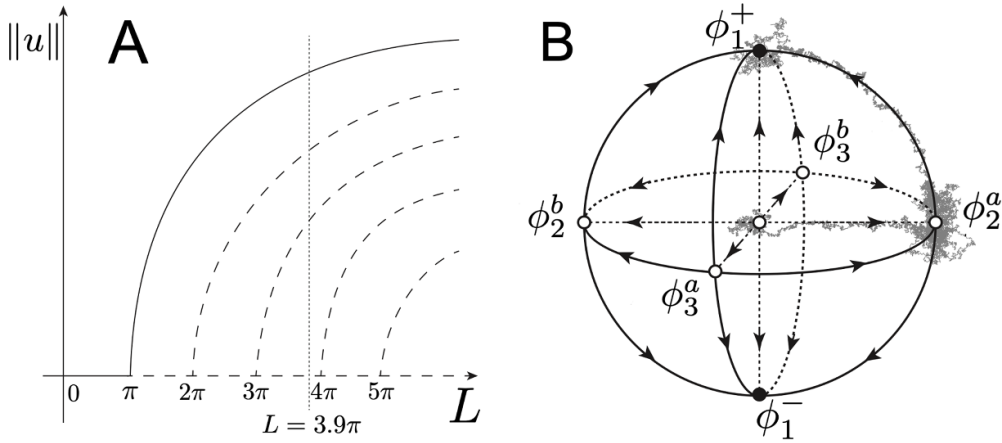
Figure 2. Panel (A): Bifurcation diagram associated with equation (76) (without noise) as L is varied. The dashed vertical line marks the parameter regime analyzed here. Panel (B): Schematic of the deterministic flow structure for this parameter regime. Are shown here, the deterministic global attractor on which is superimposed a solution path to the sACE.

connecting $u_0 = 0$ to neighborhoods of the steady states $\phi_1^\pm$ and $\phi_2^{a/b}$, but does not meanders near $\phi_3^{a/b}$ over the interval $[0, T]$ considered ($T = 40$).

Noteworthy is the visual rendering of the sojourn time of the solution path shown in figure 2(B). It is more substantial in the neighborhood of ϕ_2^a than in that of ϕ_1^+. It is a depiction of a solution path to equation (76) over $[0, T]$: most of the trajectory spends time near a sign-change metastable state (ϕ_2^a) rather than to the constant-sign one (ϕ_1^+). Large ensemble simulations of the sACE for the parameter regime (78) reveals that what is shown for a single path here is actually observed at the final time $t = T$, when the noise path is varied: the sign-change states $\phi_2^{a/b}$ are the rule whereas the constant-sign states $\phi_1^\pm$ are the exceptions, corresponding here to rare events. Figure 3 shows two solution paths in the space-time domain connecting $u_0 = 0$ to either a sign-changing metastable state (typical) or a constant-sign metastable state (rare event). In section 5.3 below we shall discuss the mechanisms behind these patterns and their transition phenomenology, and the extent to which non-Markovian reduced models described below are able to reproduce it.

5.2. Non-Markovian optimal reduced models

Recall that our general goal is the derivation of reduced models able to reproduce the emergence of noise-driven spatiotemporal patterns and predict their transitions triggered by the subtle coupling between noise and the nonlinear dynamics.

For the sACE (76) at hand, this goal is declined into the following form: To derive reduced models not only able to accurately estimate ensemble statistics (averages across many realizations) of the states reached at a given time $t = T$, but also to anticipate the system's typical metastable states as well as those reached through the rarest transition paths pointed above.

To address this issue, we place ourselves in the challenging situation for which the cutoff scale is larger than the scales forced stochastically; i.e. the case where noise acts only in the unresolved part of the system. We show below that stochastic parameterization framework of section 3 allows us to derive such an effective reduced model. This model is built
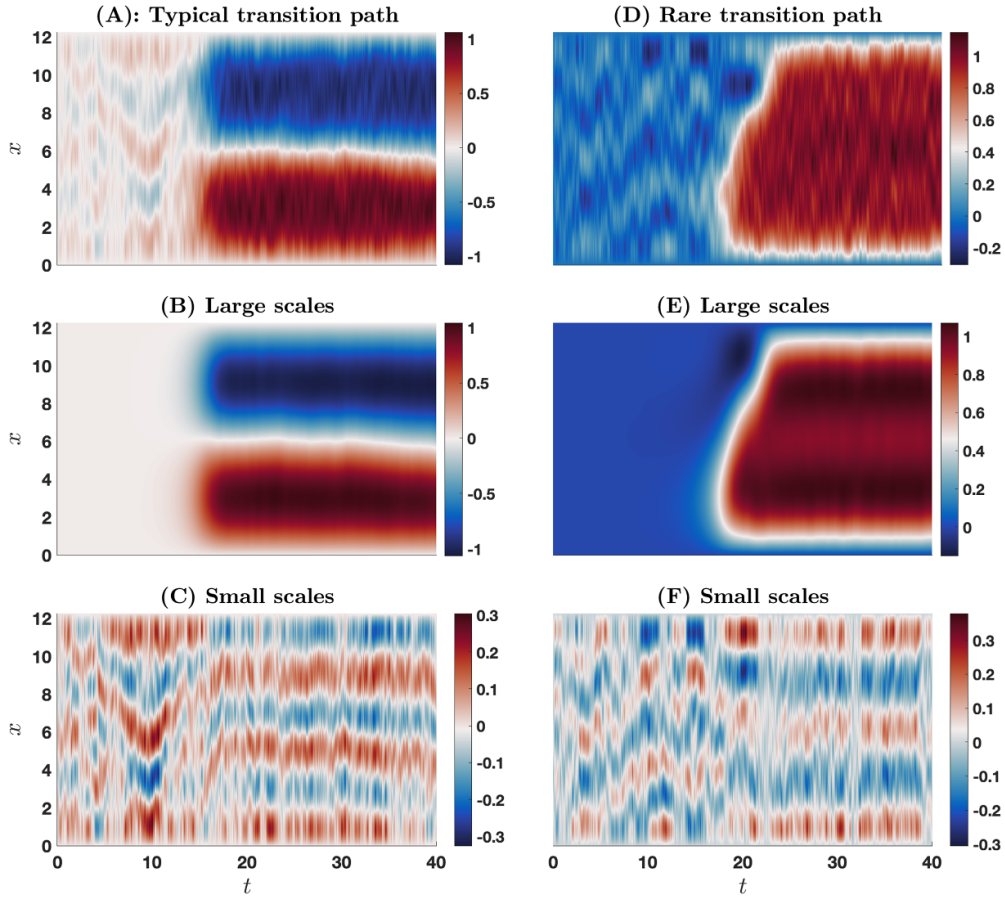
Figure 3. Typical and rare transition paths for equation (76). One path connecting $u_0 = 0$ to a sign-changing metastable state (typical, left panel), and one path connecting $u_0 = 0$ to a constant-sign metastable state (rare event, right panel). Are shown also the large ($u_{\mathfrak{c}}$) and small ($u_{\mathfrak{f}}$) scale contributions.

from the stochastic parameterization (40) obtained from the BF system (39*a*)–(39*c*) applied to equations (76) and (79), that is optimized—on a *single training path*—following section 3.3. We describe below the details of this derivation.

Thus, we take the reduced state space to be spanned by the q first unforced eigenmodes, i.e.

$$H_{\mathfrak{c}} = \mathrm{span}\{\boldsymbol{e}_1, \cdots, \boldsymbol{e}_q\}. \tag{81}$$

In other words, the reduced state space is spanned by low modes that are not forced stochastically; i.e. $q = m_c$ in (79). The subspace $H_{\mathfrak{f}}$ is taken to be the orthogonal complement of $H_{\mathfrak{c}}$ in $L^2(0,L)$ and contains the stochastically forced modes; see (79).

Equation (76) fits into the stochastic parameterization framework of section 3 (see equation (36)) with $Au = \partial_x^2 u + u$ subject to homogenous Dirichlet boundary conditions, $G_2(u) = 0$ and with G_3 denoting the cubic *Nemytski'i operator* defined as $G_3(u,v,w)(x) = -u(x)v(x)w(x)$ for any u, v and w in $L^2(0,L)$, with x in $(0,L)$.

The resulting stochastic parameterization Φ_n given by (40) (with $Y = 0$) takes then the following form:

$$\Phi_n(\tau, X, t; \omega) = \sigma_n \left(W_t^n(\omega) - e^{\tau \lambda_n} W_{t-\tau}^n(\omega) \right) + Z_\tau^n(t; \omega)$$
$$+ \sum_{i,j,k=1}^{q} \left(E_{ijk}^n(\tau) C_{ijk}^n X_i X_j X_k \right), \tag{82}$$

for each n in $(q+1, N)$. Here E_{ijk}^n is given in equation (44), and $Z_\tau^n(t; \omega)$ is given by equation (41).

Since the linear operator A is here self-adjoint, we have $e_n^* = e_n$, and the coefficients C_{ijk}^n are given by

$$C_{ijk}^n = -\langle e_i e_j e_k, e_n \rangle = -\int_0^L e_i(x) e_j(x) e_k(x) e_n(x) \, dx, \tag{83}$$

where $1 \leqslant i,j,k \leqslant q$, and $q+1 \leqslant n \leqslant N$.

For each n in $(q+1, N)$, the free parameter τ in (82) is optimized over a *single, common, training path*, $u(t)$, solving the sACE, by minimizing $\mathcal{Q}_n(\tau)$ in (53) with Φ_n given by (82); see section 3.3. To do so, we follow the procedure described in section 3.2 to simulate efficiently the required random random coefficients (here the Z^n-terms), solving in particular the random ODEs (49).

After minimization, once $\boldsymbol{\tau}^* = (\tau_{q+1}^*, \ldots, \tau_N^*)$ is obtained, the corresponding q-dimensional optimal reduced system in the class $\mathcal{P}$ (see section 4.3), is given component-wisely by

$$\dot{y}_i = \lambda_i y_i - \left\langle \left(\sum_{j=1}^{q} y_j e_j + \sum_{n=q+1}^{N} \Phi_n(\tau_n^*, \boldsymbol{y}, t; \omega) e_n \right)^3, e_i \right\rangle, \tag{84}$$

where $i = 1, \cdots, q$ and $\boldsymbol{y} = (y_1, \cdots, y_q)^{\mathrm{T}}$.

The optimal reduced system (84), referred to as the OPM closure hereafter, can be further expanded as

$$\dot{y}_i = \lambda_i y_i + \sum_{j,k,\ell=1}^{q} C_{jk\ell}^i y_j y_k y_\ell + 3 \sum_{j,k=1}^{q} \sum_{n=q+1}^{N} C_{jkn}^i y_j y_k \Phi_n(\tau_n^*, \boldsymbol{y}, t; \omega)$$
$$+ 3 \sum_{j=1}^{q} \sum_{n,n'=q+1}^{N} C_{jnn'}^i y_j \Phi_n(\tau_n^*, \boldsymbol{y}, t; \omega) \Phi_{n'}(\tau_{n'}^*, \boldsymbol{y}, t; \omega)$$
$$+ \sum_{n,n',n''=q+1}^{N} C_{nn'n''}^i \Phi_n(\tau_n^*, \boldsymbol{y}, t; \omega) \Phi_{n'}(\tau_{n'}^*, \boldsymbol{y}, t; \omega) \Phi_{n''}(\tau_{n''}^*, \boldsymbol{y}, t; \omega). \tag{85}$$

Here also, the coefficients C_{ijk}^ℓ are those defined in (83) for $1 \leqslant i,j,k,\ell \leqslant N$. Note that the required Z^n-terms to simulate equation (85) are advanced in time by solving the corresponding equation (49). A total of $N - q$ such equations are solved, each corresponding to a single random coefficient per mode to parameterize. Thus, a total of, $N = q + (N - q)$, ODEs with random coefficients are solved to obtain a reduced model of the amplitudes of the q first modes. We call equation (85) together with its auxiliary ODEs of type given by equation (49), the optimal reduced model for the sACE.

A compact form of equation (85) can be written as follows:

$$\dot{\boldsymbol{y}} = \Pi_{\mathfrak{c}} A \boldsymbol{y} + \Pi_{\mathfrak{c}} G_3 \left(\boldsymbol{y} + \Phi_{\boldsymbol{\tau}^*}(\boldsymbol{y}, t; \omega) \right), \boldsymbol{y} \in H_{\mathfrak{c}}, \tag{86}$$

where

$$\Phi_{\boldsymbol{\tau}^*}(\boldsymbol{y},t;\omega) = \sum_{n=q+1}^{N} \Phi_n(\tau_n^*,\boldsymbol{y},t;\omega)\,\boldsymbol{e}_n, \tag{87}$$

with Φ_n defined in equation (82). We refer to appendix B for the numerical details regarding the simulation of equation (86).

Noteworthy is that an approximation of the sACE solution paths can be obtained from the solutions $\boldsymbol{y}(t,\omega)$ to the optimal reduced model. Such an approximation consists of simply lifting the surrogate low-mode amplitude $\boldsymbol{y}(t,\omega)$ simulated from equation (86), to H_{f} via the optimal parameterization, $\Phi_{\boldsymbol{\tau}^*}$, defined in equation (87). This approximation takes then the form

$$u^{\mathrm{opm}}(t,x,\omega) = \sum_{j=1}^{q} y_j(t,\omega)\,\boldsymbol{e}_j(x) + \sum_{n=q+1}^{N} \Phi_n(\tau_n^*,\boldsymbol{y}(t,\omega),t;\omega)\,\boldsymbol{e}_n(x). \tag{88}$$

The insets of figure 7(A) of section 5.4 show the approximation skills achieved by this formula obtained thus only by solving the reduced equation equation (86).

5.3. Why energy flows in a certain direction: the role of non-Markovian effects

The sACE (76) is integrated over a large ensemble of noise paths (10^6 members), from $u_0 = 0$ at $t = 0$ up to a time $t = T$. In our simulations with a fixed time horizon ($T = 40$), reaching a constant-sign state at the end ($t = T$) is a rare event, even though these states correspond to the deeper energy wells in the system's landscape (as shown in figure 4). Conversely, sign-change states, which reside in shallower wells, are more common. This might seem counterintuitive from an energy perspective. Nevertheless, this can be explained by the degenerate noise (equation (79)) used to force the sACE: only a specific group of unresolved (stable) modes is excited by noise and this may limit the system's ability to explore the deeper energy wells efficiently within the simulation timeframe.

Over time (as t increases), the noise term in the sACE interacts with the nonlinear terms, gradually affecting the unforced resolved modes. Our experiments show that the amplitude of the sACE's solution amplitude carried by $\boldsymbol{e}_2$ grows the fastest for most noise paths. Because $\phi_2^{a/b}$ is nearly aligned with $\boldsymbol{e}_2$ (as shown in table 1), this means that for most noise paths, the solution is expected to reach a metastable state close to $\phi_2^{a/b}$ when the nonlinear effects become significant (nonlinear saturation kicks in).

Understanding why energy primarily accumulates in the $\boldsymbol{e}_2$ direction is complex. It stems from how noise initially affecting unresolved modes (Modes 5 to 8) propagates through nonlinear interactions, gradually influencing larger-scale resolved modes. As the initial state ($u_0 = 0$) lacks energy, this noise-induced energy transfer to larger scales takes some time. Our analysis reveals that during the interval $I = (0,10)$, the solution is dominated by these small-scale features inherited from the noise forcing (table 2 and figure 3).

This finding has significant implications for closure methods: to accurately reproduce both rare and typical transitions, the underlying parameterization must effectively represent the system's small-scale response to stochastic forcing before the transition occurs.

Figures 5 and 6 illustrate how the optimal parameterization and reduced model (equations (86) and (87)) capture the subtle differences in the small-scale component that ultimately lead to rare or typical events in the full sACE system. Notably, for both rare (figure 5) and typical paths (figure 6), the low-mode amplitudes of the sACE solution are near zero during the interval $I = (0,10)$ (Panels (C)–(F) and table 2). Therefore, it is the spatial patterns of
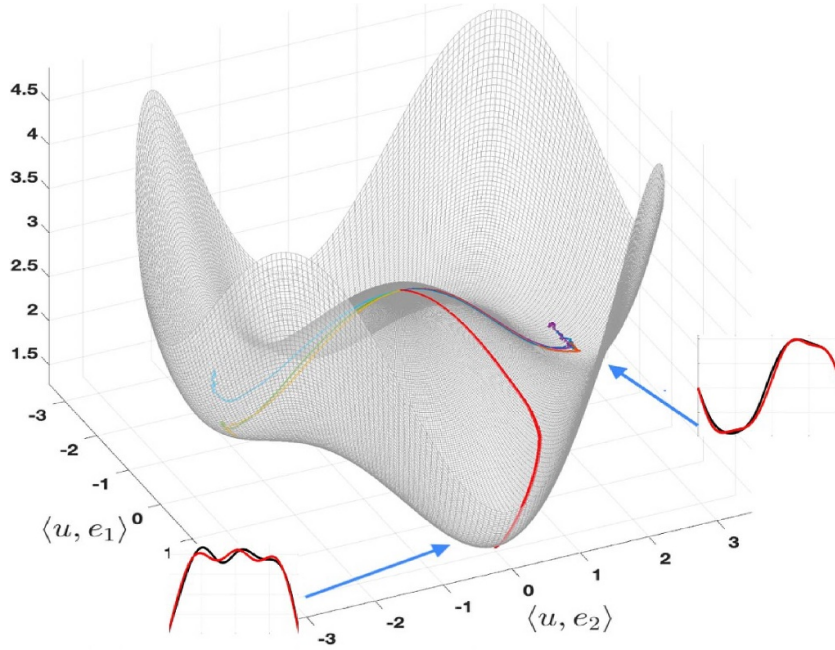
Figure 4. Ginzburg–Landau free energy functional. Here, the Ginzburg–Landau free energy functional $\mathcal{E}(u)$ given by (77) is shown over the reduced state made of the two dominant modes' amplitudes. A few solution paths navigating across this energy's landscape (over $[0, T]$) are shown. The 'shallower' local minima correspond to $\phi_2^{a/b}$ (sign-change profile), while the deeper ones correspond to $\phi_1^{\pm}$ (constant-sign profile). The insets show examples of metastable states, reached at $t = T$ by the sACE, located near these steady states. The rare paths are those trapped in the deeper wells due to the degenerate noise employed here; see Text.

the small-scale component, rather than its energy content, that control the triggering of rare or typical events (figures 3(A) and (D). The successful parameterization of these small-scale features (figures 5(B) and 6(B)) is crucial for the optimal reduced model's ability to reproduce the key behaviors of the sACE's large-scale dynamics.

The non-Markovian terms are essential for accurately capturing these small-scale patterns before a rare or typical event occurs. When large-scale mode amplitudes are near zero during interval I (table 2), the optimal parameterization (equation (87)) is dominated by the (optimized) non-Markovian field ξ_t given by:

$$\xi_t(\omega, x) = \sum_{n=q+1}^{N} M_n(t, \tau_n^*, \omega) e_n(x), \tag{89}$$

with

$$M_n(t, \tau_n^*, \omega) = \sigma_n \left(W_t^n(\omega) - e^{\tau_n^* \lambda_n} W_{t-\tau_n^*}^n(\omega) \right) + Z_{\tau_n^*}^n(t; \omega). \tag{90}$$

After this initial transient period, nonlinear effects become significant, and the large-scale mode amplitudes undergo substantial fluctuations. Panels C-F in figures 5 and 6 demonstrate this behavior and the optimal reduced model's ability to predict these large-scale fluctuations beyond interval I. The OPM reduced system's ability to capture such transitions is rooted in its
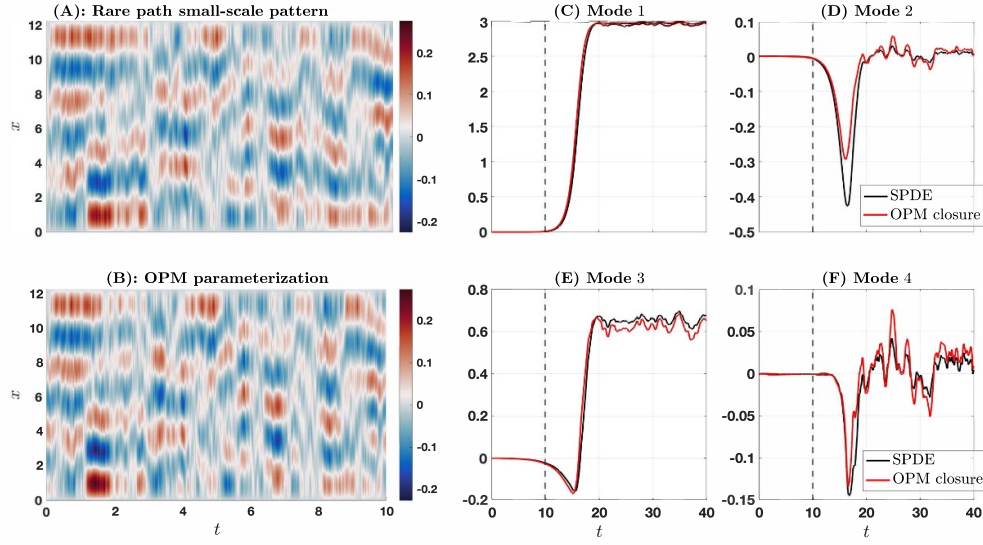
Figure 5. Non-Markovian terms and rare path's small-scale component. Panel A shows the small-scale component of the solution to the sACE, experiencing a rare transition path. Panel B shows the OPM parameterization of this small-scale component over the interval (0,10), prior the transition takes place. Panels (C)–(F) show the ability of the optimal reduced model, equation (85), to predict the large fluctuations experienced by the large-scale modes' amplitudes, beyond this interval.
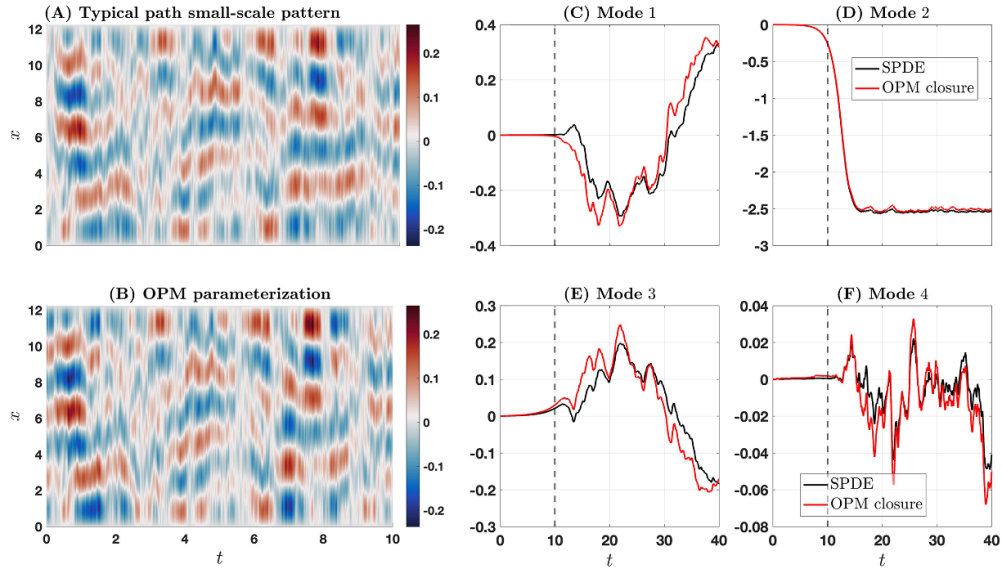
Figure 6. Non-Markovian terms and typical path's small-scale component. Same as figure 5 but for a typical path.

Table 2. Energy distribution over $(0, 10)$ for the sACE's rare and typical paths.

	e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8
Rare path	0	0.05%	0.01%	0	49.6%	26.1%	15.1%	9.2%
Typical path	0	0.22%	0	0	34.8%	33%	19.7%	12.4%

very structure. Its coefficients in equation (85) are nonlinear functionals of the non-Markovian M_n-terns within ξ_t allowing for an accurate representation of the genuine nonlinear interactions between noise and nonlinear terms in the original sACE, which drive fluctuations in the large-mode amplitudes (Modes 1 to 4).

After a transition occurs and energy is transferred to the low modes, the stochastic component ξ_t and its non-Markovian terms become secondary while the nonlinear terms in the stochastic parameterization (82) become crucial for capturing the average behavior of the sACE dynamics, as highlighted in section 5.5 below.

5.4. Rare and typical events statistics from non-Markovian reduced models

Here, we evaluate the ability of our non-Markovian reduced model (equation (86)) to statistically predict the rare and typical transitions discussed in the previous section. We test the reduced model's performance in this task using a vast ensemble of one million noise paths.

Recall that the optimal parameterization (Φ_n, defined in (82)) is trained on a single, 'typical' noise path (see figure 8(A). In other words, we minimized the parameterization defect $\mathcal{Q}_n(\tau)$ in equation (53) for this specific path. Our key objective here is to assess how well the reduced model (equation (86)) can predict the distribution of final states (metastable states at time T) across a much larger set of out-of-sample noise paths.

To do so, we look at a natural 'min+max metric." This metric consists of adding the minimum to the maximum values of the sACE's solution profile at time T. The shape of this probability distribution shows that a value close to zero corresponds to a typical sign-change metastable state, while a value close to 1 or -1, corresponds to a rare constant-sign metastable state.

Figure 7(A) demonstrates the effectiveness of the non-Markovian optimal reduced model in capturing rare events. The optimal reduced system (equation (86)) can predict noise-driven trajectories connecting an initial state ($u_0 = 0$) to rare final states (sACE solution's profile with a constant sign at time T) with statistically significant accuracy (see figure 7(A)). While these rare events occur less frequently in the reduced system compared to the full sACE (smaller red bumps compared to black ones near 1 and -1), the ability of the reduced equation equation (86) to predict them highlights the optimal parameterization's success in capturing the system's subtle interactions between the noise and nonlinear effects responsible for these rare events.

Thus, the optimal reduced system reproduces faithfully the probability distribution of both common (frequently occurring sign-change profiles) and rare (constant-sign profiles) final states at time T (figure 7(A)). This is particularly impressive considering the model was trained on just a single path, yet generalizes well to predict the behavior across a large ensemble of out-of-sample paths.

Figure 8(A) shows the training stage over a typical transition path. The optimization of the free parameter τ in the stochastic parameterization Φ_n (see equation (82)) is executed over a typical transition path, by minimizing the normalized parameterization defects $Q_n(\tau, T) = \mathcal{Q}_n(\tau, T)/\overline{|u_n(t)|^2}$ ($\mathcal{Q}_n(\tau, T)$ defined in equation (53)) for all the relevant n (here $n = 5, \cdots, 8$).
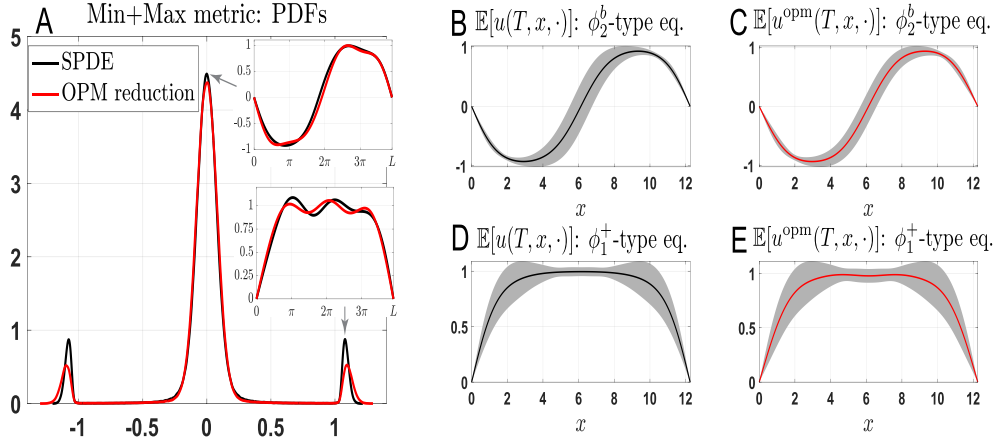
Figure 7. Comparison of metastable state statistics between the full sACE model (equation (76)) and the optimal reduced model (equation (86)) at time $t = T$. Panels B and D show the distribution of $u(T, x, \cdot)$ from the sACE solution at $t = T$. Panels C and E show the corresponding approximation $u^{\mathrm{opm}}(T, x, \cdot)$ obtained from the optimal reduced model's solutions at $t = T$, leveraging equation (88). One million test paths are used to generate these statistics. See Text for more details.
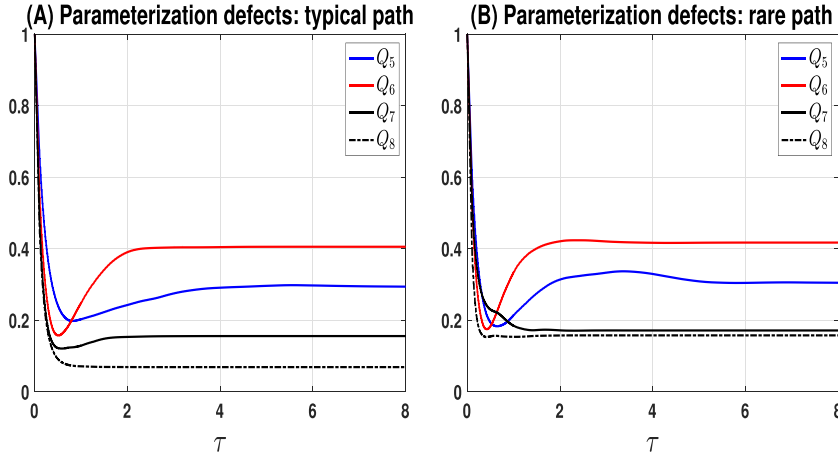
Figure 8. The (normalized) parameterization defects, $Q_n(\tau)$. Panel (A): The parameterization Φ_n given by (82) is trained offline on a (single) typical transition path taken over the interval (t_s, T), with $t_s = 10$ corresponding to the time at which nonlinear effects kick in (see figure 5). This optimized stochastic parameterization is used into the reduced system (86) to produce the online ensemble results of figure 7. Panel (B): The τ-dependence of the $Q_n(\tau)$ is shown on a rare transition path, for information purposes. See section 5.5 for more details.

We observe that Q_5 and Q_6 exhibit clear minima, whereas the other modes have their parameterization defect not substantially diminished at finite τ-values compared to its asymptotic value (as $\tau \to \infty$). Same features are observed over a rare transition path; see figure 8(B). Recall that our reduced system is an ODE system whose stochasticity comes exclusively from its random coefficients as the resolved modes in the full system are not directly forced by noise.

It is this specificity that enables the reduced equations to effectively predict noise-induced transitions, including rare events. This remarkable capability relies on the optimized memory content embedded within the path-dependent, non-Markovian coefficients.

Thus, the optimized, non-Markovian and nonlinear reduced equation (85) is able to track efficiently how the noise, acting on the 'unseen' part of the system, interacts with the resolved modes through these non-Markovian coefficients. While these coefficients effectively encode the fluctuations for a good reduction of the sACE system, they alone are insufficient to handle regimes with weak time-scale separation such as dealt with here. Understanding the average behavior of the resolved dynamics is equally important. The next section explores how our parameterization, particularly its nonlinear terms, play a crucial role in capturing this average behavior.

5.5. Weak time-scale separation, nonlinear terms and conditional expectation

We assess the ability of our reduced system to capture this average dynamical behavior in the physical space. In that respect, figures 7(B) and (D) show the expected profiles of the sACE's states obtained at $t = T$ (solid black curves) from the sACE (76), along with their standard deviation (grey shaded areas), for the ensemble used for figure 7(A). Similarly, figures 7(C) and (E) show the expected profiles (solid red curves) with their standard deviation as obtained at $t = T$ from $u^{\mathrm{opm}}(T, x, \cdot)$ defined in equation (88), once the low-mode amplitude is simulated by the optimal reduced model (86) and lifted to H_{f} by using the stochastic parameterization Φ_{τ^*}.

By conducting large-ensemble (online) simulations of our reduced model (equation (86)), we accurately reproduce the mean motion and second-order statistics of the original sACE. This success stems from our parameterization's ability to effectively represent the average behavior of the unresolved variables during the offline training phase, as defined by the probability measure μ in (58). Figure 9 illustrates this point for the fifth and sixth modes' amplitudes of the sACE system. Notably, the nonlinear nature of the Markovian optimal parameterization, evident after taking the expectation (see equation (91) below), highlights the crucial role of nonlinear terms in our stochastic parameterization (equation (82)) in achieving this accurate representation.

The manifold shown in figure 9(A) (resp. figure 9(B) is obtained as the graph of the following Markovian OPM for the 6th (resp. 5th) mode's amplitude, given by the expectation of equation (82) with $n = 6$ (resp. $n = 5$) after optimization (figure 8),

$$(X_1, X_2) \mapsto \mathbb{E}\left[\Phi_n\left(\tau_n^*, X\right)\right] = \sum_{i,j,k=1}^{q} \left(E_{ijk}^n\left(\tau_n^*\right) C_{ijk}^n X_i X_j X_k\right), \tag{91}$$

for $X = (X_1, X_2, X_3, X_4)$ with $X_3 = p_3$ and $X_4 = p_4$, where p_3 and p_4 denote the most probable values of the 3rd and 4th solution components, $u_3(t)$ and $u_4(t)$, to enable visualization as a 2D surface while favoring a certain 'typicalness' of the visualized manifold. Note that the Markovian OPM corresponds simply to the deterministic cubic term in (82) after replacing τ by the optimal τ_n^*, since the remaining stochastic terms equal to $\sigma_n \int_{t-\tau_n^*}^{t} e^{\lambda_n(t-s)} \, \mathrm{d}W_t^n(\omega)$ (see remark 3.1), whose expectation is zero as stochastic integral in the sense of Itô.

Without the optimization stage, the parameterizations of section 2 experience deficient accuracy when the memory content is taken to be $\tau = \infty$, due to small values that the spectral
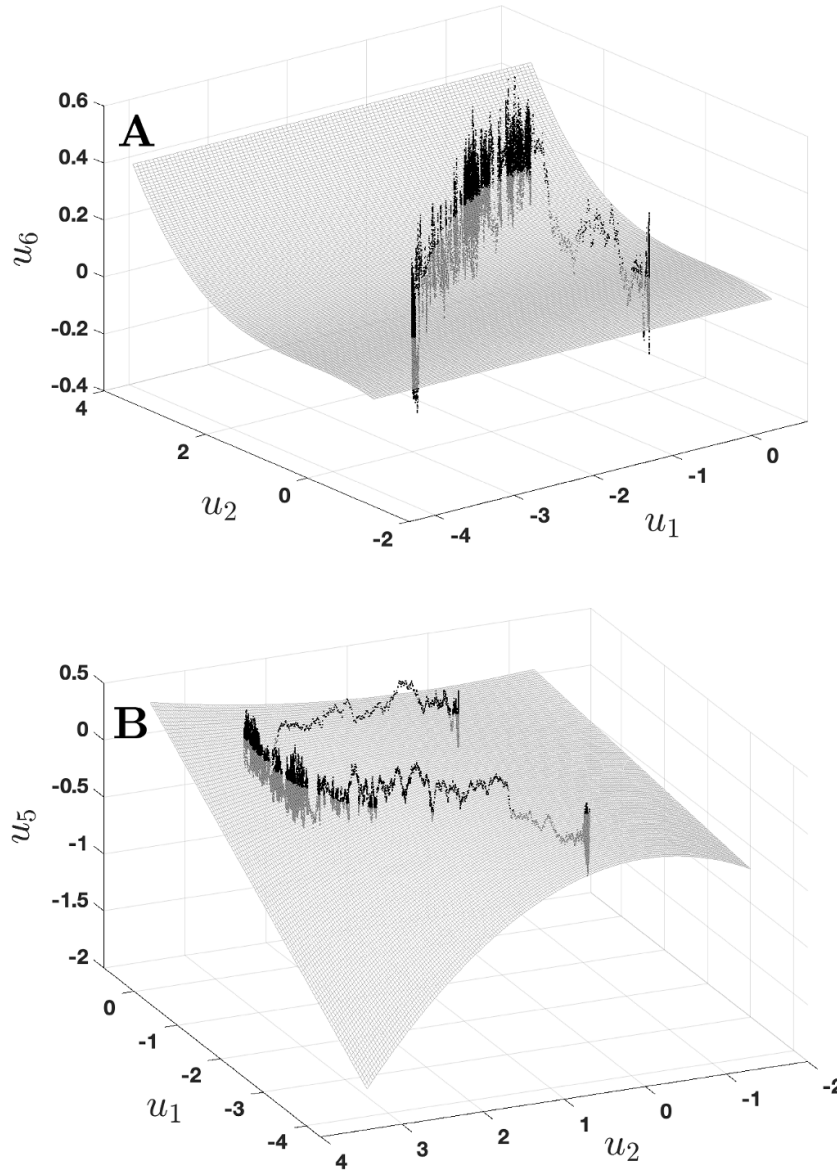
Figure 9. Markovian optimal parameterization for the 5th and 6th modes' amplitudes. Panel (A): The black curve shows the time evolution of the 6th mode's amplitude, $u_6(t)$, obtained by solving the governing equation, equation (76), over the time interval 0 to 400. This 6th mode's amplitude is shown against the corresponding first and second modes' amplitudes, $u_1(t)$ and $u_2(t)$. For visualization purposes, the other solution components in the four-dimensional subspace $H_\mathfrak{c}$, $u_3(t)$ and $u_4(t)$, are fixed at their most probable values, p_3 and p_4, respectively. The manifold shown is the underlying Markovian OPM for the 6th mode's amplitude (equation (91) with $n = 6$). Panel (B): Same for the 5th mode's amplitude $u_5(t) = \langle u(t), e_5 \rangle$ and its corresponding Markovian OPM (equation (91) with $n = 5$).
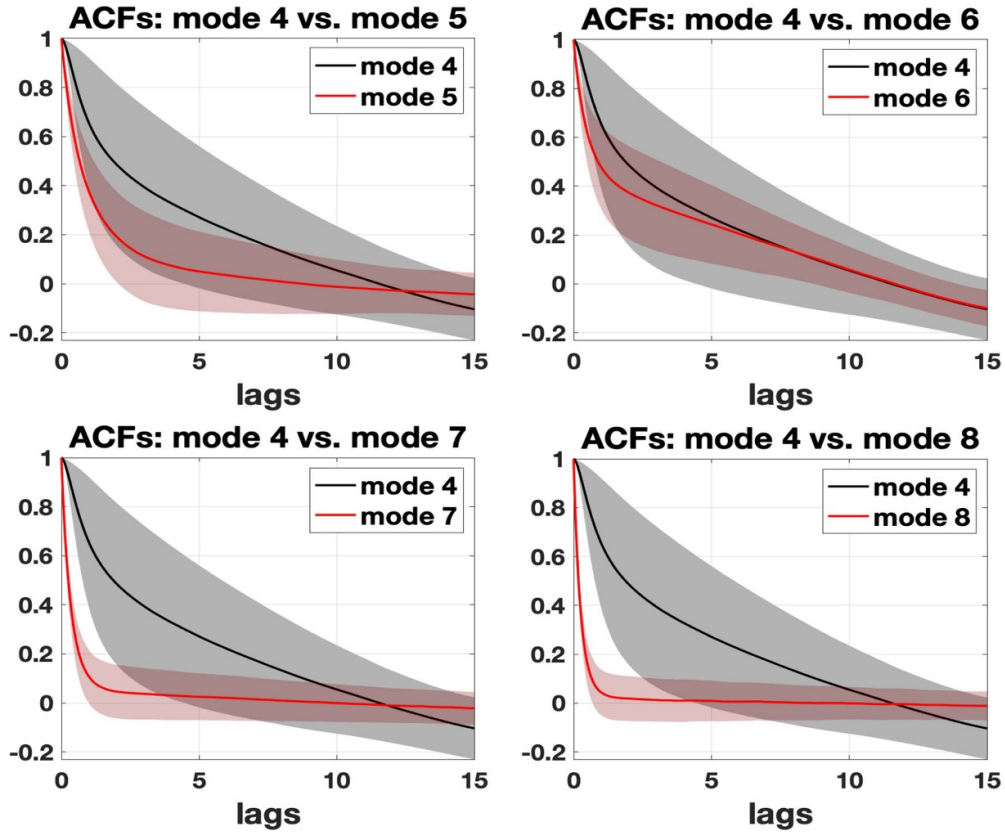
Figure 10. Autocorrelation functions (ACFs). Ensemble averages of the ACFs for the sACE's solution amplitude carried by the resolved mode e_4 (black curve) and those carried by the unresolved modes $e_5, \ldots, e_8$ (red curves). The gray (resp. pinky) shaded area around the black (resp. red) curve shows the ACF's standard deviation for e_4 (resp. the unresolved modes). These statistics are computed from the same ensemble of sACE solution paths as in figure 7.

gaps δ_{ijk}^n take in denominators. As a reminder, small spectral gaps are indeed a known limitation of traditional invariant manifold techniques, often leading to inaccurate results in this case. Optimizing the parameter τ helps address this issue. It introduces corrective terms like $(1 - \exp(-\delta_{ijk}^n)\tau)$ in the cubic coefficients (E_{ijk}^n, equation (44)) that effectively compensate for these small gaps. This is especially important for fifth and sixth (unresolved) modes whose parameterization defects show a clear minimum during optimization (figure 8).

Small spectral gaps often translate into a more observable dynamical effect: weak time-scale separation between the resolved and unresolved variables near the cutoff scale. This is where the optimization process plays a particularly important role. In our sACE system, we observe indeed that the timescales of the unresolved modes near cutoff have a greater impact on the optimization of the backward integration time τ (figure 8) from our BF system (equations (39a)–(39c)) which underpins the parameterization (equation (82)). Modes like the fifth and sixth (unresolved) have correlation decay rates closer to the fourth resolved mode, compared to the much faster decay rates of the seventh and eighth modes (figure 10). Notably, the fourth and sixth modes exhibit very similar decay rates (except for short time lags). This

similarity makes finding the optimal value for τ crucial. Interestingly, the minimum in the optimization is more pronounced for the sixth mode compared to the fifth (figure 8). This difference could be caused by non-linear interactions between these modes or their relative energy content. The seventh and eighth modes exhibit much faster correlation decay (figure 10) and consequently their parameterization defects exhibit marginal optimization at finite τ-values compared to their asymptotic values.

Thus, the closure results shown in figure 7 demonstrate our reduction approach's effectiveness in handling weak time-scale separation scenarios, while also suggesting potential for further improvement (see section 9). Our data-driven optimization of analytical parameterization formulas offers a compelling alternative to traditional multiscale methods or those based on invariant manifolds [15, 18, 27, 142, 170]. These methods, while providing rigorous error estimates, often require strong time-scale separation for the derivation of their amplitude equations. Additionally, they can achieve higher-order nonlinearities in their reduced equations too but through a different route: the Itô formula is applied after exploiting a slow timescale of the system (e.g. [18, equation (12)] and [128, equation (15)]). Integrating optimization concepts into these alternative frameworks could be a promising avenue to explore how it performs in weak time-scale separation regimes.

6. Predicting jump-induced transitions

6.1. Spatially extended system with double-fold bifurcation

We consider the following spatially extended system,

$$\partial_x^2 u + \lambda \left(1 + u^2 - \epsilon u^3 \right) = 0, \text{ on } (0, L), \tag{92a}$$

$$u(0) = u(L) = 0. \tag{92b}$$

This system serves as a basic example of a spatially extended system exhibiting a double-fold bifurcation (also known as an S-shaped bifurcation). In this bifurcation, two equilibrium points collide and vanish, leading to tipping phenomena within a hysteresis loop. Similar elliptic problems to equations (92a) and (92b) with these properties emerge in various applications such as gas combustion and chemical kinetics [8, 76], plasma physics and Grad–Shafranov equilibria in Tokamaks [30, 88, 157], and gravitational equilibrium of polytropic stars [30, 74, 99]. Beyond polynomial nonlinearities, other spatially extended systems are known to exhibit S-shaped bifurcation diagrams. Examples include: Earth's energy balance models [83, 97, 131], and oceanic models based on hydrostatic primitive or Boussinesq equations employed in the study of the Atlantic meridional overturning circulation (AMOC) [67, 165, 172]. We also note that the OPM approach has already been successfully applied to predict tipping phenomena in a prototype Stommel–Cessi model of the AMOC driven by Gaussian noise [39].

For equation (92a), the bifurcation parameter is λ. For $L = 2$, it is known that the solution curve $\lambda \mapsto u_\lambda$ exhibits an S-shaped bifurcation when $\epsilon < \epsilon^* \approx 0.3103$ [107, theorem 3.2]. This critical constant ϵ^* is derived from a general analytic formula applicable to a broader class of equations where u^3 is replaced by u^p with a fractional exponent $p > 2$ [107]. More precisely, for ϵ satisfying

$$\epsilon^{\frac{p}{p-2}} < \frac{1}{3} \left[\frac{2}{p(p-1)} \right]^{\frac{p}{p-2}} - \frac{p-1}{p+1} \left[\frac{2}{p(p-1)} \right]^{\frac{p}{p-2}}. \tag{93}$$

Theorem 3.2 of [107] ensures that the solution curve to equations (92a) and (92b) (with u^3 replaced by u^p) is exactly S-shaped. We focus here on the case $p = 3$.
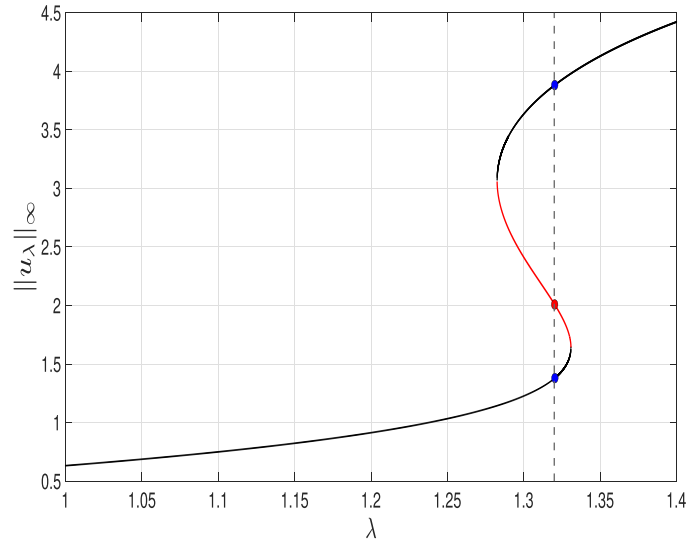
Figure 11. Double-fold bifurcation in equations (92*a*) and (92*b*). The red curve shows unstable steady states for $\epsilon = \epsilon^*/2$. The black curves represent stable steady states. The double-fold bifurcation occurs at the two turning points where the unstable states (red curve) collide with the stable states (black curves). The vertical dashed line indicates the value $\lambda = 1.32$ used in table 3 for stochastic simulation in section 6.2, and prediction experiments in section 6.3.

To locate the interval of λ over which an exact multiplicity of three steady states occurs (as predicted by the S-shape) we compute the bifurcation diagram of equations (92*a*) and (92*b*) for $\epsilon = \epsilon^*/2$; see appendix C for details. The result is shown in figure 11. The vertical dashed line at $\lambda = 1.32$ marks the λ-value at which stochastic perturbations are included below to trigger jumps across these steady states.

6.2. System's sensitivity to jumps, and jump-driven dynamics

In this section, we consider stochastic perturbations of jump type carried by some of the eigenmodes of the linearized problem at the unstable steady state U_λ^*. The latter problem is given by

$$\partial_x^2 \psi + \lambda \left(2U_\lambda^* - 3\epsilon U_\lambda^{*\,2} \right) \psi = \beta \psi, \text{ on } (0, L), \tag{94a}$$

$$\psi(0) = \psi(L) = 0, \tag{94b}$$

and is solved below using a Chebyshev spectral method [167, chapters 6 & 7] for the numerical experiments. Hereafter we denote by $(\beta_n, \boldsymbol{e}_n)$ the eigenpairs solving equations (94*a*) and (94*b*). Note that the eigenvalues are real and simple for this problem.

Thus, the jump-driven SPDE at the core of our reduction study is given by:

$$\partial_t u - \partial_x^2 u - \lambda \left(1 + u^2 - \epsilon u^3 \right) = \lambda \eta_t, \tag{95a}$$

$$u(0) = u_0 \in H, \tag{95b}$$

supplemented with Dirichlet boundary conditions. The ambient Hilbert space is $H = L^2(0, L)$ endowed with its natural inner product denoted by $\langle \cdot, \cdot \rangle$.

Table 3. Parameter values for equation (95*a*).

ϵ	f_r	Δt	σ	λ	L
$\frac{\epsilon^*}{2}$	0.35	1	300	1.32	2

Here, η_t is a jump noise term that takes the form

$$\eta_t(x) = \sigma\zeta_t f(t)\left(\boldsymbol{e}_3(x) + \boldsymbol{e}_5(x)\right), \; x \in (0, L), \tag{96}$$

where $\sigma \geqslant 0$, $\boldsymbol{e}_3$ and $\boldsymbol{e}_5$ are the third and fifth eigenmodes for equations (94*a*) and (94*b*), ζ_t is a random variable uniformly distributed in $(-1, 1)$, and $f(t)$ is a square signal, whose activation is randomly distributed in the course of time. More precisely, given a firing rate f_r in $(0, 1)$, and duration $\Delta t > 0$, we define the following real-valued jump process:

$$f(t) = \mathbb{1}_{\{\xi_n \leqslant f_r\}}, \quad n\Delta t \leqslant t < (n+1)\Delta t, \tag{97}$$

where ξ_n is a uniformly distributed random variable taking values in $[0, 1]$ and $\mathbb{1}_{\{\xi \leqslant f_r\}} = 1$ if and only if $0 \leqslant \xi \leqslant f_r$. We refer to [66, section 3.4] for existence results in Hilbert spaces of mild solutions to SPDEs driven by jump processes, that apply to equations (95*a*) and (95*b*). This type of jump process is also known as a dichotomous Markov noise [9], or a two-state Markov jump process in certain fields [160]. It is encountered in many applications [101].

The specific values for the noise parameters (frequency rate, f_r, time scale, Δt, and intensity, σ) are provided in table 3. The random forcing in our simulations is chosen with two key considerations:

- **Time scale (Δt)**: The time scale Δt is large enough to allow the system to gradually relax towards its stable steady states over time.
- **Frequency rate (f_r)**: The frequency rate, f_r, is adjusted to ensure this relaxation process can occur effectively.

This configuration ensures that if a perturbation is applied for a duration of Δt and then removed for several subsequent Δt intervals, the system naturally return to its closest stable steady state.

This relaxation behavior after a perturbation is clearly visible in the first mode's amplitude, $y_1(t) = \langle u(t), \boldsymbol{e}_1 \rangle$ (where u solves equations (95*a*) and (95*b*)). The top panel of figure 12 illustrates this concept. Shortly after the vertical dashed magenta line at $t = 21.4$, we observe $y_1(t)$ relaxing towards the constant value $\langle U_\lambda^M, \boldsymbol{e_1} \rangle$. Here, U_λ^M denotes the stable steady state with the largest energy for $\lambda = 1.32$ (represented by the top blue dot in figure 11). Similar relaxation episodes occur throughout the simulation, with $y_1(t)$ sometimes relaxing towards $\langle U_\lambda^m, \boldsymbol{e}_1 \rangle$, involving the stable steady state U_λ^m with the smallest energy (refer again to figure 11).

While the noise intensity ($\sigma = 300$) in table table 3 may seem large, this value is necessary because we specifically target modes $\boldsymbol{e}_3$ and $\boldsymbol{e}_5$ used in the stochastic forcing defined in equation (96). These modes contribute less than 1% of the total energy in the unforced parabolic equation. When the jump noise is activated with this intensity, the amplitudes of $y_3(t)$ and $y_5(t)$ fluctuate significantly, becoming comparable in magnitude to $y_1(t)$. This eliminates the previous order-of-magnitude difference.

As a result, the random perturbations introduced by modes $\boldsymbol{e}_3$ and $\boldsymbol{e}_5$, through the system's nonlinearities, can now drive the SPDE solution across the potential barriers associated with the steady states. This explains the multimodal behavior observed in the time evolution of $y_1(t)$ as shown in figure 12.
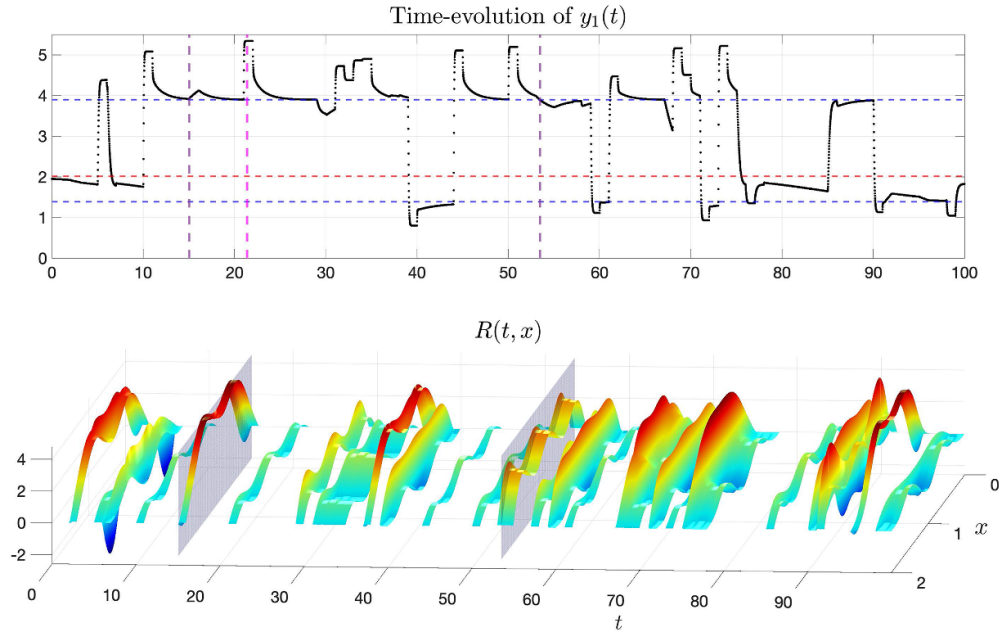
Figure 12. Jump-driven dynamics. Top panel: The top (resp. bottom) horizontal blue dashed line shown corresponds to $\langle U_\lambda^M, e_1 \rangle$ (resp. $\langle U_\lambda^m, e_1 \rangle$) with U_λ^M (resp. U_λ^m) denoting the stable steady state to equations (92a) and (92b) with maximum (resp. minimum) energy. The horizontal dashed red line corresponds to $\langle U_\lambda^*, e_1 \rangle$ with U_λ^* denoting the unstable steady state. The black dotted curve going across these horizontal dashed lines (and the associated potential barriers) is showing $y_1(t) = \langle u(t), e_1 \rangle$ with u solving equations (95a) and (95b) for the parameter values of table 3. Bottom panel: Here, is shown $R(t,x) = u(t,x)/\mathtt{std}(u(t)) - \eta_t(x)/\mathtt{std}(\eta_t)$ when $\eta_t \neq 0$. The shaded planes mark two different patterns exhibited by the system's response, for two close y_1-values (vertical purple dashed lines), illustrating the difficulty of closing accurately the y_1-variable. See Text for more details.

A crucial question remains: with such a strong random force ($\sigma = 300$), are the observed dynamics in the system solely a reflection of this forcing, or is there more to it? To address this, we calculate a quantity called $R(t,x)$. This quantity compares the normalized solution, $u(t,x)/\mathtt{std}(u(t))$, with the normalized forcing term, $\eta_t(x)/\mathtt{std}(\eta_t)$, at times when forcing is present ($\eta_t \neq 0$); see caption of figure 12. Here, $\mathtt{std}$ denotes standard deviation.

The key point is that the forcing term, $\eta_t(x)$, consists of a time-dependent coefficient multiplied by a specific spatial pattern defined by $p(x) = e_3(x) + e_5(x)$. If the system's response were strictly proportional to the forcing, then $R(t,x)$ would simply match this spatial pattern ($p(x)$ with three peaks and two valleys).

However, the bottom panel of figure 12 tells a different story. We see that $R(t,x)$ deviates significantly and in a complex way from $p(x)$ as time progresses. This implies that the system's response is not simply a mirror image of the forcing term. The nonlinear terms in the equation play a crucial role in shaping the response dynamics.

Developing reduced models for this system faces another significant hurdle: the system's high sensitivity to small changes, particularly in the first mode's amplitude (y_1). Here is the key point: even tiny variations in y_1, especially near the potential barrier of a steady state, can lead to vastly different spatial patterns in the full SPDE solution. Figure 12 illustrates this

sensitivity. For example, consider points $t_1 = 15.04$ and $t_2 = 53.49$ marked by the two vertical purple dashed lines in the top panel of figure 12. The corresponding values of y_1 are very close ($y_1(t_1) = 3.92$ and $y_1(t_2) = 3.89$). However, the SPDE solutions at these times (shown in the vertical shaded planes) exhibit dramatically different spatial patterns.

The significant difference between the response of the first mode's amplitude (y_1) and the full SPDE solution (u) in equation equations (95a) and (95b) underscores the critical need for an accurate parameterization of the neglected variables. Capturing this kind of sensitive behavior with a reduced model that only tracks y_1 requires a comprehensive representation of the influence from these excluded variables. We demonstrate in the next section that our optimal parameterization framework effectively tackles this challenging parameterization issue.

6.3. Non-Markovian optimal reduced model

We describe below how our BF framework introduced in this work to derive stochastic parameterizations of SPDEs driven by white noise can be readily adapted to SPDEs driven by jump processes. The stochastic model equations (95a) and (95b), driven by the jump process defined in equation (96), serves as a concrete example. Our objective is to develop a reduced model for the first mode's amplitude in equation (95a). In other words, we aim to create a simplified, one-dimensional stochastic model that can accurately reproduce the complex dynamics discussed in section 6.2, including the stochastic transitions.

To do so, we rewrite equation (95a) for the fluctuation variable

$$v = u - U_\lambda^*, \tag{98}$$

where U_λ^* denotes the unstable steady state of equations (92a) and (92b) (for $\lambda = 1.32$) corresponding to the red dot shown in figure 11. The fluctuation equation in the v-variable reads then as follows:

$$\partial_t v - \partial_x^2 v - \lambda \left(2U_\lambda^* - 3\epsilon U_\lambda^{*2} \right) v - \lambda \left(1 - 3\epsilon U_\lambda^* \right) v^2 + \lambda \epsilon v^3 = \lambda \eta_t, \text{ on } (0,L) \times (0,T). \tag{99}$$

Each nth eigenmode, e_n, solving equations (94a) and (94b) turns out to be very close to the Fourier sine mode $\sqrt{2/L}\sin(n\pi x/L)$, and the eigenvalues decay quickly. The first few eigenvalues are $\beta_1 = 0.1815$, $\beta_2 = -7.4966$, $\beta_3 = -19.9665$, $\beta_4 = -37.2840$, and $\beta_5 = -59.5108$. Here only the first mode e_1 is unstable while the others are all stable.

Our reduced state space is thus taken to be

$$H_\mathfrak{c} = \text{span}\{e_1\}. \tag{100}$$

For the parameter setting of table 3, most of the energy (99.9%) is distributed among the unstable mode e_1, on one hand, and the forced modes, e_3 and e_5, on the other, with energy carried by the third mode's amplitude representing up to 60% depending on the noise path. An accurate parameterization of these forced modes along with their nonlinear interactions with the unstable mode is thus key to achieve for a reliable reduced model of the first mode's amplitude.

To do so, we consider the BF parameterization framework of section 3.1 for cubic SPDEs driven by white noise that we adapt to the jump noise case. If one replaces the Brownian motion in the BF system (39a)–(39c) by the jump process term, $\sigma\zeta_t f(t)$ (from equation (96)), one gets the following parameterization of the nth mode's amplitude ($n \in \{3,5\}$):

$$\Phi_n(\tau_n, X, t; \omega) = e^{\beta_n \tau}Y + J_\tau^n(t; \omega) + D_{11}^n(\tau)B_{11}^nX^2 + E_{111}^n(\tau)C_{111}^nX^3, \tag{101}$$

where the coefficients B_{11}^n, C_{111}^n, D_{11}^n, and E_{111}^n are those defined in section 3.1, while the stochastic term is now given by the non-Markovian coefficient

$$J_\tau^n(t;\omega) = \sigma\lambda \int_{t-\tau}^t e^{\beta_n(t-s)} \zeta_s f(s)\,ds. \tag{102}$$

This stochastic term is the counterpart of the Z^n-term defined in (41) for the white noise case. As for the Z^n-term, the J^n-term is efficiently generated by solving the following ODE with random coefficients:

$$\frac{dJ}{dt} = \beta_n J + \sigma\lambda \left(\zeta_t f(t) - e^{\beta_n\tau}\zeta_{t-\tau} f(t-\tau)\right), \tag{103}$$

for $n=3$ and $n=5$; cf equation (49). Note that the parameterization equation (101) takes the same structural form as equation (82) in the case of white noise forcing (see remark 3.1), with the bounded jump noise replacing the latter here. This is due to the underlying BF systems at the core of the derivation of these formulas where only the nature of the forcing changes; see also section 8 below.

Then, after optimization over a single training path of the parameterization defects $Q_n(\tau)$ (see section 6.4 below), an analogue to the optimal reduced model (equation (86)) becomes in our case the following 1-D ODE with path-dependent coefficients:

$$\dot{y} = \beta_1 y + \left\langle (\lambda - 3\lambda\epsilon U_\lambda^*)(y(t)\boldsymbol{e}_1 + \mathcal{P}(t))^2, \boldsymbol{e}_1\right\rangle - \lambda\epsilon\left\langle (y(t)\boldsymbol{e}_1 + \mathcal{P}(t))^3, \boldsymbol{e}_1\right\rangle, \tag{104}$$

with

$$\mathcal{P}(t) = \Phi_3(\tau_3^*, y(t), t)\boldsymbol{e}_3 + \Phi_5(\tau_5^*, y(t), t)\boldsymbol{e}_5. \tag{105}$$

Here, $\mathcal{P}(t)$ is the optimized stochastic parameterization in the class $\mathcal{P}$ (see section 4.3) for the SPDE (99) driven by the jump process η_t defined in equation (96). Note that the required non-Markovian J^n-terms to simulate equation (104) are obtained by solving equation (103) in the course of time. So a total of three ODEs with random coefficients are solved to obtain a reduced model of the first mode's amplitude. Thus, equation (104) along with its auxiliary J-equations, form the optimal reduced system.

The inner product in equation (104) leverages the analytical expression of $\mathcal{P}(t)$ provided in the same equation. This inner product is computed once offline, resulting in a degree-9 polynomial with random coefficients after expanding the Φ_3 and Φ_5 terms as described in equation (101). However, this expression can be further simplified. By analyzing the order of magnitude of the terms in equation (101) for the parameter regime of table 3 (σ large), we observe that $\mathcal{P}(t)$ simplifies to

$$\mathcal{P}(t) \approx J_{\tau_3^*}^3(t;\omega)\boldsymbol{e}_3 + J_{\tau_5^*}^5(t;\omega)\boldsymbol{e}_5, \tag{106}$$

setting $Y=0$ in Φ_n ($n=3,5$) defined in equation (101). The remaining unresolved modes, which are also unforced, contribute less than 0.1% of the total energy and are therefore neglected in the parameterization $\mathcal{P}(t)$. In any event, the approximation in (106) expresses a prominence of the stochastic terms over the nonlinear ones in the parameterization (101), unlike for the reduction of the sACE analyzed in section 5.4. The parameterization becomes then purely stochastic without nonlinear terms. Yet, the optimal reduced system, equation (104), exhibits nonlinear products of stochastic coefficients.

To clarify this point, we introduce a few more notations. We denote by $a_\lambda^\epsilon(x)$ the spatial coefficient, $\lambda - 3\lambda\epsilon U_\lambda^*(x)$, involved in the quadratic term of equation (104) (as inherited from equation (99)). To simplify the notations we denote also by J_3^* (resp. J_5^*) the term $J_{\tau_3^*}^3(t;\omega)$ (resp. $J_{\tau_5^*}^5(t;\omega)$), and introduce the bilinear operator $G_2(u,v)(x) = u(x)v(x)$, for x in $(0,L)$ and
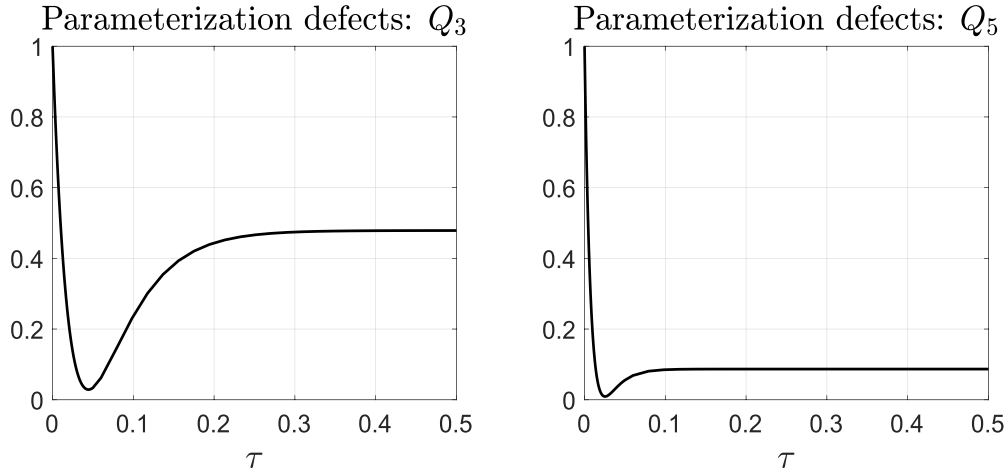
Figure 13. The normalized parameterization defects, Q_n for $n = 3$ and $n = 5$. The optimization is performed on a single path. See Text.

any functions u, v in $L^2(0, L)$. We have then, using the approximation formula (106), that the quadratic term in equation (104) expands as follows

$$
\begin{aligned}
\left\langle a_\lambda^\epsilon \left(y(t)\boldsymbol{e}_1 + \mathcal{P}(t)\right)^2, \boldsymbol{e}_1 \right\rangle = {} & y^2(t) \left\langle a_\lambda^\epsilon G_2\left(\boldsymbol{e}_1, \boldsymbol{e}_1\right), \boldsymbol{e}_1 \right\rangle + \left(J_3^*(t)\right)^2 \left\langle a_\lambda^\epsilon G_2\left(\boldsymbol{e}_3, \boldsymbol{e}_3\right), \boldsymbol{e}_1 \right\rangle \\
& + \left(J_5^*(t)\right)^2 \left\langle a_\lambda^\epsilon G_2\left(\boldsymbol{e}_5, \boldsymbol{e}_5\right), \boldsymbol{e}_1 \right\rangle + 2\left\langle a_\lambda^\epsilon G_2\left(\boldsymbol{e}_3, \boldsymbol{e}_5\right), \boldsymbol{e}_1 \right\rangle J_3^*(t)J_5^*(t) \\
& + 2\left\langle a_\lambda^\epsilon G_2\left(\boldsymbol{e}_1, \boldsymbol{e}_3\right), \boldsymbol{e}_1 \right\rangle y(t)J_3^*(t) + 2\left\langle a_\lambda^\epsilon G_2\left(\boldsymbol{e}_1, \boldsymbol{e}_5\right), \boldsymbol{e}_1 \right\rangle y(t)J_5^*(t),
\end{aligned}
\tag{107}
$$

with a similar expression obtained from the cubic terms in equation (104). Already from equation (107) one observes the production of stochastic constant terms driven by $(J_3^*(t))^2$, $(J_5^*(t))^2$ and $J_3^*(t)J_5^*(t)$ as well as linear terms driven by $J_3^*(t)$ and $J_5^*(t)$ contributing to the dynamics of the surrogate first mode's amplitude y. These stochastic terms turn out to be essential for a good emulation of the first mode's amplitude. This point is illustrated by the following experiment.

### 6.4. Training data, parameterization defects, and reduced model's skills

The training data for the optimal reduced model is comprised of $N = 40\,000$ snapshots. These sACE snapshots are generated over a training interval from 0 to T, with $T = 400$, for a single, noise path driving the sACE. The corresponding normalized parameterization defects (quantified by Q_n) are visualized in figure 13.

Here, it is important to note that the distinct minima observed in Q_3 and Q_5 are not caused by weak time-scale separations (as discussed in section 5.5 for the sACE). Instead, they are primarily a consequence of the high noise intensity (σ) used in this case, which leads to solutions with large amplitudes. This is further supported by the observation that these minima become less pronounced as σ is reduced (verified for arbitrary noise paths).

Figure 14 shows the optimal parameterization's ability to reproduce first mode's complex behavior. This figure showcases the effectiveness of the optimal reduced model (equation (104) with $\mathcal{P}(t)$ from equation (106)) in capturing the bimodal nature of the first mode's amplitude.
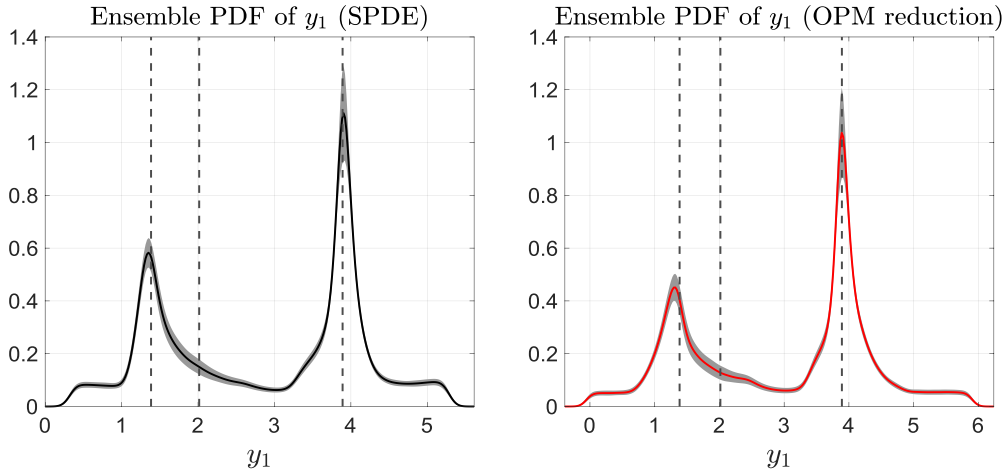
Figure 14. Ensemble PDF. Here, 10^5 out-of-sample test paths are used to estimate these ensemble PDFs. Each underlying solution path is made of $N = 2 \times 10^5$ iterations for a time-step of $\delta t = 10^{-2}$.
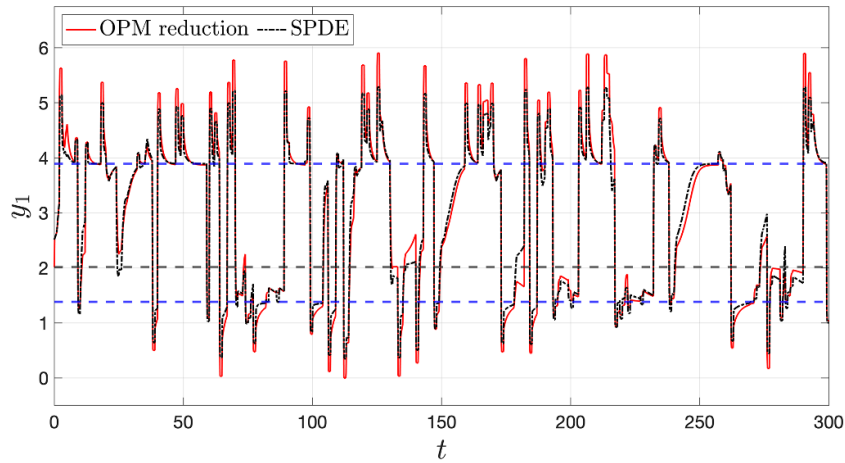
Figure 15. First's mode amplitude dynamics: Full system vs optimal reduction.

The model's performance is evaluated using a vast set of unseen test paths. When compared to the actual SPDE solution's amplitude for the first mode (across these test paths), the optimal reduced model closely follows the various transitions between metastable states, crossing as needed the potential barriers. However, it is worth noting that this reduced model tends to slightly overestimate the intensity of these larger excursions (as shown in figure 15).

7. Towards the reduction of more general Lévy-driven dynamics

We prolong the discussion and results of section 6 towards more general Lévy noise. For the sake of simplicity we focus on finite-dimensional SDEs driven by Lévy processes.

There is a vast literature on Lévy processes which roughly speaking are processes given as the sum of a Brownian motion and a jump process [3, 50, 108, 133, 139]. Unlike the case of diffusion processes (SDEs driven by Brownian motions), the representation of Kolmogorov operators is non-unique in the case of Lévy-driven SDEs, and may take the form of operators involving fractional Laplacians or singular integrals among other representations [111]. We adopt here the definition commonly used in probability theory [4, theorem 3.5.14] which presents the interest of being particularly intuitive as recalled below. We consider SDEs of the form

$$\mathrm{d}X_t = \boldsymbol{F}(X_t)\,\mathrm{d}t + \boldsymbol{\Sigma}(X_t)\,\mathrm{d}\boldsymbol{W}_t + \mathrm{d}\boldsymbol{L}_t, \tag{108}$$

where $\boldsymbol{W}_t$ is a Brownian motion in $\mathbb{R}^d$, and $\boldsymbol{L}_t$ is a Lévy process on $\mathbb{R}^d$ independent from $\boldsymbol{W}_t$. Here, we take $\boldsymbol{F}$ to be a smooth vector field on $\mathbb{R}^d$, and $\boldsymbol{\Sigma}(\boldsymbol{x})$ to be, for each $\boldsymbol{x}$, a $d \times d$ matrix with smooth entries, such that $(a_{ij}(\boldsymbol{x})) = \boldsymbol{\Sigma}(\boldsymbol{x})\boldsymbol{\Sigma}^T(\boldsymbol{x})$ is a positive definite matrix for each $\boldsymbol{x}$ in $\mathbb{R}^d$.

Roughly speaking, a Lévy process $\boldsymbol{L}_t$ on $\mathbb{R}^d$ is a non-Gaussian stochastic process with independent and stationary increments that experience sudden, large jumps in any direction. The probability distribution of the these jumps is characterized by a non-negative Borel measure ν defined on $\mathbb{R}^d$ and concentrated on $\mathbb{R}^d\backslash\{0\}$ that satisfies the property $\int_{\mathbb{R}^d\backslash\{0\}} \min(1,\boldsymbol{y}^2)\nu(\mathrm{d}\boldsymbol{y}) < \infty$. This measure ν is called the jump measure of the Lévy process $\boldsymbol{L}_t$. Sometimes X_t itself is referred to as a Lévy process with triplet $(\boldsymbol{F}, \boldsymbol{\Sigma}, \nu)$. Within this convention, we reserve ourselves the terminology of a Lévy process to a process with triplet $(0, 0, \nu)$. We refer to [3] and [139] for the mathematical background on Lévy processes.

Under suitable assumptions on $\boldsymbol{F}$, $\boldsymbol{\Sigma}$, and the Lévy measure ν, the solution X_t to equation (108) is a Markov process (e.g. [141]) and even a Feller process [109], with associated Kolmogorov operator taking the following integro-differential form for (e.g.) ψ in $C_c^\infty(\mathbb{R}^d)$ (Courrège theorem [53, 108]):

$$\mathcal{L}_K\psi(\boldsymbol{x}) = \boldsymbol{F}(\boldsymbol{x})\cdot\nabla\psi + \sum_{i,j=1}^d a_{ij}(\boldsymbol{x})\,\partial_{ij}\psi + \Gamma\psi(\boldsymbol{x}),\,\text{with} \tag{109a}$$

$$\Gamma\psi(\boldsymbol{x}) = \int_{\mathbb{R}^n\backslash\{0\}} \left[\psi(\boldsymbol{x}+\boldsymbol{y}) - \psi(\boldsymbol{x}) - \boldsymbol{y}\cdot\nabla\psi(\boldsymbol{x})\mathbb{1}_{\{\|\boldsymbol{y}\|<1\}}\right]\nu(\mathrm{d}\boldsymbol{y}), \tag{109b}$$

where $\mathbb{1}_{\{\|\boldsymbol{y}\|<1\}}$ denotes the indicator function of the (open) unit ball in $\mathbb{R}^n$; see also [3, theorem 3.3.3].

The first-order term in equation (109a) is the drift term caused by the deterministic, non-linear dynamics. The second-order differential operator represents the diffusion part of the process X_t. It is responsible for the continuous component of the process.

The Γ-term, involving the integral, represents the jump part of the process. It captures the discontinuous jumps that the process experience due to the sudden changes caused by the Lévy process $\boldsymbol{L}_t$. Its intuitive interpretation breaks down as follows. The term, $\psi(\boldsymbol{x}+\boldsymbol{y}) - \psi(\boldsymbol{x})$, calculates the difference in the test function value before and after the potential jump, capturing the change in the test function due to the jump.

The term $-\boldsymbol{y}\cdot\nabla\psi(\boldsymbol{x})\mathbb{1}_{\{\|\boldsymbol{y}\|<1\}}$ represents a first-order correction for small jumps. It aims to account for the fact that a small jump might not land exactly on the grid point $(\boldsymbol{x}+\boldsymbol{y})$, but somewhere in its vicinity. This term is often referred to as the Girsanov correction. Thus, the integral term Γ accounts for all possible jump sizes ($\boldsymbol{y}$) within a unit ball, as weighted by the Lévy measure $\nu(\mathrm{d}\boldsymbol{y})$. The notion of Kolmogorov operators can be extended to SPDEs driven by cylindrical Wiener processes, with $\boldsymbol{x}$ lying in a Hilbert space; see [65]. Building up on such a functional framework, it is possible to provide a rigorous meaning of the Kolmogorov operator

$\mathcal{L}_K$ in the case of an SPDE driven by a general Lévy process, forcing possibly infinitely many modes.

In the case of SPDE driven by a scalar jump process (forcing only two modes) studied in section 6 (see equation (96)), the jump measure is simple to describe. It corresponds to the measure $\nu(\mathrm{d}s)$ (s real) associated with the real-valued jump process $f(t)$ given by equation (97). This measure is the Dirac delta $\lambda\delta_{s=1}$, where λ is the intensity of the Poisson process associated with the random comb signal. Recall that this intensity is the limit of the probability of a single event occurring in a small interval divided by the length of that interval as the interval becomes infinitesimally small. Therefore, λ is equal to the firing rate f_r of $f(t)$.

In the case of SDEs driven by α-stable Lévy processes,

$$\nu(\mathrm{d}\boldsymbol{y}) = c_{d,\alpha}\|\boldsymbol{y}\|^{-d-\alpha}\,\mathrm{d}\boldsymbol{y}, \tag{110}$$

where the non-Gaussianity index α lies in $(0,2)$, and $c_{d,\alpha}$ is a constant that depends on the dimension and involves the gamma function of Euler [111]. A 2-stable ($\alpha = 2$) process is simply a Brownian motion.

The rigorous theory of slow and center manifolds are in their infancy for SDEs driven by α-stable Lévy processes [174, 175], but the main formulas share the same ingredients as recalled in section 2.2, with analogous BF interpretations of those discussed in section 2.3 for the Lyapunov–Perron integrals involved therein ([174, section 4]). The non-Markovian parameterizations of section 3 and their data-driven optimization, can thus be extended, at least formally, for the case of SDEs driven by α-stable Lévy processes and will be pursued elsewhere on concrete applications.

8. A chart summarizing the approach

To streamline the application of the OPM reduction approach for stochastically forced systems presented in this paper, we provide a concise overview of its key steps, objectives, and characteristics in the chart shown in figure 16. This summary is tailored to the class of SPDEs discussed in section 2.1, with the primary goal of performing a reduction with a cutoff scale just below the forcing scale. This reduction involves a solution splitting where only the disregarded 'small' scales are stochastically forced (figures 16(A) and (B)).

Given this splitting, where the disregarded scales are those directly forced by stochasticity, our approach reveals the importance of non-Markovian terms in the parameterization of these scales. These terms depend on the noise's path history, enabling the tracking of small-scale fluctuations. The accuracy of this tracking is primarily governed by the memory content (τ) of the parameterization, which is optimized from data by minimizing a natural loss function (figure 16(E)) over a single noise realization ω.

Once optimized, the resulting non-Markovian parameterization, the OPM, allows us to capture the average behavior of small-scale fluctuations as a functional of large scales by averaging over noise paths (see inset below figure 16(B)). To efficiently optimize the memory content (τ) and avoid the computation of cumbersome quadratures, we decompose the J-term in figure 16, $J(t,\boldsymbol{x}) = \sum J_n(t)\boldsymbol{e}_n(\boldsymbol{x})$, into random coefficients J_n that solve simple auxiliary ODEs dependent on τ and the noise path ω. The resulting OPM closure, exemplified by equation (85), takes the form of an SDE system with coefficients that are nonlinear functionals of these J_n terms (figure 16(F)) reflecting the nonlinear interactions between the noise and nonlinear terms in the original system.
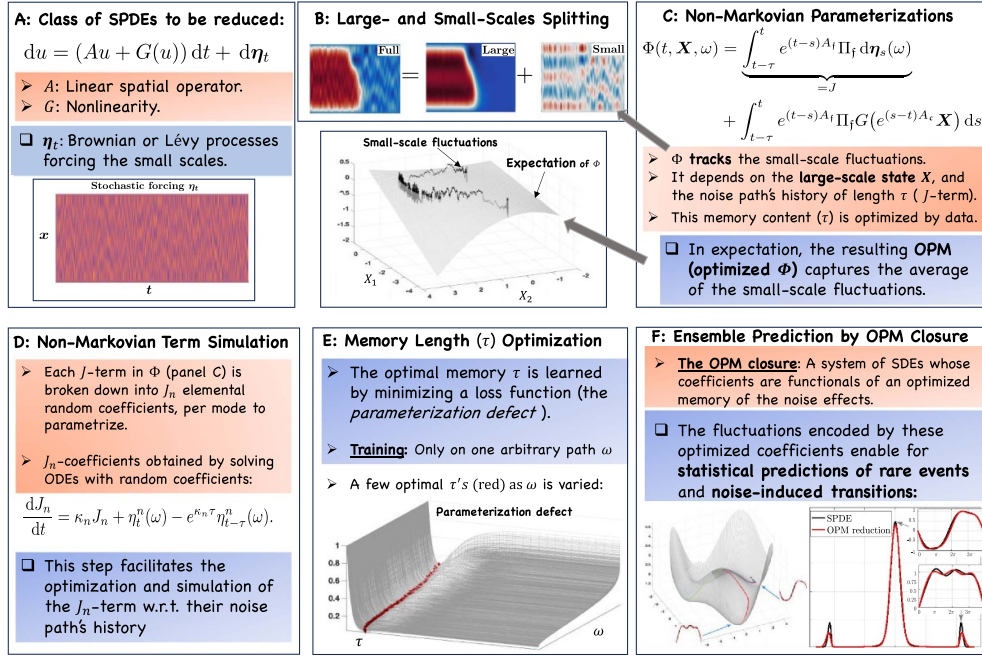
Figure 16. Chart presenting the main steps, objectives, and characteristics of the OPM reduction approach.

As demonstrated with SPDEs driven by Gaussian noise (section 5) or jump processes (section 6), the OPM closure effectively predicts noise-induced transitions and rare event statistics. This success is attributed to our hybrid approach, which combines analytical insights for constructing the parameterizing manifold Φ (figure 16(C) and section 3.1) with data-driven learning of the optimal memory content τ (section III C).

9. Discussion and perspectives

Thus, this work demonstrates that OPMs constructed from BF systems and optimized using a single solution path can accurately reproduce noise-induced transitions across a broad range of out-of-sample, test paths. This approach is effective for non-Gaussian noise with jumps. A remarkable feature of the approach is that, even though these two different types of noise typically lead to different dynamical responses, the OPMs in these two settings take exactly the same structural form with the corresponding non-Markovian terms depending on the past history of the corresponding noise forcing the original SPDE; cf equations (40) and (41) and cf equations (101) and (102).

Interestingly, the optimal parameterization formulas align with the approximation of stochastic invariant manifolds when they exist. However, when these manifolds are no longer applicable, the optimal parameterizations transcends the invariance constraint. In such cases, it

provides the optimal stochastic parameterization with respect to the random invariant measure ρ_ω (defined in section 4.1), conditioned on the resolved variable X (as proven in theorem 4.1).

The formulas derived from the BF systems in section 3 enable practical approximation of the theoretical optimal parameterization, particularly its non-Markovian random coefficients. These coefficients become crucial, especially when noise affects unresolved modes. They essentially encode how noise propagates to the resolved modes through nonlinear interactions.

We have shown, through various examples, that training on a single noise path equips the coefficients with an optimized historical memory (of the noise) that plays a key role in the accurate prediction of noise-induced excursion statistics, even when applied to a diverse set of test paths. This demonstrates the strength of our hybrid framework, which combines analytical understanding with data-driven insights. This approach allows us to overcome the 'extrapolation problem', a significant challenge for purely data-driven machine-learned parameterizations.

Section 2.3 discusses the possibility of constructing more intricate parameterizations using an iterative scheme based on more general BF systems. However, these elaborations require more computational effort due to the repeated stochastic convolutions. Implementing them efficiently necessitates deriving auxiliary ODEs (analogues to equation (49)) for the coefficients (a_n and b_{in}) in equations (31) and (32). Chekroun *et al*[43], section 7.3, provides examples of such auxiliary ODEs for handling repeated stochastic convolutions.

These more general stochastic parameterizations have the potential to significantly enhance the parameterization capabilities of Φ (equation (33)). This is because they encode higher-order interaction terms with respect to the noise amplitude, as shown in equation (30). Resolving these interactions is potentially important for improving the prediction of rare event statistics by non-Markovian reduced systems, especially when the noise intensity is low. This opens doors to potential applications using large deviation or rare event algorithms (e.g. [28, 80, 89, 143, 147, 158, 169, 171]). Additionally, iterating the BF system (equations (25a) and (25b)) beyond $\ell = 2$ can also introduce higher-order terms involving the low-mode amplitude X.

The inclusion of higher-term in X is non-unique though. For instance, inspired by [43, section 7.2.1], stochastic parameterizations accounting for higher-order terms in X are produced by solving analytically the following BF system:

$$\mathrm{d}p^{(1)} = A_{\mathfrak{c}}p^{(1)}\,\mathrm{d}s, \tag{111a}$$

$$\mathrm{d}p^{(2)} = \left(A_{\mathfrak{c}}p^{(2)} + \Pi_c G_2\left(p^{(1)}\right)\right)\mathrm{d}s, \tag{111b}$$

$$\mathrm{d}q_n = \left(\lambda_n q_n + \Pi_n G_2\left(p^{(2)}\right)\right)\mathrm{d}s + \sigma_n\,\mathrm{d}W_s^n, \tag{111c}$$

$$p^{(1)}(t) = p^{(2)}(t) = X \in H_{\mathfrak{c}},\; q_n(t-\tau) = Y, \tag{111d}$$

for which the two first equations for $p^{(1)}$ and $p^{(2)}$ are simultaneously integrated backward over $[t-\tau, t]$, followed by the forward integration of the q_n-equation over the same interval. Compared to equations (39a)–(39c), this BF system has two-layer in the low-mode variable p resulting into parameterizations of order 4 in X; see [35, theorem 2] for an example. These higher-order terms come from the nonlinear self-interactions between the low modes as brought by the term $\Pi_c G_2\left(p^{(1)}\right)$. Accounting for additional interactions either through (111a) or the iterative procedure of section 2.3 can help improve the parameterization skills (see also [18, 142]), and should be considered if accuracy is key to resolve.

An alternative route to enhance accuracy would involve to parameterize stochastically the parameterization residual from the loss functions in (53), post-minimization. This technique has proven effective in modeling the fast motion transversal to OPM manifolds by means of networks of nonlinear stochastic oscillators [38, 117], as diagnosed by Ruelle–Pollicott resonances [45, 163]. Similarly, rapid dynamical fluctuations, transversal to (Markovian) OPM manifolds (figure 9), could benefit from refined parameterizations through the incorporation of additional stochastic components following [38], potentially leading to more accurate representation of rare event statistics by the corresponding closures (figure 7).

Extending beyond SPDEs driven by Markov processes, our framework can naturally accommodate SPDEs driven by non-Markovian stochastic processes. A prime example is when the stochastic forcing is a fractional Brownian motion (fBm) [122], for which the existence of unstable manifolds has been demonstrated [82]. By leveraging Lyapunov–Perron techniques, as employed in the construction of these manifolds (section 2.2), our extension program to parameterization can proceed analogously to section 3. This involves here as well, finite-time integration of relevant BF systems to derive optimizable parameterizations from the governing equation. The random coefficients in these parameterizations would then become again functionals of the past history of the forcing noise, albeit this time incorporating a higher degree of memory compared to the stochastic process Z_τ^n used in the parameterizations section 3, due to the original memory content of the original forcing. Such generalized parameterizations should provide valuable insights into the interplay between nonlinearity and noise in physical systems exhibiting long-term memory effects [75, 138, 150].

Regardless of the Markovian or non-Markovian nature of the stochastic forcing, our reduction approach and its potential extensions offer valuable tools for addressing critical questions. By elucidating the intricate interplay between noise and nonlinearity, we can gain deeper dynamical insights. For example, we anticipate applying this reduction program to investigate early warning indicators of tipping phenomena, from a dynamical perspective, by tracing the energy flow from forced scales to physically relevant modes [118, 168]. Climate tipping elements provide a compelling domain for exploring these concepts and applying our methods [112, 113, 124].

The proposed stochastic parameterization method offers a new framework for understanding noise-nonlinearity interactions in stochastic multiscale systems. It is important to note that alternative techniques based on stochastic time scale separation, such as the stochastic path perturbation approach, have been employed to study related phenomena, including the random escape of stochastic fields from criticality in explosive systems [24] and first passage time problems in nonlocal Fisher equations [25, 79]. Future research directions could explore the potential of our method to obtain analytical approximations for first passage times in nonlocal SPDEs near criticality. Such an extension could provide valuable insights into pattern formations for a wide range of physical and biological problems such as in infectious disease modeling [149], nonlinear optics [71], or vegetation models [84].

Data availability statement

The data cannot be made publicly available upon publication because they are not available in a format that is sufficiently accessible or reusable by other researchers. The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgment

Appendix A. Proof of theorem 2.1

Proof. Step 1. Let us remark that for any n

$$\Pi_n G_k \left(e^{(s-t)A_{\mathfrak{c}}} X + z_{\mathfrak{c}} \right) = \sum_{(j_1, \cdots, j_k) \in I^k} \prod_{\ell=1}^{k} \left(\exp\left((s-t)\lambda_{j_\ell}\right) X_{j_\ell} + z_{j_\ell} \right) G_{j_1 \cdots j_k}^n,$$

where $I = \{1, \cdots, m_c\}$ and $G_{j_1 \cdots j_k}^n = \langle G_k(e_{j_1}, \cdots, e_{j_k}), e_n^* \rangle$.

Note that the product term above can be re-written as

$$\prod_{\ell=1}^{k} \left(\exp\left((s-t)\lambda_{j_\ell}\right) X_{j_\ell} + z_{j_\ell} \right) = \sum_{K_z} \left(\prod_{p \in K \setminus K_z} \exp\left((s-t)\lambda_{j_p}\right) X_{j_p} \right) \left(\prod_{q \in K_z} z_{j_q} \right), \quad \text{(A.1)}$$

where $K = \{1, \cdots, k\}$ and K_z runs through the subsets (possibly empty) of disjoint elements of K.

Thus,

$$
\begin{aligned}
&\mathfrak{J}(t, X, \omega) \\
&= \sum_{n \geqslant m_c + 1} \left(\int_{-\infty}^{t} e^{(t-s)\lambda_n} \Pi_n G_k \left(e^{(s-t)A_{\mathfrak{c}}} X + z_{\mathfrak{c}}(s, \omega) \right) ds \right) e_n \\
&= \sum_{n \geqslant m_c + 1} \sum_{(j_1, \cdots, j_k) \in I^k} \left(\int_{-\infty}^{t} \prod_{\ell=1}^{k} \left(\exp\left((s-t)\lambda_{j_\ell}\right) X_{j_\ell} + z_{j_\ell} \right) \exp\left((t-s)\lambda_n\right) ds \right) G_{j_1 \cdots j_k}^n e_n,
\end{aligned}
$$

and by using (A.1), $\mathfrak{J}(t, X, \omega)$ exists if and only if for each $(j_1, \cdots, j_k) \in I^k$ with $G_{j_1 \cdots j_k}^n \neq 0$, the integrals

$$I_{K_z}(t) = \int_{-\infty}^{t} \left(\prod_{q \in K_z} z_q(s, \omega) \right) \exp\left((t-s) \left[\lambda_n - \sum_{p \in K \setminus K_z} \lambda_{j_p} \right] \right) ds, \quad \text{(A.2)}$$

exist as K_z runs through the subsets (possibly empty) of disjoint elements of $K = \{1, \cdots, k\}$.

Since the OU process $z_q(s,\omega)$ has sublinear growth (see e.g. [42, lemma 3.1]), that is $\lim_{s\to\pm\infty}\frac{z_q(s,\omega)}{s}=0$ for all ω, one can readily check that the integral I_{K_z} defined in (A.2) is finite for all t provided that

$$\mathrm{Re}\left(\lambda_n - \sum_{p\in K\setminus K_z}\lambda_{j_p}\right) < 0. \tag{A.3}$$

Thus, the Lyapunov–Perron integral $\mathfrak{J}$ defined in equation (17) is well-defined if the non-resonance condition (18) holds.

Step 2. Thus $\mathfrak{J}$ is well defined under the condition (18) of the Main Text. By direct calculation, we get

$$\frac{\partial\mathfrak{J}(t,X,\omega)}{\partial t} = \Pi_{\mathfrak{f}}G_k\left(X+z_{\mathfrak{c}}\left(\theta_t\omega\right)\right) - A_{\mathfrak{f}}\mathfrak{J}(t,X,\omega)$$
$$- \int_{-\infty}^t \mathrm{e}^{(t-s)A_{\mathfrak{f}}}\Pi_{\mathfrak{f}}DG_k\left(\mathrm{e}^{(s-t)A_{\mathfrak{c}}}X+z_{\mathfrak{c}}\left(\theta_s\omega\right)\right)\mathrm{e}^{(s-t)A_{\mathfrak{c}}}A_{\mathfrak{c}}X\mathrm{d}s, \tag{A.4}$$

and that

$$D_X\mathfrak{J}(t,X,\omega)A_{\mathfrak{c}}X = \int_{-\infty}^t \mathrm{e}^{(t-s)A_{\mathfrak{f}}}\Pi_{\mathfrak{f}}DG_k\left(\mathrm{e}^{(s-t)A_{\mathfrak{c}}}X+z_{\mathfrak{c}}\left(\theta_s\omega\right)\right)\mathrm{e}^{(s-t)A_{\mathfrak{c}}}A_{\mathfrak{c}}X\mathrm{d}s,$$

where DG_k denote the Fréchet derivative of G_k. As a result, $\mathfrak{J}(t,X,\omega)$ solves (19).

Step 3. The solution to equations (20*a*)–(20*c*) is given explicitly

$$p(s) = \mathrm{e}^{(s-t)A_{\mathfrak{c}}}X, \tag{A.5}$$

$$q(s,\omega) = \int_{t-\tau}^s \mathrm{e}^{(s-s')A_{\mathfrak{f}}}\Pi_{\mathfrak{f}}G_k\left(\mathrm{e}^{(s'-s)A_{\mathfrak{c}}}X+z_{\mathfrak{c}}\left(s',\omega\right)\right)\mathrm{d}s'. \tag{A.6}$$

In particular, the value of $q(s,\omega)$ at $s=t$, denoted by $q_\tau(t,X,\omega)$, is given by

$$q_\tau(t,X,\omega) = \int_{t-\tau}^t \mathrm{e}^{(t-s')A_{\mathfrak{f}}}\Pi_{\mathfrak{f}}G_k\left(\mathrm{e}^{(s'-t)A_{\mathfrak{c}}}X+z_{\mathfrak{c}}\left(s',\omega\right)\right)\mathrm{d}s'.$$

We get then

$$\mathfrak{J}(t,X,\omega) - q_\tau(t,X,\omega) = \int_{-\infty}^{t-\tau} \mathrm{e}^{(t-s)A_{\mathfrak{f}}}\Pi_{\mathfrak{f}}G_k\left(\mathrm{e}^{(s-t)A_{\mathfrak{c}}}X+z_{\mathfrak{c}}\left(s,\omega\right)\right)\mathrm{d}s. \tag{A.7}$$

The boundedness of $\mathfrak{J}$ implies that the $H_{\mathfrak{f}}$-norm of the integral on the RHS of (A.7) converges to zero as τ goes to infinity, and (21) follows. □

Appendix B. Solving the sACE and its optimal reduced system

For the numerical integration of the sACE (76), we adopt a semi-implicit Euler method, where the nonlinearity $-u^3$ and the noise term $\mathrm{d}W_t$ are treated explicitly, and the linear term $\partial_x^2 u + u$ is treated implicitly; the spatial discretization is handled with a pseudo-spectral method; see [43, section 6.1] for a detailed description of this scheme applied to a stochastic Burgers-type equation. We have set the time-step size to be $\delta t = 10^{-2}$ and the spatial resolution to be $\delta x = L/201$.

The optimal reduced system (equation (85)) is also integrated with a semi-implicit Euler method where the linear term is treated implicitly and the nonlinear terms explicitly, still

with $\delta t = 10^{-2}$. The ODEs with random coefficients such as (49), used to simulate the non-Markovian coefficients, $Z_{\tau_n^*}^n$, involved in the computation of $\Phi_n(\tau_n^*, X, t)$, are simply integrated by a forward Euler scheme, for the same time-step $\delta t = 10^{-2}$. The initial condition in equation (52) for integrating the RDE, as defined in equation (49), is determined using the trapezoidal rule.

Appendix C. Computations of the double-fold bifurcation diagram

To compute the double-fold bifurcation diagram shown in figure 11 for equations (92*a*) and (92*b*), we adopt a Fourier sine series expansion of the solutions, in which *u* solving equations (92*a*) and (92*b*) is approximated by $u_N(x) = \sum_{n=1}^N a_n \sqrt{2/L} \sin(n\pi x/L)$. By doing so, the elliptic boundary value problem equations (92*a*) and (92*b*) is reduced to a nonlinear algebraic system for the coefficients $\boldsymbol{a}_N = (a_1, \ldots, a_N)^T$, which is solved by the Matlab built-in solver `fsolve`. The results shown in figure 11 are obtained from a six-dimensional algebraic system (i.e. with six sine modes, $N = 6$), which turns out to provide high-precision approximations.

We use a simple continuation approach to compute the bifurcation diagram based on this algebraic system, benefiting from the knowledge that the solution curve is S-shaped. Starting from $\lambda = 1$ with a random initial guess, an increment of $\Delta\lambda = 10^{-3}$ is used to compute the lower branch $\lambda \mapsto U_\lambda^m$ of solutions to equations (92*a*) and (92*b*) (lower black curve) with minimal energy. In our continuation procedure, the initial guess for $U_{\lambda+\Delta\lambda}^m$ is taken to be U_λ^m computed at the previous step. As λ is further increased, the procedure is stopped when the associated Jacobian becomes nearly singular which indicates that λ is approaching the lower turning point of the S-shaped curve.

The location of this turning point is estimated to be $\lambda^* \approx 1.3309$ as obtained by using finer mesh as one gets close to singularity. We then select a λ-value to be just a few $\Delta\lambda$-increments below λ^* (to ensure multiplicity of solutions) and pick as many different random initial guesses as needed to get the two other solutions: the unstable one, U_λ^*, and the stable one, U_λ^M, with maximal energy. From each of these newly computed solutions, we then adopt the same continuation method to trace out the upper branch $\lambda \mapsto U_\lambda^M$ made of solutions with maximal energy (upper black curve) and the 'in between' branch $\lambda \mapsto U_\lambda^*$ made of unstable solutions (red curve).

Appendix D. Numerical solutions to the jump-driven model and its optimal closure

The numerical integration of the stochastic equation (99) is performed in the same way as for the sACE outlined in appendix B using a semi-implicit Euler method. Here also the nonlinearity and the noise term are treated explicitly, while the linear term is treated implicitly. The spatial discretization is handled with a pseudo-spectral method; see again [43, section 6.1]. We have set the time-step size to be $\delta t = 10^{-2}$ and the spatial resolution to be $\delta x = L/257$. The initial data is set to be $u_0 = 0.5\boldsymbol{e}_1$ for the results shown in figures 12 and 15.

The optimal reduced system (equation (104)) is integrated with a forward Euler method still with $\delta t = 10^{-2}$. The initial condition $J_{\tau_n^*}^n(t_0; \omega)$ (cf (102)) for integrating (103) is computed using the trapezoidal rule, and $t_0 = \max\{\tau_3^*, \tau_5^*\}$. The initial data for equation (104) is set to be the projection of the SPDE solution onto $\boldsymbol{e}_1$ at time $t = t_0$.

ORCID iDs

Mickaël D Chekroun ● https://orcid.org/0000-0002-4525-5141
Honghu Liu ● https://orcid.org/0000-0002-9226-0744
James C McWilliams ● https://orcid.org/0000-0002-1237-5008

References

[1] Ahlers G, Meyer C W and Cannell D S 1989 Deterministic and stochastic effects near the convective onset *J. Stat. Phys.* **54** 1121–31
[2] Allen S M and Cahn J W 1979 A microscopic theory for antiphase boundary motion and its application to antiphase domain coarsening *Acta Metall.* **27** 1085–95
[3] Applebaum D 2009 *Levy Processes and Stochastic Calculus* (*Cambridge Studies in Advanced Mathematics* vol 116)(Cambridge University Press)
[4] Applebaum D 2019 *Semigroups of Linear Operators: With Applications to Analysis, Probability and Physics* (*London Mathematical Society Student Texts*) (Cambridge University Press)
[5] Arnold L 1998 *Random Dynamical Systems* (Springer)
[6] Bai F, Spence A and Stuart A M 1993 The numerical computation of heteroclinic connections in systems of gradient partial differential equations *SIAM J. Appl. Math.* **53** 743–69
[7] Baurmann M, Gross T and Feudel U 2007 Instabilities in spatially extended predator–prey systems: spatio-temporal patterns in the neighborhood of Turing–Hopf bifurcations *J. Theor. Biol.* **245** 220–9
[8] Bebernes J and Eberly D 2013 *Mathematical Problems From Combustion Theory* (*Applied Mathematical Sciences* vol 83) (Springer)
[9] Bena I 2006 Dichotomous Markov noise: exact results for out-of-equilibrium systems *Int. J. Mod. Phys.* B **20** 2825–88
[10] Bensoussan A and Flandoli F 1995 Stochastic inertial manifold *Stochastics Stochastics Rep.* **53** 13–39
[11] Berglund N and Gentz B 2013 Sharp estimates for metastable lifetimes in parabolic SPDEs: Kramers' law and beyond *Electron. J. Probab.* **18** 1–58
[12] Berloff P S and McWilliams J C 1999 Large-scale, low-frequency variability in wind-driven ocean gyres *J. Phys. Oceanogr.* **29** 1925–49
[13] Berner J *et al* 2017 Stochastic parameterization: toward a new view of weather and climate models *Bull. Am. Met. Soc.* **98** 565–88
[14] Blömker D 2007 *Amplitude Equations for Stochastic Partial Differential Equations* (*Interdisciplinary Mathematical Science* vol 3) (World Scientific Publishing Co. Pte. Ltd)
[15] Blömker D and Fu H 2020 The impact of multiplicative noise in SPDEs close to bifurcation via amplitude equations *Nonlinearity* **33** 3905
[16] Blömker D and Hairer M 2005 Amplitude equations for SPDEs: approximate centre manifolds and invariant measures *Probability and Partial Differential Equations in Modern Applied Mathematics* (*IMA Vol. Math. Appl.*) vol 140 (Springer)
[17] Blömker D, Hairer M and Pavliotis G A 2007 Multiscale analysis for stochastic partial differential equations with quadratic nonlinearities *Nonlinearity* **20** 1721–44
[18] Blömker D and Mohammed W W 2013 Amplitude equations for SPDEs with cubic nonlinearities *Stochastics Int. J. Probab. Stoch. Process.* **85** 181–215
[19] Boers N 2021 Observation-based early-warning signals for a collapse of the atlantic meridional overturning circulation *Nat. Clim. Change* **11** 680–8
[20] Boers N and Rypdal M 2021 Critical slowing down suggests that the western greenland ice sheet is close to a tipping point *Proc. Natl Acad. Sci.* **118** e2024192118
[21] Brand H, Kai S and Wakabayashi S 1985 External noise can suppress the onset of spatial turbulence *Phys. Rev. Lett.* **54** 555
[22] Bray A 2002 Theory of phase-ordering kinetics *Adv. Phys.* **51** 481–587
[23] Brézis H 2010 *Functional Analysis, Sobolev Spaces and Partial Differential Equations* (Springer)
[24] Cáceres M O and Fuentes M A 1999 Stochastic escape processes from a non-symmetric potential normal form III: extended explosive systems *J. Phys. A: Math. Gen.* **32** 3209

[25] Cáceres M O and Fuentes M A 2015 First-passage times for pattern formation in nonlocal partial differential equations *Phys. Rev.* E **92** 042122

[26] Cahn J and Hilliard J 1958 Free energy of a nonuniform system. I. Interfacial free energy *J. Chem. Phys.* **28** 258–67

[27] Caraballo T, Crauel H, Langa J A and Robinson J C 2007 The effect of noise on the Chafee-Infante equation: a nonlinear case study *Proc. Am. Math. Soc.* **135** 373–82

[28] Cérou F and Guyader A 2007 Adaptive multilevel splitting for rare event analysis *Stoch. Anal. Appl.* **25** 417–43

[29] Chafee N and Infante E F 1974 A bifurcation problem for a nonlinear partial differential equation of parabolic type *Appl. Anal.* **4** 17–37

[30] Chandrasekhar S 1957 *An Introduction to the Study of Stellar Structure* (Dover Publications)

[31] Chekroun M D, Dijkstra H, Sengül T and Wang S 2022 Transitions of zonal flows in a two-layer quasi-geostrophic ocean model *Nonlinear Dyn.* **109** 1887–904

[32] Chekroun M D, Ghil M, Liu H and Wang S 2016 Low-dimensional Galerkin approximations of nonlinear delay differential equations *Discrete Contin. Dyn. Syst.* A **36** 4133–77

[33] Chekroun M D, Koren I, Liu H and Liu H 2022 Generic generation of noise-driven chaos in stochastic time delay systems: Bridging the gap with high-end simulations *Sci. Adv.* **8** eabq7137

[34] Chekroun M D, Lamb J S W, Pangerl C J and Rasmussen M 2023 A Girsanov approach to slow parameterizing manifolds in the presence of noise (arXiv:1903.08598)

[35] Chekroun M D and Liu H 2015 Finite-horizon parameterizing manifolds and applications to sub-optimal control of nonlinear parabolic PDEs *Acta Appl. Math.* **135** 81–144

[36] Chekroun M D and Liu H 2024 Effective reduced models from delay differential equations: Bifurcations, tipping solution paths and ENSO variability *Physica* D **460** 134058

[37] Chekroun M D, Liu H and McWilliams J C 2017 The emergence of fast oscillations in a reduced primitive equation model and its implications for closure theories *Comput. Fluids* **151** 3–22

[38] Chekroun M D, Liu H and McWilliams J C 2021 Stochastic rectification of fast oscillations on slow manifold closures *Proc. Natl Acad. Sci.* **118** e2113650118

[39] Chekroun M D, Liu H and McWilliams J C 2023 Optimal parameterizing manifolds for anticipating tipping points and higher-order critical transitions *Chaos* **33** 093126

[40] Chekroun M D, Liu H, McWilliams J C and Wang S 2023 Transitions in stochastic non-equilibrium systems: efficient reduction and analysis *J. Differ. Equ.* **346** 145–204

[41] Chekroun M D, Liu H and McWilliams J 2020 Variational approach to closure of nonlinear dynamical systems: Autonomous case *J. Stat. Phys.* **179** 1073–160

[42] Chekroun M D, Liu H and Wang S 2015 *Approximation of Stochastic Invariant Manifolds: Stochastic Manifolds for Nonlinear Spdes I* (*Springer Briefs in Mathematics*) (Springer)

[43] Chekroun M D, Liu H and Wang S 2015 *Stochastic Parameterizing Manifolds and Non-Markovian Reduced Equations: Stochastic Manifolds for Nonlinear Spdes II (Springer Briefs in Mathematics)* (Springer)

[44] Chekroun M D, Park E and Temam R 2016 The Stampacchia maximum principle for stochastic partial differential equations and applications *J. Differ. Equ.* **260** 2926–72

[45] Chekroun M D, Tantet A, Dijkstra H and Neelin J D 2020 Ruelle-Pollicott resonances of stochastic systems in reduced state space. Part I: theory *J. Stat. Phys.* **179** 1366–402

[46] Chen N and Majda A J 2017 Simple stochastic dynamical models capturing the statistical diversity of El Ni no Southern Oscillation *Proc. Natl Acad. Sci.* **114** 1468–73

[47] Chen N, Mou C, Smith L M and Zhang Y 2024 A stochastic precipitating quasi-geostrophic model (arXiv:2407.20886)

[48] Chorin A J, Hald O H and Kupferman R 2002 Optimal prediction with memory *Physica* D **166** 239–57

[49] Chueshov I 2002 *Monotone Random Systems Theory and Applications* (*Lecture Notes in Mathematics* vol 1779) (Springer)

[50] Cont R and Tankov P 2003 *Financial Modelling With Jump Processes* (Chapman and Hall/CRC)

[51] Corwin I 2012 The Kardar–Parisi–Zhang equation and universality class *Random Matrices Theory Appl.* **1** 1130001

[52] Coullet P H, Elphick C and Tirapegui E 1985 Normal form of a Hopf bifurcation with noise *Phys. Lett.* A **111** 277–82

[53] Courrège P 1965 Sur la forme intégro-différentielle des opérateurs de C_k^∞ dans C satisfaisant au principe du maximum *Sém. Brelot-Choquet-Deny. Théor. Poten.* **10** 1–38

[54] Crauel H 1991 Markov measures for random dynamical systems *Stochastics Int. J. Probab. Stoch. Process.* **37** 153–73

[55] Crauel H 2002 *Random Probability Measures on Polish Spaces* (*The Stochastic Monographs: Theory and Applications of Stochastic Process Series* vol 11) (Taylor and Francis)

[56] Crauel H 2008 Measure attractors and Markov attractors *Dyn. Syst.* **23** 75–107

[57] Crauel H and Flandoli F 1994 Attractors for random dynamical systems *Probab. Theory Relat. Fields* **100** 365–93

[58] Crawford J 1991 Introduction to bifurcation theory *Rev. Mod. Phys.* **63** 991

[59] Crawford J and Knobloch E 1991 Symmetry and symmetry-breaking bifurcations in fluid dynamics *Ann. Rev. Fluid Mech.* **23** 341–87

[60] Cross M and Hohenberg P 1993 Pattern formation outside of equilibrium *Rev. Mod. Phys.* **65** 851

[61] Da Prato G and Debussche A 1996 Construction of stochastic inertial manifolds using backward integration *Stoch. Stoch. Rep.* **59** 305–24

[62] Da Prato G and Debussche A 1996 Stochastic Cahn-Hilliard equation *Nonlinear Anal. Theory Methods Appl.* **26** 241–63

[63] Da Prato G and Zabczyk J 1996 *Ergodicity for Infinite Dimensional Systems* (Cambridge University Press)

[64] Da Prato G and Zabczyk J 2008 *Stochastic Equations in Infinite Dimensions* (*Encyclopedia of Mathematics and its Applications* vol 44) (Cambridge University Press)

[65] Da Prato G 2004 *Kolmogorov Equations for Stochastic PDEs* (Springer)

[66] Debussche A, Högele M and Imkeller P 2013 *The Dynamics of Nonlinear Reaction-Diffusion Equations With Small Lévy Noise* (*Lecture Notes in Mathematics* vol 2085) (Springer)

[67] Dijkstra H and Moleamaker M 1997 Symmetry breaking and overturning oscillations in thermohaline-driven flows *J. Fluid Mech.* **331** 169–98

[68] Ditlevsen P 1999 Observation of α-stable noise induced millennial climate changes from an ice-core record *Geophys. Res. Lett.* **26** 1441–4

[69] Eisenman I, Yu L and Tziperman E 2005 Westerly wind bursts: ENSO's tail rather than the dog? *J. Clim.* **18** 5224–38

[70] Engel K-J and Nagel R 2000 *One-Parameter Semigroups for Linear Evolution Equations* (*Graduate Texts in Mathematics* vol 194) (Springer)

[71] Fernandez-Oto C, Clerc M, Escaff D and Tlidi M 2013 Strong nonlocal coupling stabilizes localized structures: an analysis based on front dynamics *Phys. Rev. Lett.* **110** 174101

[72] Flandoli F, Gess B and Scheutzow M 2017 Synchronization by noise for order-preserving random dynamical systems *Ann. Probab.* **45** 1325–50

[73] Forgoston E and Schwartz I B 2009 Escape rates in a stochastic environment with multiple scales *SIAM J. Appl. Dyn. Syst.* **8** 1190–217

[74] Fowler R H 1931 Further studies of Emden's and similar differential equations *Q. J. Math.* **2** 259–88

[75] Fraedrich K, Blender R and Zhu X 2009 Continuum climate variability: long-term memory, scaling and 1/f-noise *Int. J. Mod. Phys.* B **23** 5403–16

[76] Frank-Kamenetskii D A 2015 *Diffusion and Heat Exchange in Chemical Kinetics* (Princeton University Press)

[77] Franzke C L E, O'Kane T J, Berner J, Williams P D and Lucarini V 2015 Stochastic climate theory and modeling *Wiley Interdiscip. Rev.- Clim. Change* **6** 63–78

[78] Freidlin M I and Wentzell A D 2012 *Random Perturbations of Dynamical Systems* (*Grundlehren der Mathematischen Wissenschaften* vol 260) 3rd edn (Springer) Translated from the 1979 Russian original by Joseph Szücs

[79] Fuentes M, Kuperman M and Kenkre V 2003 Nonlocal interaction effects on pattern formation in population dynamics *Phys. Rev. Lett.* **91** 158104

[80] Gálfi V M, Lucarini V, Ragone F and Wouters J 2021 Applications of large deviation theory in geophysical fluid dynamics and climate science *La Rivista del Nuovo Cimento* **44** 291–363

[81] García-Ojalvo J, Hernández-Machado A and Sancho J M 1993 Effects of external noise on the Swift-Hohenberg equation *Phys. Rev. Lett.* **71** 1542

[82] Garrido-Atienza M J, Lu K and Schmalfuß B 2010 Unstable invariant manifolds for stochastic PDEs driven by a fractional Brownian motion *J. Differ. Equ.* **248** 1637–67

[83] Ghil M 1976 Climate stability for a Sellers-type model *J. Atmos. Sci.* **33** 3–20

[84] Gilad E, von Hardenberg J, Provenzale A, Shachak M and Meron E 2004 Ecosystem engineers: from pattern formation to habitat creation *Phys. Rev. Lett.* **93** 098105

J. Phys. A: Math. Theor. **58** (2025) 045204

M D Chekroun *et al*

[85] Ginzburg V and Landau L 1950 On the theory of superconductivity *Zh. Eksp. Teor. Fiz.* **20** 1064–82

[86] Givon D, Kupferman R and Stuart A 2004 Extracting macroscopic dynamics: model problems and algorithms *Nonlinearity* **17** R55–R127

[87] Goldys B and Maslowski B 2005 Exponential ergodicity for stochastic Burgers and 2D Navier–Stokes equations *J. Funct. Anal.* **226** 230–55

[88] Grad H and Rubin H 1958 Hydromagnetic equilibria and force-free fields *J. Nucl. Energy* **7** 284–5

[89] Grafke T, Grauer R and Schäfer T 2015 The instanton method and its numerical implementation in fluid mechanics *J. Phys. A: Math. Theor.* **48** 333001

[90] Hairer M 2002 Exponential mixing properties of stochastic PDEs through asymptotic coupling *Probab. Theory Relat. Fields* **124** 345–80

[91] Hairer M and Mattingly J 2006 Ergodicity of the 2D Navier-Stokes equations with degenerate stochastic forcing *Ann. Math.* **164** 993–1032

[92] Hairer M and Mattingly J 2011 A theory of hypoellipticity and unique ergodicity for semilinear stochastic PDEs *Electron. J. Probab.* **16** 658–738

[93] Hairer M, Mattingly J and Scheutzow M 2011 Asymptotic coupling and a general form of Harris theorem with applications to stochastic delay equations *Probab. Theory Relat. Fields* **149** 223–59

[94] Hairer M and Ohashi A 2007 Ergodic theory for SDEs with extrinsic memory *Ann. Probab.* **35** 1950–77

[95] Hasselmann K 1976 Stochastic climate models. Part I. Theory *Tellus* **28** 473–85

[96] Henry D 1981 *Geometric Theory of Semilinear Parabolic Equations* (*Lecture Notes in Mathematics* vol 840) (Springer)

[97] Hetzer G 1997 S-shapedness for energy balance climate models of Sellers-type *The Mathematics of Models for Climatology and Environment* (Springer) pp 253–87

[98] Hohenberg P C and Halperin B I 1977 Theory of dynamic critical phenomena *Rev. Mod. Phys.* **49** 435

[99] Hopf E 1931 On Emden's differential equation *Mon. Not. R. Astron. Soc.* **91** 653–63

[100] Horenko I 2011 Nonstationarity in multifactor models of discrete jump processes, memory and application to cloud modeling *J. Atmos. Sci.* **68** 1493–506

[101] Horsthemke W and Lefever R 1984 *Noise-Induced Transitions in Physics, Chemistry and Biology* (Springer)

[102] Izhikevich E M 2010 *Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting* (MIT Press)

[103] Jetschke G 1986 On the equivalence of different approaches to stochastic partial differential equations *Math. Nachr.* **128** 315–29

[104] Kai S, Fukunaga H and Brand H R 1989 Structure changes induced by external multiplicative noise in the electrohydrodynamic instability of nematic liquid crystals *J. Stat. Phys.* **54** 1133–52

[105] Khouider B, Majda A and Katsoulakis M 2003 Coarse-grained stochastic models for tropical convection and climate *Proc. Natl Acad. Sci. USA* **100** 11941–6

[106] Klepel K, Blömker D and Mohammed W W 2014 Amplitude equation for the generalized Swift–Hohenberg equation with noise *Z. Angew. Math. Phys.* **65** 1107–26

[107] Korman P and Li Y 1999 On the exactness of an S-shaped bifurcation curve *Proc. Am. Math. Soc.* **127** 1011–20

[108] Kühn F 2017 *Lévy Matters. VI* (*Lecture Notes in Mathematics*) vol 2187

[109] Kühn F 2018 Solutions of Lévy-driven SDEs with unbounded coefficients as Feller processes *Proc. Am. Math. Soc.* **146** 3591–604

[110] Kuksin S and Shirikyan A 2012 *Mathematics of Two-Dimensional Turbulence* (*Cambridge Tracts in Mathematics* vol 194) (Cambridge University Press)

[111] Kwaśnicki M 2017 Ten equivalent definitions of the fractional Laplace operator *Fract. Calc. Appl. Anal.* **20** 7–51

[112] Lenton T M, Rockström J, Gaffney O, Rahmstorf S, Richardson K, Steffen W and Schellnhuber H J 2019 Climate tipping points—too risky to bet against *Nature* **575** 592–5

[113] Lenton T M, Held H, Kriegler E, Hall J W, Lucht W, Rahmstorf S and Schellnhuber H J 2008 Tipping elements in the earth's climate system *Proc. Natl Acad. Sci.* **105** 1786–93

[114] Lin J W-B and Neelin J D 2000 Influence of a stochastic moist convective parameterization on tropical climate variability *Geophys. Res. Lett.* **27** 3691–4

[115] Liu X 2021 Random invariant manifolds of stochastic evolution equations driven by Gaussian and non-Gaussian noises *J. Math. Phys.* **62** 112702
[116] Lord G J, Powell C E and Shardlow T 2014 *An Introduction to Computational Stochastic PDEs* (Cambridge University Press)
[117] Lucarini V and Chekroun M D 2023 Theoretical tools for understanding the climate crisis from Hasselmann's programme and beyond *Nat. Rev. Phys.* **5** 744–65
[118] Lucarini V and Chekroun M D 2024 Detecting and attributing change in climate and complex systems: Foundations, Green's functions and nonlinear fingerprints *Phys. Rev. Lett.* **133** 224201
[119] Ma T and Wang S 2005 *Bifurcation Theory and Applications* (*World Scientific Series on Nonlinear Science. Series A: Monographs and Treatises* vol 53) (World Scientific Publishing Co. Pte. Ltd)
[120] Ma T and Wang S 2019 *Phase Transition Dynamics* 2nd edn (Springer)
[121] Majda A J, Timofeyev I and Vanden-Eijnden E 2001 A mathematical framework for stochastic climate models *Commun. Pure Appl. Math.* **54** 891–974
[122] Mandelbrot B B and Van Ness J W 1968 Fractional Brownian motions, fractional noises and applications *SIAM Rev.* **10** 422–37
[123] Margazoglou G, Grafke T, Laio A and Lucarini V 2021 Dynamical landscape and multistability of a climate model *Proc. R. Soc.* A **477** 20210019
[124] McKay D I A, Staal A, Abrams J F, Winkelmann R, Sakschewski B, Loriani S, Fetzer I, Cornell S E, Rockström J and Lenton T M 2022 Exceeding 1.5 °C global warming could trigger multiple climate tipping points *Science* **377** eabn7950
[125] McWilliams J 2019 A survey of submesoscale currents *Geosci. Lett.* **6** 1–15
[126] Meron E 2018 From patterns to function in living systems: dryland ecosystems as a case study *Annu. Rev. Condens. Matter Phys.* **9** 79–103
[127] Moberg A, Sonechkin D M, Holmgren K, Datsenko N M and Karlén W 2005 Highly variable Northern Hemisphere temperatures reconstructed from low-and high-resolution proxy data *Nature* **433** 613–7
[128] Mohammed W W, Blömker D and Klepel K 2014 Multi-scale analysis of SPDEs with degenerate additive noise *J. Evol. Equ.* **14** 273–98
[129] Murray J 2003 *Mathematical Biology: II: Spatial Models and Biomedical Applications* vol 3 (Springer)
[130] Namachchivaya N S and Lin Y K 1991 Method of stochastic normal forms *Int. J. Non-linear Mech.* **26** 931–43
[131] North G R, Cahalan R F and Coakley J A Jr 1981 Energy balance climate models *Rev. Geophys.* **19** 91–121
[132] Øksendal B 2013 *Stochastic Differential Equations: An Introduction With Applications* (Springer)
[133] Øksendal B and Sulem A 2019 Stochastic control of jump diffusions *Applied Stochastic Control of Jump Diffusions* (Springer) pp 93–155
[134] Palmer T N 2001 A nonlinear dynamical perspective on model error: a proposal for non-local stochastic-dynamic parametrization in weather and climate prediction models *Q. J. R. Meteorol. Soc.* **127** 279–304
[135] Pazy A 1983 *Semigroups of Linear Operators and Applications to Partial Differential Equations* (*Applied Mathematical Sciences* vol 44) (Springer)
[136] Pearson J 1993 Complex patterns in a simple system *Science* **261** 189–92
[137] Penland C and Sardeshmukh P 2012 Alternative interpretations of power-law distributions found in nature *Chaos* **22** 023119
[138] Pereira M, Gissinger C and Fauve S 2019 1/f noise and long-term memory of coherent structures in a turbulent shear flow *Phys. Rev.* E **99** 023106
[139] Peszat S and Zabczyk J 2007 *Stochastic Partial Differential Equations With Lévy Noise: An Evolution equation Approach* (*Encyclopedia of Mathematics and Its Applications* vol 113) (Cambridge University Press)
[140] Pierini S 2010 Coherence resonance in a double-gyre model of the Kuroshio Extension *J. Phys. Oceanogr.* **40** 238–48
[141] Protter P E 2005 Stochastic integration and differential equations *Stochastic Modelling and Applied Probability* (Springer)
[142] Qu S and Gao H 2024 The high-order approximation of SPDEs with multiplicative noise via amplitude equations *Commun. Nonlinear Sci. Numer. Simul.* **132** 107937
[143] Ragone F, Wouters J and Bouchet F 2018 Computation of extreme heat waves in climate models using a large deviation algorithm *Proc. Natl Acad. Sci.* **115** 24–29

[144] Rehberg I, Rasenat S, De La Torre Juarez M, Schöpf W, Hörner F, Ahlers G and Brand H 1991 Thermally induced hydrodynamic fluctuations below the onset of electroconvection *Phys. Rev. Lett.* **67** 596

[145] Roberts A J 2006 Resolving the multitude of microscale interactions accurately models stochastic partial differential equations *LMS J. Comput. Math.* **9** 193–221

[146] Roberts A J 2003 A step towards holistic discretisation of stochastic partial differential equations *ANZIAM J.* **45** C1–C15

[147] Rolland J, Bouchet F and Simonnet E 2016 Computing transition rates for the 1-D stochastic Ginzburg-Landau-Allen-Cahn equation for finite-amplitude noise with a rare event algorithm *J. Stat. Phys.* **162** 277–311

[148] Roulston M S and Neelin J D 2000 The response of an ENSO model to climate noise, weather noise and intraseasonal forcing *Geophys. Res. Lett.* **27** 3723–6

[149] Ruan S 2007 Spatial-temporal dynamics in nonlocal epidemiological models *Mathematics for Life Science and Medicine* ed Y Takeuchi, Y Iwasa and K Sato (Springer) pp 97–122

[150] Rypdal M and Rypdal K 2014 Long-memory effects in linear response models of Earth's temperature and implications for future global warming *J. Clim.* **27** 5240–58

[151] Rypdal M and Rypdal K 2016 Late Quaternary temperature variability described as abrupt transitions on a 1/f noise background *Earth Syst. Dyn.* **7** 281–93

[152] Sagués F, Sancho J and García-Ojalvo J 2007 Spatiotemporal order out of noise *Rev. Mod. Phys.* **79** 829

[153] Sapsis T P and Dijkstra H A 2013 Interaction of additive noise and nonlinear dynamics in the double-gyre wind-driven ocean circulation *J. Phys. Oceanogr.* **43** 366–81

[154] Saravanan R and McWilliams J C 1998 Advective ocean–atmosphere interaction: an analytical stochastic model with implications for decadal variability *J. Clim.* **11** 165–88

[155] Schulz M and Mudelsee M 2002 REDFIT: estimating red-noise spectra directly from unevenly spaced paleoclimatic time series *Comput. Geosci.* **28** 421–6

[156] Serdukova L, Zheng Y, Duan J and Kurths J 2017 Metastability for discontinuous dynamical systems under Lévy noise: case study on Amazonian Vegetation *Sci. Rep.* **7** 9336

[157] Shafranov V 1958 On magnetohydrodynamical equilibrium configurations *Sov. J. Exp. Theor. Phys.* **6** 545–54

[158] Simonnet E, Rolland J and Bouchet F 2021 Multistability and rare spontaneous transitions in barotropic β-plane turbulence *J. Atmos. Sci.* **78** 1889–911

[159] Solomon T, Weeks E R and Swinney H L 1993 Observation of anomalous diffusion and Lévy flights in a two-dimensional rotating flow *Phys. Rev. Lett.* **71** 3975

[160] Stechmann S N and Neelin J D 2011 A stochastic model for the transition to strong convection *J. Atmos. Sci.* **68** 2955–70

[161] Stinis P 2006 A comparative study of two stochastic mode reduction methods *Physica* D **213** 197–213

[162] Sura P and Penland C 2002 Sensitivity of a double-gyre ocean model to details of stochastic forcing *Ocean Model.* **4** 327–45

[163] Tantet A, Chekroun M D, Dijkstra H and Neelin J D 2020 Ruelle-Pollicott resonances of stochastic systems in reduced state space. Part II: stochastic Hopf bifurcation *J. Stat. Phys.* **179** 1403–48

[164] Täuber U C 2014 *Critical Dynamics: a Field Theory Approach to Equilibrium and Non-Equilibrium Scaling Behavior* (Cambridge University Press)

[165] Thual O and Mcwilliams J 1992 The catastrophe structure of thermohaline convection in a two-dimensional fluid model and a comparison with low-order box models *Geophys. Astrophys. Fluid Dyn.* **64** 67–95

[166] Thual S, Majda A J, Chen N and Stechmann S N 2016 Simple stochastic model for El Nino with westerly wind bursts *Proc. Natl Acad. Sci.* **113** 10245–50

[167] Trefethen L N 2000 *Spectral Methods in Matlab* (SIAM)

[168] van Westen R M, Kliphuis M and Dijkstra H A 2024 Physics-based early warning signal shows that AMOC is on tipping course *Sci. Adv.* **10** eadk1189

[169] Vanden-Eijnden E and Westdickenberg M G 2008 Rare events in stochastic partial differential equations on large spatial domains *J. Stat. Phys.* **131** 1023–38

[170] Wang W and Roberts A 2013 Macroscopic reduction for stochastic reaction–diffusion equations *IMA J. Appl. Math.* **78** 1237–64

[171] Weinan E, Ren W and Vanden-Eijnden E 2004 Minimum action method for the study of rare events *Commun. Pure Appl. Math.* **57** 637–56

[172] Weijer W *et al* 2019 Stability of the atlantic meridional overturning circulation: a review and synthesis *J. of Geophys. Res. Oceans* **124** 5336–75

[173] Wunderling N, Donges J F, Kurths J and Winkelmann R 2021 Interacting tipping elements increase risk of climate domino effects under global warming *Earth Syst. Dyn.* **12** 601–19

[174] Yuan S, Hu J, Liu X and Duan J 2019 Slow manifolds for dynamical systems with non-Gaussian stable Lévy noise *Anal. Appl.* **17** 477–511

[175] Yuan S, Zeng Z and Duan J 2021 Stochastic bifurcation for two-time-scale dynamical system with α-stable Lévy noise *J. Stat. Mech.* **2021** 033204

[176] Yuan S and Bloemker D 2022 Modulation and amplitude equations on bounded domains for nonlinear SPDEs driven by cylindrical α-stable Lévy processes *SIAM J. Appl. Dyn. Syst.* **21** 1748–77

[177] Zakine R and Vanden-Eijnden E 2023 Minimum-action method for nonequilibrium phase transitions *Phys. Rev.* X **13** 041044

[178] Zelnik Y R, Meron E and Bel G 2015 Gradual regime shifts in fairy circles *Proc. Natl Acad. Sci.* **112** 12327–31

[179] Zheng Y, Yang F, Duan J, Sun X, Fu L and Kurths J 2020 The maximum likelihood climate change for global warming under the influence of greenhouse effect and Lévy noise *Chaos* **30** 13132