Licentiate Thesis in Theoretical Physics

A Serendipitous Journey through Stochastic Processes

Ludovico Theo Giorgini

Stockholm University

NORDITA
Nordic Institute for
Theoretical Physics

Department of Physics
Stockholm University
SE-106 91 Stockholm

Stockholm, Sweden, 2021

Abstract

English

In this licentiate thesis I will present some new insights in different problems in the field of stochastic processes.

A stochastic resonance system is studied using path integral techniques, originally developed in quantum field theory, to recover the optimal means through which noise self-organises before a rare transition from one potential well to the other. These results allow one to determine precursors to a rare events in such system.

I then study the survival probability of an autonomous Ornstein-Uhlenbeck process using the asymptotic matching techniques developed in fluid dynamics. Here, I obtain a simple analytical expression for this quantity that exhibits a good agreement with numerical determination.

Next, rare events in similar systems are studied using a recurrent neural network to model the noisy part of the signal. The neural network facilitates the prediction of future noise realisations and hence rare transitions.

Using a combination of analytical and numerical techniques a low-dimensional model is constructed and it is able to predict and to reproduce the main dynamical and equilibrium features of the El Niño and Southern Oscillation (ENSO), the largest inter-annual variability phenomenon in the tropical Pacific which has a global impact on climate.

Finally, using the results obtained for the survival probability of the Ornstein-Uhlenbeck process, an approximate analytical solution for the probability density function and the response is derived for a stochastic resonance system in the non-adiabatic limit.

Svenska

I denna licentiatavhandling kommer jag att presentera några nya insikter i olika problem relaterade till stokastiska processer.

Ett stokastiskt resonanssystem kommer att studeras med hjälp av banintegraltekniker, som ursprungligen utvecklats i kvantfältsteori, för att återställa det optimala sättet för bruset att självorganisera innan en sällsynt övergång från en potentialbrunn till en annan. Dessa resultat gör det möjligt att bestämma föregångare till en sällsynt händelse i ett sådant system.

Jag kommer sedan att studera överlevnadssannolikheten för en autonom Ornstein-Uhlenbeck-process med användning av asymptotiska matchningstekniker lånade från vätskedynamik som gör det möjligt att fåett enkelt analytiskt uttryck för denna kvantitet som uppvisar ett bra överensstämmelse med dess numeriska uppskattning.

Sällsynta händelser i liknande system kommer senare att studeras med ett återkommande neuralt nätverk för att modellera den bullriga delen av signalen. Neurala nätverket lyckas förutsäga framtida brusförverkligande vilket möjliggör förutspår sällsynt övergång vissa bullerkorrelationstider i framtiden.

Senare kommer en kombination av analytiska och numeriska tekniker att användas för att bygga en lågdimensionell modell som kan förutsäga och reproducera de viktigaste dynamiska och jämviktsfunktionerna i ENSO, det största variationen mellan olika år i det tropiska Stilla havet med hög inverkan påglobala klimatförutsägelser.

Slutligen, med hjälp av de resultat som erhållits för överlevnadssannolikheten för Ornstein-Uhlenbeck-processen, kommer en ungefärlig analytisk lösning för sannolikhetsdensitetsfunktionen och svaret att härledas för ett stokastiskt resonanssystem i den icke-adiabatiska gränsen.

List of Papers

The following papers are included in the thesis. They are referred to by their Roman numerals in the text.

I **L.T. Giorgini**, S.H. Lim, W. Moon, and J.S. Wettlaufer, Precursors to rare events in stochastic resonance, *Europh. Lett.* **129**, 40003 (2020) (Editor's choice).

II **L.T. Giorgini**, W. Moon, and J.S. Wettlaufer, Analytical Survival Analysis of the Ornstein-Uhlenbeck Process, *J. Stat. Phys.* **181**, 2404 (2020).

III S.H. Lim, **L.T. Giorgini**, W. Moon, and J.S. Wettlaufer, Predicting rare events in multiscale dynamical systems using machine learning, *Chaos* **30**, 123126 (2020).

IV **L.T. Giorgini**, W. Moon, N. Chen and J.S. Wettlaufer, A Non-Gaussian Stochastic Model for the El Niño Southern Oscillation, arXiv: 2012.09858 (2021).

V W. Moon, **L.T. Giorgini**, and J.S. Wettlaufer, Nonadiabatic stochastic resonance: An analytical solution, arXiv: 2104.03426 (2021).

The papers are listed in chronological order.

The first author who appears in each paper above carried out most of the analytical and numerical computations and he wrote the first draft of the manuscript. The other coauthors contributed to the development of the research, they checked and corrected the analytical and numerical computations and contributed to the final writing of the paper. John Wettlaufer gave a remarkable contribution to the writing of the final version of each paper, in particular when I was the first author.

The following paper is not included in my thesis, but was published during my PhD.

VI **L.T. Giorgini**, U.D. Jentschura, E.M. Malatesta, G. Parisi, T. Rizzo, and J. Zinn-Justin, Two-loop corrections to the large-order behavior of correlation functions in the one-dimensional N-vector model, *Phys. Rev. D* ***101**, 125001* (2020).

Contents

PART I THEORETICAL BACKGROUND

PART II PAPERS

Chapter 1

Introduction

A stochastic process refers to a system that evolves in time under the influence of chance fluctuations. The random nature of such a process makes impossible the exact prediction of its future state and a probabilistic approach is needed. Randomness represents the complex processes that we have incomplete access to, but nonetheless control the fate of our system. Complexity, unfortunately, goes hand in hand with reality, since almost every natural or engineered system has inaccessible facets that prevent a purely deterministic description of them. For this reason, stochastic processes are ubiquitous non only in physics, but also in many other disciplines, such us biology, chemistry, computer science, engineering, finance, neuroscience and signal processing.

Because of the countless ways and settings in which stochastic processes can be found, a unique way to treat them does not exist. However, depending on the problem, different tools and techniques, often borrowed form other disciplines, are used. In this work I will address different problems that can all be connected to the general area of stochastic processes using different analytical and numerical tools, such as path integrals, a core approach in quantum field theory, asymptotic matching, originally developed in fluid dynamics, and the new area of machine learning.

Because most of the results that will be discussed in this work have already been published or submitted to a journal, I will avoid simply rewriting them here. All the papers on which I have worked during my PhD are attached at the end of this thesis.

Chapter 2

Methods

In this chapter I will describe the main mathematical tools that will be applied in different settings inside the papers attached at the end of my licentiate thesis. In the title of each section I will specify for which paper it is relevant for. If no paper will be specified, the section will be relevant to all papers.

2.1 From the Langevin equation to the Kolmogorov Forward and Backward Equations

In this work we will deal with non-equilibrium systems, that are systems where the temporal variable plays a fundamental role in their treatment. The simplest approximate way to treat the dynamics of non-equilibrium systems is probably the theory of Browninan motion. It was originally developed by observing the random motion of particles of size $\sim \mu m$ fluctuating inside a solution, but has been later extended to situations where the fluctuating object is not a real particle at all, but instead some collective porperty of a macroscopic system.

The fundamental equation governing the dynamics of the Brownian motion is called the Langevin equation

$$m\ddot{X}(t) = -\gamma \dot{X}(t) + R(t) \tag{2.1}$$

where m is the particle mass, $X(t)$ is the position of the particle and R is a random force describing the rapidly fluctuating interactions between the particle and the molecules of the solution. The relationship between the frictional and random forces in Eq. (2.1) is given by the fluctuation-dissipation theorem, which is obtained by solving the previous equation for $\dot{X}$ and taking the average of its square. For large values of t it becomes

$$\lim_{t\to\infty} \langle \dot{X}^2(t) \rangle = \frac{1}{2\gamma m} \int_{-\infty}^{\infty} ds\, e^{-\frac{\gamma}{m}s} \langle R(0)R(s) \rangle, \tag{2.2}$$

where the translational invariance of $\langle R(0)R(s)\rangle$ has been assumed. By invoking the equipartition theorem $\langle \dot{X}^2\rangle = kT/m$, with k the Boltzmann constant and T the thermodynamic temperature, and assuming for the random force a correlation time $\tau \ll \gamma/m$, Eq. (2.2) becomes

$$2\gamma kT = \int_{-\infty}^{\infty} ds\, \langle R(0)R(s)\rangle \tag{2.3}$$

which gives the searched relationship between T, γ and R.

In this work I will mostly deal with delta-correlated random forces with Gaussian distribution and I will consider overdamped Browninan motions, that is the case where the inertia effects $m\ddot{X}$ in Eq. (2.1) can be neglected, inside a potential $U(X)$. The Langevin equation to which I will restrict myself during this work becomes then

$$\dot{X}(t) = -\partial_X U(X) + \sqrt{2}b\xi(t) \tag{2.4}$$

where $\xi(t)$ is Gaussian white noise with correlation $\langle \xi(s)\xi(t)\rangle = \delta(t-s)$ and $\sqrt{2}b = \sqrt{2\gamma kT}$.

Associated with the one-dimensional diffusion process $X(t)$ given by Eq. (2.4) is the probability density function (PDF) $\rho(y,t|x,s)$ such that

$$\int_A dy(dx)\, \rho(y,t|x,s) = \mathcal{P}[X(t)(X(s)) \in A | X(s) = x(X(t) = y)] \quad \text{for} \quad t \geq s, \tag{2.5}$$

where the y-integral (x-integral) over the set A gives the probability of finding $X(t)(X(s))$ in the set A given the condition that it started as $X(s) = x$ (ended as $X(t) = y$). Depending on whether we are interested in the PDF at final or initial time, we consider boundary conditions $\rho(y = \pm\infty, t|x,s) = 0$ for the former case and $\rho(y,t|x = \pm\infty, s) = 0$ for the latter, with the decay to zero being sufficiently fast to ensure

$$\int_{-\infty}^{\infty} dy(dx)\, \rho(y,t|x,s) = 1, \tag{2.6}$$

which means $\lim_{|y|\to\infty} |y|\rho(y,t|x,s) = 0$ ($\lim_{|x|\to\infty} |x|\rho(y,t|x,s) = 0$). Moreover, if $t = s$, then $x = y$ and the following initial (final) condition must holds

$$\rho(y, s(t)|x, s(t)) = \delta(y-x). \tag{2.7}$$

The PDF of the final (initial) position of the particle at final (initial) time knowing its initial (final) position at initial (final) time is determined by solving the Kolmogorov forward (backward) equation. We start deriving the Kolmogorov forward equation (KFE) also known as Fokker-Plank equation.

Using the Chapman-Kolmogorov equation, we can write

$$\rho(z, t+\epsilon|x, s) = \int_{-\infty}^{\infty} dy\, \rho(z, t+\epsilon|y, t)\rho(y, t|x, s) \tag{2.8}$$

with $\epsilon \ll 1$. After an ϵ-expansion of the left term in Eq. (2.8) and using the fact that the noise is Gaussian distributed, we get

$$\rho(z, t|x, s) + \epsilon\partial_t(z, t|x, s) = \int_{-\infty}^{\infty} dy\, \frac{e^{(z-y+\partial_y U(y)\epsilon)^2/2b^2\epsilon}}{\sqrt{2\pi b^2\epsilon}}\rho(y, t|x, s). \tag{2.9}$$

Expanding in ϵ also the right term we can observe that the $O(1)$ term cancels with $\rho(z, t|x, s)$ on the left hand side. Considering only the $O(\epsilon)$ terms we finally get the KBE

$$\partial_t\rho(y, t) = \mathcal{L}_y\rho(y, t) = \partial_y[\partial_y U(y) + b^2\partial_{y,y}\rho(y, t)], \tag{2.10}$$

where the dependency of ρ on x, s has been dropped and $\mathcal{L}_y$ is the generator of the stochastic process viz., $[\mathcal{L}_y f](y, t) = \partial_y[\partial_y U(y)f(y, t)] + b^2\partial_{y,y}\rho(y, t)f(y, t)$.

The Kolmogorov backward equation (KBE) also known as Feynman-Kac equation, can be derived using the same method, but perturbing the initial time $s \to s - \epsilon$ instead of the final one. We get

$$-\partial_s\rho(x, s) = \mathcal{L}_x^*\rho(x, s) = \partial_x[-\partial_x U(x) + b^2\partial_{x,x}\rho(x, s)], \tag{2.11}$$

where the operator $\mathcal{L}_x^*$ is the adjoint of $\mathcal{L}_y$.

2.1.1 Analytical Solution of the Kolmogorov Equations for an Ornstein-Uhlenbeck Process

I report below the analytical solution of the forward and backward Kolmogorov equations for a non autonomous stochastic process, that is a system which does explicitly depend on the time. I will restrict myself to the non autonomous Ornstein-Uhlenbeck process, i.e. I will consider a potential $U(X, t) = \frac{1}{2}a(t)X^2$ and a time dependent diffusion coefficient $b(t)$. This particular process has been studied in many different settings during my PhD.

2.1.1.1 Kolmogorov Forward Equation

We start solving the KFE given in Eq. (2.10). We go in the momentum space through a Fourier transform

$$\partial_t\tilde{\rho}(k, t) + a(t)k(t)\partial_k\tilde{\rho}(k, t) = -b^2(t)k^2(t)\tilde{\rho}(k, t), \tag{2.12}$$

where we have defined

$$\tilde{\rho}(k,t) = \int e^{-iky}\rho(y,t)dy. \tag{2.13}$$

In order to solve the KFE in momentum space we use the method of charactristics. We write the right hand side of the KFE as a total derivative with respect to the time

$$\frac{d\tilde{\rho}(k,t)}{dt} = \partial_t\tilde{\rho}(k,t) + \frac{dk(t)}{dt}\partial_k\tilde{\rho}(k,t) = -b^2(t)k^2(t)\tilde{\rho}(k,t), \tag{2.14}$$

that implies

$$\frac{dk(t)}{dt} = a(t)k(t) \;\rightarrow\; k(t) = k(s)e^{\int_s^t a(r)\,dr}. \tag{2.15}$$

We can then write the KFE as a total derivative along each of the characteristic curves $k(t,s)$

$$\frac{d\tilde{\rho}(k,t)}{dt} = -b^2(t)k^2(s)e^{2\int_s^t a(r)\,dr}\tilde{\rho}(k,t). \tag{2.16}$$

The solution becomes

$$\begin{aligned}
\tilde{\rho}(k,t) =& \tilde{\rho}(k,s)\,\exp\left[-k^2(s)\int_s^t drb^2(r)e^{2\int_s^r a(u)\,du}\right] = \\
=& \exp\left[-ik(s)x - k^2(s)\int_s^t drb^2(r)e^{2\int_s^r a(u)\,du}\right] = \\
=& \exp\Bigg[-ik(t)xe^{-\int_s^t a(r)\,dr} - k^2(t)e^{-2\int_s^t a(r)\,dr} \\
&\times \int_s^t drb^2(r)e^{2\int_s^r a(u)\,du}\Bigg] = \\
=& \exp\left[-ik(t)xm_F(t,s) - \frac{\sigma_F^2(t,s)}{2}k^2(t)\right],
\end{aligned} \tag{2.17}$$

where we have defined

$$\begin{aligned}
m_F(t,s) &= e^{-\int_s^t a(r)\,dr}, \\
\sigma_F^2(t,s) &= 2e^{-2\int_s^t a(r)\,dr}\int_s^t drb^2(r)e^{2\int_s^r a(u)\,du}.
\end{aligned} \tag{2.18}$$

This is the Fourier transform of a Gaussian with mean $x\,m_F(t)$ and standard deviation $\sigma_F^2(t)$. Therefore, the solution of the KFE becomes

$$\rho(y,t;x,s) = \frac{1}{\sqrt{2\pi\sigma_F^2(t,s)}}\exp\left[-\frac{(y - xm_F(t,s))^2}{2\sigma_F^2(t,s)}\right]. \tag{2.19}$$

2.1.1.2 Kolmogorov Backward Equation

Now we will solve the KBE (2.11), i.e. the PDF governing the time evolution of the probability distribution backward in time. We can notice that if we differentiate both sides of Eq. (2.11) with respect to x we can get an expression similar to the KFE discussed previously. We now define $\phi(x,s) = \partial_x \rho(x,s)$ and we write

$$-\partial_s \phi(x,s) = -a(s)\partial_x(x\phi(x,s)) + b^2(s)\partial_{xx}\phi(x,s). \tag{2.20}$$

In Fourier space it becomes

$$-\partial_s \tilde{\phi}(x,s) - a(s)k(s)\partial_k(\tilde{\phi}(x,s)) = -b^2(s)k^2(s)\tilde{\phi}(x,s). \tag{2.21}$$

We solve the previous equation along the characteristics

$$\frac{dk(s)}{ds} = a(s)k(s) \;\rightarrow\; k(s) = k(t)e^{-\int_s^t a(r)\,dr}, \tag{2.22}$$

and we get

$$\frac{d\tilde{\phi}(x,s)}{ds} = b^2(s)k^2(t)e^{-2\int_s^t a(r)\,dr}\tilde{\phi}(x,s), \tag{2.23}$$

that has solution

$$\begin{aligned}\tilde{\phi}(k,s) =& \tilde{\phi}(k,t)\,\exp\left[-k^2(t)\int_s^t drb^2(r)e^{-2\int_r^t a(u)\,du}\right] = \\ =& - ik(s)e^{\int_s^t a(r)\,dr}\exp\Bigg[- ik(s)e^{\int_s^t a(r)\,dr}y - \\ & - k^2(s)e^{2\int_s^t a(r)\,dr}\int_s^t drb^2(r)e^{-2\int_s^r a(u)\,du}\Bigg] = \\ =& - ik(s)m_B(t,s)\exp\Bigg[- ik(s)ym_B(t,s) - \\ & - \frac{\sigma_B^2(t,s)}{2}k^2(s)\Bigg],\end{aligned} \tag{2.24}$$

where we have defined

$$m_B(t,s) = e^{\int_s^t a(r)\,dr}, \tag{2.25}$$

$$\sigma_B^2(t,s) = 2e^{2\int_s^t a(r)\,dr}\int_s^t drb^2(r)e^{-2\int_r^t a(u)\,du}, \tag{2.26}$$

and we have used

$$\begin{aligned}\tilde{\phi}(k,t) &= \int dxe^{-ikx}\phi(x,t) = \\ &= \int dxe^{-ikx}\partial_x\delta(x-y) = ike^{-iky}.\end{aligned} \tag{2.27}$$

Performing the inverse Fourier transform we get

$$\phi(x,s) = -\frac{m_B(t,s)(x - y m_B(t,s))}{\sqrt{2\pi\sigma_B^2(t,s)}\sigma_B^2(t,s)} \exp\left[-\frac{(x - y m_B(t,s))^2}{2\sigma_B^2(t,s)}\right]. \tag{2.28}$$

We can finally get the probability density integrating $\phi(x,s)$ with respect to x

$$\begin{aligned} \rho(x,s) = &- \frac{m_B(t,s)}{\sqrt{2\pi\sigma_B^2(t,s)}\sigma_B^2(t,s)} \\ &\int (x - y m_B(t,s)) \exp\left[-\frac{(x - y m_B(t,s))^2}{2\sigma_B^2(t,s)}\right] dx = \\ =&\frac{m_B(t,s)}{\sqrt{2\pi\sigma_B^2(t,s)}} \exp\left[-\frac{(x - y m_B(t,s))^2}{2\sigma_B^2(t,s)}\right], \end{aligned} \tag{2.29}$$

where the condition

$$\lim_{s\to t} \int_{-\infty}^{\infty} \rho(y,t|x,s)dx = 1 \tag{2.30}$$

has been imposed.

2.2 Survival analysis of the Ornstein-Uhlenbeck process (papers II,V)

In this section I will introduce the survival probability of a non autonomous Ornstein-Uhlenbeck (OU) process. The survival probability is associated with the probability for the system X to reach a defined threshold β for the first time starting at $t = 0$ from a given initial position x, that is

$$S(t,x) = \mathcal{P}(t < T), \tag{2.31}$$

where T is a random variable denoting the time when the system reaches the boundary for the first time

$$T \equiv \inf\{t \,:\, X(t) > \beta | X(0) = x\}. \tag{2.32}$$

I will restrict myself to an Ornstein-Uhlenbeck process defined by the Langevin equation (2.4) with $U(X,t) = \frac{1}{2}a(t)X^2$. I will consider the survival probability of this process inside the closed set $[-\alpha,\beta]$ $\alpha,\beta > 0$. Using the definition of the KBE in Eq. (2.11), we can write the survival probability as the integral of ρ over the interval $[-\alpha,\beta]$

$$S(t;x,s) = \int_{-\alpha}^{\beta} \rho(y,t;x,s)dy, \tag{2.33}$$

with initial condition

$$S(t=s;x,s)=\Theta(x+\alpha)\Theta(\beta-x), \tag{2.34}$$

and boundary conditions

$$S(t;x=-\alpha,s)=S(t;x=\beta,s)=0 \quad \forall\, s\leq t, \tag{2.35}$$

where $\Theta(.)$ is the Heaviside theta function.

The survival probability will then satisfies

$$-\partial_s S(t;x,s)=-a(s)x\partial_x S(t;x,s)+b^2(s)\partial_{xx}S(t;x,s), \tag{2.36}$$

with conditions (2.34, 2.35).

We can notice that the survival probability satisfies the same equation of the KBE with different boundary and initial conditions. This difference, however, prevents it to be solved in the same manner. In fact, because the support of $S(t;x,s)$ is $x\in[-\alpha,\beta]$ we can not anymore take its Fourier transform and solve Eq. (2.36) in momentum space. We can notice however that, if $\alpha,\beta \gg \sqrt{\frac{b^2(t)}{a(t)}}$ $\forall t$, the survival probability can be factorized. In fact, if we now discretize the stochastic process, the survival probability after n time steps can be written as

$$\begin{aligned} S_n &= \prod_{i=1}^{n}(1-P_i^\alpha-P_i^\beta)=\prod_{i=1}^{n}[(1-P_i^\alpha)(1-P_i^\beta)-P_i^\alpha P_i^\beta] \\ &\simeq \prod_{i=1}^{n}(1-P_i^\alpha)\prod_{i=1}^{n}(1-P_i^\beta)=S_n^\alpha S_n^\beta. \end{aligned} \tag{2.37}$$

We have called $P_i^\alpha(P_i^\beta)$ the probability of hitting for the first time $-\alpha(\beta)$ after i time steps and $S^\alpha(S^\beta)$ the survival probability with an absorbing boundary only in $-\alpha(\beta)$. Because α and β are distant from the equilibrium position, we can neglect the term $P_i^\alpha P_i^\beta$ in Eq. (2.37) and write the survival probability in the interval $[-\alpha,\beta]$ as the product of the two survival probabilities in the two intervals $(-\infty,\beta]$ and $[-\alpha,\infty)$. Thanks to this factorization, we can study the two survival probabilities separately and we will restrict ourselves only to that one defined in $(-\infty,\beta]$.

Also if the Fourier transform is still impossible, because we have now extended the support of $S(t;x,s)$ to $-\infty$, we can solve Eq. (2.36) in Laplace space. However, in this case additional terms coming from the finite boundary at $x=\beta$ will come out and they prevent a practicable solution for the survival probability. For this reason we have developed a different approach which is explained in details in II,V.

2.3 Path integrals in stochastic processes (paper I)

In this section I will provide a theoretical background on the use of path integrals in stochastic process, a mathematical tool used in I. Here I will deal with a more general case than that one described in Eq. (2.4) since I will take into account a stochastic process with multiplicative noise.

We start form the following overdamped Langevin equation

$$\dot{X}(t) = a(X) + b(X)\xi(t) \tag{2.38}$$

with $\xi(t)$ delta-correlated Gaussian white noise.

We write the process $X(t)$ as the continuum limit of a sequence of X_n evaluated on discrete time steps t_n, with time step Δt. Each X_n will then satisfy

$$X_n \overset{\eta}{=} X_{n-1} + a(X_{n-1} + \eta(X_n - X_{n-1}))\Delta t + b(X_{n-1} + \eta(X_n - X_{n-1}))\sqrt{\Delta t}\xi_{n-1}, \tag{2.39}$$

where the term $\eta \in [0, 1]$ is due to the arbitrariness in the evaluation of a, b between $[X_{n-1}, X_n]$.

Since

$$\dot{X}(t) \overset{\eta}{=} a(X) + b(X)\xi(t) \tag{2.40}$$

can be written as

$$\dot{X}(t) \overset{0}{=} a(X) - \eta b'(X)b(X) + b(X)\xi(t), \tag{2.41}$$

when a stochastic process has additive noise the Ito discretization ($\eta = 0$) can be considered without doing approximations.

Each trajectory $X(t)$ with $t \in [0, T]$ and $X(0) = X_0$ strictly depends on the single noise realizations and its probability can be written as

$$P(X(\xi_0, \xi_1, \ldots, \xi_{N-1})) = \int \prod_{i=0}^{N-1} \frac{d\xi_i}{\sqrt{2\pi}} e^{-\frac{\xi_i^2}{2}} \tag{2.42}$$

where $N = T/\Delta t$.

Since the ξ_i are related to the X_i through

$$\xi_i = \frac{X_{i+1} - X_i - a_i \Delta t}{b_i \sqrt{\Delta t}} \tag{2.43}$$

where $a_i = a(X_i + \eta(X_{i+1} - X_i))$ and $b_i = b(X_i + \eta(X_{i+1} - X_i))$, we can write the integral in Eq. (2.42) in terms of X_i. To this end, we have to evaluate the

Jacobian

$$J = |\det \frac{\partial \xi}{\partial X}| = |\prod_{j=0}^{N-1} \frac{\partial \xi_j}{\partial X_{j+1}}| = |\prod_{j=0}^{N-1} \frac{1}{b_j\sqrt{\Delta t}}\left[1-\eta a'_j\Delta t-\eta\frac{b'_j}{b_j}(X_{j+1}-X_j-a_j\Delta t)\right]|, \tag{2.44}$$

end Eq. (2.42) becomes

$$\begin{aligned} P(X_0, X_1, \ldots, X_{N-1}) &= \int \left(\prod_{i=1}^{N} dX_i\right)\frac{J}{(2\pi)^{N/2}}\exp\left[-\frac{1}{2}\sum_{j=0}^{N-1}\left(\frac{X_{j+1}-X_j-a_j\Delta t}{b_j\sqrt{\Delta t}}\right)^2\right] \\ &= \int \left(\prod_{i=1}^{N} dX_i\right)\left(\prod_{j=0}^{N-1} d\tilde{X}_j b_j\sqrt{\Delta t}\right)\frac{J}{(2\pi)^{N}}\exp\left[\sum_{j=0}^{N-1}\left(-\frac{b_j^2\tilde{X}_j^2\Delta t}{2}-i\tilde{X}_j(X_{j+1}-X_j-a_j\Delta t)\right)\right] \end{aligned} \tag{2.45}$$

Let's go back to the Jacobian. Using its expression reported in Eq. (2.44), we can write

$$\begin{aligned} \left(\prod_{j=0}^{N-1} b_j\sqrt{\Delta t}\right) J &= |\prod_{j=0}^{N-1}\left[1-\eta a'_j\Delta t-\eta\frac{b'_j}{b_j}(X_{j+1}-X_j-a_j\Delta t+i\tilde{X}b_j^2\Delta t-i\tilde{X}b_j^2\Delta t)\right]| \\ &= |\prod_{j=0}^{N-1}\left[1-\eta a'_j\Delta t-\eta\frac{b'_j}{b_j}\left(\frac{\partial}{\partial(i\tilde{X})}+i\tilde{X}b_j^2\Delta t\right)\right]| \\ &\simeq \exp\left[-\sum_{j=0}^{N-1}\left(\eta a'_j\Delta t+i\tilde{X}\eta b'_j b_j\Delta t\right)\right], \end{aligned} \tag{2.46}$$

where we used $\frac{\partial}{\partial(i\tilde{X})}\exp[\ldots] = (X_{j+1}-X_j-a_j\Delta t-i\tilde{X}b_j^2\Delta t)\exp[\ldots]$ and the fact that $\exp[\ldots]$ vanishes at the extremes of integration.

Using the expression of the Jacobian of Eq. (2.46) inside Eq. (2.48) we get

$$\begin{aligned} P(X_0, X_1, \ldots, X_{N-1}) &= \int \left(\prod_{i=1}^{N} dX_i\right)\left(\prod_{j=0}^{N-1}\frac{dp_j}{2\pi i}\right) \\ \times \exp&\left[-\sum_{j=0}^{N-1}\left(-\frac{b_j^2p_j^2\Delta t}{2}+p_j(X_{j+1}-X_j-a_j\Delta t)+\eta a'_j\Delta t+\eta p_j b'_j b_j\Delta t\right)\right] \end{aligned} \tag{2.47}$$

where we have defined $p_j = i\tilde{X}$.

Taking the continuum limit and defining $\mathcal{D}p(t) = \prod_{j=0}^{N-1}\frac{dp_j}{2\pi i}$, $\mathcal{D}X(t) = \prod_{i=1}^{N} dX_i$ we obtain

$$P(X(t)) = \int \mathcal{D}X(t)\mathcal{D}p(t)\, e^{-S[X,p]}, \tag{2.48}$$

with

$$S[X,p] = \int_0^T dt \left(-\frac{b(X)p(t)^2}{2} + p(t)(\dot{X}(t) - a(X)) + \eta a'(X) + \eta p(t) b'(X) b(X)) \right). \tag{2.49}$$

The action $S[X,p]$ is commonly called Martin-Siggia-Rose-Janssen-De Dominicis action.

Let's consider now a stochastic process with additive noise, that is with constant diffusion term b in the low noise limit, i.e. $b \ll 1$. After redefining $p(t) \to \frac{p(t)}{b}$, Eq. (2.48) becomes

$$\begin{aligned} P(X(t)) &= \int \mathcal{D}X(t) \frac{\mathcal{D}p(t)}{b} \exp\left[-\frac{1}{b} \int_0^T dt \left(-\frac{p(t)^2}{2} + p(t)(\dot{X}(t) - a(X)) + b\eta a'(X) \right) \right] \\ &\simeq \int \mathcal{D}X(t) \frac{\mathcal{D}p(t)}{b} \exp\left[-\frac{1}{b} \int_0^T dt \left(-\frac{p(t)^2}{2} + p(t)(\dot{X}(t) - a(X)) \right) \right]. \end{aligned} \tag{2.50}$$

We are interested in the in the most probable paths (instantons) through which the system state passes between two given points $X(t_i) = X_i$ and $X(t_f) = X_f$ with $t_f > t_i$. Because we are considering $b \ll 1$, the integral in Eq. (2.50) can be evaluated using the saddle point method by integrating only over those paths that minimize the action and thus satisfy

$$\begin{cases} \frac{\partial S[p,X]}{\partial p} = 0, \\ \frac{\partial S[p,X]}{\partial X} = 0. \end{cases} \tag{2.51}$$

The instantons then satisfy the following coupled system of equations

$$\begin{cases} \dot{X}(t) = a(X) + p(t), \\ \dot{p}(t) = p(t) a'(X), \end{cases} \tag{2.52}$$

with initial conditions

$$\begin{aligned} X(t_i) &= X_i, \\ p(t_i) &= p_i, \end{aligned} \tag{2.53}$$

end final condition

$$X(t_f) = X_f. \tag{2.54}$$

The probability of the instanton trajectory becomes then

$$P(X(t)) \sim \exp\left[-\frac{1}{b} \int_0^T dt p_i^2 e^{2\int_0^t ds a'(X(s))} \right]. \tag{2.55}$$

Chapter 3

Outlook

Currently, I'm following three different directions of research together with my supervisor and his collaborators.

- We are studying the survival probability a non autonomous Ornstein-Uhlenbeck process with time periodic drift and diffusion coefficients. We managed to extend the results of II, V to a defined range of time dependent coefficients and we are trying to obtain more general results.
- We are using Neural Ordinary Differential Equations, a neural network architecture similar to the Recurrent Neural Network used in III, to stochastic model dynamical systems in order to infer their dimension and to perform reliable predictions. We are applying this architecture to study ENSO, one of the most relevant global climate phenomenon, which we studied using a different technique in IV.
- We are combining the survival analysis developed in II,V and the instanton calculation used in I to obtain an analytical demonstration of the Landauer principle, which provides a lower theoretical limit of energy consumption in any irreversible logical operation.

Paper I

L.T. Giorgini, S.H. Lim, W. Moon, and J.S. Wettlaufer, Precursors to rare events in stochastic resonance, *Europh. Lett.* **129**, 40003 (2020).

A Letters Journal Exploring
the Frontiers of Physics

February 2020

EPL, **129** (2020) 40003
doi: 10.1209/0295-5075/129/40003

www.epljournal.org

Precursors to rare events in stochastic resonance

L. T. Giorgini[1], S. H. Lim[1], W. Moon[1,2] and J. S. Wettlaufer[1,3]

[1] *Nordita, Royal Institute of Technology and Stockholm University - Stockholm 106 91, Sweden*
[2] *Department of Mathematics, Stockholm University - Stockholm 106 91, Sweden*
[3] *Yale University - New Haven, CT 06520, USA*

received 28 November 2019; accepted in final form 18 February 2020

PACS 05.40.Jc – Brownian motion
PACS 05.40.Ca – Noise
PACS 02.50.Ey – Stochastic processes

Abstract – In stochastic resonance, a periodically forced Brownian particle in a double-well potential jumps between minima at rare increments, the prediction of which poses a major theoretical challenge. Here, we use a path-integral method to find a precursor to these transitions by determining the most probable (or "optimal") space-time path of a particle. We characterize the optimal path using a direct comparison principle between the Langevin and Hamiltonian dynamical descriptions, allowing us to express the jump condition in terms of the accumulation of noise around the stable periodic path. In consequence, as a system approaches a rare event these fluctuations approach one of the deterministic minimizers, thereby providing a precursor for predicting a stochastic transition. We demonstrate the method numerically, which allows us to determine whether a state is following a stable periodic path or will experience an incipient jump with a high probability. The vast range of systems that exhibit stochastic resonance behavior insures broad relevance of our framework, which allows one to extract precursor fluctuations from data.

Introduction. – Rare events, which frequently accompany fluctuations or phase transitions, arise in a wide range of natural and social systems, such as infectious disease outbreaks, earthquakes, stock market crashes, and many others (*e.g.*, [1–3]). Of particular interest are dynamical systems that have bifurcations, at which sudden transitions to distinct dynamical regimes occur [4,5]. Even before reaching a bifurcation, noise-induced transitions can occur with low probability [6]. In consequence, a system experiences a large-magnitude change resulting in significant positive or deleterious consequences. Hence, it is important to understand the mechanism leading to the occurrence of such events, and to seek precursors to anticipate them [7].

The desire to predict these rare events in advance has fueled studies, to simulate [8], classify [9], analyze [10] and predict [11,12] their properties. Although the existence of early warning signals for rare events has been suggested, there are few results determining reliable and robust indicators for noise-induced transitions [13]. Because most systems are inherently noisy, understanding the role of noise in inducing these transitions is critical for their quantitative prediction well in advance. Here we describe a theory quantifying the role of noise in rare events, which underlies probabilistic forecast models.

We study noise-induced transitions using a class of periodically forced low-dimensional stochastic dynamical systems and identify a novel early warning indicator for the jumps from one stable state of the system to another. Periodically forced stochastic systems are ubiquitous in nature. For example, periodic forcing and background noise are the main ingredients of stochastic resonance [14–17] (see [18,19] for reviews), wherein the response to a weak signal is magnified by noise-induced fluctuations that drive hopping from one stable state to the other in a double-well potential with two minima. Settings of relevance range from the human cardiovascular system [20] to the seasonal variability of the Earth's climate [21].

This paper is organized as follows. In the following section, we provide an outline of the mathematical formulation, with the details provided in the Supplementary Material `Supplementarymaterial.pdf` (SM). In the third section, we discuss the task of finding precursors for the occurrence of a rare event. We propose a data-driven strategy to study the problem in the fourth section. This strategy constitutes the main contribution of our paper and is presented as a five-step procedure. We test this strategy with an example in the fifth section using two different numerical simulations before concluding in the last section.

Outline of the mathematical formulation. – In order to insure our treatment is reasonably self-contained, here we outline the principal waypoints of the path-integral treatment of stochastic processes. For readers not intimate with this approach we have provided details in the SM.

The state or position, x, of the system is modeled by the following nonautonomous one-dimensional overdamped Langevin equation:

$$\dot{x}(t) = F(x(t),t) + \sqrt{2}\sigma\xi(t), \tag{1}$$

in which

$$F(x,t) = -U'(x) + A\cos(\omega t), \tag{2}$$

where the dot (prime) denotes differentiation with respect to time (position), $U(x)$ is a multi-well potential, $A\cos(\omega t)$ is external periodic forcing and $\xi(t)$ is zero mean Gaussian white noise with correlation function

$$\langle \xi(t)\xi(s)\rangle = \delta(t-s). \tag{3}$$

We study systems described by eqs. (1)–(3) with the (constant) noise intensity $\sigma \ll A$ by employing a path-integral formulation [22,23]. The use of this formulation allows us to identify the most probable (optimal) trajectories (also called instantons) among all the possible trajectories that the system state follows to go from a point with the space-time coordinates (x_i, t_i) to another point with the coordinates (x_f, t_f). These optimal paths can be derived by studying large deviations from the unperturbed deterministic dynamics of the system in the weak noise regime (see [24] for the details of sample-path large deviation theory for stochastic differential equations).

We are interested in the behavior of the system *shortly before* its state jumps from one potential well to another. Since $t_f - t_i$ is finite, there exists a finite number of optimal paths, of the order of $(t_f - t_i)/T$, where $T = 2\pi/\omega$ is the period of the external periodic forcing. In fact, it can be shown that these optimal paths, denoted as $x_k(t)$ (with the subscript k denoting a particular path), satisfy the following system of first-order differential equations [25]:

$$\begin{aligned} \dot{x}_k(t) &= 2p_k(t) + F(x_k(t),t), & (4)\\ \dot{p}_k(t) &= -p_k(t)F'(x_k(t),t), & (5) \end{aligned}$$

with the boundary conditions

$$x_k(t_i) = x_i \qquad \text{and} \qquad x_k(t_f) = x_f. \tag{6}$$

We have introduced the conjugate momenta $p_k(t)$ relative to the optimal paths $x_k(t)$. These momenta are defined as $p_k(t) := \frac{1}{2}[\dot{x}_k(t) - F(x_k(t),t)]$ and they measure the deviation from the deterministic unperturbed dynamics. Each path $x_k(t)$ starts at $t = t_i$ and first follows a stable periodic orbit $x_s(t)$, defined as the solution of eq. (1) with $\sigma = 0$:

$$\dot{x}_s(t) = F(x_s(t),t) \qquad \text{and} \qquad x_s(t) = x_s(t+T). \tag{7}$$

The path begins to deviate from this periodic orbit at a random time t_0 and then transitions to a path that closely follows another stable periodic orbit. This random time t_0, which also denotes the time at which the $p_k(t)$ begin to deviate from zero, differs for different realizations of system paths described by eq. (1).

Formally, the probability distribution that the process $x(t)$ reaches a point x_f at time t_f, given that it started at a point x_i at time t_i can be written as

$$P(x_f,t_f|x_i,t_i) = \sum_{k=1}^{n} P_k(x_f,t_f|x_i,t_i), \tag{8}$$

with $n = \lfloor (t_f - t_i)/T \rfloor$, where $\lfloor\cdot\rfloor$ denotes the floor function. Each optimal path x_k gives the contribution

$$P_k(x_f,t_f|x_i,t_i) = \frac{1}{\sqrt{4\pi\sigma^2 Q_k(t_f)}} e^{-S[x_k]/\sigma^2} \tag{9}$$

to the series defining eq. (8). Here

$$S[x_k(t)] = \int_{t_i}^{t_f} p_k^2(t)\mathrm{d}t, \tag{10}$$

where the Q_k satisfy the following second-order initial value problem [25]:

$$\frac{\ddot{Q}_k(t)}{2} - \partial_t[Q_k(t)F'(x_k(t),t)] + Q_k(t)p_k(t)F''(x_k(t),t) = 0, \tag{11}$$

with

$$Q_k(t_i) = 0, \qquad \dot{Q}_k(t_i) = 1. \tag{12}$$

The early warning indicator. – For a long time interval with high probability the system will follow a stable periodic orbit, $x_s(t)$, around one of the local minima of the potential with fluctuations of order σ. Rarely, however, the system will jump from one minimum to the other, in which case the most probable path is described by eqs. (4)–(6), with the jump beginning at time t_i and ending at time t_f. Our principal goal is to obtain quantitative precursors of such rare events and, combined with knowledge of the optimal path, estimate the most probable time of the rare event.

The key observation is as follows. We compare eq. (4) with the Langevin equation (1) and observe that the optimal condition for the system to jump from one potential well to the other is when the fluctuations around the stable periodic path,

$$\xi(t) = \frac{1}{\sqrt{2}\sigma}[\dot{x}(t) - F(x(t),t)], \tag{13}$$

accumulate to $\sqrt{2}p(t)/\sigma$, where $p(t)$ is one of the $p_k(t)$'s satisfying eq. (5). Namely, up to a multiplicative factor of $\sqrt{2}/\sigma$, as the system approaches a rare event, the fluctuations around the stable state approach one of the deterministic minimizers $\sqrt{2}p_k(t)/\sigma$. Therefore, it is crucial to extract such fluctuations from data in order to determine whether a state is simply following a stable periodic path, or begins to follow eqs. (4)–(6), describing the most probable path that can lead the system to jump.

Clearly, $p(t)$ acts as a forcing for $x(t)$ and hence the former "anticipates" the latter. Thus, although when $p \gg \sigma$ and $|x - x_s| \gg \sigma$ the instanton and its conjugate momentum can be resolved, the former condition is satisfied before the latter condition. Therefore, $p(t)$ is a better early warning indicator than $x(t)$. Now, despite the momentum being observable when $p \gg \sigma$, the influence of the oscillatory forcing term with amplitude A in eq. (4) is to delay the effect of $p(t)$ on $x(t)$ until $p(t) = O(A)$. Hence, there will be a time window τ such that $\sigma < p(t) < A$ for $t \in \tau$ in which the noise accumulates prior to the appearance of a large deviation. The system begins to follow eqs. (4), (5) at $t = t_0$ and after $t - t_0 > T$ the momenta behave as

$$p(t) = p(t_0)\exp\left[-\int_{t_0}^{t} \mathrm{d}sF'(x(s),s)\right] \simeq p(t_0)e^{-\lambda_s(t-t_0)}, \tag{14}$$

where $\lambda_s < 0$ is the Lyapunov exponent of the stable periodic orbit defined as $\lambda_s = \frac{1}{T}\int_t^{t+T}\mathrm{d}zF'(x_s(z),z)$. Therefore, during a "warning time" $\tau_W \sim -\frac{1}{\lambda_S}\ln[\frac{A}{p(t_0)}]$, eq. (14) describes the noise accumulation before $x(t)$ exhibits a transition. Clearly, because the warning time is inversely proportional to $\lambda_s < 0$, in less stable orbits we can determine a jump precursor earlier.

The rare event momentum precursor is demonstrated numerically in fig. 1, which shows the transition from one potential well to the other. We used the instanton dynamics described in the fourth section to force the appropriate accumulation of noise in the case where $F(x,t) = x - x^3 + 0.7\cos(2\pi t)$ and $\sigma = 0.01$ in eq. (1), which we evolve for ten periods with initial condition $x(0) = -1$. We then use eqs. (4)–(6) to simulate the jump. Comparing figs. 1(a) and (b) one observes the "momentum anticipation" of the deviation of the trajectory $x_F(t)$ from the stable periodic orbit $x_s(t)$.

We can obtain an accurate estimate of the time interval in which the jump will occur from eqs. (4), (5), (9) and (10). Within each period there are only a few highly probable paths and thus upon observation of optimal noise accumulation, we can determine which path the system is following. Hence, we can estimate the corresponding jump time, t_j, when the system shifts to the other stable basin (fig. 2(b)). Therefore, by studying the fluctuations around the stable periodic orbit to determine when they begin to behave as $p(t)$, we can predict if the system is approaching a jump by computing the jump probability and time. Next we provide a systematic outline of our prediction strategy.

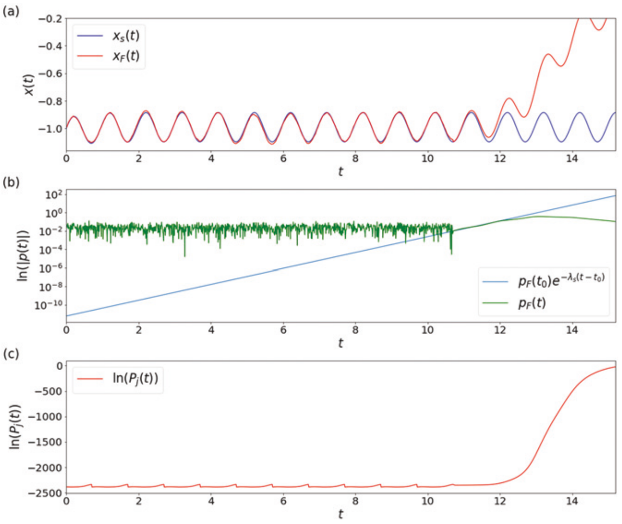


Fig. 1: (a) Time evolution of the position $x_F(t)$, compared to the stable periodic orbit $x_S(t)$. (b) A semi-log plot of $|p(t)|$. (c) Time behavior of the jump probability computed from eq. (9).

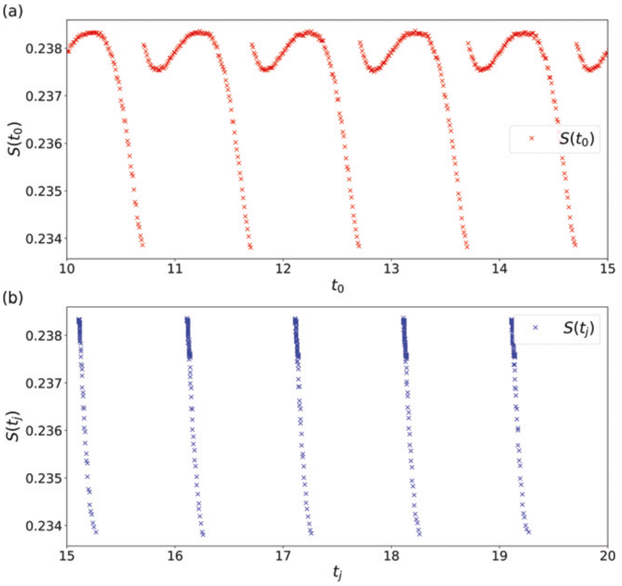


Fig. 2: Values of the action corresponding to different values of (a) t_0 and (b) t_j. The system has been evolved for 10 periods setting $p(t) = 0$ in order to be sure that it follows the stable periodic orbit, after which eq. (5) was used for $p(t)$.

Prediction scheme. – Our program for the prediction and study of rare events in stochastic resonance consists of the following five main steps:

1) We start with the nonautonomous Langevin equation (1) describing the time evolution of the system,

$$\dot{x}(t) = F(x(t),t) + \sqrt{2}\sigma\xi(t),$$

assuming we know $F(x(t),t)$ and σ. See [21] and [26] regarding the construction of these expressions from the data (the latter paper treating the autonomous case).

2) We determine the instantons as follows. First, we evolve

$$\dot{x}_s(t) = F(x_s(t), t),$$

with $x_s(t = 0)$ chosen inside one potential well. After an initial transient, the system state evolves following the stable periodic orbit. When this condition is satisfied at $t = t_0$, we modify the earlier equation to

$$\begin{aligned} \dot{x}(t) &= 2p(t) + F(x(t), t), \\ \dot{p}(t) &= -p(t)F'(x(t), t), \end{aligned}$$

with

$$x(t_0) = x_s(t_0) \qquad \text{and} \qquad p(t_0) = p_0 > \sigma. \tag{15}$$

We evolve the system many times until we observe a shift to another stable basin at $t = t_j$, using the same value of p_0 but with different values of t_0 chosen inside one period. For each of these paths we compute the relative action

$$S[x_k(t)] = \int_{t_0}^{t_j} p_k^2(t)\mathrm{d}t.$$

Because $S(t_0)$ is periodic, there will be only one instanton in every period. We find the instanton as the path that minimizes the action in every period (see fig. 2).

3) We isolate the noise from the data using the Langevin equation as in eq. (13),

$$\xi(t) = \frac{1}{\sqrt{2\sigma}}[\dot{x}(t) - F(x(t), t)],$$

which is related to the conjugate momentum as $p(t) = \sigma\xi(t)/\sqrt{2}$.

4) Prior to the jump the conjugate momentum is expected to increase exponentially as $p(t) = p_0 e^{-\lambda_s(t-t_0)}$. Thus, we scrutinize the behavior of $p(t)$ obtained from the data of $x(t)$ through eq. (13) until it ceases to exhibit fluctuations of order σ near zero and begins to grow. We compare its behavior to the conjugate momentum of the instanton and if the noise structure differs from optimality we are unable to make predictions; because the noise structure is not optimal, the jump is more rare. However, when the noise structure is optimal, we can accurately estimate the jump probability and time.

Next we demonstrate this scheme in a numerical example.

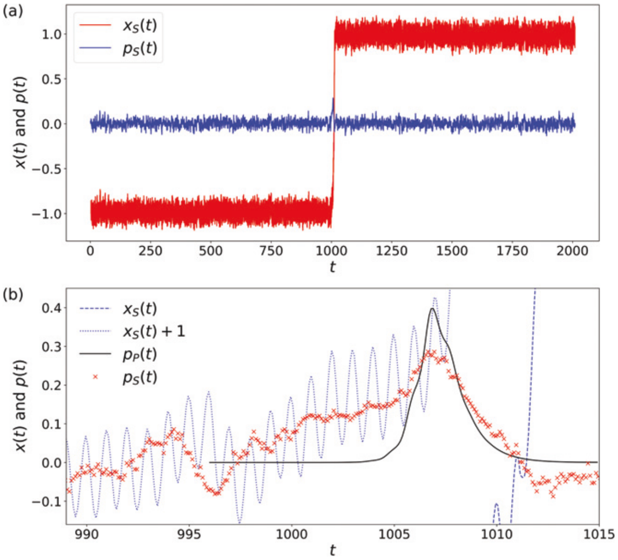


Fig. 3: (a) Plot of the position of the system $x(t)$ and of $p(t) = \frac{1}{2}[\dot{x}(t) - F(x(t), t)]$ computed numerically (using a leap frog algorithm) from eq. (1), and denoted with the subscript S. (b) Expansion of the previous plot and comparison with $p(t)$ computed from eqs. (4), (5), denoted with the subscript P. Note that $x(t)$ first hits the origin at approximately $t = 1011.5$, which is preceded with a peak in $p(t)$ by a time of 4 periods. The data are smoothed using a moving average.

Numerical demonstration. – In order to demonstrate this strategy, we evolve eq. (1) numerically for a very long time, until the system jumps from one potential well to the other. We use a quartic potential $U(x) = -x^2/2 + x^4/4$ with $A = 0.7$, $\omega = 2\pi$ and $\sigma = 0.0727$. These parameters are chosen to maximize the difference between A and σ and yet still yield a jump in a tractable simulation time.

In the absence of noise and periodic forcing, the resulting Langevin equation has two stable periodic solutions separated by an unstable one. We apply our prediction scheme to study the transition between the two stable periodic solutions in the regime $\sigma < A$. The results are discussed in the following.

Figure 3(a) shows the time evolution of the system state $x(t)$ and of the deviation $p(t)$ from deterministic flow (the noise) over a time of 2000 periods. We observe that $p(t)$ exhibits small oscillations around zero, showing a peak near the jump. Since the system is driven by white noise, $p(t)$ has no temporal structure in the time frame considered, except for the increment very near the jump. Indeed, as described above, this jump can only be observed if the noise accumulates in an optimal way. Namely, in this region the noise *does not behave randomly*. Rather, in order to drive the system to a jump, the noise should form a specific structure that depends on the shape of the potential.

Figure 3(b) shows in detail the behavior of $p(t)$ close to the jump. When the value of $p(t)$ is near zero, we expect that the noise, and thus $p(t)$, will accumulate randomly. Importantly, in that region, because there are many ways for the noise to accumulate with equal probability, finding an optimal path is meaningless. However, when $p(t)$ deviates from zero we find a substantially different cumulative influence of the noise. Namely, because the probability differences between paths increase exponentially, only a single path becomes relevant. This path is very near the optimal path satisfying eqs. (4), (5). Indeed, the solution of eq. (5) and numerical shapes of $p(t)$ agree well near the transition; increasing exponentially rather far from zero as $e^{-\lambda_s t}$ and then, when approaching the unstable periodic orbit around the maximum of the potential, it begins to decrease as $e^{-\lambda_u t}$, with λ_u the Lyapunov exponent of the unstable periodic orbit defined as $\lambda_u = \frac{1}{T}\int_t^{t+T} \mathrm{d}z F'(x_u(z), z)$. Clearly the unstable periodic orbit will not persist and, after reaching the maximum of the potential, the system will immediately fall into the stable periodic orbit around the minimum of the other well. However, the asymmetric influence of the noise as the system transitions is responsible for the deviation of the numerical and the analytical prediction near the peak.

Figures 4(a)–(c) show $x(t)$ in three different simulations and $p(t)$ during the interval in which noise is accumulating shortly before the transition. In each simulation we have modified the shape of the potential in order to change the value of the Lyapunov exponent of the stable orbit and the noise amplitude to make the waiting time for the jump comparable in each simulation. This is accomplished by varying the value of a in $-U'(x) = a\,x - x^3$, with $a = 1.5$, 1 and 0.5 in figs. 4(a)–(c), respectively, as well as the values of σ, with $\sigma = 0.1233$, 0.0727 and 0.0632 in figs. 4(a)–(c), respectively. In all cases, $A = 0.7$ and $\omega = 2\pi$. These figures demonstrate the optimal nature of the noise accumulation described by the instanton equations (4), (5). In fig. 4(d) we show the mean standard error (MSE) between the realizations of $p(t)$ constructed from the realizations of the noise and its optimal behavior described in eq. (14). This is defined as

$$MSE(t) = \frac{1}{n} \min_{p_0} \left\{ \sum_{i=m}^{m+n} \left(p_i - p_0 e^{\lambda_s t_i} \right) \right\}. \qquad (16)$$

We computed the MSE using the noise from fig. 4(c) taken over a 200 period time window spanning an interval exhibiting linear behavior on a logarithmic scale. We used a moving time window of size $n = 22 = 2.2\mathrm{d}t(\sim\tau_W)$, in each of which we chose the value of p_0 that minimizes the MSE between the values of $p(t)$ constructed from the realizations of the noise and the exponential slope given in eq. (14). We find that the MSE assumes larger values — ten times larger at least— for the time interval in which the noise deviates from $p_0 e^{-\lambda_s t}$. Thus, these simulations show that the time window during which the noise acquires a specific structure increases thereby decreasing the value of the Lyapunov exponent, consistent with the discussion in the second section. Namely, a smaller Lyapunov exponent implies a slower optimal accumulation of the noise. Therefore, the system will take more time to shift to another stable basin so that the warning time of a rare event increases for less stable potentials.

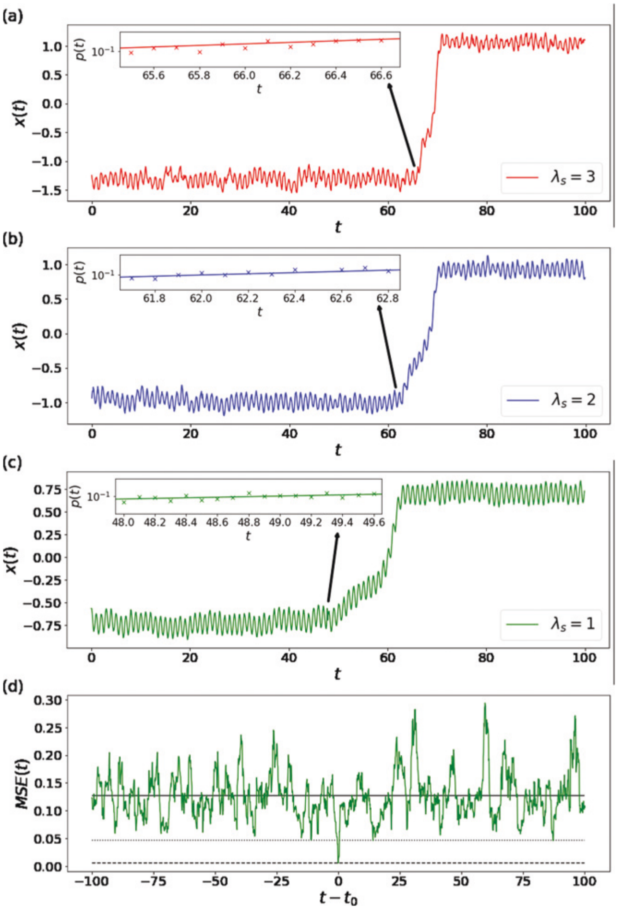


Fig. 4: The trajectories of $x(t)$ for 100 periods and $p(t)$ immediately prior to the jump (along with the analytical prediction) for three different simulations associated with three different Lyapunov exponents are shown in (a)–(c). (d) The MSE of the noise realizations with respect to the optimal behavior predicted in eq. (14), using the data from the simulation shown in (c) displayed over a larger time window. The solid line is the average MSE, the dashed line is the MSE corresponding to the values of $p(t)$ in the time window highlighted in the top left of (c) and the dotted line is the minimum value reached by the MSE if these values of $p(t)$ are removed.

When $\sigma \ll A$ and $\sigma \ll \Delta U$ the jump probability is extremely small and thus poses a substantial numerical challenge. Having shown that our strategy works well under less extreme cases we have thereby quantified how the conjugate momentum of the instanton organizes the noise prior to the jump. Therefore, we expect that the same noise organization process will be operative in the case where $\sigma \ll A$ and $\sigma \ll \Delta U$ and the system behavior shown in fig. 1 will be recovered.

Conclusion. – We have developed a theory to study and find precursors to noise-induced rare events within the general framework of stochastic resonance. In stochastic resonance, a periodically and noise forced system in a double-well potential jumps between minima, but the time-scale separation of these forcings insures that the system oscillates for a long time about one of the local minima of the potential and only very rarely jumps to the other minima. The ubiquity of such transitions underlies the importance of trying to predict when they will occur.

We have used a path-integral method to determine the particular manner in which the fluctuations around the unperturbed deterministic flow must organize prior to the system jump. We have showed how to predict the time within a period when the system will transit to another minimum, and have harnessed the signature of this fluctuation behavior as an advanced indicator of a potential jump, as well as computing the probability of such rare events. The method provides a framework to examine data in a manner that facilitates predictions across a broad spectrum of stochastic systems. Finally, the approach identifies a short well-defined structure immediately prior to the rare event. The detection of such structures in a prediction setting is a central aspect of many machine learning approaches to rare event predictions (see, *e.g.*, [27]), which provide a test bed for a wide range of applications.

* * *

The authors acknowledge the support of Swedish Research Council grant No. 638-2013-9243.

REFERENCES

[1] Ghil M., Yiou P., Hallegatte S., Malamud B., Naveau P., Soloviev A., Friederichs P., Keilis-Borok V., Kondrashov D., Kossobokov V. *et al.*, *Nonlinear Process. Geophys.*, **18** (2011) 295.

[2] Forgoston E. and Moore R. O., *SIAM Rev.*, **60** (2018) 969.

[3] Farazmand M. and Sapsis T., *Appl. Mech. Rev.*, **71** (2019) 050801.

[4] Scheffer M., Bascompte J., Brock W. A., Brovkin V., Carpenter S. R., Dakos V., Held H., Van Nes E. H., Rietkerk M. and Sugihara G., *Nature*, **461** (2009) 53.

[5] Dematteis G., Grafke T., Onorato M. and Vanden-Eijnden E., *Phys. Rev. X*, **9** (2019) 041057.

[6] Hänggi P., Talkner P. and Borkovec M., *Rev. Mod. Phys.*, **62** (1990) 251.

[7] Faranda D., Lucarini V., Manneville P. and Wouters J., *Chaos, Solitons Fractals*, **64** (2014) 26.

[8] Grafke T. and Vanden-Eijnden E., *Chaos*, **29** (2019) 063118.

[9] Ashwin P., Wieczorek S., Vitolo R. and Cox P., *Philos. Trans. R. Soc. A: Math., Phys. Eng. Sci.*, **370** (2012) 1166.

[10] Romano F. and Kuehn C., *Int. J. Bifurcat. Chaos*, **28** (2018) 1850103.

[11] Ritchie P. D. L. jr., PhD Thesis, University of Exeter (2016).

[12] Kuehn C., Malavolta G. and Rasmussen M., *J. Math. Anal. Appl.*, **464** (2018) 58.

[13] Chen Y., Gemmer J. A., Silber M. and Volkening A., *Chaos*, **29** (2019) 043119.

[14] Benzi R., Sutera A. and Vulpiani A., *J. Phys. A: Math. Gen.*, **14** (1981) L453.

[15] Benzi R., Parisi G., Sutera A. and Vulpiani A., *Tellus*, **34** (1982) 10.

[16] Herrmann S., Imkeller P., Pavlyukevich I. and Peithmann D., *Stochastic Resonance*, in *Mathematical Surveys and Monographs*, Vol. **194** (American Mathematical Society) 2013.

[17] Lucarini V., *Phys. Rev. E*, **100** (2019) 062124.

[18] Gammaitoni L., Hänggi P., Jung P. and Marchesoni F., *Rev. Mod. Phys.*, **70** (1998) 223.

[19] Wellens T., Shatokhin V. and Buchleitner A., *Rep. Prog. Phys.*, **67** (2003) 45.

[20] Stefanovska A., *Contemp. Phys.*, **40** (1999) 31.

[21] Moon W. and Wettlaufer J. S., *Sci. Rep.*, **7** (2017) 44228.

[22] Onsager L. and Machlup S., *Phys. Rev.*, **91** (1953) 1505.

[23] Dürr D. and Bach A., *Commun. Math. Phys.*, **60** (1978) 153.

[24] Freidlin M., Szucs J. and Wentzell A., *Random Perturbations of Dynamical Systems*, in *Grundlehren der mathematischen Wissenschaften*, Vol. **260** (Springer, New York) 2012.

[25] Lehmann J., Reimann P. and Hänggi P., *Phys. Rev. E*, **62** (2000) 6282.

[26] Hasselmann K., *Tellus*, **28** (1976) 476.

[27] Lim S. H., Giorgini L. T., Moon W. and Wettlaufer J. S., arXiv: 1908.03771 (2019).

Paper II

L.T. Giorgini, W. Moon, and J.S. Wettlaufer, Analytical Survival Analysis of the Ornstein-Uhlenbeck Process, *J. Stat. Phys.* **181**, 2404 (2020).

Journal of Statistical Physics (2020) 181:2404–2414
https://doi.org/10.1007/s10955-020-02669-y

Analytical Survival Analysis of the Ornstein–Uhlenbeck Process

L. T. Giorgini[1] · **W. Moon**[1,2] · **J. S. Wettlaufer**[1,3]

Received: 2 July 2020 / Accepted: 28 October 2020 / Published online: 5 November 2020

Abstract
We use asymptotic methods from the theory of differential equations to obtain an analytical expression for the survival probability of an Ornstein–Uhlenbeck process with a potential defined over a broad domain. We form a uniformly continuous analytical solution covering the entire domain by asymptotically matching approximate solutions in an interior region, centered around the origin, to those in boundary layers, near the lateral boundaries of the domain. The analytic solution agrees extremely well with the numerical solution and takes into account the non-negligible leakage of probability that occurs at short times when the stochastic process begins close to one of the boundaries. Given the range of applications of Ornstein–Uhlenbeck processes, the analytic solution is of broad relevance across many fields of natural and engineering science.

Keywords Survival probability · Ornstein–Uhlenbeck Process · Fokker–Planck equation · Asymptotics

1 Introduction

The generalized Ornstein–Uhlenbeck (OU) model describes a stochastic process with at least one equilibrium point. It provides a framework for a wide range of physical, biological and social systems, wherein stabilization is viewed in terms of a potential minimum, characterized by a negative Lyapunov exponent, and high-frequency fluctuations are interpreted in terms of

---

Communicated by Gregory Schehr.

J. S. Wettlaufer
john.wettlaufer@yale.edu

L. T. Giorgini
ludovico.giorgini@su.se

W. Moon
woosok.moon@su.se

[1] Nordita, Royal Institute of Technology and Stockholm University, 106 91 Stockholm, Sweden

[2] Department of Mathematics, Stockholm University, 106 91 Stockholm, Sweden

[3] Yale University, New Haven, USA

specific noise forcing. For example, an OU process is used to study neuronal activity [1] and the time-evolution of trait values towards their evolutionary optima [2]. In a clinical setting the health of the hepatic dynamic equilibrium is fit to an OU process using maximum likelihood estimation [3]. Stochastic volatility, crucial for deducing stock returns or option pricing, is treated in terms of an OU process [4], as are the noise spectra of climate observations [5–7].

A particular stochastic model is commonly studied in terms of the *survival probability*, which is associated with the probability of one or more events occurring, or for the system to reach a defined threshold for the first time starting from a given initial position. Due to the generality of the question it addresses, survival analysis has been widely used in science and engineering. Examples include Feshbach resonances and the quantum Zeno effect (for example [8–10]), engineering reliability analysis [11], financial risk management [12], and event history analysis in sociology [13]. Moreover, in the specific case of an OU process survival analyses from neuroscience [14] and epidemiology [15,16] to quantitative finance [17–19] and extreme value statistics of correlated random variables [20] demonstrate the ubiquity of the approach. Our survival analysis is broadly relevant to all systems that can be described by an OU process. For example, it can be shown [21] that the survival probability of Brownian motion with an absorbing boundary that moves in time t as $\propto \sqrt{t}$ can be recast as an OU process with a fixed absorbing boundary using Lamperti's (or Doob's) transformation.

Let $\{X(t),\ t \geq 0\}$ be an OU process starting at x. The time it takes for the state of a system to encounter a threshold $X(t) = \beta$ for the first time is variously called the first hitting time or first passage time, whose probability distribution function $\zeta(t, x)$, is defined as

$$\zeta(t, x) = \frac{\partial}{\partial t} \text{Prob}\{T \leq t\}, \tag{1}$$

where

$$T \equiv \inf\{t\ :\ X(t) > \beta;\, X(0) = x\}. \tag{2}$$

The time integral of the probability distribution of the first passage time is the survival probability (see for example [22–25] for reviews).

Despite the broad applicability of the survival probability, a *simple* accurate analytical expression has been lacking, which is the primary motivation of our work. The exact mathematical form is constructed using the Laplace transform and its inverse, which results in a series expansion of special functions [26]. However, its complexity confounds practical implementation. In particular, when the initial data are close to the boundary of the confining potential [27], or when the boundary itself is near the equilibrium point [28], one must retain a considerable number of terms in the expansion.

Our treatment of survival probability is shown in the schematic potential of Fig. 1, which contains a particle governed by a one-dimensional Ornstein–Uhlenbeck process through an overdamped autonomous Langevin equation given by

$$\dot{X}(t) = -aX(t) + \sqrt{2b}\xi(t), \tag{3}$$

where a and b are positive constants and $\xi(t)$ is Gaussian white noise with a zero mean and $\langle \xi(t)\xi(s) \rangle = \delta(t - s)$.

The survival probability commonly characterizes the anomalous or abnormal behavior of a system. This motivates our consideration of a threshold state ($X = \beta$) far from the equilibrium state ($X(t) = 0$), so that it is rare that the system will reach the threshold.

The next section is devoted to our overall approach. In Sect. 2.1 we summarize our method. In Sect. 2.2 we briefly review previous results on the survival probability of an OU process that will be used in Sect. 2.3 to derive a simple analytical expression that reproduces the survival probability for large values of the domain boundaries and when the initial data are close to the boundary. In Sect. 2.4 we compare the analytical and numerical results.

2 Using Matched Asymptotic Methods to Determine the Survival Probability

2.1 Summary of Analytical Method

Our approach is as follows. As shown in Fig. 1 we divide the domain into two regions [29]: a broad $O(1)$ region (I) containing the minimum of the potential, $X = 0$, and a narrow $O\left(1/\beta\right)$ boundary layer near $X = \beta$. We solve the limiting differential equations in these regions, from which we develop a uniform composite solution for the probability density of the survival probability using asymptotic matching, which is a highly accurate general analytical method [30], recently used to obtain the solution to the related problem of stochastic resonance [31].

2.2 Previous Results on the Survival Probability of an OU Process

The canonical approach of finding the survival probability of this OU process is to determine the probability distribution for the first hitting time in Laplace space

$$\tilde{\zeta}(\lambda, x) = \int_0^\infty e^{-\lambda t}\zeta(t, x)dt, \tag{4}$$

which leads to the following analytical expression [32–34],

$$\tilde{\zeta}(\lambda, x) = e^{\frac{a(x^2-\beta^2)}{4b}}\frac{D_{-\lambda/a}(-x\sqrt{a/b})}{D_{-\lambda/a}(-\beta\sqrt{a/b})}, \tag{5}$$

where $D_\lambda(z)$ is the parabolic cylinder function, β is the boundary position and x is the initial position.

This equation can be written as a spectral decomposition [35], wherein the countable number of eigenvalues correspond to the zeros of the denominator of Eq. (5). The final expression for the probability distribution of the first hitting time can be obtained by inverting each term of this spectral decomposition, which can be written as a weighted sum of

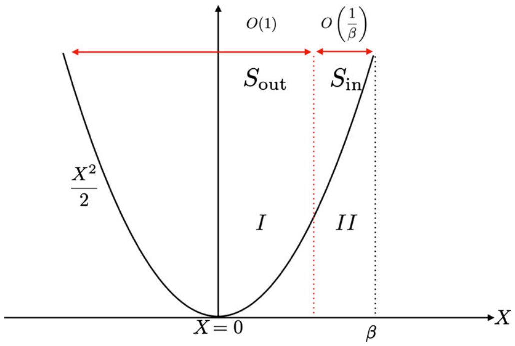


Fig. 1 Schematic of the problem under study; what is the probability that a particle reaches the edge of the potential well? We divide the potential into an "outer" region I and a boundary layer II, or the "inner" region, of the potential, in each of which the asymptotically dominant solutions are determined and then matched

exponential functions as

$$\zeta(t,x) = \sum_{p=0}^{\infty} e^{\lambda_p(\beta)t} \frac{\Phi(\lambda_p(\beta)/a, x\sqrt{a/b})}{\Phi'(\lambda_p(\beta)/a, \beta\sqrt{a/b})}, \tag{6}$$

where $\lambda_p(\beta)$ is the solution of

$$\Phi(\lambda_p(\beta)/a, \beta\sqrt{a/b}) = 0, \tag{7}$$

and Φ and Φ' are defined in [35] in terms of Kummer and digamma functions.

For large values of t and $\beta - x$, Eq. (6) simplifies considerably. In fact, because all the eigenvalues, $\lambda_p(\beta)$, are negative, only that with the smallest magnitude will contribute significantly as $t \to \infty$. Moreover, for large values of $\beta - x$, Eq. (6) becomes

$$\begin{aligned} \lim_{\beta - x \to \infty} \zeta(t,x) &= \frac{1}{t_1(\beta;0)} \exp\left[-\frac{t}{t_1(\beta;0)}\right] \quad \text{with} \\ t_1(\beta;0) &= \frac{1}{a\beta}\sqrt{\frac{2\pi b}{a}} \exp\left[\frac{a\beta^2}{2b}\right], \end{aligned} \tag{8}$$

where $t_1(\beta;0)$ is the mean first passage time of the process from $X(t=0)=0$ to the boundary β [36].

Importantly, however, the asymptotic expression in Eq. (8) is not valid when $x \approx \beta$. Indeed, as we show here, when the process starts in the neighborhood of $x = \beta$ there is a non-trivial leakage of probability. This leakage is not taken into account by Eq. (8) and transpires very rapidly, on a time scale of order $1/\beta^2$. Therefore, taking this approach requires that one calculate an enormous number of eigenvalues in Eq. (6), which is computationally inefficient. Our approach avoids this problem.

2.3 Detailed Development of the Approach

We derive an asymptotic expression for the survival probability, which is the time integral of the probability distribution of the first passage time, and in the large β limit it is trivial to evaluate when $x \approx \beta$.

The probability density, $\rho(y,t;x,s)$, of the OU process in Eq. (3) is described by the Kolmogorov forward (KFE) and backward (KBE) equations, the former of which is

$$\begin{aligned} \partial_t \rho(y,t;x,s) &= a\,\partial_y[y(t)\rho(y,t;x,s)] \\ &\quad + b\,\partial_{yy}\rho(y,t;x,s) \equiv \mathcal{L}_y \rho(y,t;x,s), \end{aligned} \tag{9}$$

where the operator $\mathcal{L}_y$ is the generator of the OU process viz., $[\mathcal{L}_y f](y,t) = a\,\partial_y[yf(y,t)] + b\,\partial_{yy} f(y,t)$. The KBE follows by replacing $\mathcal{L}_y$ with its adjoint, $\mathcal{L}_x^*$, defined as $[\mathcal{L}_x^* f](x,s) = -ax\,\partial_x f(x,s) + b\,\partial_{xx} f(x,s)$.

The KFE gives the evolution of the probability density of the process when the *initial* position and time, (x,s), are known, while the KBE treats the evolution when the *final* position and time, (y,t), are known. Both equations have initial condition

$$\rho(y,u;x,u) = \delta(y-x), \tag{10}$$

and boundary conditions

$$\rho(\pm\infty,t;x,s) = 0 \quad \text{and} \quad \rho(y,t;\pm\infty,s) = 0 \tag{11}$$

for the KFE and the KBE respectively.

The survival probability within the interval (α, β) is defined in terms of the KBE density, ρ_{KBE}, as

$$S(t; x, s) = \int_{\alpha}^{\beta} \rho_{KBE}(y, t; x, s)dy, \tag{12}$$

which satisfies

$$-\partial_s S(t; x, s) = -a\, x(s)\partial_x S(t; x, s) + b\, \partial_{xx} S(t; x, s), \tag{13}$$

with initial condition

$$S(t = s; x, s) = \Theta(x - \alpha)\Theta(\beta - x), \tag{14}$$

and boundary conditions

$$S(t; x = \alpha, s) = S(t; x = \beta, s) = 0 \quad \forall\, s \leq t, \tag{15}$$

where $\Theta(\cdot)$ is the Heaviside theta function.

Let P_i^{α} (P_i^{β}) be the probability of hitting $x = \alpha(\beta)$ for the first time after i time steps and let S^{α} (S^{β}) be the survival probability with an absorbing boundary at $x = \alpha(\beta)$. Hence, if we discretize the stochastic process, the survival probability after n time steps is

$$\begin{aligned} S_n &= \prod_{i=1}^{n}(1 - P_i^{\alpha} - P_i^{\beta}) = \prod_{i=1}^{n}[(1 - P_i^{\alpha})(1 - P_i^{\beta}) - P_i^{\alpha} P_i^{\beta}] \\ &\simeq \prod_{i=1}^{n}(1 - P_i^{\alpha})\prod_{i=1}^{n}(1 - P_i^{\beta}) = S_n^{\alpha} S_n^{\beta}. \end{aligned} \tag{16}$$

In the second line of Eq. (16), we have neglected the term $P_i^{\alpha} P_i^{\beta}$ for $|\alpha| \gg 1$ and $|\beta| \gg 1$, allowing us to write the survival probability in the interval (α, β) as the product of the two survival probabilities in the two intervals $(-\infty, \beta)$ and (α, ∞) with $\alpha < \beta$. From this point on we will only consider the survival probability in the interval $(-\infty, \beta)$, and note that the derivation for the interval (α, ∞) follows in straightforward analogy.

We now rewrite Eq. (13) in a rescaled form,

$$\partial_t S(x, t) = -x(t)\partial_x S(x, t) + \partial_{xx} S(x, t), \tag{17}$$

with the new variables,

$$x \to x\sigma, \quad \beta \to \beta\sigma \;\text{ and }\; s \to -\frac{t}{a}, \tag{18}$$

in which the spatial coordinate is expressed in terms of the standard deviation $\sigma = \sqrt{b/a}$ of a stationary OU process obtained from the solution of Eq. (9) in the limit $t \to \infty$. The new initial and boundary conditions are

$$\begin{cases} S(x, t = 0) = \Theta(\beta - x), \\ S(x = \beta, t) = 0, \text{ and} \\ S(x = -\infty, t) = 1. \end{cases} \tag{19}$$

We solve the limiting differential equations within the two regions, from which we construct an approximate uniform solution by asymptotic matching. We denote $S_{\text{out}}(x, t)$ and $S_{\text{in}}(x, t)$ as the solutions in region I and II respectively.

Region I. The outer solution is obtained by imposing $\beta - x \gg 1$. In this limit the probability distribution of the first hitting time in Eq. (8) is valid, and its time integral gives the survival probability as

$$S_{\text{out}}(t) \simeq \mathrm{e}^{-\left(\frac{\beta}{\sqrt{2\pi}}\mathrm{e}^{-\frac{\beta^2}{2}}\right)t}, \tag{20}$$

where the rescaled β and t of Eq. (18) have been used.

Region II. In the boundary layer, or the inner region, we have $x \sim \beta$, where the approximation of Region I is no longer valid. We let $\epsilon \equiv 1/\beta \ll 1$ and introduce the following stretched coordinates,

$$\eta = \frac{x - \frac{1}{\epsilon}}{\epsilon} \text{ and } \theta = \frac{t}{\epsilon^2}, \tag{21}$$

that we use to rewrite Eq. (17) as

$$\begin{aligned} \frac{1}{\epsilon^2}\partial_\theta S_{\text{in}}(\eta,\theta) = &-\left[\frac{1}{\epsilon^2} + \eta(\theta)\right]\partial_\eta S_{\text{in}}(\eta,\theta) \\ &+ \frac{1}{\epsilon^2}\partial_{\eta\eta} S_{\text{in}}(\eta,\theta), \end{aligned} \tag{22}$$

which at leading-order becomes

$$\partial_\theta S_{\text{in}}(\eta,\theta) = -\partial_\eta S_{\text{in}}(\eta,\theta) + \partial_{\eta\eta} S_{\text{in}}(\eta,\theta). \tag{23}$$

Clearly, Eq. (23) is a diffusion equation for $S_{\text{in}}(\eta,\theta)$ along the characteristics

$$\frac{d\eta}{d\theta} = 1 \text{ and } \frac{d\rho}{d\theta} = 1, \tag{24}$$

which we can then write as

$$\partial_\rho S_{\text{in}}(\mu,\rho) = \partial_{\mu\mu} S_{\text{in}}(\mu,\rho), \tag{25}$$

wherein $\mu \equiv \eta - \theta$ and $\rho \equiv \theta$, so that the boundary condition becomes $S_{\text{in}}(\mu = -\rho, \rho) = 0$.

We solve Eq. (25) by first finding its Green's function, $G(\rho,\mu;\nu)$, which satisfies

$$\partial_\rho G(\mu,\rho;\nu) - \partial_{\mu\mu} G(\mu,\rho;\nu) = \delta(\mu - \nu)\delta(\rho). \tag{26}$$

This Green's function is associated with the probability density ρ_{KBE}, satisfying the KBE, and hence the survival probability through Eq. (12). Hence, this density satisfies the following conditions;

$$\begin{cases} \rho_{KBE}(\mu, \rho = 0; \nu) = \delta(\mu - \nu), \\ \rho_{KBE}(\mu = 0, \rho; \nu) = \rho_{KBE}(\mu = -\infty, \rho; \nu) = 0. \end{cases} \tag{27}$$

The Green's function is (e.g., [37])

$$\begin{aligned} G(\rho,\mu;\nu) = &\frac{1}{\sqrt{4\pi\,\rho}}\left(\exp\left[-\frac{(\nu-\mu)^2}{4\,\rho}\right]\right. \\ &\left. - \exp\left[-\frac{(\nu+\mu)^2}{4\,\rho} - \nu\right]\right), \end{aligned} \tag{28}$$

which, as noted above, now allows us to write the solution of Eq. (25) as

$$\begin{aligned} S_{\mathrm{in}}(\mu, \rho) &= \int_{-\infty}^{0} d\nu \int_{-\infty}^{\infty} d\phi\, G(\mu, \rho; \phi)\delta(\phi - \nu) \\ &= \int_{-\infty}^{0} d\nu\, G(\mu, \rho; \nu). \end{aligned} \tag{29}$$

Upon integration and reversion to the original variable t and to the stretched coordinate η, we find

$$\begin{aligned} S_{\mathrm{in}}(\eta, t) \simeq \frac{K}{2} \mathrm{e}^{\eta} \Bigg(&- \mathrm{erfc}\left[\frac{1}{2}\sqrt{\frac{1}{t}}(-t - \eta)\right] \\ &+ \mathrm{e}^{-\eta}\mathrm{erfc}\left[\frac{1}{2}\sqrt{\frac{1}{t}}(-t + \eta)\right]\Bigg) \equiv\ K\, S'_{\mathrm{in}}(\eta, t). \end{aligned} \tag{30}$$

We determine the constant K by requiring the outer limit of the inner solution to equal the outer solution;

$$K = \lim_{\eta \to -\infty} S_{\mathrm{in}}(\eta, t) = S_{\mathrm{out}}(t) \simeq \mathrm{e}^{-\left(\frac{\beta}{\sqrt{2\pi}}\mathrm{e}^{-\frac{\beta^2}{2}}\right)t}. \tag{31}$$

Therefore, the uniformly valid approximate composite analytical solution for the survival probability is

$$\begin{aligned} S(x, t) &= S_{\mathrm{in}}(x, t) + S_{\mathrm{out}}(t) - K \\ &\simeq \frac{1}{2}\mathrm{e}^{-\left(\frac{\beta}{\sqrt{2\pi}}\mathrm{e}^{-\frac{\beta^2}{2}}\right)t}\, \mathrm{e}^{\beta(x-\beta)} \\ &\quad \times \left(- \mathrm{erfc}\left[\frac{1}{2}\sqrt{\frac{1}{t}}(-t - \beta(x - \beta)\right)\right) \\ &\quad + \mathrm{e}^{-\beta(x-\beta)}\mathrm{erfc}\left[\frac{1}{2}\sqrt{\frac{1}{t}}(-t + \beta(x - \beta))\right]\Bigg) \\ &\equiv S'{}_{\mathrm{in}}(x, t) S_{\mathrm{out}}(t), \end{aligned} \tag{32}$$

written in terms of the original spatial coordinate x rather that the stretched coordinate η.

2.4 Comparing Analytical and Numerical Results

We compare Eq. (32) with the numerical solution of Eq. (17) for different values of the parameter β in Fig. 2, and find excellent agreement even when β is *not* asymptotically large. Indeed, the smallest value of the distance to the boundary is $\beta = 1.645\sigma$, corresponding to a 90% probability that an unbounded stationary OU process remains below the boundary.

When $t \gg 1/\beta^2$ the left hand side of Eq. (23) is negligible and Eq. (32) becomes $S_{\mathrm{out}}(t)[1 - e^{\beta(x-\beta)}]$ and depends on time solely through $S_{\mathrm{out}}(t)$, which is the prefactor of $S_{\mathrm{in}}(x, t)$. Thus, depending on x the rate of the decrease in the survival probability is controlled by $S_{\mathrm{in}}(x, t)$, decaying more rapidly near the boundary for early times.

The accuracy of the asymptotic solutions over a wide range of the β facilitates simple and wide ranging applications. For example, determining the input parameters of a leaky integrate-and-fire (LIF) neural model (see e.g., [38] and refs therein) based on experimentally

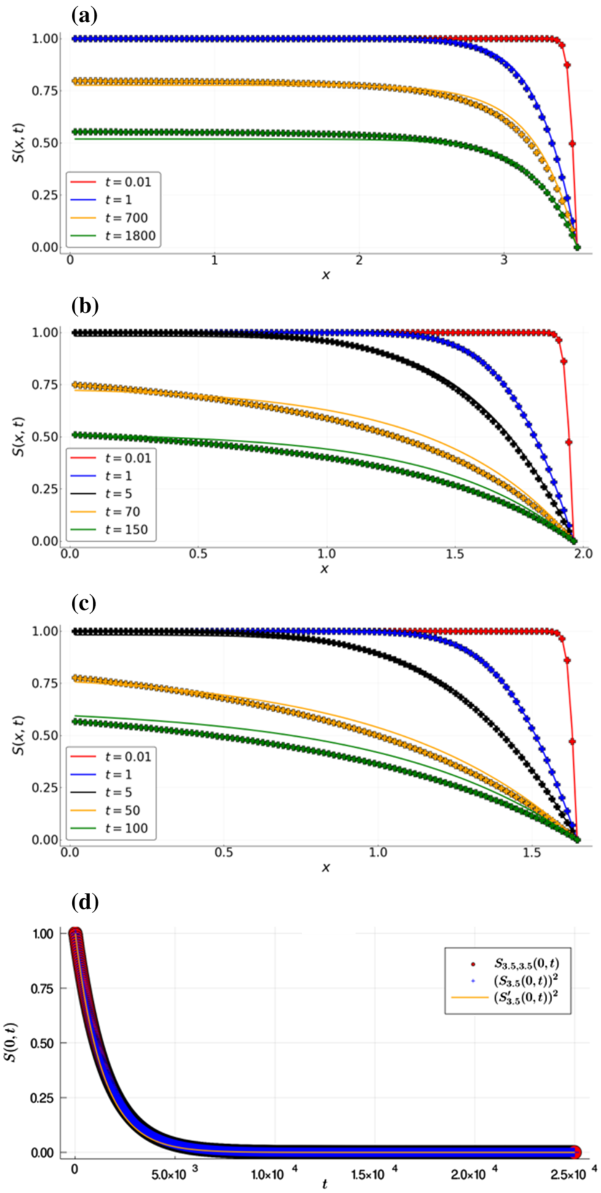


Fig. 2 **a**–**c** Plot of Eq. (32), $S(x, t)$, versus $x \in [0, \beta]$ for different values of t (solid lines) compared to the numerical solution of Eq. (17) (crosses). We use three different values of the boundary position β; **a** 3.5, **b** 1.96, and **c** 1.645, corresponding to the probability of finding the system below $x = \beta$ as $t \to \infty$ and without absorbing boundaries of **a** 99.95%, **b** 99%, and **c** 90% respectively. **d** The numerical solution of Eq. (17) with two boundaries (red circles), $S_{\alpha,\beta}(x, t)$, $S_\alpha(x, t)S_\beta(x, t)$ (blue crosses), and the approximate analytical solution in Eq. (33), $S'_\alpha(x, t)S'_\beta(x, t)$. Note that $S_{\alpha,\beta}(x, t)$ overlaps exactly with $S_\alpha(x, t)S_\beta(x, t)$. Here $\alpha = \beta = 3.5$ and $x = 0$

observable interspike intervals [28]. Of particular contemporary relevance is deducing the "critical community size" in disease epidemiology [27], or the population threshold below which infections do not persist.

Finally, we appeal to Eq. (16) to form the asymptotic expression for the survival probability of an Ornstein–Uhlenbeck process with two absorbing boundaries as

$$S_{\alpha,\beta}(x,t) = S'^{\alpha}_{\text{in}}(x,t)S^{\alpha}_{\text{out}}(t)S'^{\beta}_{\text{in}}(x,t)S^{\beta}_{\text{out}}(t)\ \forall t > s. \tag{33}$$

We note that only one of the boundary regions will contribute significantly viz.,

$$\begin{aligned} S'^{\alpha,\beta}_{\text{in}}(x,t) &= S'^{\alpha}_{\text{in}}(x,t)S'^{\beta}_{\text{in}}(x,t) = S'^{\gamma}_{\text{in}}(x,t), \\ \gamma &= \min_{|h_\alpha|,|h_\beta|}(\alpha,\beta), \end{aligned} \tag{34}$$

where $h_{\alpha(\beta)} = x - \alpha(\beta)$.

In Fig. 2d we show that the numerical solution with two boundaries matches that obtained by considering the two boundaries separately and with the analytical solution in Eq. (33).

3 Conclusion

We have obtained an asymptotic analytical solution for the survival probability of an Ornstein–Uhlenbeck process for large potentials. We divide the potential into two layers near the boundaries and a broad region between, which contains the origin, where we use the solution of Ricciardi and Sato [35]. The uniformly continuous solution is obtained by matching the two approximate solutions in the boundary layers with that in the broad region between. The solution agrees extremely well with both the numerical solution and with the more restricted asymptotic expression for the survival probability known in literature. Importantly, our analysis remains valid even when the initial position of the stochastic process is close to one of the boundaries, and furthermore it takes into account the non-negligible leakage of the probability early in the time evolution.

Despite the analysis using the assumption of asymptotically large boundaries, we showed that it agrees well with the numerical solution, even when the boundary positions are the same order of magnitude as the standard deviation of the stationary probability distribution function of the OU process. Indeed, we demonstrate consistency when there is a 90% probability that the system without absorbing boundaries is found at a position less than β; when reaching the boundary can no longer be considered a rare event. Therefore, our method can be easily extended to more general (less restrictive) settings. Finally, our compact analytical solution provides a computationally trivial framework for survival analysis of use across the broad spectrum of stochastic systems where the Ornstein–Uhlenbeck process arises.

Acknowledgements We gratefully acknowledge support from the Swedish Research Council Grant No. 638-2013-9243.

Funding Open access funding provided by Stockholm University.

References

1. Ricciardi, L.M., Sacerdote, L.: The Ornstein–Uhlenbeck process as a model for neuronal activity. Biol. Cybern. **35**(1), 1–9 (1979)
2. O'Meara, B.C., Beaulieu, J.M.: Modelling stabilizing selection: the attraction of Ornstein–Uhlenbeck models. In: Garamszegi, L. (ed.) Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology, pp. 381–393. Springer, Berlin (2014)
3. Trost, D.C., Overman, E.A., Ostroff, J.H., Xiong, W., March, P.: A model for liver homeostasis using modified mean-reverting Ornstein–Uhlenbeck process. Comput. Math. Methods Med. **11**(1), 27–47 (2010)
4. Schöbel, R., Zhu, J.: Stochastic volatility with an Ornstein–Uhlenbeck process: an extension. Rev. Finance **3**(1), 23–46 (1999)
5. Hasselmann, K.: Stochastic climate models Part I. Theory. Tellus **28**(6), 473–485 (1976)
6. Moon, W., Wettlaufer, J.S.: A unified nonlinear stochastic time series analysis for climate science. Sci. Rep. **7**, 44228 (2017)
7. Moon, W., Agarwal, S., Wettlaufer, J.S.: Intrinsic pink-noise multidecadal global climate dynamics mode. Phys. Rev. Lett. **121**, 108701 (2018)
8. Pepe, F.V., Facchi, P., Kordi, Z., Pascazio, S.: Nonexponential decay of Feshbach molecules. Phys. Rev. A **101**, 013632 (2020)
9. Maniscalco, S., Piilo, J., Suominen, K.-A.: Zeno and anti-Zeno effects for quantum brownian motion. Phys. Rev. Lett. **97**, 130402 (2006)
10. Giacosa, F., Pagliara, G.: Measurement of the neutron lifetime and inverse quantum Zeno effect. Phys. Rev. D **101**, 056003 (2020)
11. Zacks, S.: Introduction to Reliability Analysis: Probability Models and Statistical Methods. Springer, New York (2012)
12. McNeil, A.J., Frey, R., Embrechts, P.: Quantitative Risk Management: Concepts, Techniques and Tools-Eevised Edition. Princeton University Press, Princeton (2015)
13. Blossfeld, H., Hamerle, A., Mayer, K.U.: Event History Analysis: Statistical Theory and Application in the Social Sciences. Psychology Press, New York (2014)
14. Tuckwell, H.C.: Introduction to Theoretical Neurobiology (volume 2): Nonlinear and Stochastic Theories. Cambridge University Press, Cambridge (1988)
15. Mode, C.J., Sleeman, C.K.: Stochastic Processes in Epidemiology: HIV/AIDS, Other Infectious Diseases, and Computers. World Scientific, Singapore (2000)
16. Madec, Y., Japhet, C.: First passage time problem for drifted Ornstein–Uhlenbeck process. Math. Biosci. **189**, 131–140 (2004)
17. Leblanc, B., Scaillet, O.: Path dependent options on yields in the affine term structure model. Finance Stoch. **2**(4), 349–367 (1998)
18. Linetsky, V.: Computing hitting time densities for CIR and OU diffusions: applications to mean-reverting models. J. Comp. Finance **7**, 1–22 (2004)
19. Jeanblanc, M., Rutkowski, M.: Modelling of default risk: an overview. In: Modern Mathematical Finance: Theory and Practice. Higher Ed. Press, Beijing (2000)
20. Majumdar, S.N., Pal, A., Schehr, G.: Extreme value statistics of correlated random variables: a pedagogical review. Phys. Rep. **840**, 1 (2020)
21. Gautié, T., Le Doussal, P., Majumdar, S.N., Schehr, G.: Non-crossing Brownian paths and Dyson Brownian motion under a moving boundary. J. Stat. Phys. **177**(5), 752–805 (2019)
22. Bray, A.J., Majumdar, S.N., Schehr, G.: Persistence and first-passage properties in nonequilibrium systems. Adv. Phys. **62**, 225–361 (2013)
23. Redner, S.: A Guide to First-Passage Processes. Cambridge University Press, Cambridge (2001)
24. Alili, L., Patie, P., Pedersen, J.L.: Representations of the first hitting time density of an Ornstein–Uhlenbeck process. Stoch. Models **21**(4), 967–980 (2005)
25. Martin, R.J., Kearney, M.J., Craster, R.V.: Long- and short-time asymptotics of the first-passage time of the Ornstein–Uhlenbeck and other mean-reverting processes. J. Phys. A: Math. Theor. **52**, 134001 (2019)
26. Darling, D.A., Siegert, A.J.F.: The first passage problem for a continuous Markov process. Ann. Math. Stat. **24**, 624–639 (1953)
27. Dolgoarshinnykh, R.G., Lalley, S.P.: Critical scaling for the SIS stochastic epidemic. J. App. Prob. **43**(3), 892–898 (2006)
28. Ditlevsen, S., Lansky, P.: Estimation of the input parameters in the Ornstein–Uhlenbeck neuronal model. Phys. Rev. E **71**(1), 011907 (2005)
29. We note that in the parlance of the field, the boundary layer is the *inner region* and the remainder of the domain is the *outer region*, although in this case the latter is in the interior of the potential

30. Bender, C.M., Orszag, S.A.: Advanced Mathematical Methods for Scientists and Engineers I: Asymptotic Methods and Perturbation Theory. Springer, New York (2013)
31. Moon, W., Balmforth, N.J., Wettlaufer, J.S.: Nonadiabatic escape and stochastic resonance. J. Phys. A: Math. Theor. **53**, 095001 (2020)
32. Siegert, A.J.: On the first passage time probability problem. Phys. Rev. **81**(4), 617 (1951)
33. Srinivasan, S.K., Sampath, G.: Stochastic Models for Spike Trains of Single Neurons, vol. 16. Springer, New York (2013)
34. Ricciardi, L.M.: Diffusion Processes and Related Topics in Biology, vol. 14. Springer, New York (2013)
35. Ricciardi, L.M., Sato, S.: First-passage-time density and moments of the Ornstein–Uhlenbeck process. J. App. Prob. **25**(1), 43–57 (1988)
36. Nobile, A.G., Ricciardi, L.M., Sacerdote, L.: Exponential trends of Ornstein–Uhlenbeck first passage time densities. J. App. Prob. **22**(2), 360–369 (1985)
37. Duffy, D.G.: Green's Functions with Applications. Chapman and Hall/CRC, Boca Raton (2015)
38. Lansky, P., Ditlevsen, S.: A review of the methods for signal estimation in stochastic diffusion leaky integrate-and-fire neuronal models. Biol. Cybern. **99**(4–5), 253 (2008)

Paper III

S.H. Lim, L.T. Giorgini, W. Moon, and J.S. Wettlaufer, Predicting rare events in multiscale dynamical systems using machine learning, *Chaos* **30**, 123126 (2020).

Chaos ARTICLE scitation.org/journal/cha

Predicting critical transitions in multiscale dynamical systems using reservoir computing

Cite as: Chaos **30**, 123126 (2020); doi: 10.1063/5.0023764
Submitted: 31 July 2020 · Accepted: 13 November 2020 ·
Published Online: 10 December 2020

Soon Hoe Lim,[1,a)] Ludovico Theo Giorgini,[1] Woosok Moon,[1,2] and J. S. Wettlaufer[1,3]

AFFILIATIONS

[1] Nordita, KTH Royal Institute of Technology and Stockholm University, 106 91 Stockholm, Sweden
[2] Department of Mathematics, Stockholm University, 106 91 Stockholm, Sweden
[3] Yale University, New Haven, Connecticut 06520, USA

[a)] **Author to whom correspondence should be addressed:** soon.hoe.lim@su.se

ABSTRACT

We study the problem of predicting rare critical transition events for a class of slow–fast nonlinear dynamical systems. The state of the system of interest is described by a slow process, whereas a faster process drives its evolution and induces critical transitions. By taking advantage of recent advances in reservoir computing, we present a data-driven method to predict the future evolution of the state. We show that our method is capable of predicting a critical transition event at least several numerical time steps in advance. We demonstrate the success as well as the limitations of our method using numerical experiments on three examples of systems, ranging from low dimensional to high dimensional. We discuss the mathematical and broader implications of our results.

Critical transitions are ubiquitous in nature. These transition events are often induced by a fast driving signal and are rare and random. Since such events could lead to significant effects, it is important to develop effective methods to predict signal-induced critical transitions early. Recently, many studies have been devoted to exploring early warning indicators to predict and characterize the onset of critical transitions. In this paper, we propose and test an alternative method to predict critical transitions within a class of multiscale dynamical systems. Our method is data-driven, inspired by recent advances in reservoir computing, and takes into account the multiscale nature of the systems. We demonstrate the success as well as the limitations of our method using numerical experiments on both low- and high-dimensional systems. We anticipate that our work will serve as a catalyst for further progress in tackling the problem of predicting critical transitions using scientific machine learning.

I. INTRODUCTION

The dynamics of many systems in nature are nonlinear, multiscale, and noisy, making both the theoretical and numerical modeling and prediction of their states challenging. Of particular interest are those dynamics that often lead to rare transition events. Particularly, the system under study spends very long periods of time in various metastable states and only very rarely, and at seemingly random times, does it hop between states. Such sudden changes in the dynamical behavior of complex systems are known as *critical transitions*,[43,71] occurring at so-called *tipping points*.[3]

One mechanism that explains the hopping behavior is that it is induced by a *fast signal* influencing the state of the otherwise closed system. The usually noisy driving signal comes from an external source. It is often the case that there is a large separation of time scales on which the system state and the signal evolve. Without the driving signal, the system will remain in one state forever. Understanding the dynamics of such system requires us to study the ensemble of transition paths between the different metastable states.[19,21,35]

The above signal-induced phenomenon can be modeled by a class of non-autonomous open dynamical systems, whose state $x : [0, T] \to \mathbb{R}^n$, $T > 0$, is described by

$$\dot{x}(t) = F(x(t), t) + f(t), \tag{1}$$

where $F(x(t), t)$ is a force field, providing a deterministic backbone, and $f(t)$ is a fast driving noisy signal, describing small perturbations imparted to the system. In principle, one does not have *a priori* knowledge of the mathematical model for $f(t)$ but only has access to data for x and, possibly, a mathematical model for $F(x(t), t)$ at their disposal. The unknown $f(t)$ is generally a random function, and it could be regular, chaotic, or stochastic. It is challenging to infer an accurate model for $f(t)$ using only the data for x. In fact, in many cases, one could construct several models that are capable of describing the data equally well, and often there are ambiguities in the choice of model. An example that illustrates this issue is the problem of distinguishing deterministic chaos from stochasticity.[1,40] Indeed, on one hand, when certain assumptions are satisfied, chaotic systems can be well approximated by a stochastic one.[5,16,53] On the other hand, many stochastic systems can be described by a chaotic model.[26] In the absence of a uniquely defined model for $f(t)$, one has to resort to a model-free, data-driven, approach, which is currently an extremely active area in applied dynamical systems. We refer the reader to the book by Brunton and Kutz[13] for an excellent review of data-driven methods.

An important and practical question concerning the system (1) is the following—can we predict the future evolution of the system state using only its data? Moreover, can we achieve this for a sufficiently long time and to a desired level of accuracy and certainty? Note that this is different from asking one to infer a mathematical model (which may have a very low predictive power, as for example in the case of a random walk model) from the data. Typically, the data are complex and multiscale in nature, thereby complicating analysis and prediction. *The main goal of this work is to propose and test a machine learning based solution to predict rare events in multiscale noisy nonlinear dynamical systems, making use of only the (slow) system state data and a partial knowledge of the physics of the generating system.* Since the occurrence of rare events can have significant deleterious or positive implications, it is important to quantify and predict them in advance to inform decision making[27]—this is the main motivation of this paper. To the best of our knowledge, this is the first attempt to solve this particular problem using a machine learning method. Here, we will focus on the case where the system of interest is one-dimensional. While we focus on prediction of rare events, our method can be applied to predicting future states in general multiscale systems.

By exploiting recent advances in machine learning, we construct an algorithm to solve the prediction problem. The field of machine learning is experiencing a major recent resurgence of interest, with wide ranging applications and significant implications in many areas of science and engineering.[14,55] We note, in particular, neural networks and deep learning whose industry wide applications have been made possible due to availability of large amount of data and advances in computer hardware development.[31] For instance, by training on a sufficiently large set of data, one can classify handwritten digits to unprecedented accuracy,[44] predict and analyze time series,[78] infer the Hölder exponent of stochastic processes,[74] characterize anomalous diffusion,[11] learn how to construct linear embeddings of nonlinear dynamics,[52] replicate chaotic attractors and calculate Lyapunov exponents,[61] solve high-dimensional nonlinear PDEs,[65] and others that are too numerous to list here. As powerful as they appear, we emphasize that it is by no means an easy task to apply the tools of machine learning to real world datasets and one needs to proceed with caution and avoid potential pitfalls.[66] Indeed, one of the main challenges in machine learning deals with the ability of the algorithm to be generalized to unseen data. Moreover, the empirical approach of machine learning is refined only with practices that have principally been discovered by trial and error. Our approach falls within the scope of scientific machine learning,[4] an emerging research area that is focused on the opportunities and challenges of machine learning in the context of applications across science and engineering.

This paper is organized as follows. In Sec. II, we motivate and describe the class of dynamical systems of interest from which the data are generated. They are special cases of (1) with a specified model for $f(t)$. We then introduce the problem of data-based rare event prediction. In Sec. III, we present and explain in detail a reservoir computing based method to solve the prediction problem using a variant of the echo state network (ESN). In Sec. IV, we apply and test the method to predict rare events in three different systems. We make concluding remarks in Sec. V. The Appendix contains further details on training, validation (model selection), and testing.

II. MULTISCALE NOISY SYSTEMS AND THE PREDICTION PROBLEM

We consider scenarios where the available data are generated by the following family of continuous-time slow–fast systems (parameterized by $\epsilon > 0$):[9,63,67]

$$\dot{x}(t) = F(x(t), t) + \frac{\sigma}{\epsilon} G(\xi(t)), \tag{2}$$

$$\dot{\xi}(t) = \frac{1}{\epsilon^2} H(\xi(t)), \tag{3}$$

with the initial conditions $x(0) = x_0$ and $\xi(0) = \xi_0$, where x_0 and ξ_0 can be either fixed or random. In the case when both initial conditions are fixed, (2) and (3) describe a deterministic dynamical system; otherwise, it is a random dynamical system. In the above, $x : I \to \mathbb{R}$ is a slow process, $\xi : I \to \mathbb{R}^m$ is a fast process (noise), $F : \mathbb{R} \times I \to \mathbb{R}$ is a deterministic force field, $G : \mathbb{R}^m \to \mathbb{R}$ is a functional (observable) of the fast process, $H : \mathbb{R}^m \to \mathbb{R}^m$ is a deterministic vector field describing the fast process, $\sigma > 0$ is a small constant controlling the strength of influence of the fast process on the slow one, and I is a time interval on which the system evolves. Here F, G, and H are $O(1)$ as $\epsilon \to 0$. In many applications of interest, F, G, and H may be highly nonlinear. The assumption that the data are modeled by the above systems are not too restrictive, for as is very often the case in the analysis of experimental data, there is not a unique single model for the system generating the data.

Often one is interested in the dynamics of the slow process and the case where the driving signal, $f(t) = \frac{\sigma}{\epsilon} G(\xi(t))$, is a stochastic process such as Gaussian white noise. A crucial property of dynamical systems of the form (2) and (3) is that stochastic behavior emerges as ϵ becomes smaller. Indeed, it can be rigorously shown that under appropriate assumptions on the initial conditions, F, G, and H, $x(t)$ converges in law (homogenizes) to $X(t)$ as $\epsilon \to 0$ for $t \in I$, where $X(t)$ is a diffusion process solving the stochastic differential equation (SDE),[16,56]

$$dX(t) = F(X(t), t)dt + \tilde{\sigma}\, dW(t). \tag{4}$$

Here, $\tilde{\sigma} > 0$ is a constant (involving the time integral of the correlation of the fast process) and $W(t)$ is a Wiener process. Usually, a mixing assumption is imposed on the fast flow. However, such an assumption is not necessary for homogenization. The system only needs to be sufficiently chaotic. See, for instance, Remark 2.1 in Ref. 32 and also Refs. 16, 41, and 42.

Therefore, the family of Eqs. (2) and (3) can be viewed as approximations of a stochastic model. Such a perspective has been adopted to study a variety of noisy systems.[2,10,36,57] We refer to Sec. IV for concrete examples. An important class of $\xi(t)$ are those that exhibit deterministic chaos, which has been studied and observed[26] in a range of physical systems. Note that G can be generalized to be dependent on $x(t)$, in which case the limiting SDE will have a multiplicative noise, but we will not pursue this case here.

We now formulate the prediction problem introduced in Sec. I systematically. First, we lay out our assumptions. The only data available to us are that for x, which is assumed to be generated by the systems (2) and (3), and we do not have access to the data for ξ. Let $I = [t_{init}, t_{obs}] \cup [t_{obs}, t_f]$, where $t_{init} < t_{obs} < t_f$. Here, t_{init} denotes an initial time, t_{obs} denotes an observation time, beyond which we do not have access to system state, and t_f denotes the final time at which the prediction will be made. Suppose that we are given a sufficiently long and high-frequency time series for x on $[t_{init}, t_{obs}]$. Furthermore, we are blind to the actual mathematical model for the fast process. However, a partial knowledge of the physics of the system of interest is known. In particular, it is assumed that we know the exact expression describing the force field F. This assumption is satisfied when one can reconstruct the force field from the data accurately, which may be possible in many practical situations.[25]

We are given a time series data for x, a sequence $(x_k)_{k=0,\ldots,N} = (x(t_0), x(t_1), x(t_2), \ldots, x(t_N))$, where $t_0 = t_{init}$, $t_i = t_{init} + i\Delta t$ $(i = 1, \ldots, N)$ are the sampling times, Δt is the time step size, $t_N = t_{init} + N\Delta t =: t_{obs}$, and $N + 1$ is the number of available samples. Our time series does not record the occurrence of a rare event, and we assume that a rare event will occur shortly after time t_{obs}. Any precursors for this rare event must then be hidden in the time series. We then attempt to answer the following questions.

1. Can we predict if and when a rare event will occur in a given future time window? Can we infer the characteristics of the event?
2. How far in advance can we predict the rare event?
3. With what accuracy and certainty can we achieve these goals?
4. Is it possible to answer all of these questions with a computationally inexpensive method and/or using a relatively short time series data for x?

Clearly, these are challenging questions. The degree of difficulty depends on the characteristics of the dynamical systems generating the data for x. For a given amount of data, the difficulty increases as ε becomes smaller, in which case the statistical behavior of the driving noise is closer to that of a white noise; therefore, predictability is lost in the limit. Thus, analysis of the data should be performed on a case by case basis.

III. A MACHINE LEARNING BASED PREDICTION METHOD

A. The method

We first present a three-step procedure that will allow us to investigate the questions posed at the end of Sec. II. This procedure lies at the heart of our method. We then discuss a number of heuristic issues associated with our approach.

Algorithm III.1. *Predicting rare critical transition events.*

Under the assumptions, setting and notation described in Sec. II:

(S1) *Feature extraction.* Extract the fast driving signal using the data for x and the known expression for F,

$$f(t_i) = \frac{x(t_{i+1}) - x(t_i)}{\Delta t} - F(x(t_i), t_i) \tag{5}$$

for $i = 0, 1, \ldots, N - 1$, where Δt is a uniform time step.

(S2) *Machine learning.* Using $(f(t_i))_{i=0,\ldots,N-1}$ as the training data, predict the values of f for $M + 1$ time steps into the future (i.e., beyond time $t_{obs} - \Delta t$) using a supervised learning algorithm (for instance, Algorithm III.2 in Sec. IV) that is best fit for the task to infer $(f(t_N), f(t_{N+1}), \ldots, f(t_{N+M}))$.

(S3) *Numerical integration.* Numerically evolve the system (2) up to time t_{N+M}, with $(f(t))_{t=t_0,\ldots,t_{N+M}}$ in place of $\frac{\sigma}{\varepsilon} G(\xi(t))$ in (2), using the step size Δt. The predicted values for x in the time window $[t_{obs} + \Delta t, t_f = (N + M)\Delta t]$ are then obtained from the resulting numerical solutions.

Algorithm III.1 is the method that we propose and use for predicting rare events. It allows the prediction of the system state M time steps beyond the observation time t_{obs}, the result of which can be used to investigate the questions posed in Sec. II. To be able to answer these questions with a desired level of confidence, the predicted values should be as close as possible to the target (actual) values; i.e., the generalization (out-of-sample) error should be small.

Before we discuss the details of implementation for each of the three steps above, a few remarks are in order. A natural approach is to apply a suitable machine learning algorithm directly to the data for x and attempt to predict the future states. While this seems like a sensible approach, it is unrealistic to expect an algorithm to learn the multiscale nature of the data accurately. In fact, it is not clear beforehand which algorithm is best, and in many cases, the rare event will not be detected successfully (see Sec. IV). Our method circumvents this challenge and provides an alternative route for handling multiscale data. Moreover, the simplicity of our method provides an additional advantage.

The quality and accuracy of the prediction results will rely strongly on how well each step in the algorithm is executed. Errors will accumulate as one progresses through these steps. Indeed, (S1) involves a numerical approximation of the driving signal. In (S2), errors will arise from both the use of the training data [where the numerical errors from (S1) are hidden] as well as the machine learning algorithm itself. In (S3), an additional error due to numerical integration is inevitable. Provided that the accumulation of these errors is negligibly small and well controlled, one can learn and predict the system states with reasonably good accuracy, as demonstrated with the examples in Sec. IV. Rigorous error analysis is not the focus of the present paper and, therefore, will not be presented here.

Steps (S1) and (S3) are straightforward to implement; therefore, we must discuss (S2), whose implementation is the most challenging part of the method. We will formulate this step as the problem of learning the training data $(f(t_i))_{i=0,\ldots,N-1}$ with parameterized high-dimensional nonlinear dynamical systems.

B. Echo state network (ESN) and its deep version

There are many machine learning algorithms that one can use to implement step (S2). Algorithms from deep learning include convolutional neural networks, recurrent neural networks (RNNs), and encoder–decoder networks, each of which can be implemented using various architectures and training schemes.[31] As nonlinear state-space models, RNNs have *dynamical memory*, which means that they are capable of preserving in their internal state a nonlinear transformation of the input history. They are, therefore, particularly well suited to deal with sequential data. We will implement (S2) using a type of brain-inspired RNN known as the *echo state network* (ESN). For practical introductions and technical details on ESNs, we refer the reader to Refs. 38 and 49.

ESN belongs to the paradigm of reservoir computing[84] and is computationally less costly to train than other variants of RNNs, which typically use a backpropagation through time algorithm for gradient descent based training.[39,50] Similar to other RNNs, the ESN can, under fairly mild and general assumptions, be shown to be a universal approximator of arbitrary dynamical systems.[33] In contrast to standard design and training schemes for RNNs, but conceptually similar to the kernel methods (cf. Refs. 45 and 75), the neural network (called the reservoir) in ESNs is generated randomly and only the readout from the reservoir is trained. The outputs are linear combinations of the internal states and possibly the inputs and a constant (bias). This reduces the training to solving a linear regression problem, minimizing the mean squared error between the outputs and the target values.

It is crucial that typically the number of reservoir elements in an ESN is larger than that used in RNNs trained with a gradient descent based method, resulting in an over-parameterized neural network. The key observation is that with sufficient over-parameterization, gradient descent based methods will implicitly leave some weights describing the network relatively unchanged; therefore, the optimization dynamics will behave as if those weights are essentially fixed at their initial values.[81] Fixing these weights explicitly leads to the approach of learning with random reservoir features. Therefore, we can successfully learn with gradient descent based trained neural networks whenever we can successfully learn with ESNs.

Even though not all the weights of the network are trained, it has been shown that ESNs work surprisingly well and achieve excellent performance in many benchmark tasks, including winning the NN3 financial time series prediction challenge.[37] Recently, ESNs have been shown to predict chaotic systems remarkably well.[48,60–62,83] They may outperform other machine learning algorithms in certain prediction tasks. For instance, it has been shown that the ESN substantially outperforms the deep feed-forward neural network and the RNN with long short-term memory (LSTM) for predicting short-term evolution of a multiscale spatiotemporal Lorenz-96 system[15] (see also Ref. 79). Moreover, ESNs do not suffer from the vanishing and exploding gradient problem typically encountered when training other RNNs.[38] These results motivate our choice of (a variant of) the ESN over other machine learning methods. We emphasize, however, that the ESN may not be the most optimal network for our prediction task, and we remain mindful of its shortcomings, in particular, its sensitive dependence on the hyperparameters. We leave careful comparison of different machine learning methods for our prediction problem to future work.

To achieve our goals, we also use a deep version of the ESN [see also the discussion in Sec. IV(b)] whenever needed. Our deep echo state network (DESN) consists of organized hierarchically stacked ESNs whose architecture and training algorithm will be described in Subsections III B 1 and III B 2. Such deep ESNs are more expressive than shallow ESNs in the sense that they are able to develop in their internal states a multiple time-scale representation of the temporal information.[22,23] We remark that in contrast to feed-forward neural networks, it is often not obvious how one should construct a deep RNN.[59] In particular, different variants of deep echo state networks can be constructed depending on the task at hand.

1. Architecture of the ESN and its deep version

Similar to other RNNs, the ESN is a parameterized, high-dimensional, discrete-time, non-autonomous, nonlinear state-space model, describing a dynamical input–output relation,

$$\boldsymbol{x}(t_{n+1}) = \boldsymbol{f}(\boldsymbol{x}(t_n), \boldsymbol{u}(t_n), \boldsymbol{y}(t_n), \boldsymbol{v}(t_n), \boldsymbol{\theta}_f), \tag{6}$$

$$\boldsymbol{y}(t_{n+1}) = \boldsymbol{g}(\boldsymbol{x}(t_{n+1}), \boldsymbol{u}(t_{n+1}), \boldsymbol{v}(t_{n+1}), \boldsymbol{\theta}_g) \tag{7}$$

for $n = 0, 1, \ldots, N-1$. Here, $\boldsymbol{u}(t) \in \mathbb{R}^{n_u}$ is the input at time t, $\boldsymbol{x}(t) \in \mathbb{R}^{n_x}$ is the internal/hidden state of the reservoir, $\boldsymbol{y}(t) \in \mathbb{R}^{n_y}$ is the output, $\boldsymbol{v}(t) \in \mathbb{R}^{n_v}$ is an external perturbation (noise/regularization), $\boldsymbol{f}$ and $\boldsymbol{g}$ are generally nonlinear functions, and $\boldsymbol{\theta}_f$ and $\boldsymbol{\theta}_g$ are model parameters. The network non-linearly embeds the input into a higher dimensional space where the original problem is more likely to be solved linearly.

Our ESN is a specific implementation of the above state-space model, putting constraints on a fully connected RNN and *with the inputs* $(\boldsymbol{u}(t_0), \ldots, \boldsymbol{u}(t_{N-1}))$ *set to be a training time series* during the training phase. In the following, $t_n = n\Delta t$, for $n = 0, \ldots, N-1$, where Δt is a fixed time step size. If the number of layers, $n_L := L + 1$, is chosen to be one, i.e., a shallow ESN, then we have the following update equations for the states and outputs:

$$\boldsymbol{x}^{(0)}(t_{n+1}) = \tanh(\boldsymbol{W}^{(0)}\boldsymbol{x}^{(0)}(t_n) + b\boldsymbol{W}_b + \boldsymbol{W}_{in}\boldsymbol{u}(t_n)) + \nu\boldsymbol{\xi}_n^{(0)}, \tag{8}$$

$$\boldsymbol{y}(t_{n+1}) = \boldsymbol{W}_{out}(\boldsymbol{x}^{(0)}(t_{n+1}); b) \tag{9}$$

for $n = 0, 1, \ldots, N-1$, with $\boldsymbol{x}(t_0) := \boldsymbol{x}^{(0)}(t_0) = \boldsymbol{0}$. In the above, we have used the following vector concatenation notation: $(\boldsymbol{a}; \boldsymbol{b}) := (a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_m) \in \mathbb{R}^{n+m}$ for two vectors $\boldsymbol{a} = (a_1, a_2, \ldots, a_n) \in \mathbb{R}^n$ and $\boldsymbol{b} = (b_1, b_2, \ldots, b_m) \in \mathbb{R}^m$. Also, we have used the following convention for component-wise application of the activation function: $\tanh(\boldsymbol{a}) := (\tanh(a_1), \ldots, \tanh(a_n)) \in \mathbb{R}^n$ for $\boldsymbol{a} = (a_1, \ldots, a_n) \in \mathbb{R}^n$.

Otherwise, we employ the following deep version of ESN, which we refer to as DESN:

$$\boldsymbol{x}^{(0)}(t_{n+1}) = \tanh(\boldsymbol{W}^{(0)}\boldsymbol{x}^{(0)}(t_n) + b\boldsymbol{W}_b) + \nu\boldsymbol{\xi}_n^{(0)}, \tag{10}$$

$$\boldsymbol{x}^{(l)}(t_{n+1}) = \tanh(\boldsymbol{W}^{(l)}\boldsymbol{x}^{(l)}(t_n) + \boldsymbol{V}^{(l)}\boldsymbol{x}^{(l-1)}(t_n)) + \nu\boldsymbol{\xi}_n^{(l)} \quad l = 1, \ldots, L-1, \tag{11}$$

for

$$\boldsymbol{x}^{(L)}(t_{n+1}) = \tanh(\boldsymbol{W}^{(L)}\boldsymbol{x}^{(L)}(t_n) + \boldsymbol{V}^{(L)}\boldsymbol{x}^{(L-1)}(t_n) + \boldsymbol{W}_{in}\boldsymbol{u}(t_n)) + \nu\boldsymbol{\xi}_n^{(L)}, \tag{12}$$

$$\boldsymbol{y}(t_{n+1}) = \boldsymbol{W}_{out}(\boldsymbol{x}^{(0)}(t_{n+1}); \ldots; \boldsymbol{x}^{(L)}(t_{n+1}); b) \tag{13}$$

for $n = 0, 1, \ldots, N-1$, with the initial condition $\boldsymbol{x}(t_0) := (\boldsymbol{x}^{(0)}(t_0), \ldots, \boldsymbol{x}^{(L)}(t_0)) = \boldsymbol{0}$.

In (8)–(13), the training input $\boldsymbol{u}$ and output $\boldsymbol{y}$ come from a compact subset of $\mathbb{R}^{n_u}$ and of $\mathbb{R}^{n_y}$, respectively, the vector $\boldsymbol{x}^{(i)} \in \mathbb{R}^{n_{x_i}}$ (for $i = 0, \ldots, L$) is the ith hidden state; the constant $b \in \mathbb{R}$ introduces a bias; the matrices $\boldsymbol{W}^{(i)} \in \mathbb{R}^{n_{x_i} \times n_{x_i}}$ (for $i = 0, \ldots, L$), $\boldsymbol{W}_b \in \mathbb{R}^{n_{x_0}}$, $\boldsymbol{V}^{(i)} \in \mathbb{R}^{n_{x_i} \times n_{x_{i-1}}}$ (for $i = 1, \ldots, L$), and $\boldsymbol{W}_{in} \in \mathbb{R}^{n_{x_L} \times n_u}$ are fixed internal connection weights whose entry values are set to random values; and the matrix $\boldsymbol{W}_{out} \in \mathbb{R}^{n_y \times (n_{x_0} + \cdots + n_{x_L} + 1)}$ is the read-out weight matrix whose entries are to be learned. The vectors $\boldsymbol{\xi}_n^{(k)}$, $k = 0, \ldots, L$, $n = 0, 1, \ldots, N-1$, are i.i.d. random vectors describing added noise during the sampling at each layer and ν is a noise (regularization) intensity parameter (we have taken the same noise level for each layer). The activation function in each layer is taken to be a (vectorized) hyperbolic tangent function. At each training step, the input is fed into the first layer, which is then connected to the next layer via the connection weights. The ansatz for the output at each update time is taken to be a linear combination of the elements of the hidden states and a bias value.

2. Algorithm for training the ESN and its deep version

The above implementation gives a randomly constructed RNN prior to training. It may generally develop oscillatory or chaotic behavior even in the absence of external excitation by the input; therefore, the subsequent network states, starting from an arbitrary state $\boldsymbol{x}(t_0)$, may not converge to the zero state. To ensure that the ESN/DESN converges to the desired state, the internal connection weights are scaled such that the resulting (untrained) input-driven recurrent network (or the "dynamical reservoir") is appropriately stabilized or "damped," forcing it to have the so-called "echo state property." The echo state property ensures that the current network state is uniquely determined by the history of the input provided that the RNN has been run for a sufficiently long time (see Refs. 38 and 54 and the references therein for details, subtleties, and other equivalent conditions for the echo state property). Once this initialization is made, we can proceed to the training phase and then the prediction phase, to be described in the following.

We now give a complete description of setting up, training, and using the ESN and its deep version for the prediction task. It is based closely on the techniques developed in Ref. 38. For a schematic of a general ESN/DESN architecture and the training process, we refer to Fig. 1.

Algorithm III.2. *Initializing, training, and using the ESN ($L = 0$) and DESN ($L \geq 1$).* Given a training set consisting of the input sequence $(\boldsymbol{u}(t_0), \ldots, \boldsymbol{u}(t_{N-1}))$, we find a trained ESN/DESN parameterized by $(\boldsymbol{W}^{(i)}, \boldsymbol{W}_b, \boldsymbol{V}^{(j)}, \boldsymbol{W}_{in}, \boldsymbol{W}_{out}, \nu)_{i,j=0,\ldots,L}$ whose network output $(\boldsymbol{y}(t_0), \ldots, \boldsymbol{y}(t_{N-1}))$ approximates the input sequence.

(1) *Initialize the ESN/DESN to "ensure"[58] that the echo state property is satisfied.*

  (1a) Depending on the training data (length, difficulty of task, etc.), select appropriate dimensions/sizes (i.e., $n_{x_0}, \ldots, n_{x_L}$) for the connection weight matrices. These dimensions are hyperparameters that can be tuned. The matrix elements of these matrices are then selected randomly as follows:

$$(\boldsymbol{W}_b)_{kl} \sim \text{Unif}(-1, 1), \tag{14}$$

$$(\boldsymbol{V}^{(j)})_{kl} \sim \text{Unif}(-0.5, 0.5) \text{ for } j = 1, \ldots, L, \tag{15}$$

$$(\boldsymbol{W}^{(i)})_{kl} \sim \text{Unif}(-0.5, 0.5) \text{ for } i = 0, \ldots, L, \tag{16}$$

$$(\boldsymbol{W}_{in})_{kl} \sim \text{Unif}(-1, 1). \tag{17}$$

  (1b) These matrices are rescaled as follows:

$$\boldsymbol{W}_b, \boldsymbol{V}^{(j)}, \boldsymbol{W}^{(i)}, \boldsymbol{W}_{in} \mapsto \frac{\boldsymbol{W}_b}{\|\boldsymbol{W}_b\| + 0.001}, \frac{\boldsymbol{V}^{(j)}}{\|\boldsymbol{V}^{(j)}\| + 0.001}, \times \frac{\boldsymbol{W}^{(i)}}{\|\boldsymbol{W}^{(i)}\| + 0.001}, \frac{\boldsymbol{W}_{in}}{\|\boldsymbol{W}_{in}\| + 0.001}, \tag{18}$$

  where $\| \cdot \|$ denotes the Frobenius norm.

  (1c) Set a fraction of elements (connection weights) in the matrices $\boldsymbol{W}^{(i)}$ to zero (the fraction, denoted r_i, chosen is a hyperparameter for sparsity of the matrices used for each layer) and then rescale the resulting matrices $\tilde{\boldsymbol{W}}^{(i)}$ appropriately using their spectral radius, i.e.,

$$\tilde{\boldsymbol{W}}^{(i)} \mapsto \rho_{des}^{(i)} \frac{\tilde{\boldsymbol{W}}^{(i)}}{\rho^{(i)}}, \tag{19}$$

  where $\rho^{(i)}$ is the spectral radius of $\tilde{\boldsymbol{W}}^{(i)}$, $\rho_{des}^{(i)}$ is the desired spectral radius, and $i = 0, \ldots, L$. The desired spectral radius chosen is less than one to ensure contractivity of the dynamics and is another tunable hyperparameter. The sparsity hyperparameter is chosen in such a way that sufficiently rich dynamics of different internal units/hidden states can be obtained. (1a)–(1c) then give an untrained network (dynamical reservoir) that satisfies (typically in practice) the required properties. We initialize the network state with $\boldsymbol{x}(t_0) = \boldsymbol{0}$.

(2) *Train the readout by solving a least squares linear regression problem.*

  (2a) If needed, discard an initial transient by disregarding the first $i_{transient} = \min(\text{int}(N/10), 100)$ states, where $\text{int}(x)$ is the integer part of x and $\min(x, y)$ is the minimum of x and y.

  (2b) Run the network on the entire input (training) sequence and collect the output sequence, i.e., according to (8) and (9)

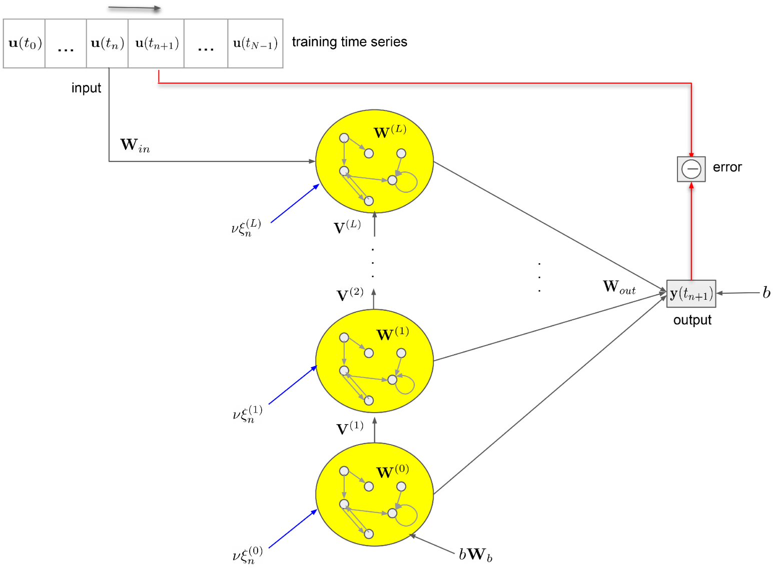


FIG. 1. A schematic of the ESN/DESN and the workflow during the training phase described in Sec. B (Algorithm III.1). In our numerical experiments, the training time series $(u(t_0), \ldots, u(t_{N-1}))$ is the fast driving signal extracted from $(x(t_0), \ldots, x(t_N))$.

for the shallow ESN and (10)–(13) for the DESN. During sampling, we have added a small amount of noise or regularization, whose intensity is determined by the hyperparameter ν, to stabilize the network and prevent overfitting.[49] We choose the elements of the i.i.d. noise vectors $\boldsymbol{\xi}_n^{(k)}$ to be distributed according to $\mathrm{Unif}(-0.5, 0.5)$.

(2c) Solve the least squares regression problem,

$$\min_{W_{out}} \frac{1}{N - i_{transient}} \sum_{i=i_{transient}}^{N-1} \|\boldsymbol{W}_{out}\tilde{\boldsymbol{x}}(t_i) - \boldsymbol{u}(t_i)\|^2, \tag{20}$$

where $\tilde{\boldsymbol{x}}(t_i) = (\boldsymbol{x}(t_i); b) \in \mathbb{R}^{n_{x_0}+\cdots+n_{x_L}+1}$. We remind the reader that $\boldsymbol{W}_{out} \in \mathbb{R}^{n_y \times (n_{x_0}+\cdots+n_{x_L}+1)}$ is the only trainable matrix in the ESN/DESN. As the above problem admits a closed form solution, the solution, denoted $\boldsymbol{W}_{out}^{opt}$, can be obtained directly by applying the Moore–Penrose pseudoinversion as follows.

Let $\tilde{\boldsymbol{X}} = [\boldsymbol{X} \quad \boldsymbol{b}] \in \mathbb{R}^{(N-i_{transient}) \times (n_{x_0}+\cdots+n_{x_L}+1)}$, with $\boldsymbol{X} \in \mathbb{R}^{(N-i_{transient}) \times (n_{x_0}+\cdots+n_{x_L})}$ the design matrix, i.e., a block matrix stacked vertically with the matrices $[\boldsymbol{x}^{(j)}(t_{i_{transient}}) \ \ldots \ \boldsymbol{x}^{(j)}(t_{N-1})]^T \in \mathbb{R}^{(N-i_{transient}) \times n_{x_j}}$ $(j = 0, \ldots, L)$, $\boldsymbol{b} = (b, b, \ldots, b) \in \mathbb{R}^{N-i_{transient}}$, and $\boldsymbol{U} = [\boldsymbol{u}(t_{i_{transient}}) \ \ldots \ \boldsymbol{u}(t_{N-1})]^T \in \mathbb{R}^{(N-i_{transient}) \times n_y}$. Then, the solution to (20) can be obtained as

$$\boldsymbol{W}_{out}^{opt} = (\tilde{\boldsymbol{X}}^{+}\boldsymbol{U})^{T}, \tag{21}$$

where $+$ and T denote the Moore–Penrose inverse and transposition, respectively. This completes the training phase.

(3) *Run the trained ESN/DESN autonomously for prediction.*
During the prediction phase, we use $\boldsymbol{y}(t_m) = \boldsymbol{W}_{out}^{opt}(\boldsymbol{x}(t_m); b)$ in place of $\boldsymbol{u}(t_m)$ for $m \geq N-1$ in the update equation (12) for DESN [or (8) for ESN]. We then propagate the trained network forward in time according to the resulting update equation. This allows us to obtain the predicted values $(\boldsymbol{y}(t_N), \ldots, \boldsymbol{y}(t_{N+M}))$, where $\boldsymbol{y}(t_i) = \boldsymbol{W}_{out}^{opt}\tilde{\boldsymbol{x}}(t_i)$ for $i = N, \ldots, N+M$.

The smaller the generalization error on the prediction time horizon, the more accurate is the prediction. The quality of prediction achieved by our ESN/DESN depends on the selection of hyperparameters, which need to be chosen following an appropriate model selection procedure, which adapts to the data set on hand. The tunable hyperparameters are $n_L := L + 1$ (number of layers in the DESN), the n_{x_i} (dimension of the connection weight matrices at layer $i+1$), the r_i (sparsity parameter), the $\rho_{des}^{(i)}$ (the desired spectral radius of these matrices at layer $i+1$), and ν (noise intensity). We remark that the above algorithm gives a specific way to initialize and train the ESN/DESN. Other variants may also be considered.[50]

To implement (S2) in Algorithm III.1, we apply Algorithm III.2 to the training data $(f(t_i))_{i=0,\ldots,N-1}$. Together with (S1) and (S3), this completes the description of our rare event prediction method.

Last, we discuss a physically motivated intuition about ESNs (as can be said for DESNs). On one hand, ESNs can be constructed by sampling from a class of continuous-time dynamical systems,[73] which satisfy a universal property in the sense that they can approximate any continuous-time system on compact time intervals to arbitrary degree of precision. On the other hand, the data itself are generated from a nonlinear chaotic system. The ESN approach amounts to learning the data by feeding the data as input (signal) into an ESN, thereby inducing an interaction between the two dynamical systems. In our case, we would like the output of the network to reproduce the input time series, and good performance could be obtained under suitable conditions (when the hyperparameters are tuned optimally). From a physical point of view, these conditions cause the two dynamical systems to synchronize; therefore, learning here corresponds to finding optimal conditions to achieve generalized synchronization of the system generating the training data and a signal-fed ESN.[64,80] Indeed, the idea of synchronization between two dynamical systems has been exploited for time series prediction.[17] It remains interesting to understand how out-of-sample performance of ESNs depends on the characteristics of input data and the noise (either those present in the data or those added to regularize the training) using the notion of generalized synchronization.

IV. NUMERICAL EXPERIMENTS

We apply the method presented in Sec. III to study the questions posed in Sec. II for three datasets generated by dynamical systems of different complexities.

From the dataset in each example, a training set (with $N+1-N_v$ data points), a validation set (with N_v data points), and a test set (with M data points) are assigned before we apply the method. We are going to perform an ensemble based prediction by utilizing multiple independently trained networks. We take weighted averaged values of the predicted values produced by these networks as the final predicted values. We refer to the Appendix, in particular, Fig. 5, for details of training and prediction procedure. We emphasize that while a comprehensive investigation of dependence of the results obtained here on the choice of hyperparameters is interesting on its own, this is not our focus here.

Throughout this section, all variables considered are real and one-dimensional.

A. Example 1 : A bi-stable system, driven by a fast Lorenz-63 system

1. Data generation

The data for x are generated by the following slow–fast system:[28]

$$\dot{x}(t) = x(t)[1 - x^2(t)] + \frac{\sigma}{\varepsilon} y_2(t), \tag{22}$$

$$\dot{y}_1(t) = \frac{10}{\varepsilon^2}[y_2(t) - y_1(t)], \tag{23}$$

$$\dot{y}_2(t) = \frac{1}{\varepsilon^2}[28y_1(t) - y_2(t) - y_1(t)y_3(t)], \tag{24}$$

$$\dot{y}_3(t) = \frac{1}{\varepsilon^2}[y_1(t)y_2(t) - \frac{8}{3}y_3(t)]. \tag{25}$$

In (22)–(25), x is the state of the system of interest and its evolution is driven by a fast chaotic signal y_2/ε, which is modeled as follows. The vector state (y_1, y_2, y_3) is described by the Lorenz-63 model with the classical parameter values that lead to chaotic behavior.[46] At these parameter values, y_2 is ergodic with an invariant measure supported on a set of zero volume. The equation for x is, therefore, an ODE driven by a fast chaotic signal with characteristic time ε^2.

To generate the data, we use a uniform time step of $\Delta t = 0.01$ to integrate (22)–(25) with $\sigma = 0.08$, $\varepsilon = 0.5$, $x(0) = -1.5$, and $y_i(0) \sim \mathrm{Unif}(-10, 10)$ for $i = 1, 2, 3$, up to time $t = 100$. The autocorrelation time [defined as the time τ at which the autocorrelation function $R(\tau) = 1/e$] of samples of the driving signal is estimated to be 0.05. Note that the time step is chosen to be small enough so that we can sample the scale on which the fast driving signal takes place. The time series generated for x is plotted in Fig. 2.

We remark that repeating the above data generation using a different numerical time step, initial conditions (or random seeds), and parameters will produce time series with different characteristics. For instance, the resulting time series will display a different number of critical transitions occurring at different times. We assume that we only have access to a segment of the *single* time series plotted in Fig. 2 up to time $t < \tau_0$.

2. Generating system: Dynamics and applications

Applying the discussion in Sec. II, the family of systems (22)–(25) (parameterized by ε) can be viewed as an approximation

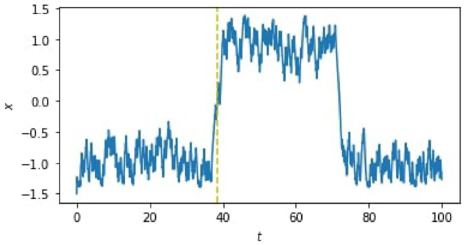


FIG. 2. Time series data for x in Example 1 up to time $t = 100$, of which only a segment prior to time $t = 37 < \tau_0 = 38.52$ (i.e., at most $N + 1 = 3701$ data points) will be available for training. Note that the sample path of x first crosses zero at $t = \tau_0$ (indicated by the yellow dashed line), where a critical transition occurs.

to the Markovian system,

$$dX(t) = X(t)[1 - X^2(t)]dt + \tilde{\sigma}\, dW(t), \qquad (26)$$

where $\tilde{\sigma}$ is an effective diffusion constant and $W(t)$ is a Wiener process in the sense that as ε becomes smaller, $x(t)$ converges in law to the process $X(t)$, solving the SDE above. Recall that we have chosen $\varepsilon = 0.5$ for data generation; therefore, the data can be thought of coming from an approximately stochastic system.

In the absence of the driving signal, the equation for x has two stable fixed points, at $x = -1$ and $x = 1$, and an unstable one at $x = 0$. Starting from an initial state, the system will eventually evolve toward a nearby stable state. The presence of noise alters this dynamics, causing an occasional transition of the system between stable states. In the case where the noise amplitude is small, such a transition is a rare event, occurring at a seemingly unpredictable time (see Fig. 2). In our case, x starts near the fixed point $x = -1$ and will eventually jump to that at $x = 1$, the prediction of which is of great interest. From the data, we find that the first crossing time of the sample path to zero is $\tau_0 := \inf\{t \in [0, 100] : x(t) > 0\} = 38.52$. Our goal is then to predict an approaching rare event using only (a segment of) the data consisting of time series up to time $t < \tau_0$.

From a statistical mechanical point of view, the system (22)–(25) describes an overdamped Brownian particle moving in a symmetric double-well potential. In this case, x is the position of the particle and time-integrated y_2 models the fluctuations due to its interaction with the environment. The above model for x is also often used in climate physics, an example of which is to view x as the sea-surface temperature anomaly and $\xi = \frac{\sigma}{\varepsilon} y_2$ as the impact of small-scale atmospheric variability.[10,36,57]

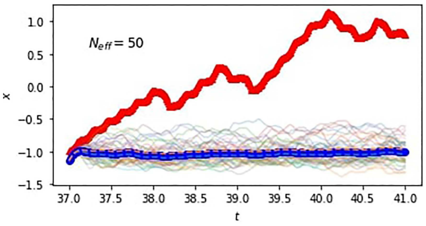
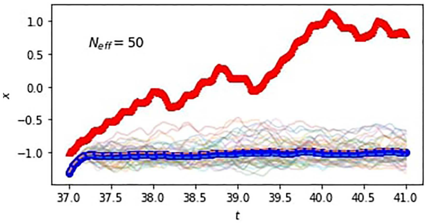
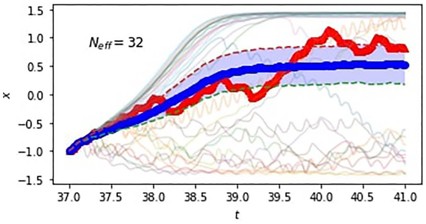
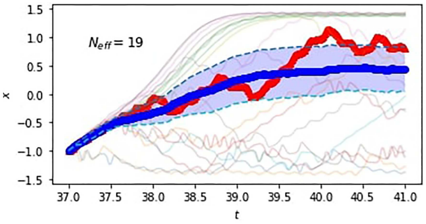
3. Results and discussion

The prediction results for different training scenarios using Algorithm III.1 are displayed in the figures of Tables I–III. For each case where a fixed number of training data points is used, we present two prediction models, each of which depends on the size (N_v) of the validation set used (see the Appendix for details).

Figures 1(a) and 1(b) and 2(a) and 2(b) in Table I show that our method is capable of quite confidently predicting the rare event and its transition path at least 162 numerical time steps in advance, i.e., prior to first crossing of the slow process to zero at $\tau_0 = 38.52$ or equivalently, at least about 32 autocorrelation times of the fast driving signal. In particular, the target future trajectory lies principally within the 90% confidence interval of the predicted trajectory. Given that a relatively short time series was used for training, it would appear that the accuracy of these predictions is remarkable.

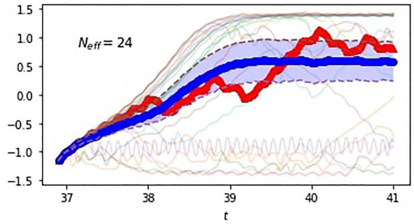
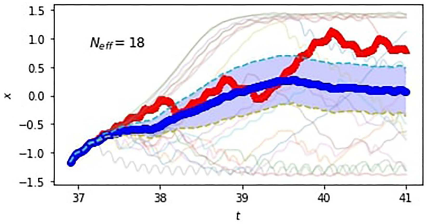
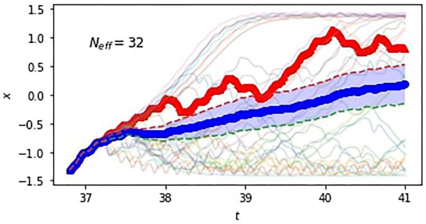
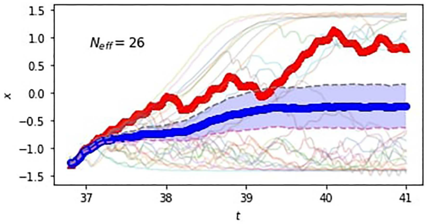
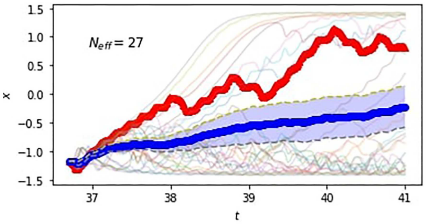
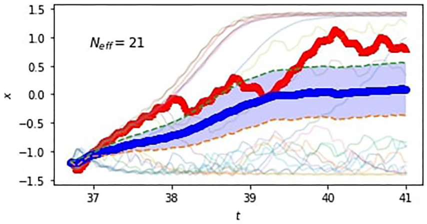
Figures 3(a)–4(b) in Table I show how the inability of the ESN to make accurate long-term predictions makes accurate prediction of the *entire* transition path (i.e., up to time $t = 40$) from an earlier time more difficult. Indeed, the longer into the future the prediction is, the less confident the results are. This is shown by the growth of the standard deviation of the prediction errors, which seems to

TABLE I. Prediction results using the dataset in Example 1 under various scenarios. N is the number of data points in the training set and N_v is the number of data points in the validation set (see the Appendix). In each figure: the target trajectory for x (in red), N_{eff} predicted trajectories from N_{eff} independently trained and selected models, and the averaged predicted trajectory (in thick blue) up to several time steps into the future. The shaded light blue region represents the 90% confidence interval of the predicted values.

Scenarios	$N_v = 10$		$N_v = 20$
Baseline (0a)		(0b)	
$N = 3700$ (1a)		(1b)	

Continued.

TABLE I. Continued.

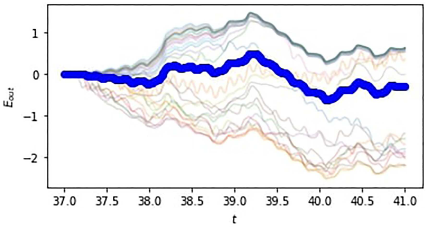
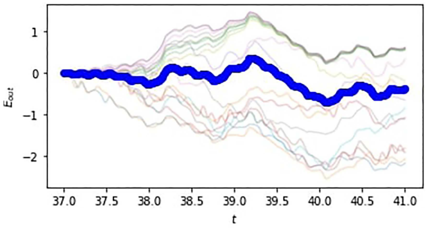
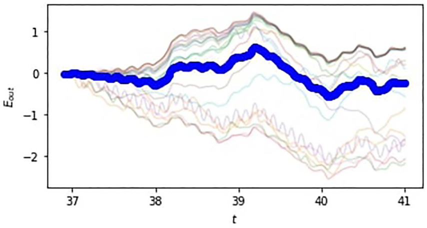
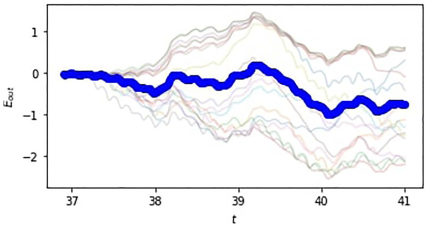
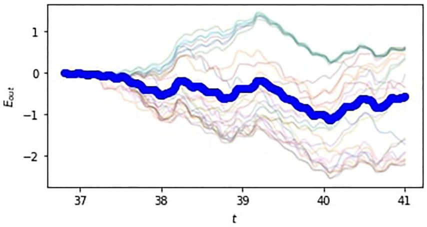
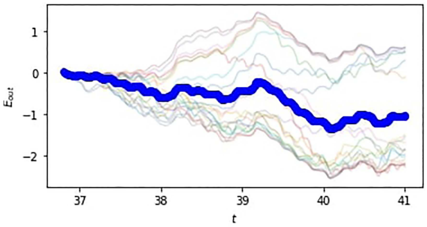
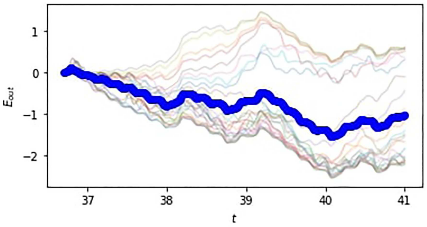
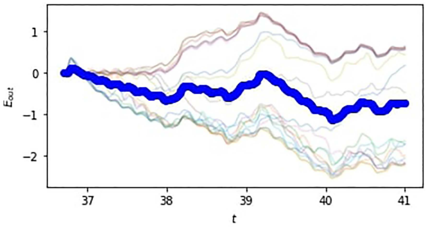
Scenarios	$N_v = 10$	$N_v = 20$
$N = 3690$ (2a)		(2b) 
$N = 3680$ (3a)		(3b) 
$N = 3670$ (4a)		(4b) 

saturate about a maximum value that increases with fewer training data points. However, the results in Figs. 3(a) and 4(b) of Table I are still respectable, as they indicate that the trajectory will cross the origin during the prediction time window, albeit not able to trace the target transition path precisely.

These figures also depict the difference in the prediction results obtained with different values of N_v used [notably in Figs. 2(a)–4(b) of Table I], showing intricacies of the machine learning based method. However, the qualitative behavior of the mean value of the error of the predicted results is essentially the same for different N_v (see the figures in Table II). The same can be said for the standard deviation of these errors, which grows rapidly on the prediction horizon where the rare transition occurs and then approaches a plateau after that (see the figures in Table III). This plateau is slightly higher in all the cases where $N_v = 20$ than those where $N_v = 10$, indicating the higher variance of the prediction models that set aside more data points for validation. However, for the case of $N = 3670$, the model with the higher variance gives better predictive performance than that with the lower variance and is able to predict a crossover to the origin during the prediction interval.

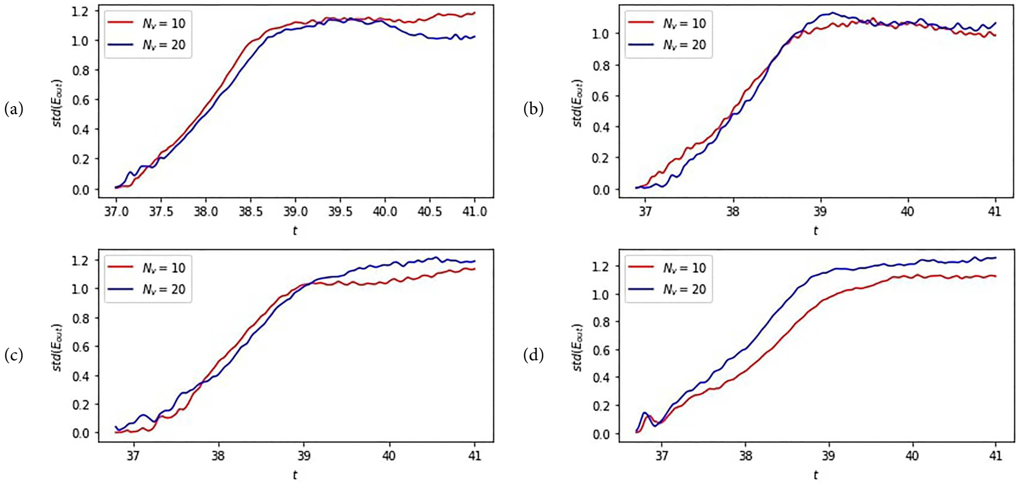
Figures 0(a) and 0(b) in Table I confirm that our method is far better than the direct method of applying an ESN to the data for x, which we take as a baseline result for comparison. We emphasize that one can also apply other machine learning algorithms such as the gradient descent based RNN and convolutional neural network to implement the direct method.[6] However, after some experiments, we found that the predictive performance is similar to the baseline result; i.e., they fail to predict the approaching rare event, rendering the prediction task almost impossible using the direct method. This comparison study enhances the veracity of our method, which exploits the crucial idea of appropriately taking into account the multiscale nature of the system generating the data. The success of our method in predicting the rare transition event lies in its ability to separate the slow and fast components of the data with the help of some physical knowledge about the generating system.

TABLE II. Error of the predicted results and its mean value (in thick blue) for Example 1.

Scenarios	$N_v = 10$	$N_v = 20$
$N = 3700$ 1(a)		(1b) 
$N = 3690$ (2a)		(2b) 
$N = 3680$ (3a)		(3b) 
$N = 3670$ (4a)		(4b) 

We remark that if we run the trained networks much further into the future, the predicted trajectory, not surprisingly, fails to capture the second transition that occurs around $t = 71$. However, we will be able to predict the second transition confidently if more data points are used for training. All the results discussed so far are obtained by training on a trajectory that starts from the initial time ($t_{init} = 0$) and ends at a time before a rare event occurs. A natural question is whether one can achieve prediction results of comparable quality using a shorter trajectory that comes sufficiently close to the rare event but starts at a later time $t_{init} > 0$. Given the chaoticity of the generating system with a known predictability horizon determined by the Lyapunov exponent, one expects the answer to this

TABLE III. Standard deviation of the errors for Example 1. (a) $N = 3700$, (b) $N = 3690$, (c) $N = 3680$, and (d) $N = 3670$.



question is affirmative and can indeed be supported by numerical experiments.

Last, we emphasize that our problem is more challenging than that of predicting the second component of a Lorenz-63 system from numerical data generated by the system itself. Even though we expect the extracted driving signal here is representative of the second component of the Lorenz-63 system (up to a scaling factor) and therefore, one would expect that good prediction results are achievable (in lights of recent results[61]), there are inherent errors in the extraction process in step (S1) and consequently, here, we are really dealing with a noise corrupted version of the data. In this case, the prediction could be highly non-trivial (and perhaps impossible if the errors are large enough) due to possible dominance of noise in certain segments of the reconstructed time series.

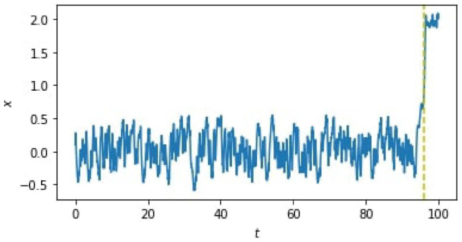


FIG. 3. Time series data for x in Example 2 up to time $t = 100$, of which only a segment prior to $t = 94.1 < \tau_1 = 96.16$ (i.e., at most $N + 1 = 9411$ data points) will be available for training. Note that the sample path of x first crosses one at $t = \tau_1$ (indicated by the yellow dashed line), where a critical transition occurs.

B. Example 2: A tri-stable system, with periodic forcing and a fast Ornstein–Uhlenbeck-like process

1. Data generation

The data for x are generated by the following slow–fast system:

$$\begin{aligned}\dot{x}(t) = {} & x(t)[1 - x(t)][1 + x(t)][x(t) - 2][x(t) + 2] \\ & + A\cos(\omega t) + \sigma_0 z(t),\end{aligned} \tag{27}$$

$$\dot{z}(t) = -\alpha_1 z(t) + \frac{\sigma_1}{\varepsilon} y_2(t), \tag{28}$$

where x describes the state of the system of interest whose evolution is driven by a fast Ornstein–Uhlenbeck like signal z and y_2 is the second component of the Lorenz-63 system (23)–(25). For data generation, we use a uniform time step of $\Delta t = 0.01$ to integrate (27) and (28) with $x(0) = 0.1$, $z(0), y_i(0) \sim \mathrm{Unif}(-10, 10)$ (for $i = 1, 2, 3$), $A = 0.5$, $\omega = 2\pi$, $\varepsilon = 0.5$, $\sigma_0 = 0.2$, $\alpha_1 = 1000$, and $\sigma_1 = 1000\varepsilon$, up to time $t = 100$. The autocorrelation time of samples of the driving signal is estimated to be 0.07. The time step is small enough to sample the scale on which the fast signal takes place. The time series generated for x is plotted in Fig. 3. We assume that we only have access to a segment of the *single* time series plotted in Fig. 3 up to time $t < \tau_1$.

2. Generating system: Dynamics and applications

The system (27) and (28) is more complex than that in Example 1, of which it is an extended version, in the sense that the force field is generalized to include a time-dependent external force and the driving signal is described by a higher dimensional system. One can view (27) and (28) as a family of systems (parameterized by ε) approximating the following SDE system:

$$dX(t) = X(t)[1 - X(t)][1 + X(t)][X(t) - 2][X(t) + 2]dt \\ + A\cos(2\pi t)dt + \sigma_0 Z(t)dt, \tag{29}$$

$$dZ(t) = -\alpha_1 Z(t)dt + \tilde{\sigma}_1 dW(t), \tag{30}$$

where $\tilde{\sigma}_1$ is an effective diffusion constant and $W(t)$ is a Wiener process. As ε (chosen to be 0.5 for data generation here) becomes smaller, $(x(t), z(t))$ converges in law to $(X(t), Z(t))$ solving the non-autonomous SDE system (29) and (30).

In contrast to Example 1, in the absence of the driving signal $z(t)$, Eq. (27) for x has three stable periodic orbits centered at $x = -2, 0, 2$ and two unstable ones centered at $x = -1, 1$. Here, our system state starts in the middle potential well and will, due to influence of the driving signal as well as the periodic forcing, transits to one of the left or right nearby stable orbits at a random time (see Fig. 3). It is thus natural to ask which potential well will the system transit to and at what time will the transition occur. From the data, we find that the first crossing time of the sample path to one is $\tau_1 := \inf\{t \in [0, 100] : x(t) > 1\} = 96.16$.

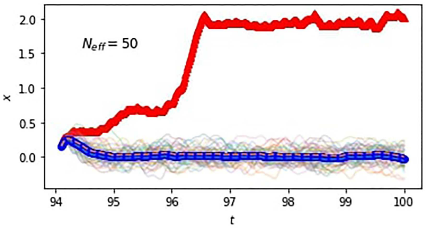
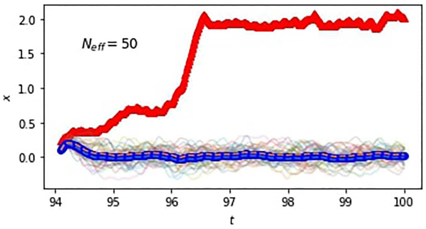
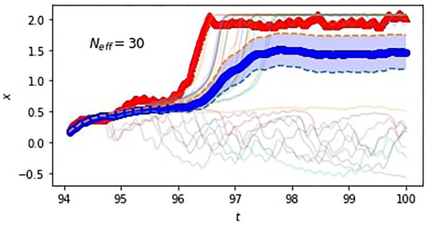
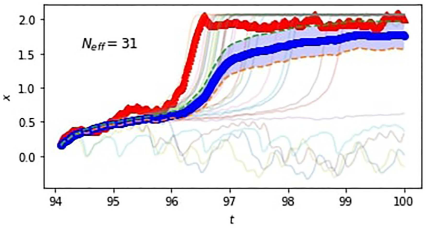
The addition of the periodic forcing $A\cos(2\pi t)$ introduces another time scale into the system. When this time scale is of the same order of the mean exit time from the potential (Kramers' time), a resonance-like mechanism where the noise can lead to the amplification of the periodic signal takes place. This resonance is induced by a chaotic signal and is closely related to stochastic resonance, a noise-induced phenomenon first introduced in the context of climate modeling,[7,8] which has been found to occur in many physical and biological systems.[24]

One example of such systems describes the dynamics of an overdamped self-propelled active particle, which converts energy absorbed from the environment into a directed motion, rendering the system out of equilibrium. The position of the particle can be described by X in (29) and the active force or self-propulsion is modeled by Z in (30), in which case the particle is trapped in a triple-well potential and subject to periodic forcing. This is a variant of the model of an active Ornstein–Uhlenbeck process widely used to study active matter.[69]

3. Results and discussion

The prediction results for different training scenarios using Algorithm III.1 are displayed in the figures of Tables IV–VI. For

TABLE IV. Prediction results using the dataset in Example 2 under various scenarios. N is the number of data points in the training set and N_v is the number of data points in the validation set (see the Appendix). In each figure, the target trajectory for x (in red), N_{eff} predicted trajectories from N_{eff} independently trained and selected models, and the averaged predicted trajectory (in thick blue) up to several time steps into the future. The shaded light blue region represents the 90% confidence interval of the predicted values.

Scenarios	$N_v = 4$	$N_v = 8$
Baseline (0a)		(0b) 
$N = 9410$ (1a)		(1b) 

Continued.

TABLE IV. Continued.

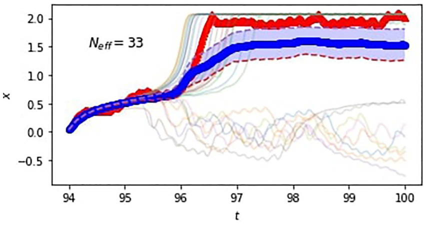
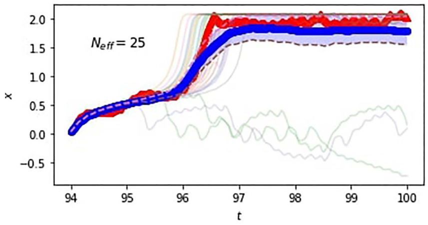
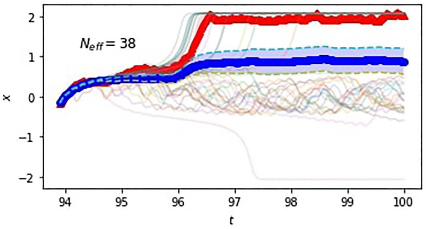
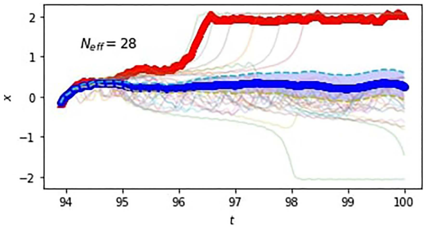
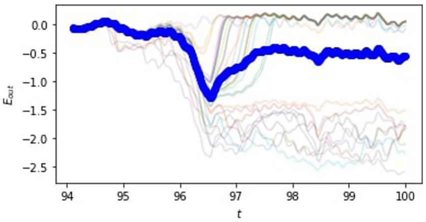
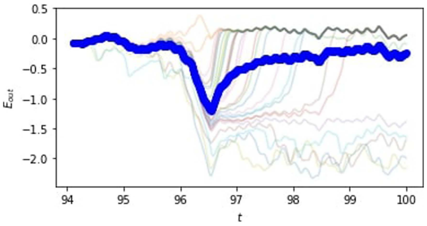
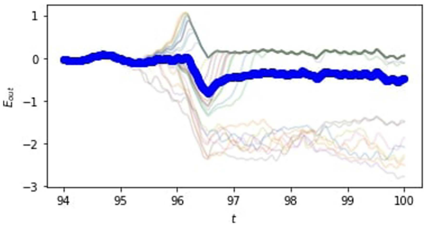
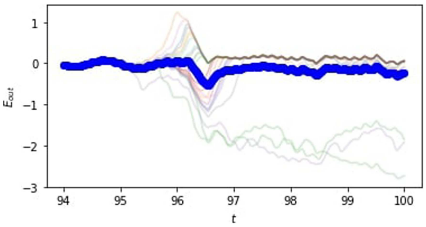
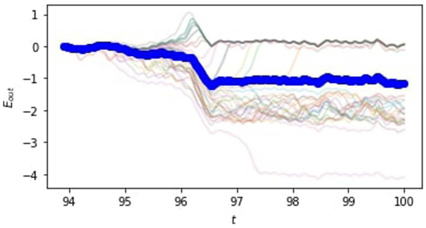
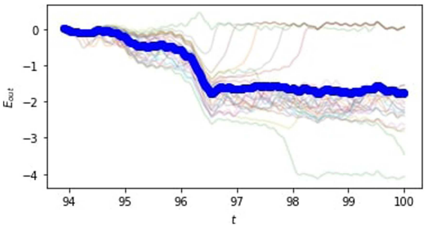
$N = 9400$ (2a)		(2b)	
$N = 9390$ (3a)		(3b)	

TABLE V. Error of the predicted results and its mean value (in thick blue) for Example 2.

Scenarios	$N_v = 4$		$N_v = 8$
$N = 9410$ (1a)		(1b)	
$N = 9400$ (2a)		(2b)	

Continued.

TABLE V. Continued.

Scenarios	$N_v = 4$		$N_v = 8$
$N = 9390$ (3a)		(3b)	

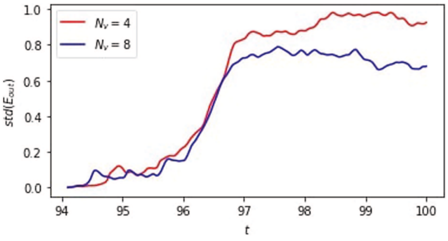
each case where a fixed number of training data points is used, we present two prediction models, each of which depends on the size (N_v) of the validation set used (see the Appendix for details).

Figures 1(b) and 2(b) in Table IV show that our method is capable of quite confidently predicting the rare event and its transition path at least 216 numerical time steps in advance or at least about 30 autocorrelation times of the fast driving signal prior to first crossover of x to zero at $t = \tau_1$. Even though the prediction results in Figs. 1(a) and 2(a) are inferior to those with the larger N_v, they are not too far behind since the crossover to the origin is successfully predicted there.

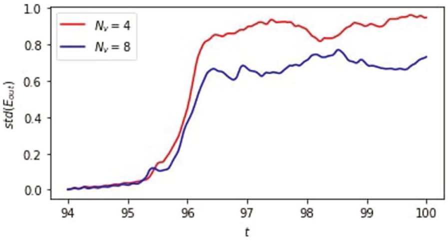
However, Figs. 3(a) and 3(b) in Table IV show how the inability of the DESN to make accurate long-term predictions again cripples the performance of the prediction method, making prediction of the rare event from an earlier time a very challenging task. In particular, neither models can confidently predict the crossover of x to the origin during the prediction interval, although the result obtained by the model with $N_v = 4$ comes closer to achieving that. We emphasize that these results are used primarily to demonstrate our method. It may be possible to improve these results by using a different training setting (with a higher number of layers and more carefully optimizing values of the hyperparameters), enabling

TABLE VI. Standard deviation of the errors for Example 2. (a) $N = 9410$, (b) $N = 9400$, and (c) $N = 9390$.

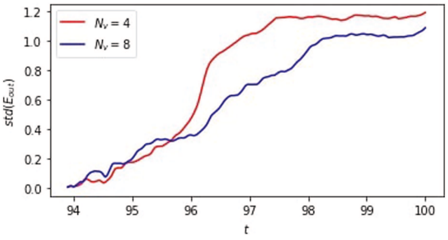
(a)



(b)



(c)



successful longer-term prediction and, therefore, prediction of a rare event from an earlier time.

These figures again depict the difference in the prediction results obtained with different values of N_v used (see the figures in Table V). Just like the case in Example 1, the qualitative behaviors of both the mean value and the standard deviation of the error of the predicted values are essentially the same for different N_v. However, the standard deviation plateau is higher in all the cases where N_v is lower (see Table VI) in contrast to the finding in the case of Example 1. This indicates that larger values of N_v should be considered here to obtain a prediction model with a lower variance.

We find that it is very difficult to obtain a comparable quality of predictions if we work with a shallow ESN instead of the three-layered variant that we have used here. This finding may be explained by the following. A closer look at the model for the driving signal $z(t)$ reveals that there are actually two widely separated time scales in the system generating it instead of only one time scale as in the case of Example 1. This difference in the multiscale behavior of the system generating the driving signal explains why a shallow ESN works well for Example 1 but not for Example 2. A deep version of the ESN is needed to handle the multiscale nature of the data for z in Example 2. This supports the intuition that increasing the depth of the ESN can lead to a better multiple time scale representation of the temporal information. This motivates our consideration of a deep version of the ESN for our method in Sec. III.

Figures (0a) and (0b) in Table IV confirm that our method is far superior to the baseline method of applying a DESN to the data for x. Indeed, the ability of our method to predict, at least 216 time steps (at least 2 periods) in advance, has significant importance for many systems exhibiting stochastic resonance. It is possible to achieve prediction results of comparable quality using a shorter but sufficiently long trajectory starting at a later time $t_{init} > 0$ and coming sufficiently close to the rare event (see the relevant discussion on this in Subsection IV A).

C. Example 3: A tri-stable system subject to periodic forcing and driven by a multiscale Lorenz-96 system

1. Data generation

The data for x are generated by the following slow–fast system:

$$\dot{x}(t) = x(t)[1 - x(t)][1 + x(t)][x(t) - 2][x(t) + 2] + B\cos(\Omega t) + \sigma\xi_1(t), \tag{31}$$

$$\dot{\xi}_1 = \frac{1}{\varepsilon}((\xi_2 - \xi_{K-1})\xi_K) - \xi_1 + F, \tag{32}$$

$$\dot{\xi}_2 = \frac{1}{\varepsilon^2}((\xi_3 - \xi_K)\xi_1 - \xi_2 + F), \tag{33}$$

$$\dot{\xi}_K = \frac{1}{\varepsilon^2}((\xi_1 - \xi_{K-2})\xi_{K-1} - \xi_K + F), \tag{34}$$

$$\dot{\xi}_n = \frac{1}{\varepsilon^2}((\xi_{n+1} - \xi_{n-2})\xi_{n-1} - \xi_n + F) \quad \text{for } n = 3, \ldots, K-1. \tag{35}$$

In (31)–(35), x is the state of the system of interest whose evolution is driven by a fast K-dimensional Lorenz-96 system whose state is denoted by the vector $(\xi_1, \xi_2, \ldots, \xi_K)$. We work with $K = 36$ and $F = 8$, which are parameter values leading to chaotic dynamics.[47] For data generation, we use a uniform time step of $\Delta t = 0.01$ to integrate the above system with $x(0) = 0$, $\xi_k(0) = F = 8$ for all k except for $k = 20$, with $\xi_{20}(0) = F + 0.01$, $B = 0.1$, $\Omega = 12\pi$, $\sigma = 0.4$, and $\varepsilon = 0.5$, up to time $t = 100$. The autocorrelation time of samples of the driving signal is estimated to be 0.11. The time step has again been chosen to be small enough to sample the scale on which the fast signal takes place. The time series generated for x is plotted in Fig. 4. We assume that we only have access to a segment of the *single* time series plotted in Fig. 4 up to time $t < \tau_1$.

2. Generating system: Dynamics and applications

The system described by (31)–(35) is similar to the one in Example 2 except that it is driven by a different periodic forcing and the fast driving signal comes from a single component of a variant (due to the scaling introduced by ε) of the Lorenz-96 system, a much higher dimensional chaotic system compared to the ones considered in the previous examples. We remark that, due to the scaling introduced, the limiting slow dynamics are deterministic (averaging) rather than stochastic (homogenization). The critical transitions in this multiscale system (31)–(35) are induced by finite ε-effects, which are fluctuations around some constant mean driver.

The Lorenz-96 model (with $\varepsilon = 1$) above is the first model introduced by Lorenz in Ref. 47. It is a model extensively used in data assimilation and parameter estimation (parameterization) research, as well as in testing machine learning algorithms for parameter learning (sub-grid parameterization).[12,79] It was originally used by Lorenz as a one-dimensional atmospheric model. The variables in the model represent values of some atmospheric quantity in K sectors of a latitude circle, giving a periodic system of K ODEs. The basic physics of the atmosphere is captured in the right hand side of the ODEs, which contains advection terms, damping terms, and external forcing. One can, therefore, view x as a sea-surface temperature anomaly, influenced by the atmospheric quantity in one of the sectors of the latitude circle in the presence of a seasonal forcing.

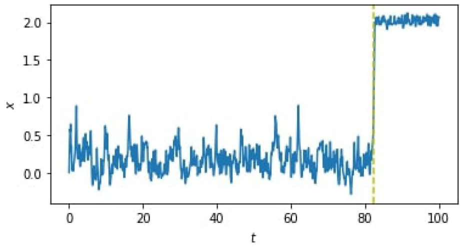


FIG. 4. Time series data for x in Example 3 up to time $t = 100$, of which only a segment prior to $t = 82.33 < \tau_1 = 82.39$ (i.e., at most $N + 1 = 8234$ data points) will be available for training. Note that the sample path of x first crosses one at $t = \tau_1$ (indicated by the yellow dashed line), where a critical transition occurs.

From the data, we find that the first crossing time of the sample path to one is $\tau_1 := \inf\{t \in [0, 100] : x(t) > 1\} = 82.39$.

3. Results and discussion

Figure 1(a) in Table VII shows that our method is capable of confidently predicting the rare event and its transition path at least six time steps in advance, prior to first crossing of the slow process to one at $\tau_1 = 82.39$. This is equivalent to about 0.55 autocorrelation times of the fast driving signal. The target future trajectory lies mostly within the 90% confidence interval of the predicted trajectory. This result is far better than the direct method of applying an ESN to the data for x [see the figures in (0a)–(0b) of Table VII], which we take as the baseline results.

Figures 1(b), 2(a), and 2(b) in Table VII show that the prediction models are unable to anticipate the rare transition event in the prediction interval. Therefore, in comparison with the results in earlier examples, here, the predictive performance of our models is far inferior. However, we still manage to produce an accurate prediction model as shown in Fig. 1(a). This is despite the fact that we are working with a set of data extracted only from one of the components of the Lorenz-96 system; moreover, these data are noise corrupted—a highly challenging task even in the case when the data are noise-free (see Ref. 79). We have also attempted to use deeper versions of network for training, but, unlike the case of Example 2, we are unable to obtain better predictions than those obtained by shallow networks.

Figures 1(a)–2(b) in Table VII depict the difference in the prediction results obtained with different values of N_v used, as discussed before. Unlike the earlier examples, the qualitative behavior of both the mean value and the standard deviation of the error of the predicted results is not the same for different N_v in the case when $N = 8233$ (see the figures in Tables VIII and IX). This indicates

TABLE VII. Prediction results using the dataset in Example 3 under various scenarios. N is the number of data points in the training set and N_v is the number of data points in the validation set (see the Appendix). In each figure: the target trajectory for x (in red), N_{eff} predicted trajectories from N_{eff} independently trained and selected models, and the averaged predicted trajectory (in thick blue) up to several time steps into the future. The shaded light blue region represents the 90% confidence interval of the predicted values.

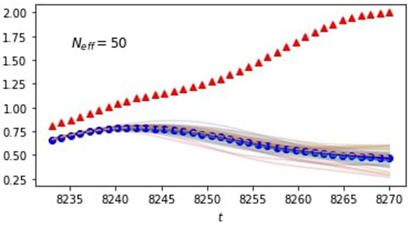
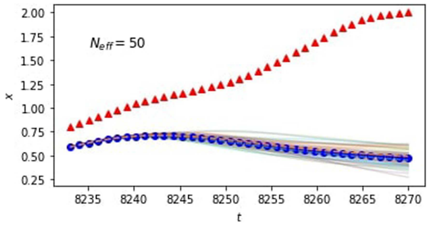
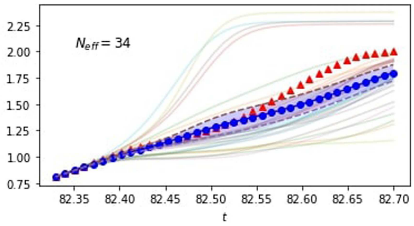
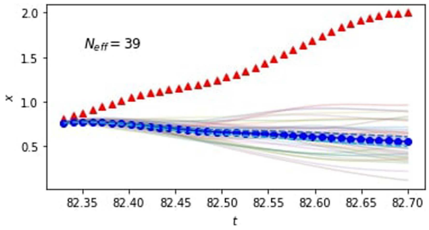
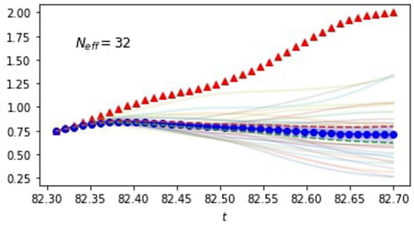
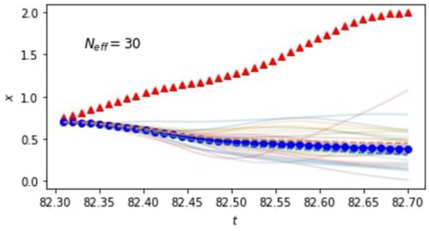
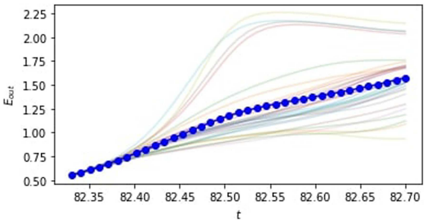
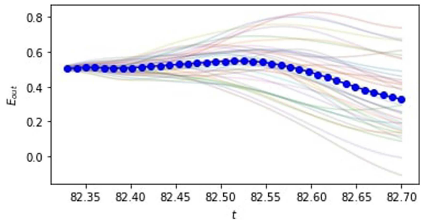
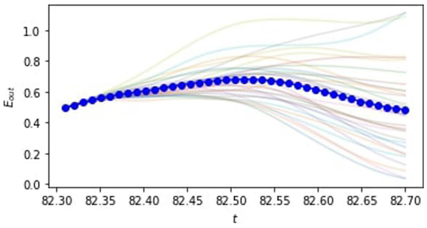
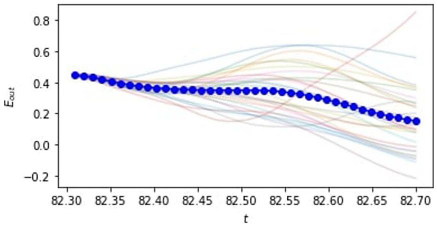
Scenarios	$N_v = 5$		$N_v = 8$
Baseline (0a)		(0b)	
$N = 8233$ (1a)		(1b)	
$N = 8231$ (2a)		(2b)	

TABLE VIII. Error of the predicted results and its mean value (in thick blue) for Example 3.

Scenarios	$N_v = 5$		$N_v = 8$
$N = 8233$ (1a)		(1b)	
$N = 8231$ (2a)		(2b)	

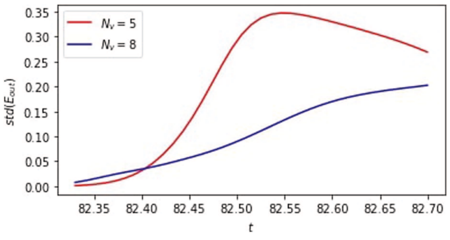
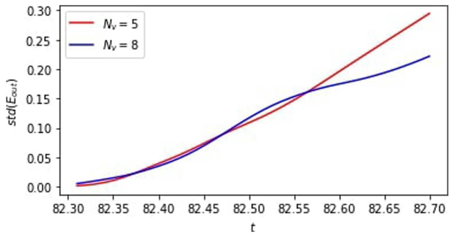
high sensitivity of the prediction models to the choice of N_v when dealing with highly complex multiscale data. We suspect that such sensitivity is due not only to the sensitivity of the echo state network to the choice of parameters, but also due to the noise corrupted nature of the extracted data. It is interesting to observe that the successful prediction model described here produces predictions whose standard deviation of the error shows a peak on the prediction interval, whereas those for the unsuccessful ones have strictly increasing standard deviation on the interval.

V. CONCLUSIONS

We have presented a data-driven method for tackling the task of predicting noise-induced critical transition events in a large class of multiscale nonlinear dynamical systems. The method was demonstrated on three examples of different complexity, each providing a model for many important physical and biological systems. For Examples 1 and 2, we obtain excellent predictions that would not be possible by using a direct method that does not take into account the multiscale nature of the problem. In particular, our method successfully predicts a rare transition event up to several numerical time steps in advance. Each run of the ESN/DESN incurs relatively low computational cost, thanks to the use of a reservoir computing based training technique, rather than a gradient descent based method. This low computational cost allows us to leverage the power of ensemble learning for the predictions.

These results demonstrate the promise of our reservoir computing based method in predicting rare events occurring in a wide

TABLE IX. Standard deviation of the errors for Example 3. (a) $N = 8233$ and (b) $N = 8231$.

(a)		(b)	

range of dynamical systems, a problem that is of substantial interest in science and engineering. We expect the accuracy of these results to improve by carefully optimizing the hyperparameters, using a more refined training method and other more sophisticated machine learning algorithms, at the expense of a higher computational cost. Importantly, rather than demonstrating only the successful results, we also highlight the limitation of the approach when dealing with highly complex multiscale datasets, such as that shown in Example 3. We also empirically show and discuss the sensitivity of the method to parameters and training data. Apart from achieving successful predictions, understanding and mitigating these issues are equally important. Therefore, our findings open up a range of interesting problems for future work.

We now discuss a few potential future directions. Thus far, we have applied the method to "toy examples," where there are few widely separated time scales. In many systems of interest, there may be many widely separated time scales and the competition between them may be crucial in triggering a critical transition event. The driving noise may also be multiplicative in nature. Therefore, it is important to extend the present work to these systems. In many realistic situations, the available time series data may be multivariate, rather than univariate, and additionally, there may be missing and/or uneven data. It would then be important to extend our method to treat these situations.

It is also of practical interest to apply the method presented here to study more non-trivial yet physically relevant datasets, such as those generated by a chaotic version of the model in Ref. 20 and/or real world data from climate science.[68,72] On the other hand, because the method is based on the use of (a deep version of) an echo state network, a firm theoretical understanding of the underpinnings behind such a network, in particular, its initialization (e.g., the role of randomness[70]) and the generalization error, will shed light on the nature of the prediction results. Therefore, it is important to carry out a systematic theoretical study to understand how the network works. This is, in fact, an important problem in the field of reservoir computing, and vigorous efforts have been made in recent years to tackle this task (see, for instance, Refs. 18, 29, 30, 34, 76, and 77).

ACKNOWLEDGMENTS

The authors acknowledge the Swedish Research Council (Grant No. 638-2013-9243). S. H. Lim is grateful to Stefano Bo for critical reading of the manuscript and many insightful discussions.

APPENDIX: DETAILS ON TRAINING, MODEL SELECTION, AND TESTING

We adopt the following approach for training, validation (model selection), and testing for the three examples studied in the paper.

From a given time series for the slow variable x (consisting of at least $N+1+M$ data points), we assign a training set, a validation set (with a total of $N+1$ data points), and a test set (with M data points). The training set and the validation set are the only accessible (in-sample) data, and we are *not allowed* to use the (out-of-sample) data in the test set, which records the occurrence of a rare event. From these accessible data, we first extract the driving signal according to (S1) in Algorithm III.1. There will be N data points for the signal. We set aside the last N_v data points of the extracted signal on the time interval $[(N-N_v)\Delta t, (N-1)\Delta t]$ to build a validation set, which will be used for hyperparameter tuning and model selection.

For each realization of the ESN/DESN, our model selection procedure is as follows. We employ the classical static validation scheme (see Ref. 51 for details) and perform a grid search over a prechosen hyperparameter space, looking for hyperparameter values that minimize the root mean squared error (RMSE) on the validation set. To choose this space, we focus on optimizing over the number of reservoir elements (n_{x_i}) in each layer, while fixing the desired spectral radius ($\rho^{(i)}_{des}$), noise intensity (ν), and sparsity parameter (r_i) to reasonable default values. This choice is informed by the practice that optimizing over n_x (the memory capacity) should be prioritized over other hyperparameters.[49] The values of all these hyperparameters are not fully optimized; i.e., an exhaustive search over the hyperparameters has not been performed. However, good average performance is typically not found in a very narrow parameter range; thus, a very detailed fine-tuning of parameters does not give a significant improvement and is not necessary.[49] An alternative model selection procedure is the more expensive automated Bayesian optimization approach proposed in Ref. 82.

The optimal ESN/DESN models selected by the above procedure will depend on the choice of N_v. The choice of N_v should be guided by the size and characteristic of the accessible data: N_v needs to be sufficiently large so that the validation error is a reliable estimate of an out-of-sample error but should be small enough, keeping in mind the size of accessible data and the short-term forecasting capability of ESN/DESN. Different choices of N_v will, therefore, give rise to different prediction models. Once an optimal model is chosen, we run the trained network to obtain predicted values for the driving signal on the time interval $[(N-N_v)\Delta t, (N+M)\Delta t]$ [see (S2) in Algorithm III.1] and proceed to (S3) in Algorithm III.1. For numerical integration in (S3), we use a Runge–Kutta method with a uniform step size of $\Delta t = 0.01$.

Our prediction and testing procedure is as follows. We perform an ensemble-like prediction by repeating the above training and model selection procedure N_{ens} times, obtaining N_{ens} independently trained and optimized prediction models, whose networks are initialized using different random seeds. The idea of this ensemble algorithm is that independent training and model selection here produce different learners (committee of networks), learning on a diverse set of features. The predictions of several base models built with the same algorithm are then aggregated in order to improve robustness over a single model; there may be predictions that are completely off, but these will partially cancel each other out. This increases stability, lowers the variance, and optimizes the overall predictive performance. A higher (lower) N_{ens} gives a more (less) confident estimator. The choice of N_{ens} should not be too large to minimize computational expense in training models and to avoid diminishing returns in performance from adding more ensemble members. We choose $N_{ens} = 50$ in all examples here. We have repeated the experiments with a larger N_{ens} of 100 but find only a slight improvement in performance.

Since some members of the ensemble will make better predictions than others, we expect to reduce the test error further if we assign greater weight to some members than to others. For this reason, we are taking weighted averaged values as the final predicted values as follows. Let N_{eff} be the number of predicted trajectories that lie entirely inside a pre-chosen confidence interval (which should also cover the trajectory obtained from the validation set) around the mean predicted trajectory on the time interval defining the validation set. The weight ascribed to each predicted trajectory is set to $1/N_{eff}$ (getting a vote) if it falls entirely inside the confidence interval and zero (getting no vote) otherwise. This filters out the outliers, leaving us with $N_{eff} \leq N_{ens}$ selected models that are expected to have reasonably good short-term forecasting capability, at least on the validation set. To compute the ensemble average in our experiments, we choose the confidence interval to be within eight standard deviations of the mean predicted trajectory on the validation time interval (if none of the predicted trajectories falls inside this region, we increase the number of standard deviations used by one iteratively until at least one falls inside the region).

To evaluate the quality of the predictions of the N_{eff} models, we study the statistics (the mean and standard deviation) of the difference between the predicted values and the target (real) values. Note that alternative measures for evaluating the predictive performance could also be considered. Our choice of measure here, in contrast with coarse-grained metrics such as the percentage of predicted trajectories that display a jump at a time close to the real one, reflects our focus on inferring the characteristics of the whole transition path. Figure 5 shows a schematic of our method.

The following are implementation details specific to each example.

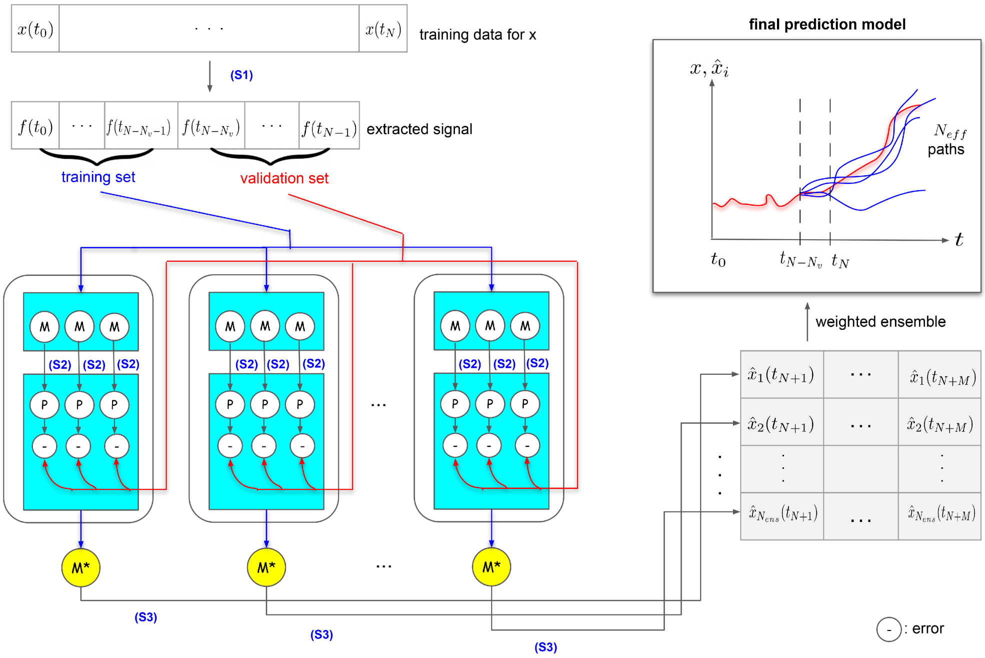


FIG. 5. A schematic summarizing the workflow of our data-driven method for a rare event prediction. In the schematic above, (S1)–(S3) denote the procedures described in Algorithm III.1, M denotes a (realization of) ESN/DESN model described in Sec. III B during the training phase, P denotes a prediction model obtained by running a ESN/DESN model autonomously during the prediction phase, M^* denotes an optimal ESN/DESN model produced by our model selection procedure, N is the number of data points in the training set, N_v is the number of data points in the validation set, N_{ens} is the number of learners in an ensemble, N_{eff} is the number of predicted paths that lie entirely inside a pre-chosen confidence interval around the mean predicted path on $[t_{N-N_v}, t_{N-1}]$, and the $\hat{x}_i$ denote the predicted values obtained by the chosen learners in the ensemble. In the plot (not drawn to scale) for the final prediction model, the target path (x) is in red and the predicted paths ($\hat{x}_i$) are in blue. We take the averaged predicted paths (over the N_{eff} blue paths) as the final prediction model.

- **Example 1.** We apply Algorithms III.1 and III.2 using a shallow ESN (i.e., $n_L = 1$ and $L = 0$) with $n_u = n_y = 1$, the bias value of $b = 1$, and $i_{transient} = 0$. The size of the validation set is chosen to be $N_v = 10$ and $N_v = 20$. For the grid search, the range of values that we use for n_{x_0} are from 600 to 740, while $\rho_{des}^{(0)} = 0.7$, $r_0 = 0.1$, and $\nu = 0.001$.
- **Example 2.** We apply Algorithms III.1 and III.2 using a three-layered DESN (i.e., $n_L = 3$, $L = 2$) with $n_u = n_y = 1$, the bias value of $b = 1$, and $i_{transient} = 0$. The size of the validation set is chosen to be $N_v = 4$ and $N_v = 8$. For the grid search, the range of values that we use for n_{x_i} (for $i = 0, 1, 2$) is from 100 to 300, while $\rho_{des}^{(0)} = 0.6$, $\rho_{des}^{(1)} = 0.7$, $\rho_{des}^{(2)} = 0.8$, $r_0 = r_1 = r_2 = 0.05$, and $\nu = 0.003$.
- **Example 3.** We apply Algorithms III.1 and III.2 using a shallow ESN (i.e., $n_L = 1$ and $L = 0$) with $n_u = n_y = 1$ and the bias $b = 1$. The first 200 data points are not used (therefore, $t_0 = 2$), and the first $i_{transient} = 100$ hidden states of the ESN are discarded during training. The size of the validation set is chosen to be $N_v = 5$ and $N_v = 8$. For the grid search, the range of values that we use for n_{x_0} is from 500 to 700, while $\rho_{des}^{(0)} = 0.95$, $r_0 = 0.1$, and $\nu = 0.003$.

We invite interested readers to experiment with the choice of hyperparameters using the codes provided at the website in the Data Availability section.

DATA AVAILABILITY

Complete codes (in Python) that reproduce all the results obtained in this paper are openly available in GitHub at https://github.com/shoelim/predicting-rare-critical-transitions-in-multiscale-systems.

REFERENCES

[1] S. Agarwal and J. S. Wettlaufer, "Maximal stochastic transport in the Lorenz equations," Phys. Lett. A **380**, 142–146 (2016).

[2] L. Arnold, "Hasselmann's program revisited: The analysis of stochasticity in deterministic climate models," in *Stochastic Climate Models* (Springer, 2001), pp. 141–157.

[3] P. Ashwin, S. Wieczorek, R. Vitolo, and P. Cox, "Tipping points in open systems: Bifurcation, noise-induced and rate-dependent examples in the climate system," Philos. Trans. R. Soc. A: Math. Phys. Eng. Sci. **370**, 1166–1184 (2012).

[4] N. Baker, F. Alexander, T. Bremer, A. Hagberg, Y. Kevrekidis, H. Najm, M. Parashar, A. Patra, J. Sethian, S. Wild *et al.*, "Workshop report on basic research needs for scientific machine learning: Core technologies for artificial intelligence," Technical Report, U.S. DOE Office of Science (SC), Washington, DC, 2019.

[5] C. Beck, "Brownian motion from deterministic dynamics," Physica A **169**, 324–336 (1990).

[6] Y. Bengio *et al.*, "Learning deep architectures for AI," Found. Trends Mach. Learn. **2**, 1–127 (2009).

[7] R. Benzi, G. Parisi, A. Sutera, and A. Vulpiani, "Stochastic resonance in climatic change," Tellus **34**, 10–16 (1982).

[8] R. Benzi, A. Sutera, and A. Vulpiani, "The mechanism of stochastic resonance," J. Phys. A: Math. Gen. **14**, L453 (1981).

[9] N. Berglund and B. Gentz, *Noise-Induced Phenomena in Slow-Fast Dynamical Systems: A Sample-Paths Approach* (Springer Science & Business Media, 2006).

[10] J. Berner, U. Achatz, L. Batté, L. Bengtsson, A. D. L. Cámara, H. M. Christensen, M. Colangeli, D. R. Coleman, D. Crommelin, S. I. Dolaptchiev *et al.*, "Stochastic parameterization: Toward a new view of weather and climate models," Bull. Am. Meteorol. Soc. **98**, 565–588 (2017).

[11] S. Bo, F. Schmidt, R. Eichhorn, and G. Volpe, "Measurement of anomalous diffusion using recurrent neural networks," Phys. Rev. E **100**, 010102 (2019).

[12] J. Brajard, A. Carrassi, M. Bocquet, and L. Bertino, "Combining data assimilation and machine learning to emulate a dynamical model from sparse and noisy observations: A case study with the Lorenz 96 model," Geosci. Model Dev. Discuss. **2019**, 1–21.

[13] S. L. Brunton and J. N. Kutz, *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control* (Cambridge University Press, 2019).

[14] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, "Machine learning and the physical sciences," Rev. Mod. Phys. **91**, 045002 (2019).

[15] A. Chattopadhyay, P. Hassanzadeh, and D. Subramanian, "Data-driven predictions of a multiscale Lorenz 96 chaotic system using machine-learning methods: Reservoir computing, artificial neural network, and long short-term memory network," Nonlinear Process. Geophys. **27**, 373–389 (2020).

[16] I. Chevyrev, P. K. Friz, A. Korepanov, I. Melbourne, and H. Zhang, "Multiscale systems, homogenization, and rough paths," in *International Conference in Honor of the 75th Birthday of SRS Varadhan* (Springer, 2016), pp. 17–48.

[17] A. B. Cohen, B. Ravoori, T. E. Murphy, and R. Roy, "Using synchronization for prediction of high-dimensional chaotic dynamics," Phys. Rev. Lett. **101**, 154102 (2008).

[18] C. Cuchiero, L. Gonon, L. Grigoryeva, J. P. Ortega, and J. Teichmann, "Discrete-time signatures and randomness in reservoir computing," arXiv:2010.14615 (2020).

[19] W. E. and E. Vanden-Eijnden, "Towards a theory of transition paths," J. Stat. Phys. **123**, 503 (2006).

[20] I. Eisenman and J. S. Wettlaufer, "Nonlinear threshold behavior during the loss of Arctic sea ice," Proc. Natl. Acad. Sci. U.S.A. **106**, 28–32 (2009).

[21] E. Forgoston and R. O. Moore, "A primer on noise-induced transitions in applied dynamical systems," SIAM Rev. **60**, 969–1009 (2018).

[22] C. Gallicchio and A. Micheli, "Deep echo state network (DeepESN): A brief survey," arXiv:1712.04323 (2017).

[23] C. Gallicchio and S. Scardapane, "Deep randomized neural networks," in *Recent Trends in Learning from Data* (Springer, 2020), pp. 43–68.

[24] L. Gammaitoni, P. Hänggi, P. Jung, and F. Marchesoni, "Stochastic resonance," Rev. Mod. Phys. **70**, 223 (1998).

[25] L. P. García, J. D. Pérez, G. Volpe, A. V. Arzola, and G. Volpe, "High-performance reconstruction of microscopic force fields from Brownian trajectories," Nat. Commun. **9**, 5166 (2018).

[26] P. Gaspard, M. Briggs, M. Francis, J. Sengers, R. Gammon, J. R. Dorfman, and R. Calabrese, "Experimental evidence for microscopic chaos," Nature **394**, 865 (1998).

[27] L. Giorgini, S. Lim, W. Moon, and J. Wettlaufer, "Precursors to rare events in stochastic resonance," Europhys. Lett. **129**, 40003 (2020).

[28] D. Givon, R. Kupferman, and A. Stuart, "Extracting macroscopic dynamics: Model problems and algorithms," Nonlinearity **17**, R55 (2004).

[29] L. Gonon, L. Grigoryeva, and J.-P. Ortega, "Risk bounds for reservoir computing," arXiv:1910.13886 (2019).

[30] L. Gonon, L. Grigoryeva, and J.-P. Ortega, "Approximation bounds for random neural networks and reservoir systems," arXiv:2002.05933 (2020).

[31] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, 2016).

[32] G. A. Gottwald and I. Melbourne, "Homogenization for deterministic maps and multiplicative noise," Proc. R. Soc. A: Math. Phys. Eng. Sci. **469**, 20130201 (2013).

[33] L. Grigoryeva and J.-P. Ortega, "Echo state networks are universal," Neural Netw. **108**, 495–508 (2018).

[34] A. Hart, J. Hook, and J. Dawes, "Embedding and approximation theorems for echo state networks," Neural Netw. **128**, 234–247 (2020).

[35] C. Hartmann, R. Banisch, M. Sarich, T. Badowski, and C. Schütte, "Characterization of rare events in molecular dynamics," Entropy **16**, 350–376 (2014).

[36] K. Hasselmann, "Stochastic climate models part I. Theory," Tellus **28**, 473–485 (1976).

[37] I. Ilies, H. Jaeger, O. Kosuchinas, M. Rincon, V. Sakenas, and N. Vaskevicius, "Stepping forward through echoes of the past: Forecasting with echo state

networks" (2007), Vol. 53, p. 76, Short report on the winning entry to the NN3 financial forecasting competition, see http://www.neural-forecasting-competition.com/downloads/NN3/methods/27-NN3HerbertJaegerreport.pdf.

[38] H. Jaeger, "Tutorial on training recurrent neural networks, covering BPPT, RTRL, EKF and the "echo state network" approach," Vol. 5, GMD-Forschungszentrum Informationstechnik Bonn, 2002.

[39] H. Jaeger and H. Haas, "Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication," Science **304**, 78–80 (2004).

[40] D. T. Kaplan and L. Glass, "Coarse-grained embeddings of time series: Random walks, Gaussian random processes, and deterministic chaos," Physica D **64**, 431–454 (1993).

[41] D. Kelly and I. Melbourne, "Deterministic homogenization for fast–slow systems with chaotic noise," J. Funct. Anal. **272**, 4063–4102 (2017).

[42] D. Kelly, I. Melbourne *et al.*, "Smooth approximation of stochastic differential equations," Ann. Probab. **44**, 479–520 (2016).

[43] C. Kuehn, "A mathematical framework for critical transitions: Bifurcations, fast–slow systems and stochastic dynamics," Physica D **240**, 1020–1035 (2011).

[44] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner *et al.*, "Gradient-based learning applied to document recognition," Proc. IEEE **86**, 2278–2324 (1998).

[45] S. H. Lim, "Understanding recurrent neural networks using nonequilibrium response theory," arXiv:2006.11052 (2020).

[46] E. N. Lorenz, "Deterministic nonperiodic flow," J. Atmos. Sci. **20**, 130–141 (1963).

[47] E. N. Lorenz, "Predictability: A problem partly solved," in *Seminar Proceedings on Predictability* (ECMWF, 1995), Vol. 1.

[48] Z. Lu, B. R. Hunt, and E. Ott, "Attractor reconstruction by machine learning," Chaos **28**, 061104 (2018).

[49] M. Lukoševičius, "A practical guide to applying echo state networks," in *Neural Networks: Tricks of the Trade* (Springer, 2012), pp. 659–686.

[50] M. Lukoševičius and H. Jaeger, "Reservoir computing approaches to recurrent neural network training," Comput. Sci. Rev. **3**, 127–149 (2009).

[51] M. Lukoševičius and A. Uselis, "Efficient cross-validation of echo state networks," in *Artificial Neural Networks and Machine Learning—ICANN 2019: Workshop and Special Sessions*, edited by I. V. Tetko, V. Kůrková, P. Karpov, and F. Theis (Springer International Publishing, Cham, 2019), pp. 121–133.

[52] B. Lusch, J. N. Kutz, and S. L. Brunton, "Deep learning for universal linear embeddings of nonlinear dynamics," Nat. Commun. **9**, 4950 (2018).

[53] M. C. Mackey and M. Tyran-Kamińska, "Deterministic Brownian motion: The effects of perturbing a dynamical system by a chaotic semi-dynamical system," Phys. Rep. **422**, 167–222 (2006).

[54] G. Manjunath and H. Jaeger, "Echo state property linked to an input: Exploring a fundamental characteristic of recurrent neural networks," Neural Comput. **25**, 671–696 (2013).

[55] P. Mehta, M. Bukov, C.-H. Wang, A. G. Day, C. Richardson, C. K. Fisher, and D. J. Schwab, "A high-bias, low-variance introduction to machine learning for physicists," Phys. Rep. **810**, 1–124 (2019).

[56] I. Melbourne and A. Stuart, "A note on diffusion limits of chaotic skew-product flows," Nonlinearity **24**, 1361 (2011).

[57] L. Mitchell and G. A. Gottwald, "Data assimilation in slow–fast systems using homogenized climate models," J. Atmos. Sci. **69**, 1359–1377 (2012).

[58] Doing so will usually, but not always, ensure that the resulting network satisfies the echo state property.[38]

[59] R. Pascanu, C. Gulcehre, K. Cho, and Y. Bengio, "How to construct deep recurrent neural networks," arXiv:1312.6026 (2013).

[60] J. Pathak, B. Hunt, M. Girvan, Z. Lu, and E. Ott, "Model-free prediction of large spatiotemporally chaotic systems from data: A reservoir computing approach," Phys. Rev. Lett. **120**, 024102 (2018).

[61] J. Pathak, Z. Lu, B. R. Hunt, M. Girvan, and E. Ott, "Using machine learning to replicate chaotic attractors and calculate Lyapunov exponents from data," Chaos **27**, 121102 (2017).

[62] J. Pathak, A. Wikner, R. Fussell, S. Chandra, B. R. Hunt, M. Girvan, and E. Ott, "Hybrid forecasting of chaotic processes: Using machine learning in conjunction with a knowledge-based model," Chaos **28**, 041101 (2018).

[63] G. Pavliotis and A. Stuart, *Multiscale Methods*, Texts in Applied Mathematics Vol. 53 (Springer, New York, 2008).

[64] L. M. Pecora and T. L. Carroll, "Synchronization of chaotic systems," Chaos **25**, 097611 (2015).

[65] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics informed deep learning (part I): Data-driven solutions of nonlinear partial differential equations," arXiv:1711.10561 (2017).

[66] P. Riley, "Three pitfalls to avoid in machine learning," Nature **572**, 27–29 (2019).

[67] C. Rödenbeck, C. Beck, and H. Kantz, "Dynamical systems with time scale separation: Averaging, stochastic modelling, and central limit theorems," in *Stochastic Climate Models* (Springer, 2001), pp. 189–209.

[68] D. Rolnick, P. L. Donti, L. H. Kaack, K. Kochanski, A. Lacoste, K. Sankaran, A. Slavin Ross, N. Milojevic-Dupont, N. Jaques, and A. Waldman-Brown, "Tackling climate change with machine learning," arXiv:1906.05433 (2019).

[69] P. Romanczuk, M. Bär, W. Ebeling, B. Lindner, and L. Schimansky-Geier, "Active Brownian particles," Eur. Phys. J. Spec. Top. **202**, 1–162 (2012).

[70] S. Scardapane and D. Wang, "Randomness in neural networks: An overview," Wiley Interdiscip. Rev. Data Min. Knowl. Discov. **7**, e1200 (2017).

[71] M. Scheffer, J. Bascompte, W. A. Brock, V. Brovkin, S. R. Carpenter, V. Dakos, H. Held, E. H. Van Nes, M. Rietkerk, and G. Sugihara, "Early-warning signals for critical transitions," Nature **461**, 53–59 (2009).

[72] S. Scher, "Toward data-driven weather and climate forecasting: Approximating a simple general circulation model with deep learning," Geophys. Res. Lett. **45**, 12616–12622, https://doi.org/10.1029/2018GL080704 (2018).

[73] A. Sherstinsky, "Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network," Physica D **404**, 132306 (2020).

[74] H. Stone, "Calibrating rough volatility models: A convolutional neural network approach," Quant. Finance **20**, 379–392 (2020).

[75] P. Tino, "Dynamical systems as temporal feature spaces," J. Mach. Learn. Res. **21**, 1–42 (2020).

[76] P. Verzelli, C. Alippi, and L. Livi, "Learn to synchronize, synchronize to learn," arXiv:2010.02860 (2020).

[77] P. Verzelli, C. Alippi, L. Livi, and P. Tino, "Input representation in recurrent neural networks dynamics," arXiv:2003.10585 (2020).

[78] P. R. Vlachas, W. Byeon, Z. Y. Wan, T. P. Sapsis, and P. Koumoutsakos, "Data-driven forecasting of high-dimensional chaotic systems with long short-term memory networks," Proc. R. Soc. A: Math. Phys. Eng. Sci. **474**, 20170844 (2018).

[79] P. R. Vlachas, J. Pathak, B. R. Hunt, T. P. Sapsis, M. Girvan, E. Ott, and P. Koumoutsakos, "Backpropagation algorithms and reservoir computing in recurrent neural networks for the forecasting of complex spatiotemporal dynamics," Neural Netw. **126**, 191–217 (2020).

[80] T. Weng, H. Yang, C. Gu, J. Zhang, and M. Small, "Synchronization of chaotic systems and their machine-learning models," Phys. Rev. E **99**, 042203 (2019).

[81] G. Yehudai and O. Shamir, "On the power and limitations of random features for understanding neural networks," in *Advances in Neural Information Processing Systems* (Curran Associates, Inc., 2019), Vol. 32, pp. 6594–6604.

[82] J. Yperman and T. Becker, "Bayesian optimization of hyper-parameters in reservoir computing," arXiv:1611.05193 (2016).

[83] R. S. Zimmermann and U. Parlitz, "Observing spatio-temporal dynamics of excitable media using reservoir computing," Chaos **28**, 043118 (2018).

[84] G. Tanaka, T. Yamane, J. B. Héroux, R. Nakane, N. Kanazawa, S. Takeda, H. Numata, D. Nakano, and A. Hirose, "Recent advances in physical reservoir computing: A review," Neural Networks **115**, 100–123 (2019).

Paper IV

L.T. Giorgini, W. Moon, N. Chen and J.S. Wettlaufer, A Non-Gaussian Stochastic Model for the El Niño Southern Oscillation, arXiv: 2012.09858 (2021).

A Non-Gaussian Stochastic Model for the El Niño Southern Oscillation

L. T. Giorgini,[1, *] W. Moon,[2, 1, †] N. Chen,[3, ‡] and J. S. Wettlaufer[4, 1, §]
[1]*Nordita, Royal Institute of Technology and Stockholm University, Stockholm 106 91, Sweden*
[2]*Department of Mathematics, Stockholm University 106 91 Stockholm, Sweden*
[3]*Department of Mathematics, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA*
[4]*Yale University, New Haven, Connecticut 06520, USA*

(Dated: December 21, 2020)

A non-autonomous stochastic dynamical model approach is developed to describe the seasonal to interannual variability of the El Niño-Southern Oscillation (ENSO). We determine the model coefficients by systematic statistical estimations using partial observations involving only sea surface temperature data. Our approach reproduces the observed seasonal phase locking and its uncertainty, as well as the highly non-Gaussian statistics of ENSO. Finally, we recover the intermittent time series of the hidden processes, including the thermocline depth and the wind bursts.

Air-sea interactions in the tropical Pacific drive the largest interannual variability process in climate called the El Niño-Southern Oscillation (ENSO), with an influence that reaches the higher latitudes via atmospheric and oceanic teleconnections [1, 2]. Due to its spatiotemporal impacts, estimating the seasonal to interannual state of ENSO is essential for predicting a wide range of regional and global climate phenomena [3, 4]. However, the complexity of ENSO, involving as it does stochastic forcing from atmospheric transients [e.g., 5] and nonlinear air-sea interactions [e.g., 6], poses great challenges in both modeling and prediction.

There are two traditional methods of modeling ENSO [see e.g., 7, for a review]. One is to use the state-of-art coupled general circulation models (GCMs), which solve the governing conservation (energy, mass, momentum) equations for the rotating planet with many sub-grid-scale parameterizations of unresolved small-scale processes. The other is to build low-order statistical models. The former treat the main physical aspects of ENSO and thus provide reasonable representations of the spatiotemporal patterns and air-sea interactions from intraseasonal to interannual time scales [8]. The latter, commonly designed to describe and predict particular ENSO indices, characterize the large-scale features of the system and are much cheaper computationally than GCMs [9]. Nonetheless, despite great progress in developing these methods, there remain limitations to advancing analysis and prediction. On the one hand, GCMs incur systematic errors originating from inaccurate state estimations and incomplete sub-grid-scale parameterizations associated with the ocean thermocline, tropical instability waves, and the Madden-Julian Oscillation (MJO) [11, 12]. On the other hand, the statistical models based on multivariate regressions cannot provide detailed physical information regarding the phase and intensity of ENSO [13]. One direction for progress is to utilize a hybrid strategy that simultaneously maximizes the advantages of both physically-oriented and statistical models.

The basic theoretical understanding of the tropical air-sea interactions underlying ENSO [14, 15] has lead to a hierarchy of simple models of the main processes that control the sea surface temperature (SST) in the tropical ocean [e.g., 16]. A prominent approach is the recharge-discharge model [17], and its extensions and generalizations, derived from the forced shallow water equations using the two-strip approximation. Despite it being a two-dimensional linear model, it (and its generalizations) can capture inter-annual variability, seasonal variability with time-periodic coefficients [18, 19], and a range of stochastic forcing representing weather and intraseasonal processes, such as westerly wind bursts (WWBs) [20].

Here we describe a two-stage stochastic model approach that captures the large-scale dynamical and statistical characteristics of ENSO. Stochastic recharge-discharge oscillator models have used observations to recover statistics on inter-annual time-scales [21, 22], and have incorporated the seasonal cycle of the Bjerknes feedback to treat ENSO phase-locking and the spring predictability barrier [18, 19, 23]. However, while our approach builds upon the recharge-discharge model, we incorporate a slowly-varying low-order deterministic component that systematically treats time-varying coefficients and multiplicative noise to accurately describe the central intraseasonal to interannual features of ENSO. In particular, our framework reproduces ENSO's observed seasonal phase locking, uncertainty and highly non-Gaussian statistics. However, we clearly depart from multivariate regression statistical models that require observations of all state variables. Rather, our non-Gaussian model incorporates a subset of observations from which we can infer unobserved quantities and quantify their uncertainties using an efficient and exact data assimilation scheme, and can thereby accurately recover the difficult to observe thermocline depth and wind bursts. Finally, by systematically determining time-dependent model coefficients, we precisely simulate the seasonal to inter-annual ENSO statistics.

---

[*] ludovico.giorgini@su.se
[†] woosok.moon@su.se
[‡] chennan@math.wisc.edu
[§] john.wettlaufer@yale.edu

We begin with the following coupled model,

$$\begin{aligned}\dot{x}(t) &= a(t)x(t) + \omega(t)h(t) + N(t)\xi_x(t) \quad \text{and} \\ \dot{h}(t) &= -\omega(t)x(t) + \lambda h(t) + \sigma\xi_h(t),\end{aligned} \tag{1}$$

in which $x(t)$ is the averaged SST anomaly in the equatorial central-eastern Pacific (the so-called Niño 3.4 region), and $h(t)$ is the thermocline depth, which is a surrogate of the heat content, averaged over the western tropical Pacific. The time-dependent function, $a(t)$, represents the seasonal evolution of the Bjerknes feedback and the noise amplitude, $N(t)$, captures the role of the relatively short time-scale forcing of the SST anomaly $x(t)$. Both $\xi_x(t)$ and $\xi_h(t)$ are independent Gaussian white noise processes. Finally, the other time-dependent function, $\omega(t)$, controls the coupling of tropical atmosphere and the upper ocean, which is one of the key parameters that determines the quasi-oscillatory behavior of the dominant ENSO modes, and in traditional low-order models is treated as a constant [18]. Here, however, treating it as time-dependent facilitates a more accurate description of the intraseasonal and interannual statistics.

We emphasize the role of the seasonality of background fields in the tropical Pacific by making all the time dependent coefficients in (1) periodic on the annual cycle. We determine the functions $a(t)$ and $N(t)$ using the monthly statistics of the Niño 3.4 index [24]. The coefficient $a(t)$ is related to the relationship between the monthly variance $\langle x^2(t)\rangle$ and the covariance between two adjacent months $\langle x(t)x(t+\Delta t)\rangle$. We use the derived $a(t)$ to estimate $N(t)$ from the statistics of the derived data; $y = x(t+\Delta t) - x(t) - a(t)x(t)\Delta t$. It has been shown previously that $a(t)$ and $N(t)$ can be accurately estimated independent of the other coefficients (see supplementary material in [24]). The coupling function, $\omega(t)$, and the constants λ and σ, are determined by an expectation-maximization algorithm [26, 27, 42], which is a useful and efficient statistical estimation method involving incomplete data. We note that only the Niño 3.4 index data is used as the partial observation for model calibration, whereas the thermocline depth, $h(t)$, is assumed to be unknown since it is not directly available from satellite observations. Therefore, as detailed in Appendix B, C, we infer $h(t)$ together with unknown coefficients $\omega(t)$, λ, and σ [28].

We evaluate the quality of the model by comparing the model statistics with those of the observed Niño 3.4 index in Fig. 5. A qualitative comparison between a random realization of the model and the observed data show comparable variability in Fig. 5(a) and quantitative comparisons of the probability density functions (PDFs) and the autocorrelation functions (ACFs) of the SST anomaly shown in Figs. 5(b) and (c) demonstrate that the model is able to accurately simulate the interannual quasi-oscillatory behavior of ENSO. Moreover, as seen in Fig. 5(d), the model reproduces monthly standard deviations of the Niño 3.4 index data with high precision. Therefore, two important seasonal features of ENSO, "phase locking" to the seasonal cycle [29] and the "spring predictability barrier" [19], are accurately captured in this two dimensional model. Phase locking is related to the activation of the positive Bjerknes feedback from summer to early winter, which leads to the concentration of abnormal ENSO events during November and December, as is clear from the maximum of the standard deviation near the end of the year. This phase locking feature is clearly seen from the model trajectory (Fig. 5a), where the peaks of nearly all of the major ENSO events occur in the boreal winter, consistent with the observations. The spring predictability barrier is related to the increase of the noise magnitude, $N(t)$, during the spring, which increases model uncertainties during the boreal summer thereby causing the loss of predictability [24]. Indeed, this is reflected by the structural similarity between the model and the observations in the persistence forecast diagrams shown in Figs. 5(e) and (f).

Despite its simplicity, the model system of Eqs. (1) recovers the monthly standard deviation and ACFs of the Niño 3.4 index with higher accuracy than coupled GCMs [30], linear and nonlinear statistical models [31], and simpler recharge-discharge based approaches [18]. Our systematic statistical estimation of coefficients enable this stochastic model to regenerate the statistics of the Niño 3.4 index. In particular, as shown in Fig. 5(c), the precise representation of the ACFs beyond 2 years suggests that the overall dynamical features of the thermocline depth, $h(t)$, are well captured in this model. Variations in $h(t)$, and thus of the equatorial warm water volume (WWV) in the western tropical Pacific, influence the SST through vertical advection of temperature anomalies by mean upwelling, a process known as the "thermocline feedback". Thus, the thermocline depth acts as a precursor for the inter-annual prediction of SST anomalies [32].

Although the linear model described above successfully captures the seasonal variability and the interannual large-scale dynamical features of ENSO, it only allows Gaussian PDFs, whereas the observed statistics are highly non-Gaussian. Importantly, a model must include atmospheric wind burst forcing, which plays a critical role in generating the extreme ENSO events responsible for non-Gaussianity [20, 34]. Therefore, we include wind bursts in the stochastic model of Eqs. (1) as follows,

$$\begin{aligned}\dot{x}(t) &= a(t)x(t) + \omega(t)h(t) + \alpha_1\tau(t) + N(t)\xi_x(t), \\ \dot{h}(t) &= -\omega(t)x(t) + \lambda h(t) + \alpha_2\tau(t) + \sigma\xi_h(t) \quad \text{and} \\ \dot{\tau}(t) &= d_\tau\tau(t) + \rho(x)\xi_\tau(t),\end{aligned} \tag{2}$$

where we adopt a simple linear process with multiplicative noise to generate the stochastic wind bursts $\tau(t)$ that drive the SST and thermocline dynamics. The parameters in the wind burst equation are chosen to accurately reproduce the observed wind stress (Fig. 6 (d)-(f)), while the coefficients of Eqs. (1) were slightly modified to take into account the additional process in the model. Despite the non-Gaussianity of the model, the thermocline depth and the wind bursts remain as Gaussian processes

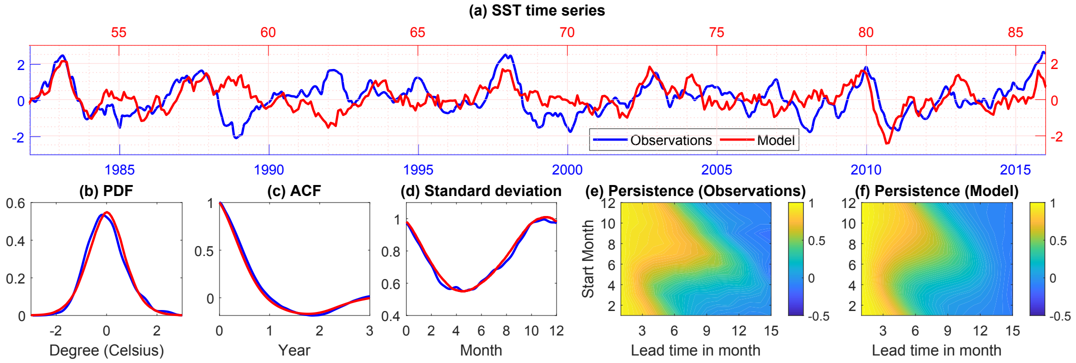


FIG. 1. Comparison of the observed SST (blue; bottom x-axis) with the model Eq. (1) (red; top x-axis). The SST time series are shown in (a), where the model result is one of the random realizations from Eq. (1) that shares the same length as the observational data from 1983 to 2016. The comparisons of the PDFs and ACFs are shown in (b) and (c). The SST standard deviation as a function of calendar month is shown in (d). In (e) and (f), the diagram of the persistence forecast from the observational data and that from the model can be compared.

conditioned on the observed SST. Moreover, by using the partial observations that involve only the Niño SST data, we have closed analytic formulae for solving the conditional distribution of the unobserved thermocline depth and the wind bursts [see 33, and Appendix B],

$$p(h(t), \tau(t)|x(s), s \leq t) \sim \mathcal{N}(\mu_f(t), R_f(t)), \qquad (3)$$

where $\mu_f(t)$ and $R_f(t)$ are the so-called filter mean and filter covariance in data assimilation.

The filter estimate (3) of the non-Gaussian model (2) can be used to carry out an online time series reconstruction of the unobserved variables and analyze the associated uncertainty, which can also serve as the forecast initialization. Note that the observational data for the equatorial heat content (the integrated WWV above the 20^oC isotherm between 5N-5S, 120E to 80W [32], which is equivalent to the thermocline depth $h(t)$ in (2)) is only available since 1983 [?], whereas SST data is available from the pre-industrial era. Therefore, the available SST data can be incorporated into the model (2) to reconstruct the thermocline data over a much longer period.

We first validate the accuracy of the recovered $h(t)$ and $\tau(t)$ in the period after 1983. Fig. 6 (a) shows the filter mean time series of the thermocline depth, $h(t)$, (red) from the non-Gaussian model (2), which compares favorably to the observational value (blue) with a pattern correlation that is nearly 0.9. Moreover, the small error between the observed and the recovered time series lies approximately within one standard deviation of the filter estimated uncertainty (pink shaded region). In particular, as expected from the physical role of the WWBs, the extreme events are accurately recovered. Similar conclusions are reached for the wind burst time series (Fig. 6e). In addition to a skillful state estimation, the reconstructed wind burst time series has the further influence of filtering out the independent white noise in observations. As shown in Figs. 6(b, f), the model PDFs capture those of the observations for both the thermocline depth and the wind bursts, particularly the significant non-Gaussian skewness and kurtosis. Moreover, similar accuracy is found in ACFs and the power spectra as shown Figs. 6(c, d, g, h). The success in recovering unobserved processes is due to the confluence of the model containing the essential physics and properly constrained stochastic forcing and coefficients.

As shown Fig. 6, the accurate nonlinear filter estimates of Eq. (3) provide an online estimation method of the thermocline depth $h(t)$ and the wind bursts $\tau(t)$ relevant for model initialization. On the other hand, reconstructing historical data can be achieved by solving a slightly different conditional distribution

$$p(h(t), \tau(t)|x(s), 0 \leq s \leq T) \sim \mathcal{N}(\mu_s(t), R_s(t)), \qquad (4)$$

where $t \in [0, T]$. This is the so-called smoother estimate. It differs from the filter in the sense that it exploits all the available information up to the current time instant. One desirable feature of the nonlinear smoother estimate in Eq. (4) associated with the non-Gaussian system in Eqs. (2) is that it can be solved via closed analytic formulae (see Appendix B). We use the smoother estimate to recover the historical realizations of the two hidden variables conditioned on the entire observational period of the Niño 3.4 index from 1870 to present. The results are shown in Fig. 3. Since the thermocline depth and the wind bursts are the most important precursors for predicting the state of ENSO, these recovered variables can be used to evaluate and improve prediction models and provide a guideline for improving parameterizations in coupled GCMs.

Understanding the complexity of climate dynamics presents a dilemma of completeness, between using solely statistical regression models lack the essential physics, to

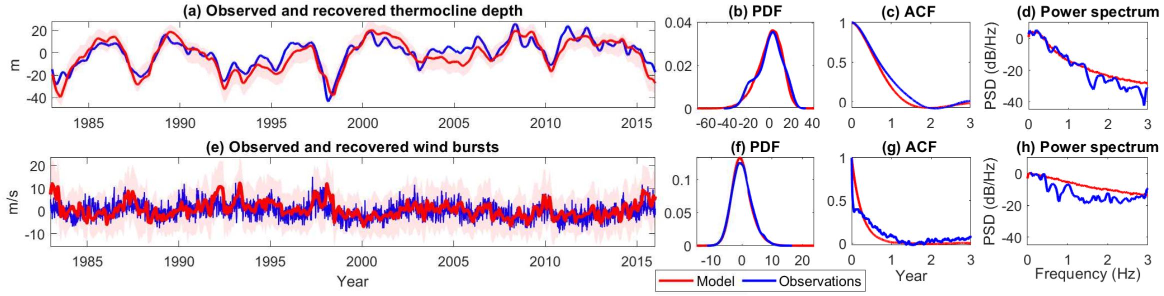


FIG. 2. Comparison of the observational value (blue) and the recovered time series from the filter estimate (red) using Eq. (3), based on the non-Gaussian model (2) for the thermocline depth (a) and the wind bursts (e). The solid red curves are the filter mean while the shaded red areas show one standard deviation in the filter estimate (the square root of the filter variance). The comparison of the PDFs, the ACFs and the power spectra between the observations and simulations from the non-Gaussian model (2) are shown in (b)–(d), for the thermocline depth and (f)–(h), for the wind bursts.

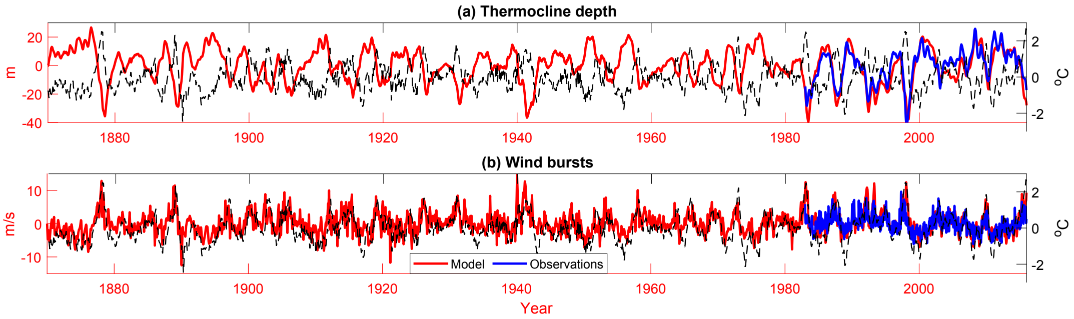


FIG. 3. The recovered time series of the hidden variables (red curves; left y-axis) for (a) the thermocline depth and (b) the wind bursts, using the non-Gaussian model, Eqs. (2), with the partial observations involving only SST data (dashed black curves; right y-axis) from 1870 to 2016. The method used here is the nonlinear smoother estimate, Eq. (4). The blue curves show the observational data from 1983 to 2016.

global climate models, which have more variables and parameterizations than observables. Our approach presents a compromise through hybridization. We combine a slowly-varying low-order dynamical system with an observationally faithful statistical representation of short-time processes. In consequence we find a realistic description of ENSO as measured by the success in (i) reproducing the observed seasonal phase locking and its uncertainty; (ii) capturing the spring predictability barrier; (iii) simulating the observed highly non-Gaussian statistics, and (iv) accurately recovering the intermittent time series of the hidden processes via a nonlinear but analytically solvable data assimilation scheme. Furthermore, we suggest that extending the approach described here to examine climate variability on a wide range of time scales [2, 36], wherein proxy data are available, will provide important insights into mechanisms and feedbacks in the climate system. Finally, given the generality of the stochastic dynamical systems in physics, it is hoped that our approach can be adopted not only for other problems in climate, but far more broadly in the physical sciences.

ACKNOWLEDGMENTS

L.T.G., W.M. and J.S.W. gratefully acknowledge support from the Swedish Research Council Grant No. 638-2013-9243. The research of N.C. is partially funded by the Office of VCRGE at UW-Madison.

Appendix A: Observational Data

All the observational data used in the main paper are plotted in Fig. (4) and described here.

(1) Monthly sea surface temperature (SST) anomalies derived from the temperature analyses of the Bureau National Operations Centre (BNOC) at the Australian Bureau of Meteorology [37]. These data have been averaged over the Niño 3.4 region (5N-5S, 170W-120W) and start from the pre-industrial era spanning the entire period 1870-2016. The SST data are the only observations used in the main pa-

per to recover the model coefficients and to reconstruct the hidden processes involved.

(2) Daily thermocline depth data, defined as the integrated warm water volume (WWV) above the 20^0C isotherm, downloaded from the NCEP/GODAS reanalysis [38] and averaged within 5N-5S and 80W-120W in the tropical Pacific Ocean. This dataset was used in the main paper to test the performance of both the two and three dimensional models in estimating the hidden state of the thermocline depth correctly, together with their long-term (or "equilibrium") statistical and dynamical features, evaluated by comparing the probability distribution functions (PDFs) and the autocorrelation functions (ACFs) obtained from the models and from observations. The thermocline depth data are available only in the post-industrial era, from 1983. Therefore, it was not possible to compare with observations the recovered hidden process over its total length, but only its realization from 1983.

(3) Daily westerly wind bursts (WWBs) at 850hPa from the NCEP/NCAR reanalysis [39]. By introducing the WWBs into our three dimensional model (Eqs. 2 in the main paper) we reproduced the non-Gaussian features observed in ENSO statistics that are not possible in a linear model. The relevant observations were only available from 1983 and thus comparisons of the mean state to the statistics of the hidden processes recovered through the model to observations began at that date.

Appendix B: State Estimation and Data Assimilation

1. Online data assimilation

Our two-dimensional model (Eqs. 1 in the main text) is a linear model with time-varying coefficients. Therefore, given the observations of the SST, the classical Kalman-Bucy filter [40] can be directly applied to obtain the state estimation of the thermocline depth. In our three-dimensional non-Gaussian model (Eqs. 2 of the main text), the multiplicative noise prevents the use of the Kalman-Bucy filter for the state estimation. Nevertheless, conditional on the observed SST, the thermocline depth and the wind bursts remain as Gaussian processes. For such a conditional Gaussian system, the conditional distribution $p(h(t),\tau(t)|x(s), s\le t)\sim\mathcal{N}(\boldsymbol{\mu}_f(t),\mathbf{R}_f(t))$ is Gaussian. Note that despite the non-Gaussianity, the conditional Gaussian distribution can be solved via closed analytic formulae [33]. This facilitates an efficient data assimilation scheme to recover the states of $h(t)$ and $\tau(t)$ and their uncertainties through the filter mean $\boldsymbol{\mu}_f(t)$ and the square root of the filter covariance $\mathbf{R}_f(t)$, which for our three dimensional model are given by

$$\mathrm{d}\boldsymbol{\mu}_f = (\mathbf{c}_0+\mathbf{c}_1\boldsymbol{\mu}_f)\,\mathrm{d}t + \mathbf{R}_f\mathbf{C}_1^T B^{-2}(\,\mathrm{d}x - (C_0+\mathbf{C}_1\boldsymbol{\mu}_f)\,\mathrm{d}t), \tag{B1a}$$

$$\mathrm{d}\mathbf{R}_f = \big(\mathbf{c}_1\mathbf{R}_f + \mathbf{R}_f\mathbf{c}_1^T + \mathbf{b}\mathbf{b}^T - (\mathbf{R}_f\mathbf{C}_1^T)B^{-2}(\mathbf{R}_f\mathbf{C}_1^T)^T\big)\,\mathrm{d}t, \tag{B1b}$$

respectively, where

$$\begin{aligned} \mathbf{c}_0 &= \begin{bmatrix}-\omega x\\ 0\end{bmatrix}, \quad \mathbf{c}_1 = \begin{bmatrix}\lambda & \alpha_2\\ 0 & d_\tau\end{bmatrix},\\ C_0 &= ax, \quad \mathbf{C}_1 = \begin{bmatrix}\omega & \alpha_1\end{bmatrix},\\ B &= N, \quad \mathbf{b} = \begin{bmatrix}\sigma & 0\\ 0 & \rho\end{bmatrix}. \end{aligned} \tag{B2}$$

The filter mean and covariance of the two dimensional model can be obtained from these expressions by setting $\alpha_1, \alpha_2 = 0$.

2. Offline smoothing

The data assimilation process exploits the observations up to the current time instant for the purpose of initialization. On the other hand, for reconstructing the unobserved variables, conditioned on the entire observational signal (such as obtaining reanalysis data), the so-called smoothing technique is preferable. For example, the smoothing of the three-dimensional non-Gaussian system aims to solve $p(h(t),\tau(t)|x(s), 0\le s\le T)\sim\mathcal{N}(\boldsymbol{\mu}_s(t),\mathbf{R}_s(t))$, where $t\in[0,T]$. The smoothing technique can be regarded as a two-step data assimilation process, with a forward pass of filtering and a backward pass of smoothing. Although there are no closed formulae for the smoother mean $\boldsymbol{\mu}_s(t)$ and covariance $\mathbf{R}_s(t)$, closed analytic formulae *are* available for both the systems studied in the main text with appeal to the conditional Gaussian structure [41]

$$\mathrm{d}(-\boldsymbol{\mu}_s) = \big(-\mathbf{c}_0 - \mathbf{c}_1\boldsymbol{\mu}_s + (\mathbf{b}\mathbf{b}^T)\mathbf{R}_f^{-1}(\boldsymbol{\mu}_f-\boldsymbol{\mu}_s)\big)\,\mathrm{d}t, \tag{B3a}$$

$$\begin{aligned}\mathrm{d}(-\mathbf{R}_s) = -\big((\mathbf{c}_1 + (\mathbf{b}\mathbf{b}^T)\mathbf{R}_f^{-1})\mathbf{R}_s\\ + \mathbf{R}_s(\mathbf{c}_1^T + (\mathbf{b}\mathbf{b}^T)\mathbf{R}_f^{-1}) - \mathbf{b}\mathbf{b}^T\big)\,\mathrm{d}t,\end{aligned} \tag{B3b}$$

where $\boldsymbol{\mu}_f, \mathbf{R}_f$ are the filter mean and covariance respectively, defined in Eq. (B1), while the coefficients $\mathbf{c}_0, \mathbf{c}_1, \mathbf{b}$ have been specified in Eq. (B2).

In (B3), the terms of the left hand side are understood as

$$\begin{aligned}\mathrm{d}(-\boldsymbol{\mu}_s) &= \lim_{\Delta t\to 0}\boldsymbol{\mu}_s(t) - \boldsymbol{\mu}_s(t+\Delta t),\\ \mathrm{d}(-\mathbf{R}_s) &= \lim_{\Delta t\to 0}\mathbf{R}_s(t) - \mathbf{R}_s(t+\Delta t).\end{aligned}$$

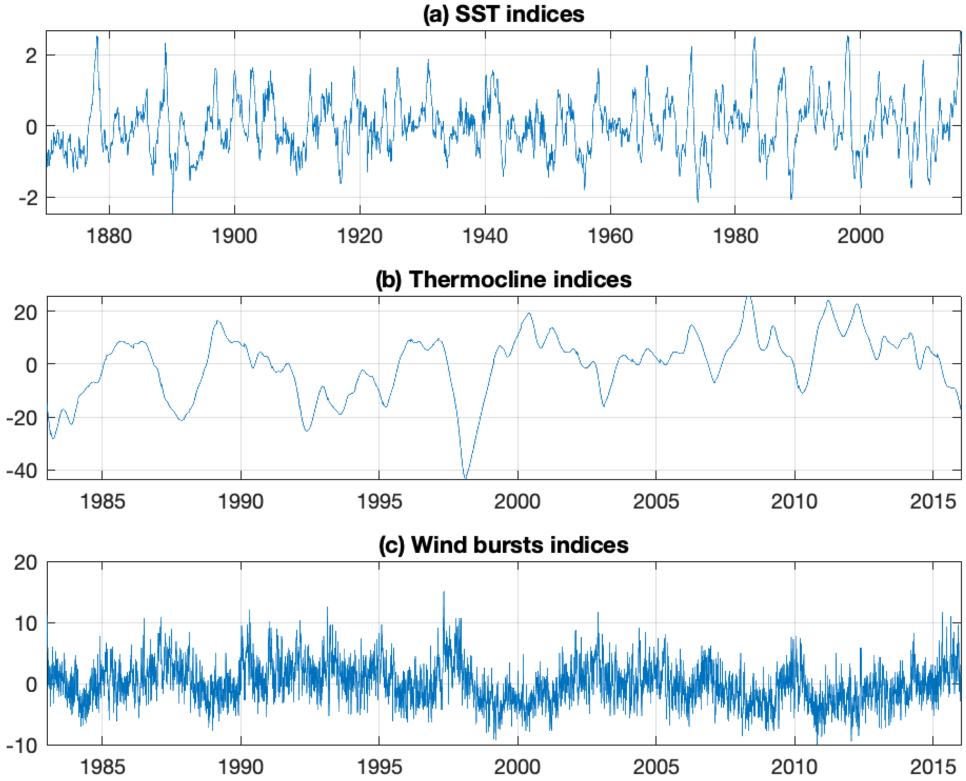


FIG. 4. Observational data used (a) the Niño 3.4 index, (b) the thermocline depth averaged over the west Pacific and (c) the westerly wind bursts.

Appendix C: Parameter Estimation

1. Parameter estimation of the time-varying parameters $a(t)$ and $N(t)$

In this section, we will use the monthly-averaged SST data to construct the periodic drift and diffusion coefficients, $a(t)$ and $N(t)$, that appear in the x-equation of Eqs. (1) in the main paper viz.,

$$\begin{aligned}\dot{x}(t) &= a(t)x(t) + \omega(t)h(t) + N(t)\xi_x(t) \quad \text{and}\\ \dot{h}(t) &= -\omega(t)x(t) + \lambda h(t) + \sigma\xi_h(t).\end{aligned} \tag{C1}$$

Here we consider a discretized version of this two dimensional model with Δt the discrete time unit. For simplicity, we set the yearly frequency of the coefficients equal to 2π. We define N_y as the total number of years of the observed SST data, each of which contains Δt^{-1} data points. We multiply both sides of the x-equation by x, and upon taking the average we obtain

$$\langle x_i x_{i+1}\rangle_i - \langle x_i^2\rangle_i = \Delta t(a_i\langle x_i^2\rangle_i - \omega_i\langle h_i\, x_i\rangle_i) \;\; \forall i \in (0, \Delta t^{-1}], \tag{C2}$$

where the average has been performed only over the quantities in the i-th time step inside each period T with $T = 1$. Using the h-equation, the last term on the right hand side of Eq. (C2) can be written as

$$\begin{aligned}\langle h_i\, x_i\rangle_i &\simeq \frac{1}{N_y}\left(\sum_{j=1}^{N_y}\sum_{s=0}^{j\Delta t^{-1}+i\Delta t} e^{-\lambda(j\Delta t^{-1}+i\Delta t-s)}\omega_s x_s x_i\Delta t\right)\\ &\simeq \frac{1}{N_y}\bar{\omega}\left(\sum_{j=1}^{N_y}\sum_{s=j\Delta t^{-1}+i\Delta t-\Lambda(\lambda)}^{j\Delta t^{-1}+i\Delta t} x_s\, x_i\Delta t\right) \simeq\\ &\simeq \bar{\omega}\langle x_{i+\frac{1}{4\Delta t}}\, x_i\rangle_i \simeq 0.\end{aligned} \tag{C3}$$

Here we have used the fact that x is periodic and oscillates around a mean of zero with an average frequency of $\pi/2$ (the natural frequency of ENSO is approximately 4 years), which is four times smaller than that of ω. Therefore, this last coefficient has been approximated in the sum with its average over one period. Moreover, we have exploited the effect of the exponential inside the sum, which provides a cutoff in the lower limit. Thus, the average of this periodic function is a phase shift of a quarter of period and we can approximate this term as zero.

We can then write

$$a_i = \frac{\langle x_i x_{i+1}\rangle_i - \langle x_i^2\rangle_i}{\Delta t\langle x_i^2\rangle_i} + \Delta a_i, \tag{C4}$$

where Δa_i is the subleading correction to a_i arising from the term $\langle h_i\, x_i\rangle_i$.

Taking the square of the x-equation in Eq. (C1) and then averaging, we can isolate N_i viz.,

$$\begin{aligned} N_i^2\Delta t =& \langle[x_{i+1} - x_i - \Delta t(a_i x_i - \omega_i h_i)]^2\rangle_i \simeq \\ \simeq& \langle[x_{i+1} - x_i - \Delta t a_i x_i]^2\rangle_i + \Delta t^2\omega_i^2\langle h_i^2\rangle_i \\ & \forall i \in (0, \Delta t^{-1}], \end{aligned} \tag{C5}$$

where all the terms proportional to $\langle h_i\, x_i\rangle_i$ have been neglected. Now we define $y_i = x_{i+1} - x_i - \Delta t a_i x_i$ and write

$$\begin{aligned} \frac{\langle y_i^2\rangle_i - \langle y_{i+1}y_i\rangle_i}{\Delta t} &= N_i^2 + \Delta t(\omega_i^2\langle h_{i+1}h_i\rangle_i - \omega_i^2\langle h_i^2\rangle_i) \simeq \\ &\simeq N_i^2 + O(\Delta t^2) \;\; \forall i \in (0, \Delta t^{-1}]. \end{aligned} \tag{C6}$$

Finally we can write

$$N_i = \sqrt{\frac{\langle y_i^2\rangle_i - \langle y_{i+1}y_i\rangle_i}{\Delta t}} + \Delta N_i. \tag{C7}$$

We have used Eqs. (C4), (C7) to reconstruct the coefficients $a(t)$ and $N(t)$ from the data and we have plotted them in Fig. (5). From Fig. (5)(a) can be noticed a good agreement between our estimation of $a(t)$ (blue) and the seasonal Bjerknes instability index as given by [18] (red) as $a(t) = -1 - \sin(2\pi t)$.

2. Parameter estimation of the constant coefficients using an expectation-maximization algorithm

An expectation-maximization (EM) algorithm is used for parameter estimation of both the two- and three-dimensional systems. We use this algorithm because we only have the SST data over the entire observational period and hence the system is only "partially observed". Therefore, the parameters and the unobserved states of the thermocline depth and the wind bursts must be estimated simultaneously. The advantage of the EM algorithm is that it iterates and updates the parameters and the hidden states alternatively. Moreover, convergence can be guaranteed. In particular, in the E-step the smoother is applied to estimate the state while in the M-step a maximum likelihood approach is used for parameter estimation. In the E-step, we have used for convenience a normalized (i.e., with unitary variance) smoother estimate of the hidden process, which is the thermocline depth for the two dimensional model. The details of the algorithm can be found in [42].

Below, the EM algorithm is applied to estimate $\omega = \omega_0 + \omega_1\sin(2\pi t) + \omega_2\cos(2\pi t)$ and σ in the linear model. These coefficients have been estimated keeping the drift term λ of the h-process constant in order to guarantee the stability of the algorithm. This procedure has been iterated many times for different values of λ choosing that which minimizes both the relative and spectral relative entropy (see [43] and references therein).

Figure 6 illustrates the iteration of the coefficients towards their optima; $\omega_0 = 1.5$, $\omega_1 = 0.6$, $\omega_2 = -0.5$ and $\sigma = 0.9$, with the optimal value of $\lambda = -0.8$. Note that because of the simplicity of the model structure, the convergence here is global.

3. Parameter estimation of the wind bursts coefficients

We introduced a three dimensional model (Eqs. 2 in the main paper) to take into account the effects of wind bursts as

$$\begin{aligned} &\dot{x}(t) = a(t)x(t) + \omega(t)h(t) + \alpha_1\tau(t) + N(t)\xi_x(t), \\ &\dot{h}(t) = -\omega(t)x(t) + \lambda h(t) + \alpha_2\tau(t) + \sigma\xi_h(t) \;\; \text{and} \\ &\dot{\tau}(t) = d_\tau\tau(t) + \rho(x)\xi_\tau(t), \end{aligned} \tag{C8}$$

with the following parameters

$$\begin{aligned} &a(t) \to 1.5a(t),\; N(t) \to 0.8N(t), \\ &\omega_0 = 1.5,\; \omega_1 = 0.6,\; \omega_2 = -0.5, \\ &\lambda = -1.5,\; \sigma = 0.8, \\ &\alpha_1 = 1,\; \alpha_2 = -0.6,\; d_\tau = -1.5, \\ &\rho(x) = 4.5(\tanh(x) + 1) + 8. \end{aligned} \tag{C9}$$

The parameters that were part of the two dimensional model have been slightly modified to take into account the additional process of wind bursts.

The coefficient $\rho(x)$ in the τ equation has two additive components. The first is a multiplicative noise effect arising from the fact that the warmer SST in the eastern tropical Pacific is usually accompanied by strong wind bursts [e.g., 44]. The second is an additive noise effect, which will increase the variability of the quiescent phases of the SST. These combined effects reproduce the observed non-Gaussian statistics of the model variables.

We let $d_\tau = \pi/2$, so that the damping time scale of the wind bursts is $1/4$ year, but we emphasize that this is not

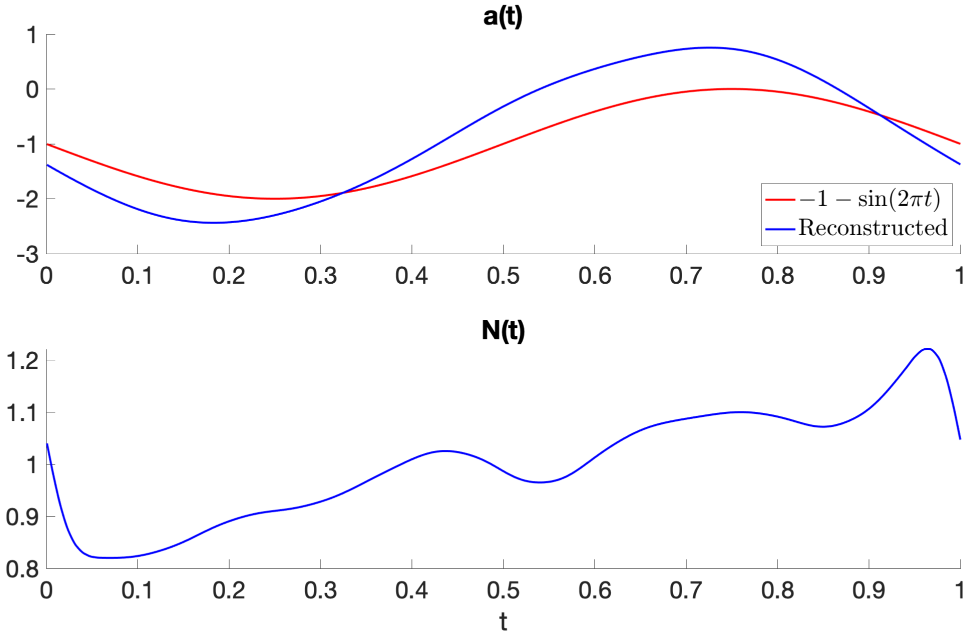


FIG. 5. Shape of $a(t)$ and $N(t)$ obtained from the data (blue) and comparison with the estimation of $a(t)$ used in [18] (red).

the time scale for the wind bursts themselves. Rather, this is the time scale for the accumulation effect of the wind bursts, tropical convection and the MJO to trigger ENSO. This choice ensures that the decorrelation timescale of the wind bursts in the model is the same as that of the observations.

[1] M. J. McPhaden, S. E. Zebiak, and M. H. Glantz, Science **314**, 1740 (2006).

[2] M. Ghil and V. Lucarini, Rev. Mod. Phys. **92**, 035002 (2020).

[3] D. S. Battisti, and E. S. Sarachik, Rev. Geophys. **33**, 1367 (1995).

[4] D. Zhou, A. Gozolchiani, Y. Ashkenazy, and S. Havlin, Phys. Rev. Lett. **115**, 268501 (2015).

[5] R. Kleeman, and A. M. Moore, J. Atmos. Sci. **54**, 753 (1997).

[6] S. I. An, and F. F. Jin, J. Clim. **17**, 2399 (2004).

[7] Y. Tang, R. H. Zhang, T. Liu, W. Duan, D. Yang, F. Zheng, ..., and M. Mu, Natl. Sci. Rev. **5**, 826 (2018).

[8] G. Jan van Oldenborgh, M. A. Balmaseda, L. Ferranti, T. N. Stockdale, and D. L. Anderson, J. Clim. **18**, 3240 (2005).

[9] D. Petrova, J. Ballester, S. J. Koopman, and X. Rodo, J. Clim. **33**, 163 (2020).

[10] Y. G. Ham , J. H. Kim, and J. J. Luo, Nature **573**, 568 (2019).

[11] D. Chen, and M. A. Cane, J. Comp. Phys. **227**, 3625 (2008).

[12] A. G. Barnston, "Evolution of ENSO Prediction over the Past 40 Years", in Science and Technology Infusion Climate Bulletin, *40^{th} NOAA Annual Climate Diagnostics and Prediction Workshop*, pages 94-101 (2015).

[13] M. K. Tippett, M. Ranganathan, M. L'Heureux, A. G. Barnston, and T. Del Sole, Clim. Dyn. **53**, 7497 (2019).

[14] M. A. Cane, and S. E. Zebiak, Science **228**, 1085 (1985).

[15] F. F. Jin, and J. D. Neelin, J. Atmos. Sci. **50**, 3477 (1993).

[16] C. Wang, C. Deser, J. Y. Yu, P. Di Nezio, and A. Clement, in *Coral Reefs of the Eastern Tropical Pacific*, edited by P. Glynn, D. Manzello, I. Enochs (Springer, Dordrecht, 2017), p. 85.

[17] F. F. Jin, J. Atmos. Sci. **54**, 811 (1997).

[18] K. Stein, N. Schneider, A. Timmermann, and F. F. Jin, J. Clim. **23**, 5629 (2010).

[19] A. F. Levine, and M. J. McPhaden, Geophys. Res. Lett. **42**, 5034 (2015).

[20] N. Chen, and A. J. Majda, Proc. Natl. Acad. Sci. U.S.A. **114**, 1468 (2017).

[21] G. Burgers, Clim. Dyn. **15**, 521-531 (1999)

[22] G. Burgers, F. F. Jin, and G. J. Van Oldenborgh, Geophys. Res. Lett. **32**, 13 (2005)

[23] K. Stein, A. Timmermann, and N. Schneider, Phys. Rev. Lett. **107**, 128501 (2011)

[24] W. Moon, and J. S. Wettlaufer, Sci. Rep. **7**, 44228 (2017).

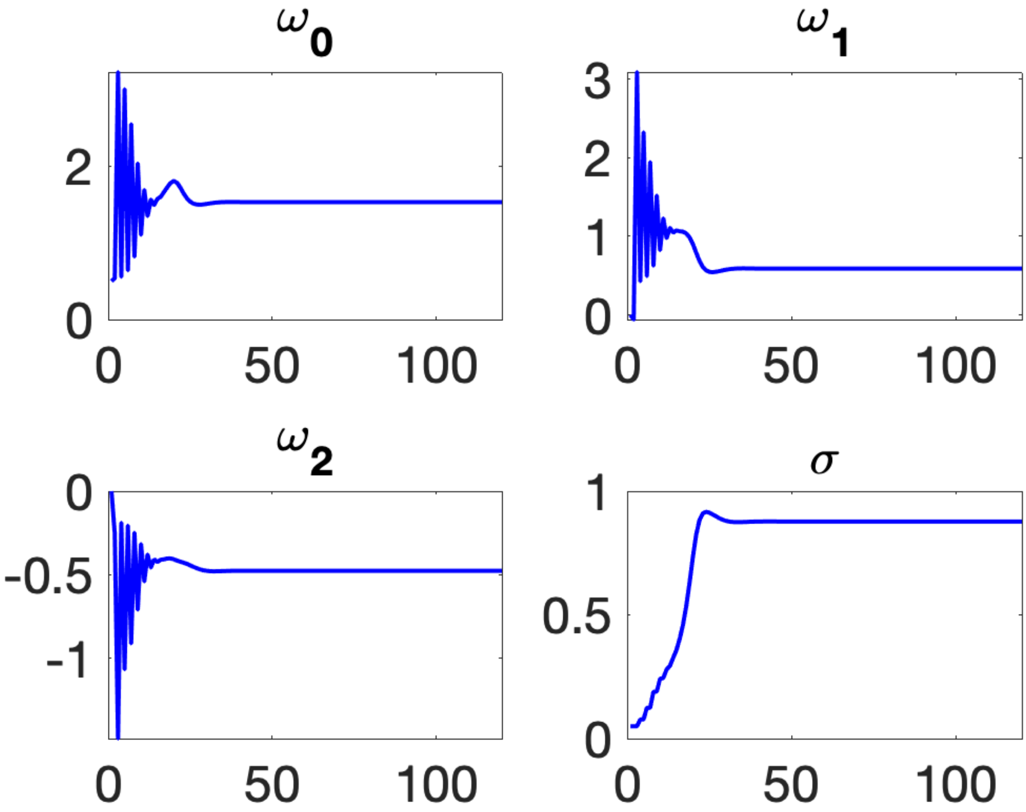


FIG. 6. Coefficients of the model extracted from real data as a function of the number of iterations. The estimated optimal values of the coefficients obtained using this method are $\omega_0 = 1.5$, $\omega_1 = 0.6$, $\omega_2 = -0.5$, $\lambda = -0.8$, $\sigma = 0.9$.

[25] A. P. Dempster, N. M. Laird, and D. B. Rubin, J. R. Stat. Soc.: Series B (Methodological) **39**, 1 (1977).

[26] R. Sundberg, Scand. J. Stat. **1**, 49 (1974).

[27] R. Sundberg, Commun. Stat. Sim. Comp. **5**, 55 (1976).

[28] N. Chen, J. Comp. Phys. **418**, 109635 (2020).

[29] E. Tziperman, M. A. Cane, S. E. Zebiak, Y. Xue, and B. Blumenthal, J. Clim. **11**, 2191 (1998).

[30] H. Bellenger, E. Guilyardi, J. Leloup, M. Lengaigne, and J. Vialard, Clim. Dyn. **42**, 1999 (2014).

[31] D. Kondrashov, S. Kravtsov, A. W. Robertson, and M. Ghil, J. Clim. **18**, 4425 (2005).

[32] C. S. Meinen, and M. J. McPhaden, J. Clim. **13**, 3551 (2000).

[33] N. Chen, and A. J. Majda, Entropy. **20**, 644 (2018).

[34] S. Thual, A. J. Majda, N. Chen, and S. N. Stechmann, Proc. Natl. Acad. Sci. U.S.A. **113**, 10245 (2016).

[35] N. R. Smith, Aust. Met. Mag. **44**, 93 (1995).

[36] W. Moon, S. Agarwal, and J. S. Wettlaufer, Phys. Rev. Lett. **121**, 108701 (2018).

[37] BNOC (Bureau National Operations Centre) Operations Bulletin Number 120 "Operational Implementation of ACCESS-S1 Seasonal Prediction System (Dec 2018)"

[38] D. W. Behringer, M. Ji, and A. Leetmaa, Mon. Weather Rev. **126**, 1013 (1998).

[39] E. Kalnay, M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, et al., Bull. Am. Met. Soc. **77**, 437 (1996).

[40] R. E. Kalman, and R. S. Bucy, J. Basic Eng. **83**, 95 (1961).

[41] N. Chen, and A. J. Majda, J. Comp. Phys. **410**, 109381 (2020).

[42] A. P. Dempster, N. M. Laird, and D. B. Rubin, J. R. Stat. Soci.: Series B (Methodological) **39**, 22, (1977).

[43] N. Chen, X. Hou, Q. Li, and Y. Li, Atmosphere **10**, 248 (2019).

[44] S. An, S. Kim, and A. Timmermann, Sci. Rep. **10**, 16282 (2020).

Paper V

W. Moon, L.T. Giorgini, and J.S. Wettlaufer, Nonadiabatic stochastic resonance: An analytical solution, arXiv: 2104.03426 (2021).

Nonadiabatic stochastic resonance: An analytical solution

Woosok Moon,[1, 2] L. T. Giorgini,[2] and J. S. Wettlaufer[3, 2]
[1]Department of Mathematics, Stockholm University 106 91 Stockholm, Sweden
[2]Nordita, Royal Institute of Technology and Stockholm University, SE-10691 Stockholm, Sweden
[3]Yale University, New Haven, USA
(Dated: April 9, 2021)

We generalize stochastic resonance to the non-adiabatic limit by treating the double-well potential using two quadratic potentials. We use a singular perturbation method to determine an approximate analytical solution for the probability density function that asymptotically connects local solutions in boundary layers near the two minima with those in the region of the maximum that separates them. The validity of the analytical solution is confirmed numerically. Free from the constraints of the adiabatic limit, the approach allows us to predict the escape rate from one stable basin to another for systems experiencing a more complex periodic forcing.

I. INTRODUCTION

It is common to consider noise as a hindrance in measurements and observations, which underlies the general idea of filtering in signal processing [1, 2]. In contrast, there are circumstances in which the presence of noise may facilitate the detection of a signal. A prominent example is *stochastic resonance*, during which noise can amplify a weak signal and drive a dramatic transition in the state of a system [3–7]. The simplest stochastic resonance configuration considers the trajectory of a Brownian particle in a double-well potential influenced by a weak periodic forcing; as the random forcing increases so too does the observed signal-to-noise ratio [6]? As the noise amplitude varies a resonance with the periodic forcing triggers transitions between the stable minima.

The term stochastic resonance originated in a series of studies by Benzi et al., [3–5] that focused on the observed periodicity of Earth's ice ages. Namely, because the 100 kyr eccentricity of the Milankovitch orbital cycles provides such a weak periodic solar insolation forcing, as the strength of random forcing varies, a resonance may drive the transition between the cold and warm states of the Earth's climate system. Whilst the original motivation was an explanation of ice-age periodicity, the generality of the framework has driven a myriad of studies across science and engineering. For example, a modest subset in which stochastic resonance is a key process includes the sensory systems of many animals, including humans, facilitating the recognition of weak signals buried in environment noise [8–10], it is used to enhance signals in blurry images [11, 12] and in detecting machine faults in mechanical engineering [13]. Despite this breadth, the theoretical foundation of stochastic resonance is based on the Kramer's escape rate from one of the mimina of a bi-stable system within the adiabatic limit [6, 7]. In particular, rather than directly solving the non-autonomous Fokker-Planck equation, the periodic forcing of the potential is treated as a constant assuming that the frequency of the periodic forcing is asymptotically small. However, in most realistic settings, a weak signal is not consistent with a single periodic function, but rather with a continuous spectrum of many frequencies [14]. Therefore, use of stochastic resonance in practical systems requires a generalization of the existing theory to the *non-adiabatic case.*

We have made a recent advance in this direction as follows [15]. We first treated the Kramers escape problem with periodic forcing within the framework of singular perturbation theory and matched asymptotic expansions. In particular, we divided the cubic potential into three regions–two boundary layers near the extrema and one connecting them–and determined the local approximate analytic solutions of the Fokker-Planck equation within each region. We constructed a uniformly valid composite global solution by systematic asymptotic matching of the local solutions. The bi-stable system was then obtained by reflection of the Kramers problem. In consequence, the non-adiabatic hopping rate was determined and the full solution was tested numerically. The analytic solution is principally reliable only when the magnitude of the periodic forcing is much smaller than the noise magnitude. Our goal here is the simplify this calculation to make more transparent and useful the general treatment of non-adiabatic stochastic resonance.

Insight for this simplification is provided from a calculation of the mean first passage time of the Ornstein-Uhlenbeck process using similar singular perturbation methods [16]. In that problem we solve the Fokker-Planck equation with an absorbing boundary condition at a point far away from the minimum of a quadratic potential. In what follows we revisit this problem and derive the probability density function from which we determine the escape rate. We will then use this method to treat the stochastic resonance problem as a periodically forced bi-stable potential by combining two quadratic potentials. We then obtain approximate analytical solutions for stochastic resonance in the non-adiabatic regime. The analytical solutions compare extremely well with the numerical solutions.

II. FIRST PASSAGE PROBLEM FOR THE ORNSTEIN-UHLENBECK PROCESS

In order to make this paper reasonably self-contained, we outline the asymptotic method previously used to solve the survival probability problem for the Ornstein-Uhlenbeck problem [16], which we relate explicitly in this section to the first passage problem. As shown in Fig. 1 we divide the domain into two regions: a broad $O(1)$ region (I) containing the minimum of the potential, $x = 0$, and a narrow $O\left(1/\beta\right)$ boundary layer (II) near $x = \beta$. In the parlance of asymptotic methods in differential equations [17], the boundary layer is referred to as the *inner region* and the remainder of the domain is the *outer region*, although in this case the latter is in the interior of the potential. We solve the limiting differential equations in these regions, from which we develop a uniform composite solution for the probability density using asymptotic matching. From this composite solution we calculate the mean first passage time.

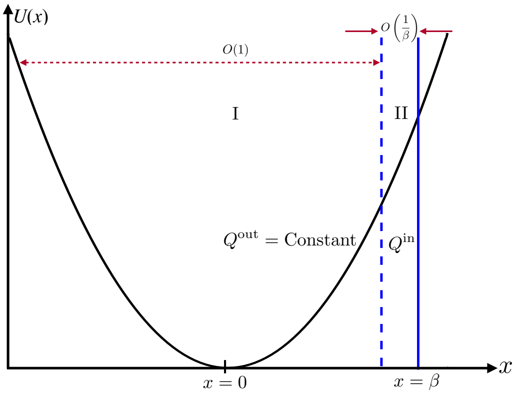


FIG. 1: Schematic of the first passage problem with a quadratic potential $U(x) = \frac{1}{2}\lambda x^2$ and Gaussian white noise of magnitude $\sqrt{2}\sigma$. The potential is divided into an "outer" region (I) around the minimum at $x = 0$, and a "boundary layer", or the "inner" region (II), near the absorbing boundary at $x = \beta$. In each region the asymptotically dominant solutions are determined and then matched throughout the entire domain. The boundary layer solution, $Q^{\rm in}$, satisfying the boundary condition $x = \beta$ converges to $Q^{\rm out}$ as the independent variable approaches the origin.

A Brownian particle resides in a quadratic potential (Fig. 1) and its position, $x(t)$, obeys the following Langevin equation,

$$\frac{dx}{dt} = -\lambda x + \sqrt{2}\sigma\xi(t), \tag{1}$$

where $\lambda > 0$ captures the stability of the quadratic potential, $\xi(t)$ is Gaussian white noise with zero mean, intensity $\sqrt{2}\sigma$, and correlation $\langle\xi(t)\xi(s)\rangle = \delta(t-s)$.

The first passage problem describes the mean time required for the Brownian particle to reach a particular location, say $x = \beta$. The Langevin equation can be transformed to a Fokker-Planck equation for the probability density, $P(x,t)$, given by

$$\frac{\partial P}{\partial t} = \lambda\frac{\partial}{\partial x}(xP) + \sigma^2\frac{\partial^2 P}{\partial x^2}, \tag{2}$$

with an absorbing boundary condition at $x = \beta$, $P(x = \beta) = 0$, and $P(x = -\infty) = 0$. Here, the absorbing boundary condition implies that Brownian particles disappear at $x = \beta$, where we seek the loss rate of probability density. We have solved this problem analytically for $\beta \gg 1$ with both λ and $\sigma = O(1)$ [16].

Region I: *The outer solution.* The probability density in the outer region (the interior of the potential), $P^{\rm out}$, satisfies

$$\frac{\partial P^{\rm out}}{\partial t} = \lambda\frac{\partial}{\partial x}\left(xP^{\rm out}\right) + \sigma^2\frac{\partial^2 P^{\rm out}}{\partial x^2}, \tag{3}$$

with boundary conditions $P^{\rm out}(\pm\infty) = 0$. The steady-state solution of Eq. (3) is ${P_{\rm ss}}^{\rm out} = \sqrt{\frac{\lambda}{2\pi\sigma^2}}\exp(-\lambda x^2/2\sigma^2)$. Now, we let $P(x,t) = {P_{\rm ss}}^{\rm out}\,Q(x,t)$ in Eq. (2) and hence $Q(x,t)$ satisfies

$$\frac{\partial Q}{\partial t} = -\lambda x\frac{\partial Q}{\partial x} + \sigma^2\frac{\partial^2 Q}{\partial x^2}. \tag{4}$$

Therefore, in the interior of the potential $Q \equiv Q^{\rm out} \approx N(t)$, where $N(t)$ is a very slowly-varying function due to the leaking of probability density at $x = \beta$. Thus, we set $N(t) = N$ to be a constant.

Region II: *The inner solution.* In the boundary layer we have $x \sim \beta$ and because $\beta \gg 1$, we have $\varepsilon \equiv 1/\beta \ll 1$, motivating the stretched coordinate $\eta = \frac{x - \frac{1}{\varepsilon}}{\varepsilon} = \beta(x-\beta)$. Thus, expressed using η, in the boundary layer Eq. (4) becomes

$$\frac{\partial Q^{\rm in}}{\partial t} = -\lambda(\eta+\beta^2)\frac{\partial Q^{\rm in}}{\partial \eta} + \beta^2\sigma^2\frac{\partial^2 Q^{\rm in}}{\partial \eta^2}. \tag{5}$$

Because $\beta \gg 1$ the leading-order balance in Eq. (17) is

$$-\lambda\frac{\partial Q^{\rm in}}{\partial \eta} + \sigma^2\frac{\partial^2 Q^{\rm in}}{\partial \eta^2} \simeq 0. \tag{6}$$

Therefore, $Q^{\rm in} \simeq K_1\exp(\lambda\eta/\sigma^2) + K_2$ with constants K_1 and K_2. To satisfy the boundary condition $P(x = \beta) = 0$, which is equivalent to $Q^{\rm in}(\eta = 0) = 0$, we must have $K_2 + K_1 = 0$.

Uniformly Valid Composite Solution: Asymptotic matching of the solutions between Regions I and II requires that $\lim_{\eta\to-\infty} Q^{\rm in} = Q^{\rm out} = N$, which gives $Q^{\rm in} = N[1 - \exp(\lambda\eta/\sigma^2)]$. A uniformly valid composite solution is then $Q \equiv Q^{\rm out} + Q^{\rm in} - \lim_{\eta\to-\infty} Q^{\rm in}$, or $Q \simeq N\left(1 - \exp[\beta\lambda(x-\beta)/\sigma^2]\right)$, which gives

$$P \simeq N\sqrt{\frac{\lambda}{2\pi\sigma^2}}\exp\left(-\frac{\lambda}{2\sigma^2}x^2\right)\left(1 - \exp\left[\frac{\beta\lambda}{\sigma^2}(x-\beta)\right]\right). \tag{7}$$

Integrating the Fokker-Planck equation (2) over the entire domain $(-\infty, \beta)$ we have

$$\frac{\partial}{\partial t}\int_{-\infty}^{\beta} P dx = J_{x=\beta}, \tag{8}$$

where $J = \lambda x P + \sigma^2 \partial P/\partial x$ and we use $J(x=-\infty) = 0$. Probability density is principally concentrated near $x = 0$, and hence

$$\int_{-\infty}^{\beta} P dx \simeq Q^{\mathrm{out}} \int_{-\infty}^{\infty} P^{\mathrm{out}}\, dx = N, \tag{9}$$

so that Eq. (8) becomes

$$\frac{dN}{dt} = -\beta\lambda\sqrt{\frac{\lambda}{2\pi\sigma^2}}\exp\left(-\frac{\lambda\beta^2}{2\sigma^2}\right) N \equiv -rN. \tag{10}$$

Therefore, the global probability density decreases with rate r and the mean first passage time is $\langle T\rangle = 1/r$, with

$$r = \beta\lambda\sqrt{\frac{\lambda}{2\pi\sigma^2}}\exp\left(-\frac{\lambda\beta^2}{2\sigma^2}\right) \tag{11}$$

III. STOCHASTIC RESONANCE IN DOUBLE QUADRATIC POTENTIALS

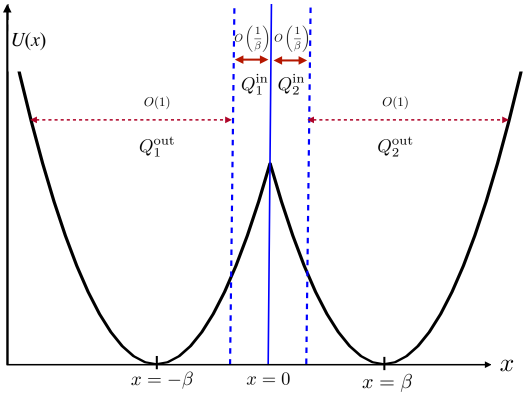


FIG. 2: Schematic of stochastic resonance in which a double-well potential is treated using two quadratic potentials, $U_1(x) = \frac{1}{2}\lambda(x+\beta)^2$ when $x < 0$ and $U_2(x) = \frac{1}{2}\lambda(x-\beta)^2$ when $x \geq 0$. There are two regions, the interior of each near the two minima $x = \pm\beta$ and boundary layers near $x = 0$.

A. Asymptotic solutions

We now combine the asymptotic methods used to solve the general problem of stochastic resonance [15] with the particular setting described in §II, to treat the double-well potential in stochastic resonance using two quadratic potentials. As shown in Fig. 2, the potential $U(x)$ is $\frac{1}{2}\lambda(x+\beta)^2$ when $x<0$ and $\frac{1}{2}\lambda(x-\beta)^2$ when $x \geq 0$. Thus, with the addition of noise, $\sqrt{2}\sigma\xi(t)$, and periodic forcing $A\cos(\omega t)$, stochastic resonance between the minima at $x = \pm\beta$ is described by the following Fokker-Planck equation for the probability density $P(x,t)$,

$$\frac{\partial P}{\partial t} = \frac{\partial}{\partial x}\left\{\left[\frac{dU}{dx} - A\cos(\omega t)\right] P\right\} + \sigma^2\frac{\partial^2 P}{\partial x^2}, \tag{12}$$

with boundary conditions $P(x=\pm\infty, t) = 0$.

As shown in the Fig. (2), there are two *outer* regions interior to each side of the potential centered upon the two minima, $x = \pm\beta$, and two *boundary layers* for each quadratic potential at $x = 0$. We now consider approximate solutions to Eq. (12) in these regions.

In the outer region of $U_1(x)$ centered on $x = -\beta$, the probability density P_1^{out} satisfies

$$\frac{\partial P_1^{\mathrm{out}}}{\partial t} = \lambda\frac{\partial}{\partial y}(yP_1^{\mathrm{out}}) - A\cos(\omega t)\frac{\partial P_1^{\mathrm{out}}}{\partial y} + \sigma^2\frac{\partial^2 P_1^{\mathrm{out}}}{\partial y^2} \tag{13}$$

with $P_1^{\mathrm{out}}(y=\pm\infty) = 0$, where $y = x+\beta$. Eq. (13) has solution

$$P_1^{\mathrm{out}} = \sqrt{\frac{\lambda}{2\pi\sigma^2}}\exp\left(-\frac{\lambda}{2\sigma^2}[y-h(t)]^2\right) \tag{14}$$

where

$$h(t) = \frac{A}{\sqrt{\lambda^2+\omega^2}}\cos(\omega t - \phi) \tag{15}$$

and $\tan\phi = \omega/\lambda$.

Now we substitute $P(x,t) = P_1^{\mathrm{out}} Q(x,t)$ into Eq. (12) which becomes

$$\frac{\partial Q}{\partial t} = [-\lambda y + A\cos(\omega t - 2\phi)]\frac{\partial Q}{\partial y} + \sigma^2\frac{\partial^2 Q}{\partial y^2}, \tag{16}$$

the outer solution of which is $Q_1^{\mathrm{out}} = N_1(t)$; a slowly-varying function of time that we approximate as a constant, N_1.

In the boundary layer we introduce the stretched coordinate $\eta = \beta x$, which leads to

$$\frac{\partial Q_1^{\mathrm{in}}}{\partial t} = [-\lambda\eta + \beta^2\lambda + \beta A\cos(\omega t - 2\phi)]\frac{\partial Q_1^{\mathrm{in}}}{\partial \eta} + \beta^2\sigma^2\frac{\partial^2 Q_1^{\mathrm{in}}}{\partial\eta^2}. \tag{17}$$

Because $\beta \gg 1$, keeping terms to $O(\beta)$ and higher, gives

$$\left[-\lambda + \frac{A}{\beta}\cos(\omega t - 2\phi)\right]\frac{\partial Q_1^{\mathrm{in}}}{\partial\eta} + \sigma^2\frac{\partial^2 Q_1^{\mathrm{in}}}{\partial\eta^2} \simeq 0, \tag{18}$$

the solution of which is

$$Q_1^{\mathrm{in}} = K_1\exp\left(\left[\frac{\lambda - \frac{A}{\beta}\cos(\omega t - 2\phi)}{\sigma^2}\right]\eta\right) + K_2, \tag{19}$$

with constants K_1 and K_2 to be determined. We note here that Eq. (17) could be solved using a regular perturbation method by setting $Q_1^{\text{in}} \simeq Q_{10}^{\text{in}} + \frac{1}{\beta}Q_{11}^{\text{in}}$, which results in $Q_1^{\text{in}} \simeq K_1 \exp\left(\frac{\lambda}{\sigma^2}\eta\right)\left(1 - \frac{A}{\beta\sigma^2}\cos(\omega t - 2\phi)\eta\right) + K_2$. However, as noted, to $O(\beta)$, the solution (19) uses $\exp\left(-\frac{A}{\beta\sigma^2}\cos(\omega t - 2\phi)\eta\right)$ instead of $1 - \frac{A}{\beta\sigma^2}\cos(\omega t - 2\phi)\eta$, simplifying the subsequent development at that order.

Asymptotic matching requires that the outer limit of the inner solution equal the inner limit of the outer solution; $\lim_{\eta\to-\infty} Q_1^{\text{in}} = \lim_{y\to\beta} Q_1^{\text{out}} = N_1$ and hence $K_2 = N_1$. Therefore, a uniformly valid composite asymptotic solution in $U_1(x)$ is $Q_1 = Q_1^{\text{out}} + Q_1^{\text{in}} - \lim_{\eta\to-\infty} Q_1^{\text{in}}$, giving

$$P_1 = \sqrt{\frac{\lambda}{2\pi\sigma^2}} \exp\left(-\frac{\lambda}{2\sigma^2}[x + \beta - h(t)]^2\right) \times \left\{N_1 + K_1 \exp\left(\frac{\beta\lambda - A\cos(\omega t - 2\phi)}{\sigma^2}x\right)\right\}. \quad (20)$$

Similarly, we obtain the solution P_2 in $U_2(x)$ as

$$P_2 = \sqrt{\frac{\lambda}{2\pi\sigma^2}} \exp\left(-\frac{\lambda}{2\sigma^2}[x - \beta - h(t)]^2\right) \times \left\{N_2 + D_1 \exp\left(-\frac{\beta\lambda + A\cos(\omega t - 2\phi)}{\sigma^2}x\right)\right\}. \quad (21)$$

Now we determine the constants K_1 and D_1 from the continuity of probability density $P(x)$ and flux $J(x)$ at $x = 0$, where $J(x) = [dU/dx - A\cos(\omega t)]P + \sigma^2 \partial P/\partial x$. Firstly, continuity of $P(x)$ at $x = 0$ is $P_1(x = 0) = P_2(x = 0)$, which results in

$$D_1 = (N_1 + K_1)\exp\left(\frac{2\beta\lambda}{\sigma^2}h(t)\right) - N_2. \quad (22)$$

The fluxes at the origin from both sides are

$$J_1(x = 0) = [(\beta\lambda + \lambda h(t) - 2A\cos(\phi)\cos(\omega t - \phi))K_1 + (\lambda h(t) - A\cos(\omega t))N_1] \sqrt{\frac{\lambda}{2\pi\sigma^2}} \exp\left(-\frac{\lambda}{2\sigma^2}[\beta - h(t)]^2\right)$$

and

$$J_2(x = 0) = [(-\beta\lambda + \lambda h(t) - 2A\cos(\phi)\cos(\omega t - \phi))D_1 + (\lambda h(t) - A\cos(\omega t))N_2] \sqrt{\frac{\lambda}{2\pi\sigma^2}} \exp\left(-\frac{\lambda}{2\sigma^2}[\beta + h(t)]^2\right) \quad (23)$$

respectively. Now, imposing $J_1(x = 0) = J_2(x = 0)$ and $P_1(x = 0) = P_2(x = 0)$, viz., Eq. (22), gives

$$K_1 = \left[\frac{1}{2} + \frac{A}{2\beta\lambda}\cos(\omega t - 2\phi)\right] \times \left[-N_1 + N_2 \exp\left(-\frac{2\beta\lambda}{\sigma^2}h(t)\right)\right]$$

$$D_1 = \left[\frac{1}{2} - \frac{A}{2\beta\lambda}\cos(\omega t - 2\phi)\right] \times \left[-N_2 + N_1 \exp\left(\frac{2\beta\lambda}{\sigma^2}h(t)\right)\right]. \quad (24)$$

Integration of Eq. (12) from $x = -\infty$ to $x = 0$,

$$\frac{\partial}{\partial t}\int_{-\infty}^{0} P_1 dx = J_1|_{x=0}, \quad (25)$$

gives

$$\frac{dN_1}{dt} = -r_1 N_1 + r_2 N_2, \quad (26)$$

with escape rates

$$r_1 = \left[\frac{1}{2} - \frac{A}{2\beta\lambda}\cos(\omega t - 2\phi)\right][\beta\lambda - \lambda h(t) + A\cos(\omega t) + A\cos(\omega t - 2\phi)] \sqrt{\frac{\lambda}{2\pi\sigma^2}} \exp\left(-\frac{\lambda}{2\sigma^2}[\beta - h(t)]^2\right)$$

and

$$r_2 = \left[\frac{1}{2} + \frac{A}{2\beta\lambda}\cos(\omega t - 2\phi)\right][\beta\lambda + \lambda h(t) - A\cos(\omega t) - A\cos(\omega t - 2\phi)] \sqrt{\frac{\lambda}{2\pi\sigma^2}} \exp\left(-\frac{\lambda}{2\sigma^2}[\beta + h(t)]^2\right). \quad (27)$$

Similarly, upon integration of Eq. (12) from $x = 0$ to $x = \infty$, we obtain

$$\frac{dN_2}{dt} = r_1 N_1 - r_2 N_2. \quad (28)$$

Therefore, r_1 and r_2 are the escape rates from $x = \pm\beta$ over the barrier at the origin and Eqs. (26) and (28) are simplified forms of the two-state Master equations derived for general quartic potentials [15].

B. The validity of asymptotic solutions

We test the validity of the asymptotic solutions by comparison with full numerical solutions of the Fokker-Planck equation (12) using the implicit finite difference method of Chang and Cooper [18]. To facilitate this comparison, near $x = 0$ we introduce a function defined as $H(x) = P_1/P_1^{\text{out}}$ when $x < 0$ and $H(x) = P_2/P_2^{\text{out}}$ when $x \geq 0$. We study two different magnitudes of the periodic forcing, A, and the noise magnitude, σ, and combined with the range of the angular frequency ω, this puts our results in the non-adiabatic regime. In particular, A is not asymptotically smaller than σ, and ω is not trivially small. Previously we showed that the adiabatic limit requires $\omega \ll 1$ and $A \ll \sigma$ and the non-adiabatic case referred to non-trivial ω [15]. Here, we extend the non-adiabaticity with the condition $A \sim \sigma$. Figure 3 shows that the asymptotic and numerical solutions compare very well. When $A \sim \sigma$ or $A \ll 1$ (not shown), the asymptotic solutions are nearly indistinguishable from numerical solutions. However, when $A \gg \sigma$, the analytic solutions become less accurate.

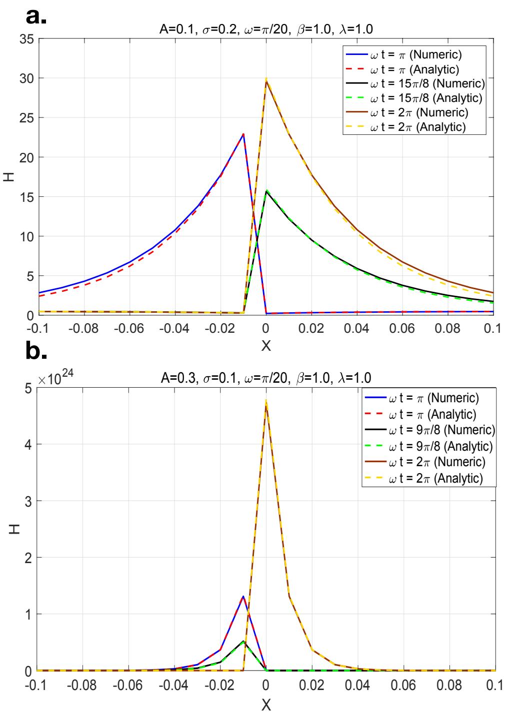


FIG. 3: The comparison between numerical simulations (thick lines) and asymptotic solutions (dashed lines) using $H(x,t)$ defined by $H(x,t) = P_1(x,t)/P_1^{\rm out}$ when $x < 0$ and $H(x,t) = P_2(x,t)/P_2^{\rm out}$. The two examples use $\omega = \pi/20$, $\beta = 1.0$, and $\lambda = 1.0$ as common variables. The first case (a) has $A = 0.1$ and $\sigma = 0.2$ and the second case (b) has $A = 0.3$ and $\sigma = 0.1$.

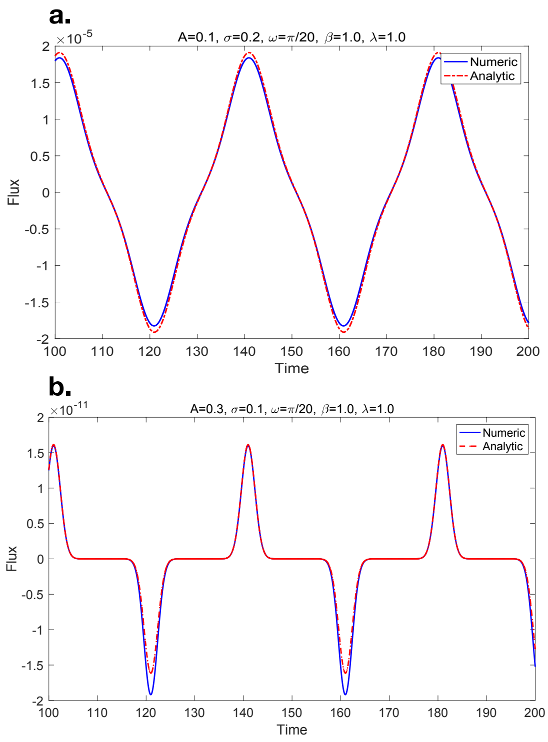


FIG. 4: The probability flux calculated at the barrier ($x = 0$) between the two wells of the potential. Comparison between numerical simulations (solid lines) and asymptotic solutions (dashed lines from Eq. 23) for the same two cases used in figure 3.

The central quantity in stochastic resonance is the flux at the barrier ($x = 0$) between the two wells of the potential, which controls the oscillatory behavior of the probability density. In Figure 4 we compare the analytical (dashed lines from Eq. 23) and numerical (solid lines) fluxes at $x = 0$. Clearly these compare very favorably.

The escape (or hopping) rates r_1 and r_2 shown in Eq. (27) are valid independent of the magnitude of ω and A, and so long as A is not much larger than σ, the analytic solutions are very accurate. Therefore, our analytic solutions can be used for the wide range of applications of stochastic resonance that appear in science and engineering, which removes the need for substantial simulations of either the Langevin or Fokker-Planck equations.

C. Weak periodic forcing $A \ll 1$

In the original treatment of stochastic resonance [3–5], one has $A \ll 1$, which implies that there is a weak signal in a noise dominated background. Here, we still seek to understand the amplification of the periodic forcing $A\cos(\omega t)$ for $A \ll 1$, however we treat ω as arbitrary so that we are have the non-adiabatic case, which is not within the corpus of the original work.

Because we are in possession of the probability density function $P(x,t)$, we can calculate the mean position of a Brownian particle as

$$\langle x \rangle = \int_{-\infty}^{0} x P_1 dx + \int_{0}^{\infty} x P_2 dx. \tag{29}$$

We note that P_1 and P_2 are principally concentrated near $x = -\beta$ and $x = \beta$, so that

$$\begin{aligned} &\int_{-\infty}^{0} x P_1 dx \\ &\simeq \sqrt{\frac{\lambda}{2\pi\sigma^2}} N_1 \int_{-\infty}^{\infty} x \exp\left(-\frac{\lambda}{2\sigma^2}[x+\beta-h(t)]^2\right) dx \\ &\simeq -N_1\beta. \end{aligned} \tag{30}$$

Similarly, because $\int_0^\infty x P_2 dx \simeq N_2\beta$, we have $\langle x \rangle = \beta(-N_1 + N_2)$. Using Eqs. (26) and (28), we write the time-evolution of $-N_1 + N_2$ as

$$\frac{d}{dt}(-N_1 + N_2) = -(r_1 + r_2)(-N_1 + N_2) + (r_1 - r_2). \tag{31}$$

We now take $A \ll 1$ to find

$$r_1 + r_2 \equiv r \simeq \beta\lambda\sqrt{\frac{\lambda}{2\pi\sigma^2}}\exp\left(-\frac{\lambda\beta^2}{2\sigma^2}\right) \text{ and}$$
$$r_1 - r_2 \simeq r\frac{\beta\lambda h(t)}{\sigma^2}, \tag{32}$$

and hence the approximate solution of Eq. (31) is

$$-N_1 + N_2 \simeq \frac{\beta A}{\sigma^2}\cos\phi\cos\psi\cos(\omega t - \phi - \psi), \tag{33}$$

where $\cos\phi = \lambda/\sqrt{\lambda^2 + \omega^2}$ and $\cos\psi = r/\sqrt{r^2 + \omega^2}$. Hence,

$$\langle x\rangle = \beta(-N_1 + N_2) = \beta^2\frac{A}{\sigma^2}\cos\phi\cos\psi\cos(\omega t - \phi - \psi). \tag{34}$$

Therefore, the original signal is $A\cos(\omega t)$ and the response, $\langle x\rangle$, is order $O(\beta^2)$.

Eq. (34) shows that the magnitude of the response, $\langle x\rangle$, depends nonlinearly on the stability factor, λ, and the noise amplitude, σ, as $\frac{\lambda A}{\sigma^2}\cos\psi$, where $\cos\psi$ is a function of λ and σ (and other parameters) as seen in Eq. (32). Figure 5 shows two contour plots of the magnitude of the response, $\frac{\lambda A}{\sigma^2}\cos\psi$ for two values of the location of the absorbing boundary (a) $\beta = 5$ and (b) $\beta = 10$ for fixed $\omega = 0.1$. Optimal values of the noise amplitude, σ, and the stability of the potential, λ, are revealed as the maxima in these plots. This optimal noise magnitude is signature of stochastic resonance.

When $A \ll 1$ our results are similar to the adiabatic limit, except for the fact that there is a phase shift, ϕ. However, as shown above and expected from the original theory, the response to the periodic forcing, $A\cos(\omega t)$, is magnified by a factor β^2. In strong contrast to the original theory, we provide an explicit mathematical expression, Eq. (34), that shows this dependence. Although we have focused upon a single forcing frequency, our method can be generalized to weak signals consisting of many frequencies.

IV. CONCLUSION

We have used asymptotic methods that are central in the theory of differential equations to derive analytical expressions for the entire suite of results in stochastic resonance. Having previously used this general methodology to analyze the Fokker Planck equation (2) for a general quartic potential in the non-adiabatic limit [15], here we have managed to further simplify the problem whilst maintaining the key features of that analysis. In particular, we approximated the quartic potential of the underlying Ornstein-Uhlenbeck process as two parabolic potentials. We derived explicit formulae for the escape rates from one basin to the other free from the constraints of the adiabatic limit, and have shown their veracity using direct numerical solutions of the dynamical equations. Our results can easily be generalized to multiple frequencies and forcing amplitudes and the ease of use of explicit formulae free a practitioner interested in stochastic resonance from the labor of numerical simulations.

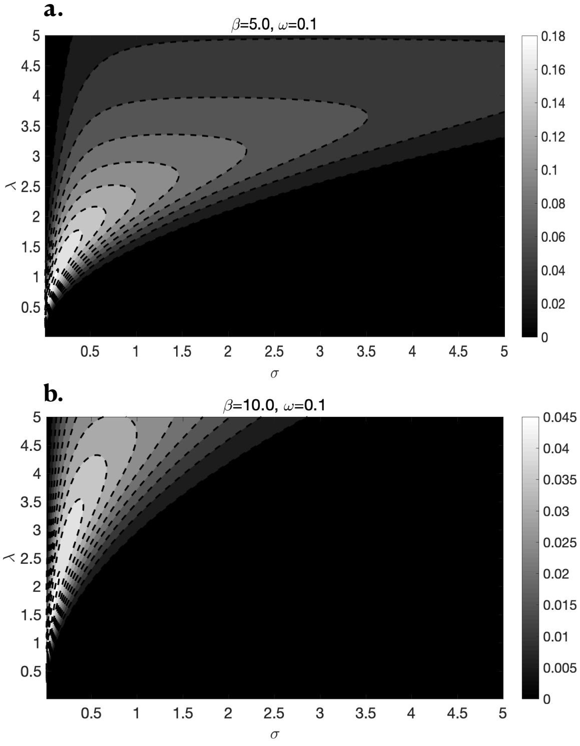


FIG. 5: Contour plot of the σ and λ dependence of the magnitude of the response $\frac{\lambda A}{\sigma^2}\cos\psi$ from Eq. (34). The location of the absorbing boundary is (a) $\beta = 5$ and (b) $\beta = 10$ for a fixed angular frequency of $\omega = 0.1$.

ACKNOWLEDGMENTS

The authors acknowledge the support of Swedish Research Council grant no. 638-2013-9243.

[1] M. Lang, H. Guo, J. E. Odegard, C. S. Burrus, and R. O. Wells, IEEE Sig. Proc. Lett. **3**, 10 (1996).

[2] J. Chen, J. Benesty, Y. Huang, and S. Doclo, IEEE Trans. Audio, speech, and lang. process. **14**, 1218 (2006).

[3] R. Benzi, A. Sutera, and A. Vulpiani, J. Phys. A: Math. Gen. **14**, L453 (1981).

[4] R. Benzi, G. Parisi, A. Sutera, and A. Vulpiani, Tellus **34**, 10 (1982).

[5] R. Benzi, G. Parisi, A. Sutera, and A. Vulpiani, SIAM J. Appl. Math. **43**, 565 (1983).

[6] B. McNamara and K. Wiesenfeld, Phys. Rev. A **39**, 4854 (1989).

[7] L. Gammaitoni, P. Hänggi, P. Jung, and F. Marchesoni, Rev. Mod. Phys. **70**, 223 (1998).

[8] J. E. Levin and J. P. Miller, Nature **380**, 165 (1996).

[9] D. F. Russell, L. A. Wilkens, and F. Moss, Nature **402**, 291 (1999).

[10] K. Kitajo, D. Nozaki, L. M. Ward, and Y. Yamamoto, Phys. Rev. Lett. **90**, 218103 (2003).

[11] V. P. S. Rallabandi and P. K. Roy, Magn. Res. Imaging **28**, 1361 (2010).

[12] R. Chouhan, R. K. Jha, and P. K. Biswas, IET Image Process. **7**, 174 (2013).

[13] S. Lu, Q. He, and J. Wang, Mech. Syst. and Sign. Proc. **116**, 230 (2019).

[14] J. J. Collins, C. C. Chow, A. C. Capela, and T. T. Imhoff, Phys. Rev. E **54**, 5575 (1996).

[15] W. Moon, N. Balmforth, and J. S. Wettlaufer, J. Phys. A: Math. Th. **53**, 095001 (2020).

[16] L. T. Giorgini, W. Moon, and J. S. Wettlaufer, J. Stat. Phys. **181**, 2404 (2020).

[17] C. M. Bender and S. A. Orszag, *Advanced Mathematical Methods for Scientists and Engineers I: Asymptotic Methods and Perturbation Theory* (Springer, New York, 2013).

[18] J. S. Chang and G. Cooper, J. Comp. Phys. **6**, 1 (1970).