Potsdam Institute for
Climate Impact Research

Multi-scaling reservoir computing learns noise-induced transitions with Lévy noise

Zequn Lin, Jürgen Kurths, Ying Tang

Document Version
Version of Record (Publisher Version)

This version is available at
https://publications.pik-potsdam.de/pubman/item/item_ 33312

Originally published as
Lin, Z., Kurths, J., Tang, Y. (2025): Multi-scaling reservoir computing learns noise-induced transitions with Lévy noise. - Chaos, 35, 7, 073132.
https://doi.org/10.1063/5.0278558

Terms of Use
This article may be downloaded for personal use only. Any other use requires prior permission of the author and AIP Publishing.

Chaos
An Interdisciplinary Journal of Nonlinear Science

RESEARCH ARTICLE | JULY 18 2025

Multi-scaling reservoir computing learns noise-induced transitions with Lévy noise

Special Collection: Nonlinear Dynamics of Reservoir Computing: Theory, Realization and Application

Zequn Lin ; Jürgen Kurths ; Ying Tang

Check for updates

Chaos 35, 073132 (2025)
https://doi.org/10.1063/5.0278558

View Online
Export Citation

Articles You May Be Interested In

Pre-asymptotic analysis of Lévy flights

Chaos (July 2024)

Transition pathways for a class of high dimensional stochastic dynamical systems with Lévy noise

Chaos (June 2021)

How close are time series to power tail Lévy diffusions?

Chaos (July 2017)

07 September 2026 09:11:39

AIP Publishing
Chaos
Special Topics Open for Submissions
Learn More

AIP Publishing

Chaos ARTICLE pubs.aip.org/aip/cha

Multi-scaling reservoir computing learns noise-induced transitions with Lévy noise

Cite as: Chaos **35**, 073132 (2025); doi: 10.1063/5.0278558
Submitted: 30 April 2025 · Accepted: 1 July 2025 ·
Published Online: 18 July 2025

View Online Export Citation CrossMark

Zequn Lin,[1,2,a)] Jürgen Kurths,[3,4,5] and Ying Tang[6,7,b)]

AFFILIATIONS

[1]Department of Physics, Fudan University, Shanghai 200433, China
[2]Center for Interdisciplinary Studies, Westlake University, Hangzhou 310030, China
[3]Potsdam Institute for Climate Impact Research, Potsdam 14412, Germany
[4]Department of Physics, Humboldt University Berlin, Berlin 12489, Germany
[5]Research Institute of Intelligent Complex Systems, Fudan University, Shanghai 200433, China
[6]Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China, Chengdu 611731, China
[7]Key Laboratory of Quantum Physics and Photonic Quantum Information, Ministry of Education, University of Electronic Science and Technology of China, Chengdu 611731, China

Note: This paper is part of the Special Topic on Nonlinear Dynamics of Reservoir Computing: Theory, Application and Realization.
[a)]**Electronic mail:** zqlin24@m.fudan.edu.cn
[b)]**Author to whom correspondence should be addressed:** jamestang23@gmail.com

ABSTRACT

Noise-induced transitions, where random fluctuations drive systems between multiple stable states, are pivotal in stochastic dynamical systems across physics, biology, and engineering. These transitions become intricate in the presence of non-Gaussian noise or when systems exhibit complex dynamics beyond fixed points. We leverage the recent multi-scaling reservoir computing framework to learn such noise-induced transitions, focusing on a bistable system driven by Lévy noise and a limit-cycle system under Gaussian noise. We find that the multi-scaling reservoir computing can generate data capturing transition statistics when trained by a trajectory with a few dozen transitions, where the predictions of transition intervals and probability distributions align closely with test data. This showcases its ability to capture stochastic transitions with abrupt noisy shifts and oscillatory dynamics in a probabilistic manner. These findings highlight the potential of reservoir computing in studying general stochastic systems.

Published under an exclusive license by AIP Publishing. https://doi.org/10.1063/5.0278558

07 September 2026 09:11:39

In stochastic systems, noise or fluctuations often trigger transitions between multiple stable states, a phenomenon known as noise-induced transitions. Far from being a theoretical nuance, this process underpins a wide array of natural and engineered systems, such as protein conformational shifts, genetic regulatory dynamics, and chemical reaction pathways. However, modeling and predicting these transitions pose significant challenges. To address these, we have previously developed a multi-scaling reservoir computing framework. The framework was shown to be effective in a set of systems with transitions between fixed points induced by Gaussian noise. Whether it is generally applicable to more challenging systems, such as those with non-Gaussian noise or complex dynamics of limit cycles, remains elusive. This study tackles these problems by using the framework to study two representative systems: a bistable system driven by Lévy noise and a limit-cycle system under Gaussian perturbations. Thus, we show that the framework can handle diverse stochastic influences, suggesting its applicability to predict and control more complex stochastic dynamical systems.

I. INTRODUCTION

In stochastic dynamical systems with multistable states, noise drives transitions between stable states.[1] This phenomenon known as noise-induced transitions[2] underlies diverse natural processes: chemical reactions with multimodal distribution,[3,4] protein conformational switching,[5,6] and noisy genetic switches.[7]

Published under an exclusive license by AIP Publishing

Understanding these dynamics is crucial for enabling accurate predictions and designing control mechanisms in stochastic systems. Previous approaches to analyzing noise-induced transitions encompass the Onsager–Machlup functional[8] or large deviation theory.[9,10] Recently, Lin *et al.*[11] have introduced a data-driven method trying to capture the transition dynamics from data. While this approach has been validated for systems with Gaussian noise, its applicability to more general systems remains an open challenging problem.

The central limit theorem justifies the successful use of Gaussian noise for modeling stochastic transitions, as demonstrated by the effectiveness of Gaussian-driven stochastic differential equation (SDE) in capturing transition dynamics.[12–15] However, many systems exhibit non-Gaussian noise, which often leads to more complex dynamics and transition behavior.[16–21] Among various non-Gaussian noise types, α-stable Lévy motions hold particular significance for modeling transition phenomena arising from jump events; its heavy-tailed distribution and discontinuous path characteristics prove particularly effective in capturing these abrupt state changes,[1] such as Dansgaard–Oeschger events in climate systems.[18]

Although noise significantly influences transition paths in stochastic systems, the underlying deterministic dynamics are crucial for understanding the mechanisms of these transitions. A paradigmatic example arises in non-gradient dynamical systems exhibiting limit cycles: closed attracting trajectories in phase space that sustain persistent oscillations.[22] Such limit-cycle dynamics are ubiquitous in biological and chemical systems.[23–26] Unlike fixed-point attractors, which represent equilibrium states, limit cycles embody periodic solutions, offering a distinct form of stability. Consequently, systems capable of transitioning between limit cycles and fixed point attractors induced by noise are of significant importance, such as in neuronal models where weak noise can induce such transitions, thereby markedly altering firing patterns.[27]

In recent years, machine learning has emerged as a powerful approach for identifying and predicting dynamics[28–30] and has a deep connection to stochastic dynamics.[31–33] An important learning framework the reservoir computing (RC)[34,35] has proven to be an effective tool for learning dynamical systems,[36–38] including chaotic systems,[39–43] stochastic resonance,[44] and associative memory.[45] Recent advances[11,46,47] based on RC have been used to capture stochastic dynamical systems, particularly the multi-scaling RC framework,[11] demonstrating that it can help capture stochastic transitions under Gaussian noise in representative gradient and non-gradient bistable systems, where some traditional methods[28–30] struggle with.

In this study, we extend the multi-scaling reservoir computing framework to tackle more complex stochastic transitions, showcasing the effectiveness of machine learning in addressing these phenomena. Specifically, we apply this framework to a one-dimensional bistable system driven by α-stable Lévy noise to model transitions triggered by rare jumps, with training data containing approximately 20–30 transitions. To evaluate whether our method remains effective for limit-cycle systems, we test it on a two-dimensional non-gradient model that exhibits transitions between limit cycles and fixed points, using a dataset comprising around 10–20 transitions. This data-driven RC approach successfully captures the dynamics of stochastic transitions under non-Gaussian fluctuations and in limit-cycle systems, providing valuable insights into these processes and reinforcing the potential of machine learning and RC to analyze complex stochastic systems.

II. METHODS

A. Reservoir computing

An RC consists of three layers:[48,49] an input layer, a dynamical reservoir, and an output layer. The input layer maps the d_{in} dimensional time series to the N dimensional space of the reservoir using the matrix $\mathbf{W}_{in} \in \mathbb{R}^{N\times d_{in}}$. The dynamical reservoir, typically a recurrent neural network, processes these signals and captures temporal dependencies. The dynamical reservoir is characterized by its adjacency matrix $\mathbf{A} \in \mathbb{R}^{N\times N}$, which defines the connections between the N reservoir neurons. The state of the reservoir at time t is represented by the state vector $\mathbf{r}_t \in \mathbb{R}^N$. The output layer reads out the d_{out}-dimensional output from the reservoir states using the output matrix $\mathbf{W}_{out} \in \mathbb{R}^{d_{out}\times N}$. The RC is trained by optimizing $\mathbf{W}_{out}$ using ridge regression to minimize the loss function. The state of the reservoir at time $t+1$ is calculated by

$$\mathbf{r}_{t+1} = (1-\hat{\alpha})\mathbf{r}_t + \hat{\alpha} f(\mathbf{A}\mathbf{r}_t + \mathbf{W}_{in}\mathbf{u}_t), \tag{1}$$

where f is the activation function (hyperbolic tangent function). The output of the reservoir at time $t+1$ is given by

$$\tilde{\mathbf{u}}_{t+1} = \mathbf{W}_{out}\mathbf{r}_{t+1}. \tag{2}$$

The other hyperparameters of the RC[48,49] include the spectral radius ρ of the adjacency matrix $\mathbf{A}$, the average degree D of $\mathbf{A}$, the leak parameter $\hat{\alpha}$, and the regularization parameter β used in ridge regression. The input scale K_{in} determines the range of values for $\mathbf{W}_{in}$, which are uniformly sampled from this interval.

The training process involves optimizing $\mathbf{W}_{out}$ to minimize the loss function L, defined as

$$L = \sum_{t=1}^{T} \|\mathbf{u}_t - \mathbf{W}_{out}\mathbf{r}_t\|^2 + \beta\|\mathbf{W}_{out}\|^2, \tag{3}$$

where $\mathbf{u}_t$ represents the target at time t and $\mathbf{r}_t$ is the state of the reservoir at time t. This ensures that the RC generalizes well to unseen data by preventing overfitting through the regularization term $\beta\|\mathbf{W}_{out}\|^2$. Here, $\top$ denotes the transpose of a matrix. The $\mathbf{W}_{out}$ is trained using ridge regression.[49] Solving for $\mathbf{W}_{out}$, we obtain

$$\mathbf{W}_{out} = \mathbf{U}\mathbf{R}^\top(\mathbf{R}\mathbf{R}^\top + \beta\mathbf{I})^{-1}, \tag{4}$$

where $\mathbf{U} = \begin{bmatrix}\mathbf{u}_1 & \cdots & \mathbf{u}_T\end{bmatrix}$ and $\mathbf{R} = \begin{bmatrix}\mathbf{r}_1 & \cdots & \mathbf{r}_T\end{bmatrix}$ are matrices containing the target outputs and reservoir states, respectively, over the training period T. The term $\mathbf{I}$ denotes the identity matrix.

B. Multi-scaling reservoir computing

The framework of multi-scaling reservoir computing[11] is a data-driven approach for capturing stochastic transitions. It can effectively model stochastic transition dynamics from time series. The framework is inspired by the empirical observation that the leak parameter $\hat{\alpha}$ corresponds to the time scale of the dynamics.[50] In the training phase, the $\hat{\alpha}$ and other parameters are adjusted, and the RC is trained to capture the deterministic dynamics from the noisy data. After the $\mathbf{W}_{out}$ is trained, the noise can be separated from the

07 September 2026 09:11:39

Published under an exclusive license by AIP Publishing

training data. This is done by subtracting the RC prediction of the deterministic dynamics ($\mathbf{W}_{out}\mathbf{r}_t$) from the actual training data ($\mathbf{u}_t$) ($t \in [0, T_{\text{training}}]$). The residual, denoted as η_t, represents the noise

$$\eta_t = \mathbf{u}_t - \mathbf{W}_{out}\mathbf{r}_t. \tag{5}$$

During the prediction phase ($t \in [T_{\text{training}}, T_{\text{prediction}}]$), the trained RC iteratively simulates the system. At each time step, a noise sample is drawn from the distribution of the previously separated noise (η). This noise is then added to the output of the RC

$$\hat{\mathbf{u}}_{t+1} = \mathbf{W}_{out}\mathbf{r}_{t+1} + \eta_{t+1} \tag{6}$$

for generating a complete stochastic trajectory $\hat{\mathbf{u}}_{t+1}$ capable of exhibiting noise-induced transitions.

The output $\hat{\mathbf{u}}_{t+1}$ is then used as the input for the next time step, and the process continues iteratively. This approach allows the RC to capture the stochastic transitions in the system while maintaining the underlying deterministic dynamics. The multi-scaling RC framework effectively separates the noise from the training data, enabling accurate predictions of stochastic transitions in complex systems.

III. RESULTS

A. The Lévy noise

Lévy noise, a generalization of Brownian motion, is a stochastic process with heavy-tailed distributions and infinite variance, characterized by the Lévy–Khintchine formula,[1] which accounts for its discontinuous sample paths due to jumps. This makes it a powerful tool for modeling complex stochastic processes.

Consider the following stochastic differential equation driven by α-stable Lévy noise,

$$\mathrm{d}\mathbf{X}_t = \mathbf{b}(\mathbf{X}_t)\,\mathrm{d}t + \sigma(\mathbf{X}_t)\,\mathrm{d}\mathbf{L}_t^{\alpha}, \tag{7}$$

where t denotes time, the lower script t indicates time dependence, the drift term $\mathbf{b}(\mathbf{X}_t)$, and the diffusion coefficient $\sigma(\mathbf{X}_t)$ are functions of the state $\mathbf{X}_t$, and $\mathbf{L}_t^{\alpha}$ represents the α-stable Lévy motion. The Lévy noise is characterized by its stability index $0 < \alpha < 2$, which determines the heavy-tailed nature of its distribution. When $\alpha = 2$, the Lévy noise reduces to standard Brownian motion, and Eq. (7) simplifies to

$$\mathrm{d}\mathbf{X}_t = \mathbf{b}(\mathbf{X}_t)\,\mathrm{d}t + \sigma(\mathbf{X}_t)\,\mathrm{d}\mathbf{B}_t, \tag{8}$$

where $\mathbf{B}_t$ denotes Brownian motion. Further details on the differences between Lévy and Brownian motions are provided in Appendixes A and B. The Lévy noise exhibits jumps and discontinuities in its sample paths, which can lead to non-trivial dynamics and transitions in complex systems.

B. A double-well potential system driven by Lévy noises

Let us consider a one-dimensional bistable system driven by α-stable noise

$$\mathrm{d}u_t = \left(u_t - u_t^3\right)\,\mathrm{d}t + \varepsilon\,\mathrm{d}L_t^{\alpha}, \tag{9}$$

where ε represents the noise intensity parameter. The system has two stable states at $u = \pm 1$, which can exhibit noise-induced transitions. We simulate Eq. (9) with $\alpha = 1.5$ and $\alpha = 2$ to compare the dynamics driven by Lévy and Brownian noises, respectively. The stochastic transitions driven by Lévy noise exhibit discontinuous transition paths [Fig. 1(a)], while Brownian noise leads to continuous transition paths [Fig. 1(b)]. The probability density distributions of the noise used to simulate the paths are shown in Fig. 1(c).

To capture the stochastic transition dynamics driven by Lévy noise, we employ the multi-scaling RC framework,[11] which has been demonstrated to effectively model stochastic transitions under Gaussian noise. The training data [Fig. 2(a)], containing approximately 20–30 transitions, are generated by simulating the system for $t \in [0, 100]$, with the Lévy noise simulated using the Python library

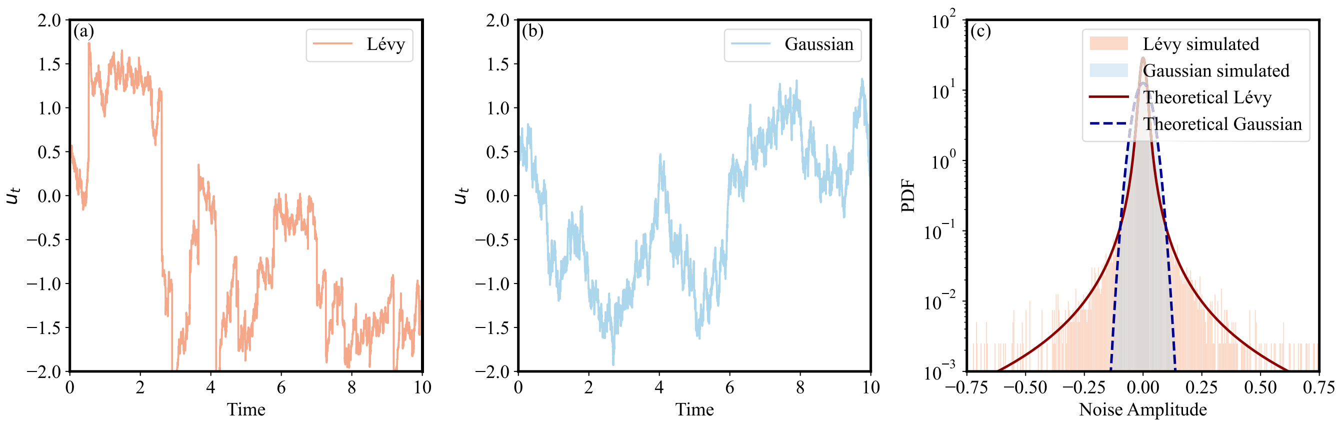

FIG. 1. Data generated from a one-dimensional bistable system with Lévy noise. (a) Sample path of Eq. (9), the stochastic transitions driven by α-stable motion with $\alpha = 1.5$ and $\varepsilon = 1$, illustrating the jump characteristics of Lévy motions, the time step $\Delta t = 0.001$. (b) Sample path of Eq. (9), the stochastic transitions driven by α-stable motion with $\alpha = 2$ and $\varepsilon = 1$, corresponding to Wiener motions. (c) Stable distributions for different α values. The probability density function (PDF) curves represent the theoretical Gaussian PDF, while the shaded areas correspond to the simulated noise PDF (red: Lévy; blue: Gaussian).

07 September 2026 09:11:39

Published under an exclusive license by AIP Publishing

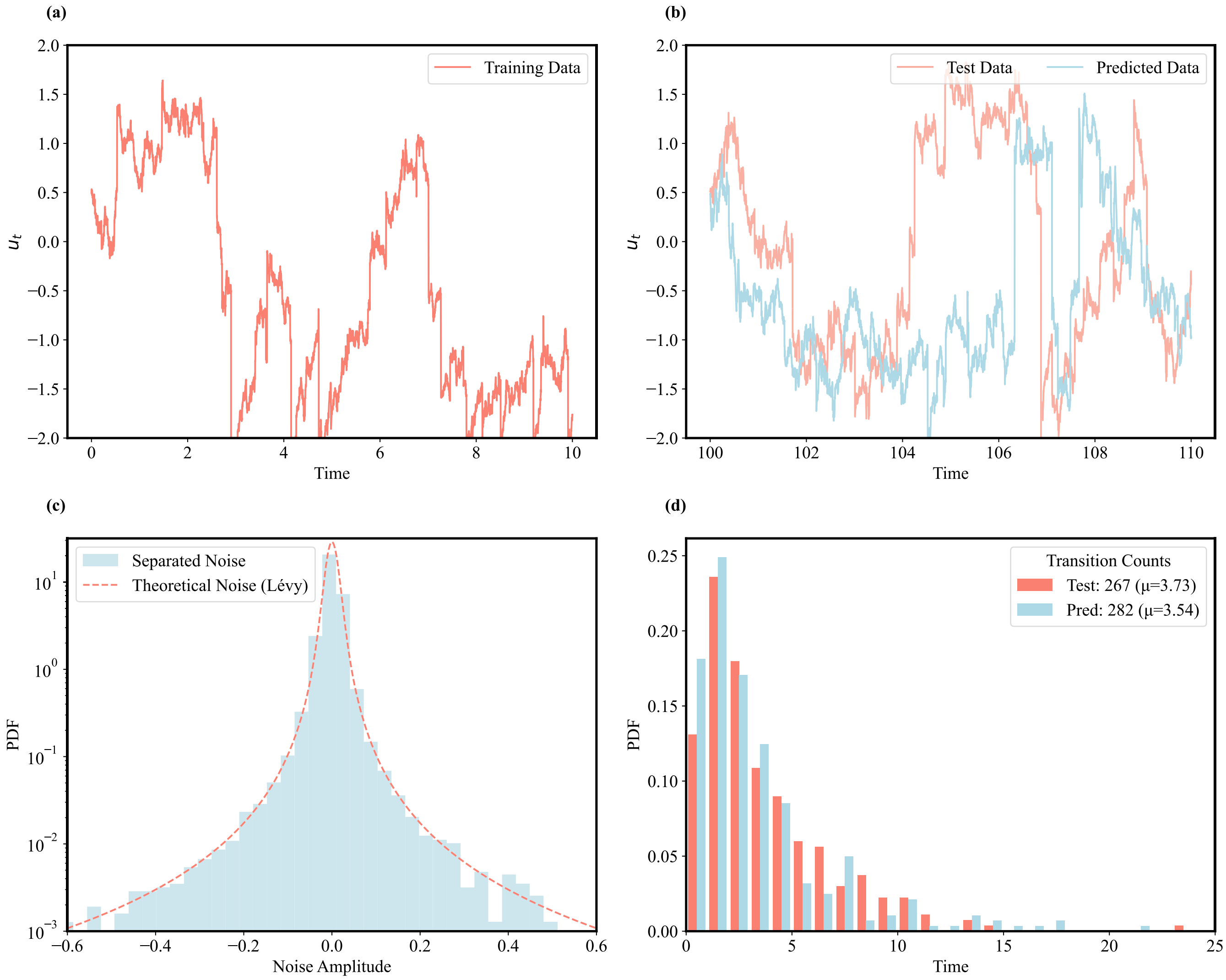

FIG. 2. Learning noise-induced transitions of the one-dimensional bistable system with Lévy noise. (a) The first 10 000 time steps of the training data generated from Eq. (9), with $\Delta t = 0.001$, as $t \in [0, 10]$. (b) Predicted data and test data for the first 10 000 time steps of the prediction phase ($\Delta t = 0.001$, $t \in [100, 110]$). (c) The noise amplitude probability distributions of both the separated noise and the theoretical noise. (d) Comparison of 267 true transitions (salmon) and 282 model-predicted transitions (light blue) when $t \in [100, 1100]$ (1×10^6 time steps), and μ represents the mean value of transition intervals. The distributions of transition times exhibit strong agreement between the test and predicted data.

SciPy.[51] Following the multi-scaling RC protocol, during the training phase, we adjust the hyperparameters (Table I) and train the RC to capture the deterministic mode using ridge regression. Additionally, the noise is separated at each time step and treated as a distribution.

In the prediction phase, as $t \in [100, 1100]$, we employ the trained RC model to simulate the deterministic dynamics and sample a noise from the separated noise distribution and added to the RC output, and use the output as the next timestep input iteratively. To validate the performance of the model, this predicted trajectory is compared against the corresponding segment of the simulation [Fig. 2(b)]. The results demonstrate that the predicted data and the test data have the same transition dynamics and discontinuous sample paths, while the probability density function of the separated noise distribution matches the theoretical Lévy distribution [Fig. 2(c)].

The distribution of transition intervals is a key characteristic of the transition dynamics. A transition interval is defined as the time elapsed between two consecutive transitions, where a transition refers to a switch from $\mathbf{u} = -1$ to $\mathbf{u} = 1$ or vice versa. We count the number of transitions in the test data (267 transitions) and the predicted data (282 transitions) and calculate the corresponding transition intervals. A comparison of the transition interval

07 September 2026 09:11:39

Published under an exclusive license by AIP Publishing

TABLE I. Hyperparameters used in the one-dimensional bistable system and the two-dimensional limit-cycle system.

Type	Hyperparameter	1D bistable system	2D limit-cycle system
RC hyperparameters	Spectral radius ρ	0.92	0.42
	Average degree D	6.00	22.00
	Leak parameter $\hat{\alpha}$	0.05	0.30
	Input scale K_{in}	4.00	4.00
	Regularization parameter β	1.00×10^{-8}	1.00×10^{-3}
	Number of neurons N	400	4000
System parameters	Noise intensity ε	1.0	0.8
	Stability index α	1.5	2
	Time step Δt	0.001	0.01
	γ_{right}	N/A	1
	k	N/A	5
	$\tilde{\mu}$	N/A	0.2

distributions is shown in Fig. 2(d). The results demonstrate that the multi-scaling RC framework effectively captures the transition dynamics driven by Lévy noise in this one-dimensional bistable system, as evidenced by the agreement between the distributions and mean values.

C. A two-dimensional non-gradient system with limit cycles under Gaussian noise

The transitions in the one-dimensional bistable system driven by Lévy noise highlight the effectiveness of the multi-scaling RC framework in capturing the stochastic dynamics induced by non-Gaussian noise. To further demonstrate the versatility of this framework, we extend its application to a more complex system: a two-dimensional non-gradient system with limit cycles. It is governed by the following stochastic differential equations:

$$\begin{aligned} \mathrm{d}x &= y\,\mathrm{d}t, \\ \mathrm{d}y &= \left[-\frac{\mathrm{d}U}{\mathrm{d}x} - \Gamma(x)y\right]\mathrm{d}t + \varepsilon\,\mathrm{d}B_t. \end{aligned} \tag{10}$$

The potential U is defined by the quartic polynomial

$$U(x) = \frac{x^4}{4} - \frac{x^2}{2}, \tag{11}$$

with a position-dependent damping coefficient featuring adaptive switching

$$\Gamma(x) = \gamma_{\text{right}}\sigma_{\text{tr}}(x) + \tilde{\mu}(x^2 - 1)[1 - \sigma_{\text{tr}}(x)]. \tag{12}$$

The switching function adopts a sigmoidal form

$$\sigma_{\text{tr}}(x) = \frac{1}{1 + \exp(-k(x))}. \tag{13}$$

This system is characterized by rotational dynamics, limit cycles, and non-gradient flow in the phase space. To illustrate the deterministic dynamics of the system, we employ a set of random initial points and simulate the trajectories for $t \in [0, 100]$ (Fig. 3). The trajectories converge to distinct attracting regions: limit cycles and stable points. This confirms the non-gradient nature of the system. Here, Gaussian noise with $\varepsilon = 0.8$ is used to induce stochastic transitions between the limit cycles and the stable points, while other parameters of the system are shown in Table I.

We first simulate this system using the Euler–Maruyama method[1] with the Python library sdeint[52] to generate training data for $t \in [0, 50]$, which includes around 10–20 transitions. Within the multi-scaling reservoir computing framework, we adjust the hyperparameters (Table I). We train the model to capture the deterministic dynamics, separate the noise distribution, and generate predictions for $t \in [50, 1100]$. The time series for each dimension of the training data and the predictions for $t \in [50, 100]$ are shown in Fig. 4(a). The transition dynamics and the oscillatory behavior of the limit cycle in the X-direction and Y-direction are effectively

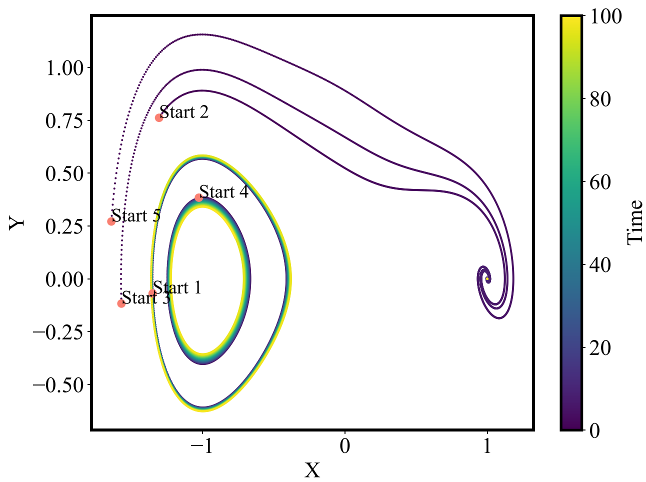

FIG. 3. Deterministic phase-space structure of a two-dimensional non-gradient system with limit cycles. Trajectories from five random initial conditions converge to distinct attracting regions: the limit cycles and the spiral stable point areas. Both the two areas have the rotational dynamics show the non-gradient system without detailed balance.

07 September 2026 09:11:39

Published under an exclusive license by AIP Publishing

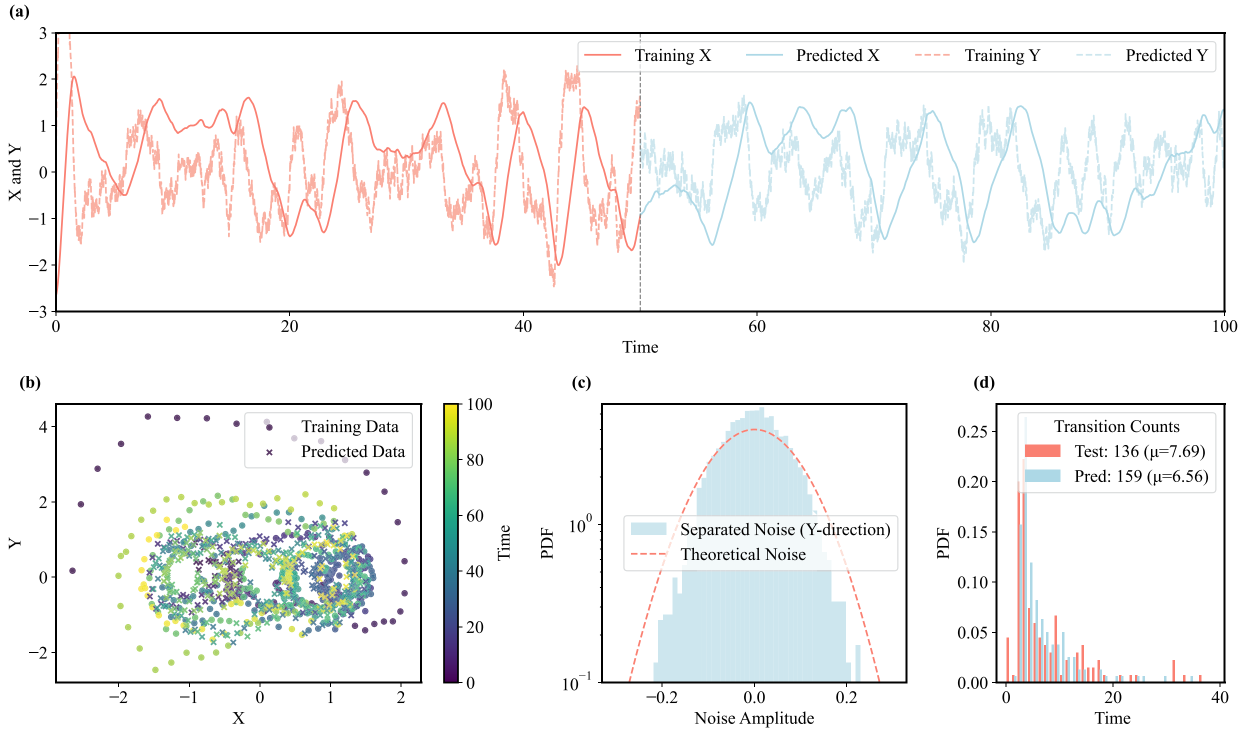

FIG. 4. Learning transition dynamics of the two-dimensional non-gradient system with limit cycles. (a) The one-dimensional time series of training data generated from Eq. (10) ($t \in [0, 50]$) and the predicted data ($t \in [50, 100]$), both in X and Y directions (time steps multiplied by $\Delta t = 0.01$). (b) The two-dimensional scatter plot of training data ($t \in [0, 50]$) and predicted data ($t \in [50, 100]$) showing the rotational and transition dynamics, where only every 10th point is plotted for clarity. (c) The noise amplitude probability distribution of both the separated noise and the theoretical noise (Y-direction). (d) The probability distributions and the mean value μ of the transition intervals of the test and predicted data for $t \in [50, 1100]$ show matching transition dynamics.

07 September 2026 09:11:39

captured. The two-dimensional scatter plot of the training and predicted data is shown in Fig. 4(b), demonstrating that the predicted data closely match the training data in terms of both limit-cycle dynamics and rotational behavior.

The separated noise distribution closely matches the theoretical Gaussian distribution ($\Delta t = 0.01$), as shown in Fig. 4(c). Its distribution is distinct from the Lévy distribution. As in the scenario of the one-dimensional bistable system, we compare the transition intervals of the test and predicted data for $t \in [50, 1100]$ to evaluate the transition dynamics. A transition occurs when the time series crosses the zero point in the X-direction without returning for $50\Delta t$.[11] The transition time is here defined as the interval between two consecutive zero crossings. The probability distributions and the mean value μ of the transition intervals of the test and predicted data for $t \in [50, 1100]$ are shown in Fig. 4(d). The results demonstrate that the multi-scaling RC framework effectively captures the transition dynamics driven by Gaussian noise in this two-dimensional non-gradient system with limit cycles.

IV. DISCUSSION

In this work, we have showcased the efficacy of the multi-scaling reservoir computing framework in modeling noise-induced transitions, focusing on a one-dimensional bistable system driven by Lévy noise and a two-dimensional limit-cycle system under Gaussian perturbations. For Lévy noise with the stability index $\alpha = 1.5$, the framework successfully captures discontinuous transition paths and the distribution of transition intervals. However, the robustness of this approach for smaller α values remains unexplored. As α decreases, the Lévy distribution's tails grow heavier, amplifying the frequency and magnitude of jumps. This shift could challenge the framework's ability to disentangle deterministic dynamics from increasingly extreme noise. Investigating the method's performance across a wider range of α values is thus essential to ascertain its applicability to systems with more pronounced heavy-tailed noise, such as those observed in climate or financial models.

Among non-Gaussian noises, we have exclusively tested the paradigmatic α-stable Lévy noise. While this choice effectively captures jump-driven transitions, other non-Gaussian forms like Poisson noise, may exhibit distinct stochastic behaviors. The adaptability of the framework to these alternatives is unknown. Future efforts should extend the multi-scaling RC approach to diverse non-Gaussian noise types to evaluate its versatility and refine its applicability across broader physical contexts.

In second-order dynamical systems, it is common practice in physics to introduce noise preferentially along the velocity

Published under an exclusive license by AIP Publishing

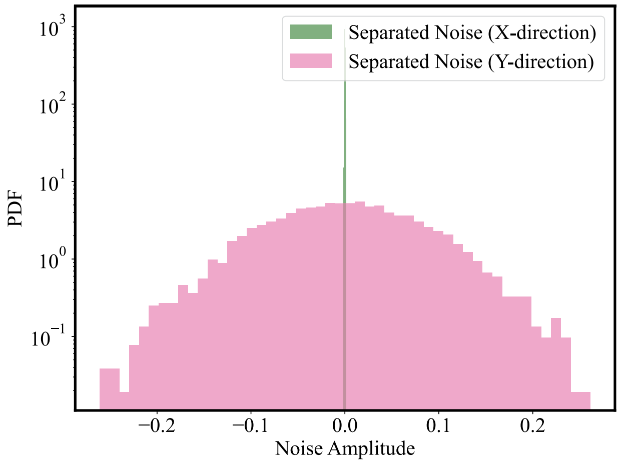

FIG. 5. Results of separated noise of different dimensions. A histogram illustrating the noise separated from a two-dimensional limit-cycle system. The separated noise exhibits a Gaussian distribution in the Y-direction and is nearly zero in the X-direction, consistent with the theoretical expectation that noise is injected solely in the Y-direction.

direction[23] (the Y-direction in our example). However, identifying the specific directional coupling of noise in two-dimensional stochastic systems presents a significant challenge, as this generally requires *a priori* knowledge of the system's underlying structure, which is a question beyond the scope of this investigation. Following physical principles, we intentionally restrict Gaussian noise injection to the Y-direction during both training data generation and prediction phase. Crucially, the multi-scaling reservoir computing framework successfully isolates the Y-direction noise distribution matches with the injected Gaussian distribution, while simultaneously recovering the X-direction noise distribution statistically indistinguishable from zero (Fig. 5). This noise-deterministic dynamics decoupling capability enables a data-driven characterization of stochastic systems lacking prior knowledge of anisotropy noise. This demonstrated the ability to decouple noise and identify the underlying deterministic dynamics that makes the framework a compelling candidate for integration with model-based estimators. Specifically, the trained reservoir could be a data-driven surrogate for the state-transition model required by filters such as the unscented Kalman filter (UKF).[53] This hybrid approach is a promising direction for future work, potentially enabling robust state estimation in systems where the governing equations are unknown *a priori*.

In conventional reservoir computing (RC), the inherent randomness of the reservoir matrices leads to performance variability across different matrix realizations, even when hyperparameters are held constant. To address this issue, we try the next-generation reservoir computing (NGRC) for the training phase, which eliminates random matrices in favor of nonlinear vector autoregression.[38] In NGRC, the time scale analogous to the leak parameter $\hat{\alpha}$ in traditional RC is governed by the adjustable time delay and stride parameters, enabling to capture the deterministic dynamics. However, the reduced hyperparameter set in NGRC limits the flexibility of manual tuning, particularly for modeling complex deterministic systems. Moreover, increasing the time delay to capture slower dynamical scales causes a sharp rise in the dimensionality of the nonlinear layer, imposing significant memory constraints. Similarly, incorporating higher order nonlinearities exacerbates these computational demands. Consequently, traditional RC may remain preferable for applications involving stochastic transitions, where such resource limitations are less restrictive.

Beyond exploring different noise profiles, the scalability of our framework to higher-dimensional systems presents an important avenue for future research. Our focus on low-dimensional systems aligns with the established tradition in both stochastic transition[7,54–56] and reservoir computing studies. However, the challenge of high dimensionality for our specific problem is often mitigated by the nature of the transition phenomenon itself. Even in a high-dimensional state space, the transition path between two stable states frequently evolves along a low-dimensional manifold or a "reaction coordinate."[57] This effectively reduces our core task to modeling a low-dimensional process. Furthermore, applying RC to higher-dimensional systems is considered tractable,[36] with the primary challenge being the need for increased computational resources rather than a fundamental algorithmic limitation.

In summary, the multi-scaling RC framework adeptly captures noise-induced transitions under Lévy noise and in limit-cycle systems, highlighting its potential as a data-driven tool for stochastic dynamics. The close agreement between predicted and test data validates its precision. Nonetheless, its dependence on $\alpha = 1.5$, exclusive focus on Lévy noise, and hyperparameter challenges signal avenues for refinement. Extending its scope to diverse noise profiles and stabilizing its implementation will solidify its role in modeling noise-induced phenomena across physics, biology, and beyond.

07 September 2026 09:11:39

ACKNOWLEDGMENTS

The authors thank Mr. Zhaofan Lu, Mr. Kaidian Wang, and Mr. Jiarui Dong for their valuable discussions. This work is supported by Project No. 12322501 of the National Natural Science Foundation of China.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Zequn Lin: Data curation (equal); Investigation (equal); Methodology (equal); Software (equal); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Jürgen Kurths:** Conceptualization (equal); Project administration (equal); Supervision (equal); Writing – original draft (equal); Writing – review & editing (equal). **Ying Tang:** Conceptualization (equal); Funding acquisition (equal); Methodology (equal); Project administration (equal); Resources (equal); Supervision (equal); Writing – original draft (equal); Writing – review & editing (equal).

Published under an exclusive license by AIP Publishing

DATA AVAILABILITY

The data produced in this research, encompassing both the training and the predicted data, are available in the GitHub repository at https://github.com/Machine-learning-and-complex-systems/NIT-RC-Levy, Ref. 58. The Python implementation code is hosted in the same GitHub repository as the data.

APPENDIX A: LÉVY AND BROWNIAN MOTION

Following the standard definitions as presented in the textbooks of the Lévy motion[59] and Brownian motion,[1] we briefly introduce these stochastic processes. The key difference between these two motions lies in their continuity properties. The Brownian motion, also known as the Wiener process, is continuous almost surely. In contrast, Lévy motion, which includes jumps, is continuous in probability, but may exhibit discontinuities in its sample paths. This fundamental distinction leads to different behaviors and applications in modeling stochastic processes.

Lévy Motion Definition: An α-stable Lévy motion, denoted as $\mathbf{L}_t^\alpha$, is a stochastic process that meets the following conditions:

1. $\mathbf{L}_0^\alpha = 0$ a.s.
2. Independent increments: For any time partition $t_1 < t_2 < \cdots < t_n$, the random variables $\mathbf{L}_{t_2}^\alpha - \mathbf{L}_{t_1}^\alpha, \ldots, \mathbf{L}_{t_n}^\alpha - \mathbf{L}_{t_{n-1}}^\alpha$ are independent.
3. Stationary increments: For $t > s$, $\mathbf{L}_t^\alpha - \mathbf{L}_s^\alpha$ and $\mathbf{L}_{t-s}^\alpha$ have the same distribution.
4. Sample paths are stochastically continuous: For all $\delta > 0$ and $s \geq 0$, $\mathbb{P}(|\mathbf{L}_t^\alpha - \mathbf{L}_s^\alpha| > \delta) \to 0$, as $t \to s$.

Brownian Motion Definition: A Brownian motion, defined on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$, is a stochastic process B_t that satisfies the following conditions:

1. $\mathbf{B}_0 = 0$ a.s.
2. The sample function $t \to \mathbf{B}_t(\omega)$ is continuous, a.s.
3. $\mathbf{B}_t$ has independent increments: For any n and $0 \leq t_1 < t_2 < \cdots < t_n$, the random variables $\mathbf{B}_{t_2} - \mathbf{B}_{t_1}, \ldots, \mathbf{B}_{t_n} - \mathbf{B}_{t_{n-1}}$ are independent.
4. $\mathbf{B}_t$ has Gaussian stationary increments, that is, for any s, t $(0 \leq s < t)$, $\mathbf{B}_t(\omega) - \mathbf{B}_s(\omega)$ follows a Gaussian distribution with mean value zero and variance $t - s$.

APPENDIX B: NUMERICAL DETAILS FOR SIMULATING GAUSSIAN NOISE AND LÉVY NOISE

To generate the stochastic transition paths used as the training and test data in the one-dimensional bistable driven by Lévy noise, the increments of the Lévy motion, $d\mathbf{L}_t^\alpha$, were generated using the SciPy library.[51] The parameters were set as follows: Stability index $\alpha = 1.5$, determining the tail heaviness of the distribution; Skewness $\hat{\beta} = 0$, ensuring a symmetric stable distribution; Location parameter $\hat{\mu} = 0$; Scale parameter $\hat{\sigma} = (\Delta t)^{1/\alpha}$ because it is the stable distribution,[1] where Δt is the time step, consistent with the self-similarity property of α-stable Lévy processes.

To mitigate numerical instability arising from extreme jumps inherent in Lévy noise, a dynamical truncation was applied. The absolute values of the increments were computed, and the 99.9th percentile (q) was determined. Increments were then clipped to the range $[-20q, 20q]$.

We employ the Python library sdeint[52] to simulate the transition trajectories of the two-dimensional non-gradient system driven by Gaussian noise. Numerical integration is performed using the `stratHeun` solver (Stratonovich interpretation) with a fixed time step $\Delta t = 0.01$. Initial conditions are set randomly, and we use a single trajectory with 110 000 time steps (matching Example 1) as the training data ($t \in [0, 50]$) and test data ($t \in [50, 1100]$).

REFERENCES

[1] J. Duan, *An Introduction to Stochastic Dynamics* (Cambridge University Press, 2015), Vol. 51.
[2] W. Horsthemke, "Noise induced transitions," in *Stochastic Nonlinear Systems in Physics, Chemistry, and Biology: Proceedings of the Workshop Bielefeld, Federal Republic of Germany, October 5–11, 1980* (Springer, 1985), pp. 116–126.
[3] P. Hänggi, P. Talkner, and M. Borkovec, "Reaction-rate theory: Fifty years after kramers," Rev. Mod. Phys. **62**, 251 (1990).
[4] Y. Tang, J. Weng, and P. Zhang, "Neural-network solutions to stochastic reaction networks," Nat. Mach. Intell. **5**, 376–385 (2023).
[5] H. Qian, "From discrete protein kinetics to continuous brownian dynamics: A new perspective," Protein Sci. **11**, 1–5 (2002).
[6] R. Tapia-Rojo, M. Mora, S. Board, J. Walker, R. Boujemaa-Paterski, O. Medalia, and S. Garcia-Manyes, "Enhanced statistical sampling reveals microscopic complexity in the talin mechanosensor folding energy landscape," Nat. Phys. **19**, 52–60 (2023).
[7] M. Assaf, E. Roberts, and Z. Luthey-Schulten, "Determining the stability of genetic switches: Explicitly accounting for mRNA noise," Phys. Rev. Lett. **106**, 248102 (2011).
[8] Y. Chao and J. Duan, "The Onsager–Machlup function as Lagrangian for the most probable path of a jump-diffusion process," Nonlinearity **32**, 3715 (2019).
[9] Z. Chen and X. Liu, "Noise induced transitions and topological study of a periodically driven system," Commun. Nonlinear Sci. Numer. **48**, 454–461 (2017).
[10] P. Bernuzzi and T. Grafke, "Large deviation minimisers for stochastic partial differential equations with degenerate noise," arXiv:2409.17839 (2024).
[11] Z. Lin, Z. Lu, Z. Di, and Y. Tang, "Learning noise-induced transitions by multi-scaling reservoir computing," Nat. Commun. **15**, 6584 (2024).
[12] E. Forgoston and I. B. Schwartz, "Escape rates in a stochastic environment with multiple scales," SIAM J. Appl. Dyn. Syst. **8**, 1190–1217 (2009).
[13] M. F. Weber and E. Frey, "Master equations and the theory of stochastic path integrals," Rep. Prog. Phys. **80**, 046601 (2017).
[14] F. Pinski and A. Stuart, "Transition paths in molecules at finite temperature," J. Chem. Phys. **132**, 184104 (2010).
[15] S. Xu, S. Jiao, P. Jiang, and P. Ao, "Two-time-scale population evolution on a singular landscape," Phys. Rev. E **89**, 012724 (2014).
[16] J.-P. Bouchaud and A. Georges, "Anomalous diffusion in disordered media: Statistical mechanisms, models and physical applications," Phys. Rep. **195**, 127–293 (1990).
[17] R. Metzler and J. Klafter, "The random walk's guide to anomalous diffusion: A fractional dynamics approach," Phys. Rep. **339**, 1–77 (2000).
[18] V. Lucarini, L. Serdukova, and G. Margazoglou, "Lévy noise versus Gaussian-noise-induced transitions in the Ghil–Sellers energy balance model," Nonlinear Process. Geophys. **29**, 183–205 (2022).
[19] Y. Zheng, L. Serdukova, J. Duan, and J. Kurths, "Transitions in a genetic transcriptional regulatory system under Lévy motion," Sci. Rep. **6**, 29274 (2016).
[20] F. Wu, X. Chen, Y. Zheng, J. Duan, J. Kurths, and X. Li, "Lévy noise induced transition and enhanced stability in a gene regulatory network," Chaos **28**, 075510 (2018).
[21] Y. Zheng, F. Yang, J. Duan, X. Sun, L. Fu, and J. Kurths, "The maximum likelihood climate change for global warming under the influence of greenhouse effect and Lévy noise," Chaos **30**, 013132 (2020).
[22] S. H. Strogatz, *Nonlinear Dynamics and Chaos with Student Solutions Manual: With Applications to Physics, Biology, Chemistry, and Engineering* (CRC Press, 2018).

07 September 2026 09:11:39

Published under an exclusive license by AIP Publishing

[23] W. Zhang, W. Xu, Y. Bai, Y. Tang, and J. Kurths, "Stochastic response analysis of a birhythmic oscillator under poisson white noise excitation via path integration method," Phys. Rev. E **110**, 044212 (2024).
[24] C. Morris and H. Lecar, "Voltage oscillations in the barnacle giant muscle fiber," Biophys. J. **35**, 193–213 (1981).
[25] S. Kar and D. Ray, "Large fluctuations and nonlinear dynamics of birhythmicity," Europhys. Lett. **67**, 137 (2004).
[26] S. Kar and D. S. Ray, "Collapse and revival of glycolytic oscillation," Phys. Rev. Lett. **90**, 238102 (2003).
[27] M. E. Yamakou and J. Jost, "Weak-noise-induced transitions with inhibition and modulation of neural oscillations," Biol. Cybern. **112**, 445–463 (2018).
[28] K. Kaheman, S. L. Brunton, and J. N. Kutz, "Automatic differentiation to simultaneously identify nonlinear dynamics and extract noise probability distributions from data," Mach. Learn.: Sci. Technol. **3**, 015031 (2022).
[29] D. Sussillo and L. F. Abbott, "Generating coherent patterns of activity from chaotic neural networks," Neuron **63**, 544–557 (2009).
[30] S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Discovering governing equations from data by sparse identification of nonlinear dynamical systems," Proc. Natl. Acad. Sci. U.S.A. **113**, 3932–3937 (2016).
[31] Y. Feng and Y. Tu, "The inverse variance–flatness relation in stochastic gradient descent is critical for finding flat minima," Proc. Natl. Acad. Sci. U.S.A. **118**, e2015617118 (2021).
[32] N. Yang, C. Tang, and Y. Tu, "Stochastic gradient descent introduces an effective landscape-dependent regularization favoring flat solutions," Phys. Rev. Lett. **130**, 237101 (2023).
[33] R. Li, H. Wang, Q. Liao, and Y. Li, "Predicting the energy landscape of stochastic dynamical system via physics-informed self-supervised learning," arXiv:2502.16828 (2025).
[34] H. Jaeger, "The 'echo state' approach to analysing and training recurrent neural networks—With an erratum note," GMD Tech. Rep. 148, German National Research Centre for Information Technology, 2001, p. 13.
[35] W. Maass, T. Natschläger, and H. Markram, "Real-time computing without stable states: A new framework for neural computation based on perturbations," Neural comput. **14**, 2531–2560 (2002).
[36] 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).
[37] H. Jaeger and H. Haas, "Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication," Science **304**, 78–80 (2004).
[38] D. J. Gauthier, E. Bollt, A. Griffith, and W. A. Barbosa, "Next generation reservoir computing," Nat. Commun. **12**, 1–8 (2021).
[39] J. Z. Kim, Z. Lu, E. Nozari, G. J. Pappas, and D. S. Bassett, "Teaching recurrent neural networks to infer global temporal structure from local examples," Nat. Mach. Intell. **3**, 316–323 (2021).
[40] L. Shi, Y. Yan, H. Wang, S. Wang, and S.-X. Qu, "Predicting chaotic dynamics from incomplete input via reservoir computing with $(D+1)$-dimension input and output," Phys. Rev. E **107**, 054209 (2023).
[41] T. L. Carroll, "Using reservoir computers to distinguish chaotic signals," Phys. Rev. E **98**, 052209 (2018).
[42] H. Tan, L. Shi, S. Wang, and S.-X. Qu, "Improving model-free prediction of chaotic dynamics by purifying the incomplete input," AIP Adv. **14**, 125225 (2024).
[43] J. Guo, Y. Du, H. Luo, X. Wang, Y. Yu, and X. Wang, "Model-free prediction of chaotic dynamics with parameter-aware reservoir computing," Chin. Phys. B **34**(4), 040505 (2025).
[44] Z.-M. Zhai, L.-W. Kong, and Y.-C. Lai, "Emergence of a resonance in machine learning," Phys. Rev. Res. **5**, 033127 (2023).
[45] L.-W. Kong, G. A. Brewer, and Y.-C. Lai, "Reservoir-computing based associative memory and itinerancy for complex dynamical attractors," Nat. Commun. **15**, 4840 (2024).
[46] S. H. Lim, L. Theo Giorgini, W. Moon, and J. S. Wettlaufer, "Predicting critical transitions in multiscale dynamical systems using reservoir computing," Chaos **30**, 123126 (2020).
[47] A. E. Hramov, N. Kulagin, A. N. Pisarchik, and A. V. Andreev, "Strong and weak prediction of stochastic dynamics using reservoir computing," Chaos **35**, 033140 (2025).
[48] M. Lukoševičius and H. Jaeger, "Reservoir computing approaches to recurrent neural network training," Comput. Sci. Rev. **3**, 127–149 (2009).
[49] K. Nakajima and I. Fischer, *Reservoir Computing* (Springer, 2021).
[50] G. Tanaka, T. Matsumori, H. Yoshida, and K. Aihara, "Reservoir computing with diverse timescales for prediction of multiscale dynamics," Phys. Rev. Res. **4**, L032014 (2022).
[51] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, and J. Bright *et al.*, "Scipy 1.0: Fundamental algorithms for scientific computing in python," Nat. Methods **17**, 261–272 (2020).
[52] M. J. Aburn, "Critical fluctuations and coupling of stochastic neural mass models," Ph.D. thesis, School of Mathematics and Physics, The University of Queensland, 2017. https://doi.org/10.14264/uql.2017.148
[53] S. J. Julier and J. K. Uhlmann, "New extension of the Kalman filter to nonlinear systems," in *Signal Processing, Sensor Fusion, and Target Recognition VI* (SPIE, 1997), Vol. 3068, pp. 182–193.
[54] H. A. Kramers, "Brownian motion in a field of force and the diffusion model of chemical reactions," Physica **7**, 284–304 (1940).
[55] W. Horsthemke and R. Lefever, *Noise-Induced Transitions in Physics, Chemistry, and Biology* (Springer, 1984).
[56] T. S. Gardner, C. R. Cantor, and J. J. Collins, "Construction of a genetic toggle switch in escherichia coli," Nature **403**, 339–342 (2000).
[57] E. Weinan and E. Vanden-Eijnden, "Towards a theory of transition paths," J. Stat. Phys. **123**, 503–523 (2006).
[58] Z. Lin (2025). GitHub. https://github.com/Machine-learning-and-complex-systems/NIT-RC-Levy
[59] D. Applebaum, *Lévy Processes and Stochastic Calculus* (Cambridge University Press, 2009).

07 September 2026 09:11:39

Published under an exclusive license by AIP Publishing