

## RESEARCH ARTICLE

## Reconstructing noisy gene regulation dynamics using extrinsic-noise-driven neural stochastic differential equations

Jiancheng Zhang<sup>1</sup>, Xiangting Li<sup>2</sup>, Xiaolu Guo<sup>3</sup>, Zhaoyi You<sup>4</sup>, Lucas Böttcher<sup>5,6</sup>, Alex Mogilner<sup>7</sup>, Alexander Hoffmann<sup>3</sup>, Tom Chou<sup>2,8\*</sup>, Mingtao Xia<sup>9\*</sup>

**1** Department of Electrical and Computer Engineering, University of California, Riverside, California, United States of America, **2** Department of Computational Medicine, University of California, Los Angeles, California, United States of America, **3** Department of Microbiology, Immunology, and Molecular Genetics (MIMG) and Institute for Quantitative and Computational Biosciences, University of California, Los Angeles, California, United States of America, **4** Ray and Stephanie Lane Computational Biology Department, School of Computer Science, Carnegie Mellon University, Pittsburgh, Pennsylvania, United States of America, **5** Department of Computational Science and Philosophy, Frankfurt School of Finance and Management, Frankfurt am Main, Germany, **6** Laboratory for Systems Medicine, Department of Medicine, University of Florida, Gainesville, Florida, United States of America, **7** Courant Institute of Mathematical Sciences, New York University, New York, New York, United States of America, **8** Department of Mathematics, University of California, Los Angeles, California, United States of America, **9** Department of Mathematics, University of Houston, Houston, Texas, United States of America

✉ These authors contributed equally to this work.

\* [xiaoluguo@g.ucla.edu](mailto:xiaoluguo@g.ucla.edu) (XG); [tomchou@ucla.edu](mailto:tomchou@ucla.edu) (TC); [mxia4@uh.edu](mailto:mxia4@uh.edu) (MX)



## OPEN ACCESS

**Citation:** Zhang J, Li X, Guo X, You Z, Böttcher L, Mogilner A, et al. (2025) Reconstructing noisy gene regulation dynamics using extrinsic-noise-driven neural stochastic differential equations. *PLoS Comput Biol* 21(9): e1013462. <https://doi.org/10.1371/journal.pcbi.1013462>

**Editor:** Michael A. Beer, Johns Hopkins University School of Medicine, UNITED STATES OF AMERICA

**Received:** April 2, 2025

**Accepted:** August 24, 2025

**Published:** September 17, 2025

**Peer Review History:** PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pcbi.1013462>

**Copyright:** © 2025 Zhang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data availability statement:** No data was created in this research. All data used in this

## Abstract

Proper regulation of cell signaling and gene expression is crucial for maintaining cellular function, development, and adaptation to environmental changes. Reaction dynamics in cell populations is often noisy because of (i) inherent stochasticity of intracellular biochemical reactions (“intrinsic noise”) and (ii) heterogeneity of cellular states across different cells that are influenced by external factors (“extrinsic noise”). In this work, we introduce an extrinsic-noise-driven neural stochastic differential equation (END-nSDE) framework that utilizes the Wasserstein distance to accurately reconstruct SDEs from stochastic trajectories measured across a heterogeneous population of cells (extrinsic noise). We demonstrate the effectiveness of our approach using both simulated and experimental data from three different systems in cell biology: (i) circadian rhythms, (ii) RPA-DNA binding dynamics, and (iii) NF $\kappa$ B signaling processes. Our END-nSDE reconstruction method can model how cellular heterogeneity (extrinsic noise) modulates reaction dynamics in the presence of intrinsic noise. It also outperforms existing time-series analysis methods such as recurrent neural networks (RNNs) and long short-term memory networks (LSTMs). By inferring cellular heterogeneities from data, our END-nSDE reconstruction method can reproduce noisy dynamics observed in experiments. In summary, the reconstruction method we propose offers a useful surrogate modeling approach for complex biophysical processes, where high-fidelity mechanistic models may be impractical.

research are publicly available at <https://www.nature.com/articles/s41467-023-39579-y> and <https://www.embopress.org/doi/full/10.1038/s44320-024-00047-4> and have been properly cited. The simulated datasets, neural SDE model code, and analysis scripts to replicate the study findings are available on GitHub at <https://github.com/JianchengZ/Neural-SDE-GeneDynamics>.

**Funding:** XG acknowledges financial support from UCLA Collaboratory Fellowship. LB acknowledges financial support from hessian. AI and the ARO through grant W911NF-23-1-0129. TC acknowledges inspiring discussions at the “Statistical Physics and Adaptive Immunity” program at the Aspen Center for Physics, which is supported by the National Science Foundation grant PHY-2210452. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. XG and AH acknowledge support from NIH R01AI173214.

**Competing interests:** The authors have declared that no competing interests exist.

## Author summary

In this work, we propose extrinsic-noise-driven neural stochastic differential equations (END-nSDE) to reconstruct noisy regulated gene expression dynamics. One of our main contributions is that we generalize a recent Wasserstein-distance-based SDE reconstruction approach to incorporate extrinsic noise (parameters that vary across different cells). Our approach can thus capture intrinsic fluctuations in gene regulatory dynamics driven by extrinsic noise (heterogeneity among cells), offering an advantage over deterministic models and outperforming other benchmarks. By inferring noise intensities from batches of experimental data, our END-nSDE can partially capture experimental noisy signaling dynamic data and provides a surrogate model for biomolecular processes that are too complex to model directly.

## 1. Introduction

Reactions that control signaling and gene regulation are important for maintaining cellular function, development, and adaptation to environmental changes, which impact all aspects of biological systems, from embryonic development to an organism’s ability to sense and respond to environmental signals. Variations in gene regulation, arising from noisy biochemical processes [1,2], can result in phenotypic heterogeneity even in a population of genetically identical cells [3].

Noise within cell populations can be categorized as (i) “intrinsic noise,” which arises from the inherent stochasticity of biochemical reactions and quantifies, *e.g.*, biological variability across cells in the same state [2,4,5], and (ii) “extrinsic noise,” which encompasses heterogeneities in environmental factors or differences in cell state across a population. A substantial body of literature has focused on quantifying intrinsic and extrinsic noise from experimental and statistical perspectives [1,2,6–13]. Experimental studies have specifically identified relevant sources of noise in various organisms, including *E. coli* (Escherichia coli), yeast, and mammalian systems [2,14–17].

Extrinsic noise is associated with uncertainties in biological parameters that vary across different cells. The distribution over physical and chemical parameters determine the observed variations in cell states, concentrations, locations of regulatory proteins and polymerases [1,2,18], and transcription and translation rates [19]. For example, extrinsic noise is the main contributor to the variability of concentrations of oscillating p53 protein levels across cell populations [20]. On the other hand, intrinsic noise, *i.e.*, inherent stochasticity of cells in the same state, can limit the accuracy of expression and signal transmission [2,5]. Based on the law of mass action [21,22], ordinary differential equations (ODEs) apply only in some deterministic or averaged limit and do not take into account intrinsic noise. Therefore, stochastic models are necessary to accurately represent biological processes, such as thermodynamic fluctuations inherent to molecular interactions within regulatory networks [1,5,18] or random event times in birth-death processes.

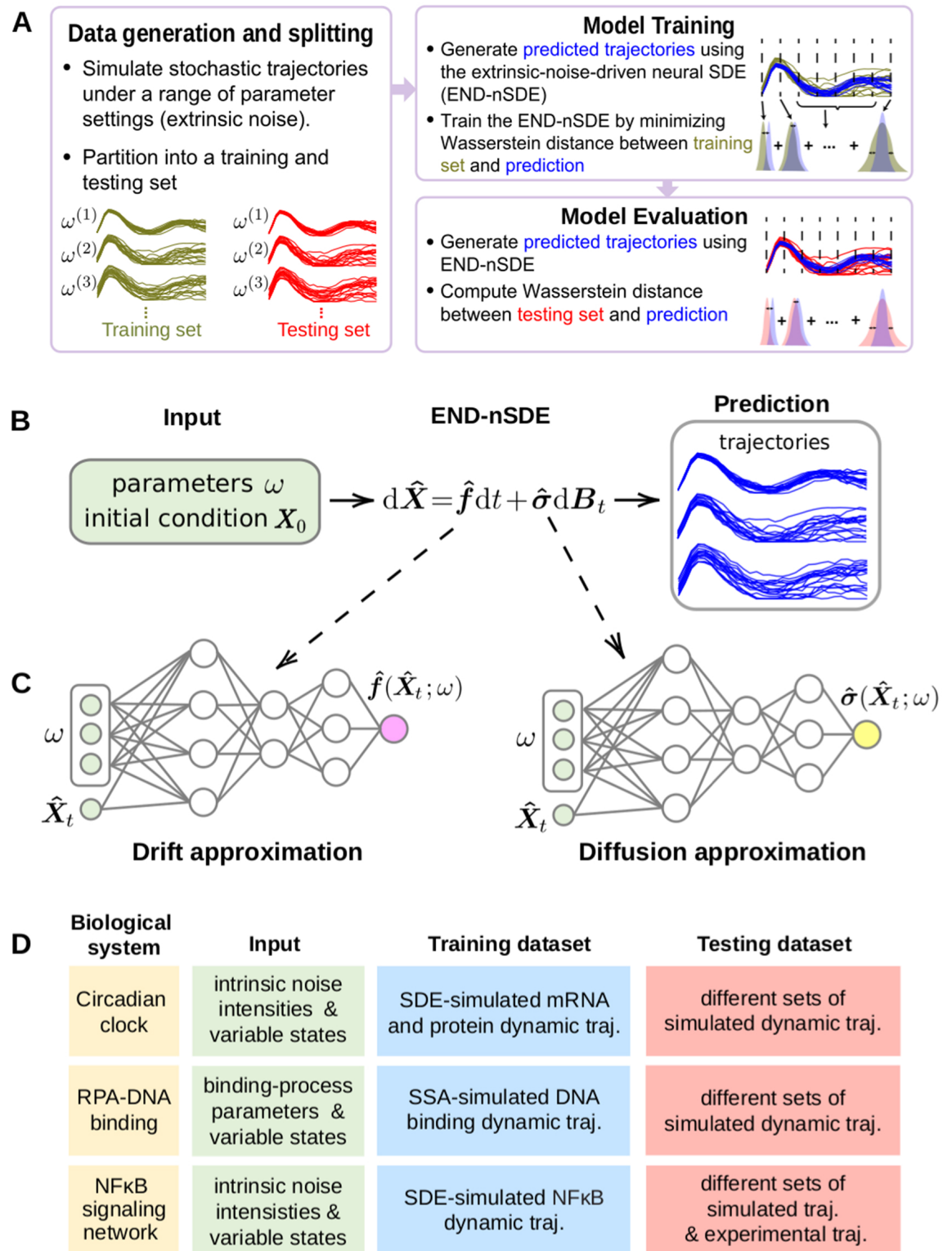
Existing stochastic modeling methods that account for intrinsic noise include Markov jump processes [23,24] and SDEs [25–27]. These approaches are applicable to different system sizes: Markov jump processes provide exact descriptions for discrete molecular systems, while SDEs serve as continuous approximations to Markov processes when molecular abundances are sufficiently high. SDE approaches may not be suitable for gene expression systems with very low copy numbers, where discrete master equation descriptions are more accurate. However, SDE approaches become more appropriate when modeling protein dynamics or when

gene regulatory interactions are modeled implicitly through Hill functions. Additionally, a hierarchical Markov model was designed in [28] for parameter inference in dual-reporter experiments to separate the contributions of extrinsic noise, intrinsic noise, and measurement error when both extrinsic and intrinsic noise are present. The described methods have been effective in the reconstruction of low-dimensional noisy biological systems. Discrete master-equation methods to model the evolution of probabilities in systems characterizing, *e.g.*, gene regulatory dynamics [29–31], can be computationally expensive and usually require specific forms of a stochastic model with unknown parameters that need to be inferred. It is unclear whether such methods and their generalizations can be applied to more complex (*e.g.*, higher-dimensional) systems for which a mechanistic description of the underlying biophysical dynamics is not available or impractical.

SDEs can capture both the mean dynamics (as ODEs do) and random fluctuations, offering a practical and scalable alternative to master equations in complex systems. Thus, we introduce an extrinsic-noise-driven neural stochastic differential equation (END-nSDE) reconstruction method that builds upon a recently developed Wasserstein distance ( $W_2$  distance) nSDE reconstruction method [32]. Our method is used to identify macromolecular reaction kinetics and cell signaling dynamics from noisy observational data in the presence of *both extrinsic and intrinsic noise*. A key question we address in this paper is how extrinsic noise that characterizes cellular heterogeneity influences the overall stochastic dynamics of the population.

The major differences between the approach presented here and prior work [32] are: (i) the inclusion of extrinsic noise into the framework allowing one to model cell-to-cell variability through parameter heterogeneity, and (ii) the ability of our method to learn the dependency of the SDE on those parameters, enabling reconstruction of a family of SDEs rather than a single SDE model. In contrast, the method developed in reference [32] focuses on reconstructing a single SDE without considering parameter variations or extrinsic noise sources. In Fig 1, we provide an overview of the specific applications that we study in this work.

Our approach employs neural networks as SDE approximators in conjunction with the `torchsde` package [33,34] for reconstructing noisy dynamics from data. Previous work showed that for SDE reconstruction tasks, the  $W_2$  distance nSDE reconstruction method outperforms other benchmark methods such as generative adversarial networks [32,35]. Compared to other probabilistic metrics such as the KL divergence, the Wasserstein distance better incorporates the metric structure of the underlying space. This geometric property makes the Wasserstein distance particularly suitable for trajectory and image data on high-dimensional manifolds, where the supports of different distributions do not always overlap [36]. Additionally, the  $W_2$ -distance-based nSDE reconstruction method can directly extract the underlying SDE from temporal trajectories without requiring specific mathematical forms of the terms in the underlying SDE model. We apply our END-nSDE methodology to three biological processes: (i) circadian clocks, (ii) RPA-DNA binding dynamics, and (iii) NF $\kappa$ B signaling to illustrate the effectiveness of the END-nSDE method in predicting how extrinsic noise modulates stochastic dynamics with intrinsic noise. Additionally, our method demonstrates superior performance compared to several time-series modeling methods including recurrent neural networks (RNNs), long short-term memory networks (LSTMs), and Gaussian processes. In summary, the reconstruction method we propose provides a useful surrogate modeling approach for complex biophysical and biochemical processes, especially in scenarios where high-fidelity mechanistic models are impractical.



**Fig 1. Workflow of our proposed END-nSDE prediction on parameters altering stochastic dynamics.** A. Workflow for training and testing of the extrinsic-noise-driven neural SDE (END-nSDE). Predicted trajectories are simulated (see B) using a range of model parameters (see Sect 2.2) before splitting into training and testing sets (see Fig E in S1 Text for details on the splitting strategy). Model parameters and state variables serve as inputs to a neural network that reconstructs drift and diffusion terms (see C). Network weights are optimized by minimizing the Wasserstein distance (Eq 8) between the training set and predicted trajectories. B. Predicted trajectories are generated by the reconstructed SDE  $d\hat{X} = \hat{f}(\hat{X}; \omega)dt + \hat{\sigma}(\hat{X}; \omega)dB_t$ . C. The drift and diffusion functions,  $\hat{f}$  and  $\hat{\sigma}$ , are

approximated using parameterized neural networks. The parameterized neural-network-based drift function  $\hat{f}(\hat{X}; \omega)$  and diffusion function  $\hat{\sigma}(\hat{X}; \omega)$  take the system state  $\hat{X}$  and biological parameters  $\omega$  as inputs. D. Table of three examples illustrating the nSDE input, along with training and testing datasets. For the last, NFκB example, a more detailed workflow for validation on experimental datasets is illustrated in Fig 8.

<https://doi.org/10.1371/journal.pcbi.1013462.g001>

## 2. Methods and models

In this work, we extend the temporally decoupled squared  $W_2$ -distance SDE reconstruction method proposed in Refs. [32,37] to reconstruct noisy dynamics across a heterogeneous cell population (“extrinsic noise”). Our goal is to not only reconstruct SDEs for approximating noisy cellular signaling dynamics from time-series experimental data, but to also quantify how heterogeneous biological parameters, such as enzyme- or kinase-mediated biochemical reaction rates, affect such noisy cellular signaling dynamics.

### 2.1. SDE reconstruction with heterogeneities in biological parameters

The  $W_2$ -distance-based neural SDE reconstruction method proposed in [32] aims to approximate the SDE

$$dX(t) = f(X(t), t)dt + \sigma(X(t), t)dB(t), X(t) \in \mathbb{R}^d, \tag{1}$$

using an approximated SDE

$$d\hat{X}(t) = \hat{f}(\hat{X}(t), t)dt + \hat{\sigma}(\hat{X}(t), t)dB(t), \hat{X}(t) \in \mathbb{R}^d, \tag{2}$$

where  $\hat{f}$  and  $\hat{\sigma}$  are two parameterized neural networks that approximate the drift and diffusion functions  $f$  and  $\sigma$  in Eq (1), respectively. These two neural networks are trained by minimizing a temporally decoupled squared  $W_2$ -distance loss function

$$\tilde{W}_2^2(\mu, \hat{\mu}) = \int_0^T \inf_{\pi \in \Pi(\mu(t), \hat{\mu}(t))} \mathbb{E}_\pi \left[ \|X(t) - \hat{X}(t)\|^2 \right] dt, \tag{3}$$

where  $\Pi(\mu(t), \hat{\mu}(t))$  denotes the set of all coupling distributions  $\pi$  of two distributions  $\mu(t), \hat{\mu}(t)$  on the probability space  $\mathbb{R}^d$ , and  $X(t)$  and  $\hat{X}(t)$  are the observed trajectories at time  $t$  and trajectories generated by the approximate SDE model Eq (2) at time  $t$ , respectively.  $\mu$  and  $\hat{\mu}$  are the probability distributions associated with the stochastic processes  $\{X(t)\}, 0 \leq t \leq T$  and  $\{\hat{X}(t)\}, 0 \leq t \leq T$ , respectively, while  $\mu(t)$  and  $\hat{\mu}(t)$  are the probability distributions of  $X(t)$  and  $\hat{X}(t)$  at a specific time  $t$ . A coupling  $\pi \in \Pi(\mu(t), \hat{\mu}(t))$  between  $\mu(t)$  and  $\hat{\mu}(t)$  is defined by

$$\pi(A, \mathbb{R}^d) = \mu(A), \pi(\mathbb{R}^d, B) = \hat{\mu}(B), \forall A, B \in \mathcal{B}(\mathbb{R}^d), \tag{4}$$

where  $\mathcal{B}(\mathbb{R}^d)$  is the Borel  $\sigma$ -algebra on  $\mathbb{R}^d$  and  $\mathbb{E}_\pi \left[ \|X(t) - \hat{X}(t)\|^2 \right]$  represents the expectation when  $(X, \hat{X}) \sim \pi$ .

The  $\tilde{W}_2^2$  term in Eq (3) is denoted as the temporally decoupled squared  $W_2$  distance loss function. For simplicity, in this paper, we shall also denote Eq (3) as the squared  $W_2$  loss. The infimum is taken over all possible coupling distributions  $\pi \in \Pi(\mu(t), \hat{\mu}(t))$  and  $\|\cdot\|$  denotes the  $\ell^2$  norm of a vector. That is,

$$\|X(t)\|^2 := \sum_{i=1}^d |X_i(t)|^2. \tag{5}$$

Across different cells, extrinsic noise or cellular heterogeneities such as differences in kinase or enzyme abundances resulting from cellular variabilities, can lead to variable, cell-specific, gene regulatory dynamics. Such heterogeneous and stochastic gene expression (both intrinsic and extrinsic noise) can be modeled using SDEs with distributions of parameter values reflecting cellular heterogeneity. To address heterogeneities in gene dynamics across different cells, we propose an END-nSDE method that is able to reconstruct a family of SDEs for the same gene expression process under different parameters. Specifically, for a given set of (biological) parameters  $\omega$ , we are interested in reconstructing

$$dX(t; \omega) = f(X(t; \omega); \omega)dt + \sigma(X(t; \omega); \omega)dB(t), \quad (6)$$

using the approximate SDE

$$d\hat{X}(t; \omega) = \hat{f}(\hat{X}(t; \omega); \omega)dt + \hat{\sigma}(\hat{X}(t; \omega); \omega)d\hat{B}(t), \quad (7)$$

in the sense that the errors  $f(X(t; \omega); \omega) - \hat{f}(\hat{X}(t; \omega); \omega)$  and  $\sigma(X(t; \omega); \omega) - \hat{\sigma}(\hat{X}(t; \omega); \omega)$  for all different values of  $\omega$  will be minimized. In Eq (7),  $\hat{f}$  and  $\hat{\sigma}$  are represented by two parameterized neural networks that take both the state variable  $\hat{X}$  and the parameters  $\omega$  as inputs. To train these two neural networks, we propose an extrinsic-noise-driven temporally decoupled squared  $W_2$  distance loss function

$$L(\Lambda) = \sum_{\omega \in \Lambda} \tilde{W}_2^2(\mu(\omega), \hat{\mu}(\omega)), \quad (8)$$

where  $\mu(\omega)$  and  $\hat{\mu}(\omega)$  are the distributions of the trajectories  $X(t; \omega)$ ,  $0 \leq t \leq T$  and  $\hat{X}(t; \omega)$ ,  $0 \leq t \leq T$ , and  $\tilde{W}$  is the temporally decoupled squared  $W_2$  loss function in Eq (3).  $\Lambda$  denotes the set of parameters  $\omega$ . Eq (8) is different from the local squared  $W_2$  loss in Refs. [38,39] since we do not require a continuous dependence of  $\{X(t; \omega)\}_{t \in [0, T]}$  on the parameter  $\omega$  nor do we require that  $\omega$  is a continuous variable. The extrinsic-noise-driven temporally decoupled squared  $W_2$  loss function Eq (8) takes into account both parameter heterogeneity and intrinsic fluctuations as a result of the Wiener processes  $B(t)$  and  $\hat{B}(t)$  in Eqs (1) and (2).

Our END-nSDE method is outlined in Fig 1A–1C. With observed noisy single-cell dynamic trajectories as the training data, we train two parameterized neural networks [40] by minimizing Eq (8) to approximate the drift and diffusion terms in the SDE. The reconstructed nSDE is a surrogate model of single-cell dynamics (see Fig 1B and 1C). The hyperparameters and settings for training the neural SDE model are summarized in Table A in S1 Text. Through the examples outlined in Fig 1D, we will show that our  $W_2$ -distance-based method can yield very small errors in the reconstructed drift and diffusion functions  $f - \hat{f}$  and  $\sigma - \hat{\sigma}$ .

## 2.2. Biological models

We consider three biological examples where stochastic dynamics play a critical role and use our END-nSDE method to reconstruct noisy single-cell gene expression dynamics under both intrinsic and extrinsic noise (also summarized in Fig 1D). In these applications, we investigate the extent to which the END-nSDE can efficiently capture and infer changes in the dynamics driven by extrinsic noise.

**2.2.1. Noisy oscillatory circadian clock model.** Circadian clocks, often with a typical period of approximately 24 hours, are ubiquitous in intrinsically noisy biological rhythms generated at the single-cell molecular level [41].

**Algorithm 1 END-nSDE training and prediction framework.**

Obtain training trajectories  $\{X(t;\omega)\}$  (simulated or experimental time-series data). Maximum training epochs =  $i_{\max}$ . Preprocess the relevant training trajectories by grouping them according to different biophysical parameters  $\omega$ .

**Phase 1: Training**

**for**  $i \leq i_{\max}$  **do**

    Input the initial state  $X_0$  and  $\omega \in \Lambda$  into the END-nSDE to generate new predictions  $\hat{X}(t;\omega)$ .

    Calculate the loss function  $L(\Lambda)$  in Eq (8) and perform gradient descent to train the END-nSDE model.

**end for**

**return** the trained END-nSDE model

**Phase 2: Prediction**

Input initial condition  $X_0$  and corresponding noise parameters  $\omega$  from testing data into the trained END-nSDE model.

Generate predicted trajectories  $\hat{X}(t;\omega)$  from the learned model.

We consider a minimal SDE model of the periodic gene dynamics responsible for *per* gene expression which is critical in the circadian cycle. Since *per* gene expression is subject to intrinsic noise [42], we describe it using a linear damped-oscillator SDE

$$\begin{aligned} dx &= -\alpha x dt - \beta y dt + \xi_{x1} dB_{1,t} + \xi_{x2} dB_{2,t} \\ dy &= \beta x dt - \alpha y dt + \xi_{y1} dB_{1,t} + \xi_{y2} dB_{2,t}, \end{aligned} \quad (9)$$

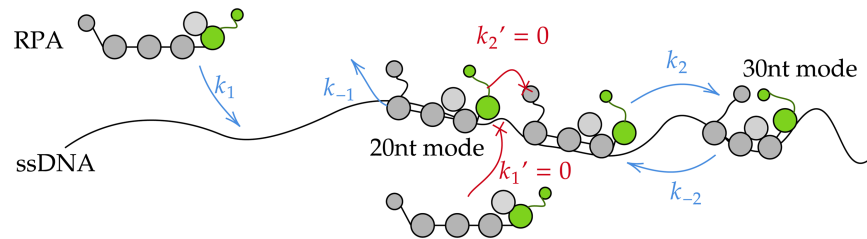
where  $x$  and  $y$  are the dimensionless concentrations of the *per* mRNA transcript and the corresponding *per* protein, respectively.  $dB_{1,t}$ ,  $dB_{2,t}$  are two independent Wiener processes and the parameters  $\alpha > 0$  and  $\beta > 0$  denote the damping rate and angular frequency, respectively. A stability analysis at the steady state  $(x, y) = (0, 0)$  in the noise-free case ( $\xi_x = \xi_y = 0$  in Eq (9)) reveals that the real parts of the eigenvalues of the Jacobian matrix  $\begin{pmatrix} -\alpha & -\beta \\ \beta & -\alpha \end{pmatrix}$  at  $(x, y) = (0, 0)$  are all negative, indicating that the origin is a stable steady state when the system is noise-free. Noise prevents the state  $(x(t), y(t))$  from remaining at  $(0, 0)$ ; thus, fluctuations in the single-cell circadian rhythm are noise-induced [42].

To showcase the effectiveness of our proposed END-nSDE method, we take different forms of the diffusion functions  $\xi_x$  and  $\xi_y$  in Eq (9), accompanied by different values of noise strength and the correlation between the diffusion functions in the dynamics of  $x, y$ .

**2.2.2. RPA-DNA binding model.** Regulation of gene expression relies on complex interactions between proteins and DNA, often described by the kinetics of binding and dissociation. Replication protein A (RPA) plays a pivotal role in various DNA metabolic pathways, including DNA replication and repair, through its dynamic binding with single-stranded DNA (ssDNA) [43–46]. By modulating the accessibility of ssDNA, RPA regulates multiple biological mechanisms and functions, acting as a critical regulator within the cell [47]. Understanding the dynamics of RPA-ssDNA binding is therefore a research area of considerable biological interest and significance.

Multiple binding modes and volume exclusion effects complicate the modeling of RPA-ssDNA dynamics. The RPA first binds to ssDNA in 20 nucleotide (nt) mode, which occupies 20nt of the ssDNA. When the subsequent 10nt of ssDNA is free, 20nt-mode RPA can transform to 30nt-mode, further stabilizing its binding to ssDNA, as illustrated in Fig 2. Occupied ssDNA is not available for other proteins to bind. Consequently, the gap size between adjacent ssDNA-bound RPAs determines the ssDNA accessibility to other proteins.

Mean-field mass-action type chemical kinetic ODE models cannot describe the process very well because they do not capture the intrinsic stochasticity. A stochastic model that



**Fig 2. A continuous-time discrete Markov chain model for multiple RPA molecules binding to long ssDNA.** The possible steps in the biomolecular kinetics of multiple RPA molecules binding to ssDNA. The RPA in the free solution can bind to ssDNA with rate  $k_1$  provided there are at least 20 nucleotides (nt) of consecutive unoccupied sites. This bound “20nt mode” RPA unbinds with rate  $k_{-1}$ . When space permits, the 20nt-mode RPA can extend and bind an additional 10nt of DNA at a rate of  $k_2$ , converting it to a 30nt-mode bound protein. The 30nt-mode RPA transforms back to 20nt-mode spontaneously with the rate  $k_{-2}$ . However, when the gap is not large enough to accommodate the RPA, the binding or conversion is prohibited ( $k'_1 = 0$  and  $k'_2 = 0$ ).

<https://doi.org/10.1371/journal.pcbi.1013462.g002>

tracks the fraction of two different binding modes of RPA, 20nt-mode ( $x_1$ ) and 30nt-mode ( $x_2$ ), has been developed to capture the dynamics of this process. A brute-force approach using forward stochastic simulation algorithms (SSAs) [48] was then used to fit the model to experimental data [47]. However, a key challenge in this approach is that the model is nondifferentiable with respect to the kinetic parameters, making it difficult to estimate parameters. Yet, simple spatially homogeneous stochastic chemical reaction systems can be well approximated by a corresponding SDE of the form given in Eq (1) when the variables are properly scaled in the large system size limit [49]. While interparticle interactions shown in Fig 2 make it difficult to find a closed-form SDE approximation, results from [49] motivate the possibility of an SDE approximation for the RPA-ssDNA binding model in terms of the variables  $x_1$  and  $x_2$ .

Here, to address the non-differentiability issue associated with the underlying Markov process, we use our END-nSDE model to construct a differentiable surrogate for SSAs, allowing it to be readily trained from data. Further details on the models and data used in this study are provided in Appendix B of S1 Text. Throughout our analysis of RPA-DNA binding dynamics, we benchmark the SDE reconstructed by our extended  $W_2$ -distance approach against those found using other time series analysis and reconstruction methods such as the Gaussian process, RNN, LSTM, and the neural ODE model. We show that our surrogate SDE model is most suitable for approximating the RPA-DNA binding process because it can capture the intrinsic stochasticity in the dynamics.

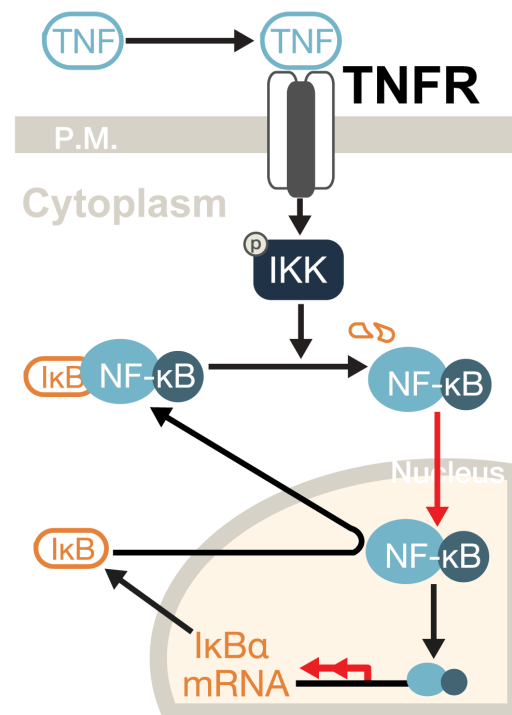
**2.2.3. NF $\kappa$ B signaling model.** Macrophages can sense environmental information and respond accordingly with stimulus-response specificity encoded in signaling pathways and decoded by downstream gene expression profiles [50]. The temporal dynamics of NF $\kappa$ B, a key transcription factor in immune response and inflammation, encodes stimulus information [51]. NF $\kappa$ B targets and regulates vast immune-related genes [52–54]. While NF $\kappa$ B signaling dynamics are stimulus-specific, they exhibit significant heterogeneity across individual cells under identical conditions [51]. Understanding how specific cellular heterogeneity (extrinsic noise) contributes to heterogeneity in NF $\kappa$ B signaling dynamics can provide insight into how noise affects the fidelity of signal transduction in immune cells.

A previous modeling approach employs a 52-dimensional ODE system to quantify the NF $\kappa$ B signaling network [51] and recapitulate the signaling dynamics of a representative cell. This ODE model includes 52 molecular entities and 47 reactions across a TNF-receptor module, an adaptor module, and a core module with and NF $\kappa$ B- $\text{IKK-}\text{I}\kappa\text{B}\alpha$  ( $\text{I}\kappa\text{B}\alpha$  is an inhibitor

of NF $\kappa$ B, while IKK is the I $\kappa$ B kinase complex that regulates the I $\kappa$ B $\alpha$  degradation) feedback loop (see Fig 3) [55]. However, such an ODE model is deterministic and assumes no intrinsic fluctuations in the biomolecular processes. Yet, from experimental data, the NF $\kappa$ B signaling dynamics fluctuate strongly; such fluctuations cannot be quantitatively described by any deterministic ODE model. Due to the system's high dimensionality and nonlinearity, it is challenging to quantify how intrinsic noise influences temporal coding in NF $\kappa$ B dynamics.

To incorporate the intrinsic noise within the NF $\kappa$ B signaling network, we introduce noise terms into the 52-dimensional ODE system to build an SDE that can account for the observed temporally fluctuating nuclear NF $\kappa$ B trajectories. While NF $\kappa$ B signaling pathways involve many variables, experimental constraints limit the number of measurable components. Among these, nuclear NF $\kappa$ B concentration is the most direct and critical experimental read-out. As a minimal stochastic model, we hypothesize that only the biophysical and biochemical processes of NF $\kappa$ B translocation (which directly affects experimental measurements) and I $\kappa$ B $\alpha$  transcription (a key regulator of NF $\kappa$ B translocation) are subject to Brownian-type noise (red arrows in Fig 3), as these processes play crucial roles in the oscillatory dynamics of NF $\kappa$ B [55].

The intensity of Brownian-type noise in the NF $\kappa$ B dynamics may depend on factors such as cell volume (smaller volumes result in higher noise intensity), or copy number (lower copy numbers lead to greater noise intensity), and is therefore considered a form of extrinsic noise. Noise intensity parameters thus capture an aspect of cellular heterogeneity. There are other sources of cellular heterogeneity, such as variations in kinase or enzyme abundances, which



**Fig 3. Simplified schematic of the NF $\kappa$ B Signaling Network.** TNF binds its receptor, activating IKK, which degrades I $\kappa$ B $\alpha$  and releases NF $\kappa$ B. The free NF $\kappa$ B translocates to the nucleus and promotes I $\kappa$ B $\alpha$  transcription. Newly synthesized I $\kappa$ B $\alpha$  then binds NF $\kappa$ B and exports it back to the cytoplasm. Red arrows indicate noise that we consider in the corresponding SDE system.

<https://doi.org/10.1371/journal.pcbi.1013462.g003>

are too complicated to model and are thus not included in the current model. For simplicity, all kinetic parameters, except for the noise intensity ( $\sigma$ ), are assumed to be consistent with those of a representative cell [55]. The 52-dimensional ODE model for describing NF $\kappa$ B dynamics is given in Refs. [51,56]. We extend this model by adding noise to the dynamics of the sixth, ninth, and tenth ODEs of the 52-dimensional ODE model. We retain 49 ODEs but convert the equations for the sixth, ninth, and tenth components to SDEs:

$$\begin{aligned} du_6 &= \left( k_{\text{basal}} + \frac{k_{\text{max}} u_{52}^{n_{\text{NF}\kappa\text{B}}}}{u_{52}^{n_{\text{NF}\kappa\text{B}}} + K_{\text{NF}\kappa\text{B}}^{n_{\text{NF}\kappa\text{B}}}} - k_{\text{deg}} u_6 \right) dt + \sigma_1 dB_{1,t} \\ du_9 &= \left( k_{\text{imp}} u_9 - k_{\text{a-I}\kappa\text{B-NF}\kappa\text{B}} u_2 u_9 - k_{\text{deg-NF}\kappa\text{B}} u_9 + v^{-1} k_{\text{exp}} u_{10} + k_{\text{d-I}\kappa\text{B-NF}\kappa\text{B}} u_4 + k_{\text{phos}} u_7 \right) dt - \sigma_2 dB_{2,t} \\ du_{10} &= \left( -k_{\text{exp}} u_{10} - k_{\text{a-I}\kappa\text{B-NF}\kappa\text{B}} u_3 u_{10} + \nu k_{\text{imp}} u_9 + k_{\text{d-I}\kappa\text{B-NF}\kappa\text{B}} u_5 \right) dt + \sigma_2 dB_{2,t}. \end{aligned} \tag{10}$$

In Eqs 10,  $u_2$  is the concentration of I $\kappa$ B $\alpha$  in the cytoplasm;  $u_3$  is the concentration of I $\kappa$ B $\alpha$  in the nucleus;  $u_4$  is the concentration of the I $\kappa$ B $\alpha$ -NF $\kappa$ B complex;  $u_5$  is the concentration of the I $\kappa$ B $\alpha$ -NF $\kappa$ B complex in the nucleus;  $u_6$  is the mRNA of I $\kappa$ B $\alpha$ ;  $u_7$  is the IKK-I $\kappa$ B $\alpha$ -NF $\kappa$ B complex;  $u_9$  is NF $\kappa$ B;  $u_{10}$  represents nuclear NF $\kappa$ B concentration; and  $u_{52}$  is the nuclear concentration of NF $\kappa$ B with RNA polymerase II that is ready to initiate mRNA transcription. A description of the parameters and their typical values are given in Table C in S1 Text. The quantities  $\sigma_1 dB_{1,t}$  and  $\sigma_2 dB_{2,t}$  are noise terms associated with I $\kappa$ B $\alpha$  transcription and NF $\kappa$ B translocation, respectively. The remaining variables are latent variables and their dynamics are regulated via the remaining 49-dimensional ODE in Refs. [51,56]. The activation of NF $\kappa$ B is quantified by the total nuclear NF $\kappa$ B concentration ( $u_5 + u_{10}$ ), which is also measured in experiments.

Within this example, we wish to determine if our proposed parameter-associated nSDE can accurately reconstruct the dynamics underlying experimentally observed NF $\kappa$ B trajectory data.

### 3. Results

#### 3.1. Accurate reconstruction of circadian clock dynamics

As an illustrative example, we use the  $W_2$ -distance nSDE reconstruction method to first reconstruct the minimal model for damped oscillatory circadian dynamics (see Eq (9)) under different forms of the diffusion function. We set the two parameters  $\alpha = 0.19$  and  $\beta = 0.21$  in Eq (9) and impose three different forms for the diffusion functions  $\xi_{x1}, \xi_{x2}, \xi_{y1}, \xi_{y2}$ : a constant diffusion function [57], a Langevin [58] diffusion function, and a linear diffusion function [59]. These functions, often used to describe fluctuating biophysical processes, are

$$\text{const: } \begin{bmatrix} \xi_{x1} & \xi_{x2} \\ \xi_{y1} & \xi_{y2} \end{bmatrix} = \sigma_0 \begin{bmatrix} 1 & c \\ c & 1 \end{bmatrix}, \tag{11}$$

$$\text{Langevin: } \begin{bmatrix} \xi_{x1} & \xi_{x2} \\ \xi_{y1} & \xi_{y2} \end{bmatrix} = \sigma_0 \begin{bmatrix} \sqrt{|x|} & c\sqrt{|y|} \\ c\sqrt{|x|} & \sqrt{|y|} \end{bmatrix}, \tag{12}$$

and

$$\text{linear: } \begin{bmatrix} \xi_{x1} & \xi_{x2} \\ \xi_{y1} & \xi_{y2} \end{bmatrix} = \sigma_0 \begin{bmatrix} x & c|y| \\ c|x| & y \end{bmatrix}. \tag{13}$$

There are two additional parameters in Eqs (11), (12), and (13):  $\sigma_0$  that determines the intensity of the Brownian-type fluctuations and  $c$  that controls the correlation of fluctuations

between the two dimensions. For each type of diffusion function, we trained a different nSDE model, each of which takes the state variables  $(x, y)$  and the two parameters  $(c, \sigma_0)$  as inputs and which outputs the values of the reconstructed drift and diffusion functions.

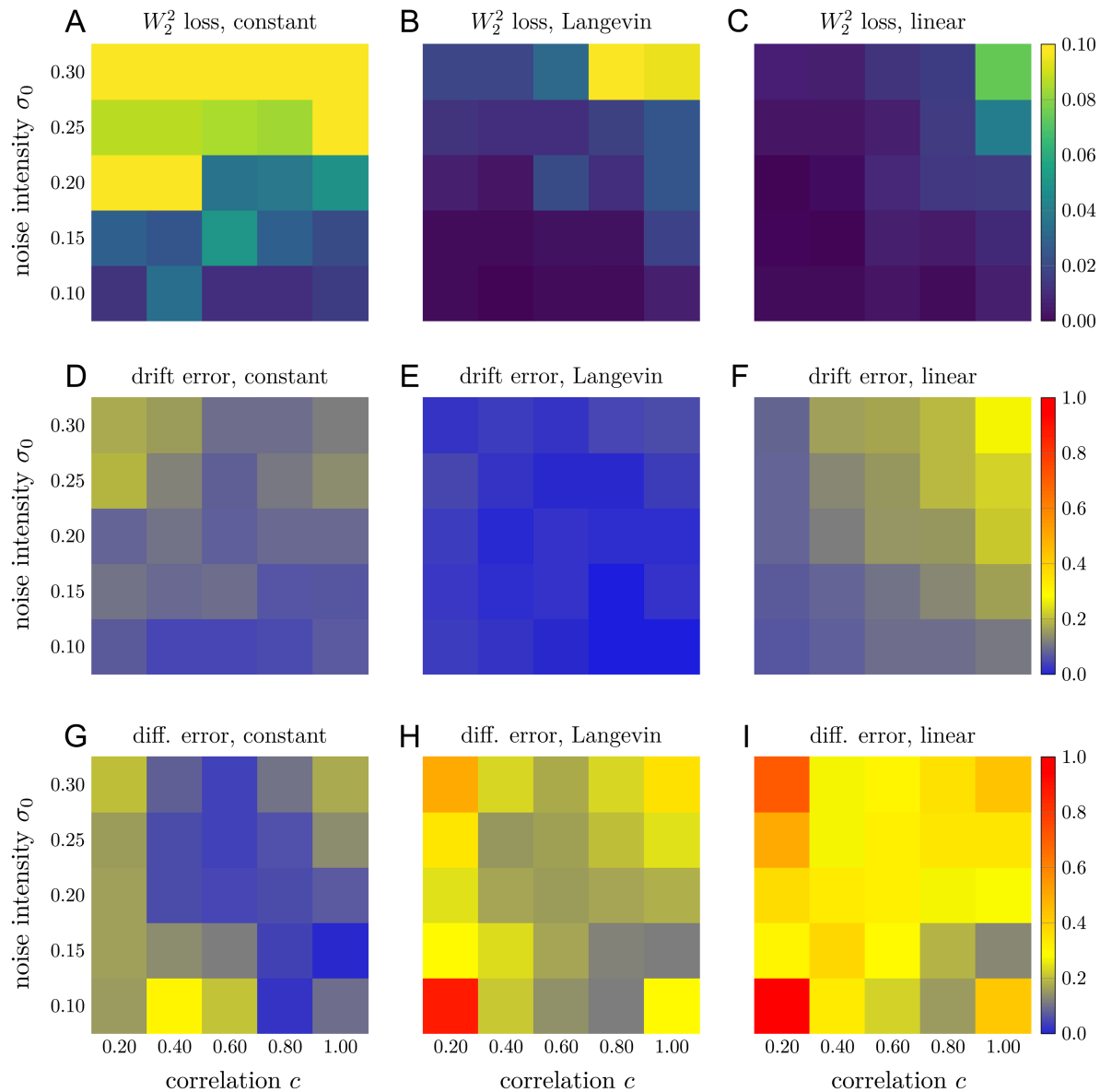
We take 25 combinations of  $(\sigma_0, c) \in \{(0.1 + 0.05i, 0.2 + 0.2j), i \in \{0, \dots, 4\}, j \in \{0, \dots, 4\}\}$ ; for each combination of  $(\xi_{x1}, \xi_{x2}, \xi_{y1}, \xi_{y2})$ , we generate 50 trajectories from the ground truth SDE (9) as the training data with  $t \in [0, 1]$ . The initial condition is set as  $(x(0), y(0)) = (0, 1)$ . To test the accuracy of the reconstructed diffusion and drift functions, we measure the following relative errors:

$$\text{Error in } \mathbf{f} := \frac{\sum_{i=1}^M \sum_{j=0}^T |f(\mathbf{X}_i(t_j; \omega); \omega) - \hat{f}(\hat{\mathbf{X}}_i(t_j; \omega); \omega)|_1}{\sum_{j=0}^T |f(\mathbf{X}_i(t_j; \omega); \omega)|_1}, \tag{14}$$

$$\text{Error in } \boldsymbol{\sigma} := \frac{\sum_{i=1}^M \sum_{j=0}^T \left\| \boldsymbol{\sigma}(\mathbf{X}_i(t_j; \omega); \omega) \boldsymbol{\sigma}^T(\mathbf{X}_i(t_j; \omega); \omega) - \hat{\boldsymbol{\sigma}}(\mathbf{X}_i(t_j; \omega); \omega) \hat{\boldsymbol{\sigma}}^T(\mathbf{X}_i(t_j; \omega); \omega) \right\|_m}{\sum_{i=0}^M \sum_{j=0}^d \left\| \boldsymbol{\sigma}(\mathbf{X}_i(t_j; \omega); \omega) \boldsymbol{\sigma}^T(\mathbf{X}_i(t_j; \omega); \omega) \right\|_m}. \tag{15}$$

Here,  $\mathbf{f} := (-\alpha x - \beta y, \beta x - \alpha y)^T$  is the vector of ground truth drift functions and  $\hat{\mathbf{f}}$  is the reconstructed drift function.  $\boldsymbol{\sigma}$  is the matrix of ground truth diffusion functions  $[\xi_{x1}, \xi_{x2}; \xi_{y1}, \xi_{y2}]$  given in Eqs (11), (12), and (13).  $M$  is the number of training samples,  $|\cdot|_1$  denotes the  $\ell^1$  norm of a vector, and the matrix norm  $|A|_m := \sum_{i=1}^m \sum_{j=1}^n |A_{ij}|$  for a matrix  $A \in \mathbb{R}^{m \times n}$ . The errors are measured separately for different parameters  $\omega := (\sigma_0, c)$ .

The errors in the reconstructed drift function  $\hat{\mathbf{f}}$  and diffusion function  $\hat{\boldsymbol{\sigma}}$  as well as the temporally decoupled squared  $W_2$  loss Eq (3) associated with different forms of the diffusion function and different values of  $(\sigma_0, c)$  are shown in Fig 4. When the diffusion function is a constant Eq (11), the mean reconstruction error of the drift function is 0.15, the mean reconstruction error of the diffusion function is 0.16, and the mean temporally decoupled squared  $W_2$  loss between the ground truth trajectories and the predicted trajectories is 0.074 (averaged over all sets of parameters  $(\sigma_0, c)$ ). When a Langevin-type diffusion function Eq (12) is used as the ground truth, the mean errors for the reconstructed drift and diffusion functions are 0.069 and 0.29, respectively, and the mean temporally decoupled squared  $W_2$  loss between the ground truth and predicted trajectories is 0.020. For a linear-type diffusion function as the ground truth, mean reconstruction errors of the drift and diffusion functions are 0.19 and 0.41, respectively, and the mean temporally decoupled squared  $W_2$  distance is 0.013. For all three forms of diffusion, our END-nSDE method can accurately reconstruct the drift function  $(-\alpha x - \beta y, \beta x - \alpha y)$  (see Fig 4D–4F). When the diffusion function is a constant, our END-nSDE model can also accurately reconstruct this constant (see Fig 4G). When the diffusion function takes a more complicated form such as the Langevin-type diffusion function Eq (12) or the linear-type diffusion function Eq (13), the reconstructed nSDE model can still approximate the diffusion function well for most combinations of  $(\sigma_0, c)$ , especially when the correlation  $c > 0.2$  (see Fig 4H–4I). Overall, our proposed END-nSDE model can accurately reconstruct the minimal stochastic circadian dynamical model Eq (9) in the presence of extrinsic noise (different values of  $(\sigma_0, c)$ ); the accuracy of the reconstructed drift and diffusion functions is maintained for most combinations of  $(\sigma_0, c)$ . While the drift function is reconstructed with high accuracy, the reconstructed diffusion function exhibits larger relative errors, particularly for models with more complex diffusion forms. How errors depend on the functional forms of the diffusion should be investigated.



**Fig 4. Reconstructing the circadian model using END-nSDE.** Temporally decoupled squared  $W_2$  losses Eq (3) and errors in the reconstructed drift and diffusion functions for different types of the diffusion function and different values of  $(\sigma_0, c)$ . A-C. The temporally decoupled squared  $W_2$  loss between the ground truth trajectories and the trajectories generated by the reconstructed nSDEs for the constant-type diffusion function Eq (11), Langevin-type diffusion function Eq (12), and the linear-type diffusion function Eq (13). D-F. Errors in the reconstructed drift function for the three different types of ground truth diffusion functions and the linear-type diffusion function Eq (13). G-I. Errors in the reconstructed diffusion function for the three different types of ground truth diffusion functions.

<https://doi.org/10.1371/journal.pcbi.1013462.g004>

To investigate how the strengths of the extrinsic and intrinsic noise and affect our reconstruction of extrinsic-noise-driven SDEs, we conduct an additional test on the reconstruction of circadian clock dynamics. We generate training trajectories from a revised version of Eq (9):

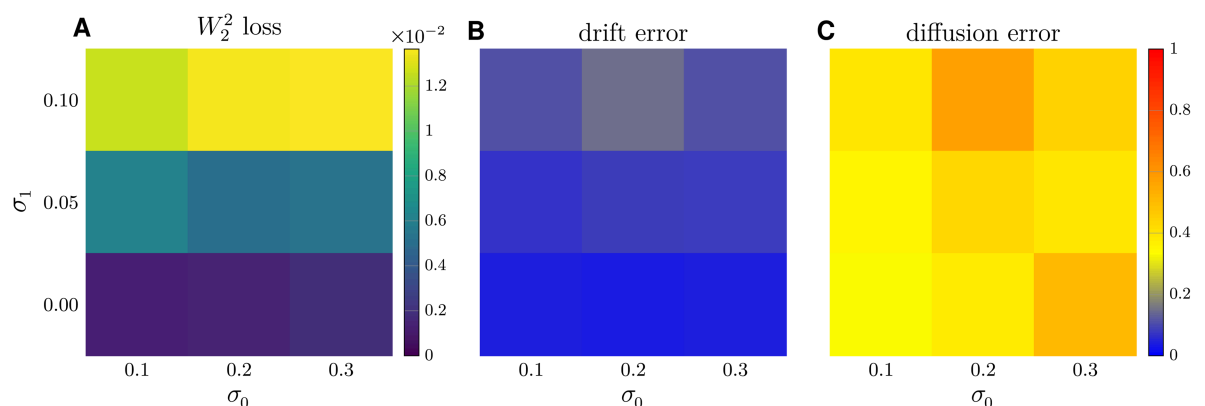
$$\begin{aligned} dx &= -\alpha xdt - \beta ydt + (\sigma_0 + \sigma_1 k_1) \sqrt{x} dB_{1,t} \\ dy &= \beta xdt - \alpha ydt + (\sigma_0 + \sigma_1 k_2) \sqrt{y} dB_{2,t}. \end{aligned} \tag{16}$$

In Eq (16), for each set of  $(\sigma_0, \sigma_1)$ , we generate 25 groups of  $(k_1, k_2) \in \{0, \pm 0.5, \pm 1\} \times \{0, \pm 0.5, \pm 1\}$  with each group containing 50 trajectories as training data. To train the neural SDE model, both the state variables  $(x, y)$  and  $(k_1, k_2)$  are input into the neural SDE.  $\sigma_0$  characterizes the average level of intrinsic noise while  $\sigma_1$  represents the strength of extrinsic noise, and we use different values of  $(\sigma_0, \sigma_1)$ . As shown in Fig 5B and 5C, errors in the reconstructed drift function and in the reconstructed diffusion function, averaged over all different sets of  $(k_1, k_2)$ , increases with both  $\sigma_0$  and  $\sigma_1$ . Specifically, an increase in the intrinsic noise level ( $\sigma_0$ ) reduces the reconstruction accuracy more than an increase in the extrinsic noise ( $\sigma_1$ ) does. More analysis on how the variation in intrinsic noise and extrinsic noise could affect the accuracy of the reconstructed drift and diffusion functions using our proposed END-nSDE method is promising.

### 3.2. Accurate approximation of interacting DNA-protein systems with different kinetic parameters

To construct a differentiable surrogate for stochastic simulation algorithms (SSAs), the neural SDE model should be able to take kinetic parameters as additional inputs. Thus, the original  $W_2$ -distance SDE reconstruction method in [32] can no longer be applied because the trained neural SDE model cannot take into account extrinsic noise [60], *i.e.*, different values of kinetic parameters. To be specific, we vary one parameter (the conversion rate  $k_2$  from 20nt-mode RPA to 30nt-mode RPA) in the stochastic model and then apply our END-nSDE method which takes the state variables and the kinetic parameter  $k_2$  as the input. We set  $k_2 \in \{10^{-4+j/10}, j = 0, \dots, 25\}$  with other parameters taken from experiments [47] ( $k_1 = 10^{-3} \text{ s}^{-1}$ ,  $k_{-1} = 10^{-6} \text{ s}^{-1}$ ,  $k_{-2} = 10^{-6} \text{ s}^{-1}$ , see Fig 2). For each  $k_2$ , we generate 100 trajectories and use 50 for the training set and the other 50 for the testing set. Each trajectory encodes the dynamics of the fraction of 20nt-mode DNA-bound RPA  $x_1(t)$  and the fraction of 30nt-mode DNA-bound RPA  $x_2(t)$ .

When approximating the dynamics underlying the RPA-DNA binding process, we compare our SDE reconstruction method with other benchmark time-series analysis or reconstruction approaches, including the RNN, LSTM, Gaussian process, and the neural ODE model [61,62]. These benchmarks are described in detail in Appendix C in S1 Text.



**Fig 5.** Average temporally decoupled squared  $W_2$  losses Eq (3) and errors in the reconstructed drift and diffusion functions for different choices of intrinsic noise strength and extrinsic noise strength  $(\sigma_0, \sigma_1)$  in Eq (16).

<https://doi.org/10.1371/journal.pcbi.1013462.g005>

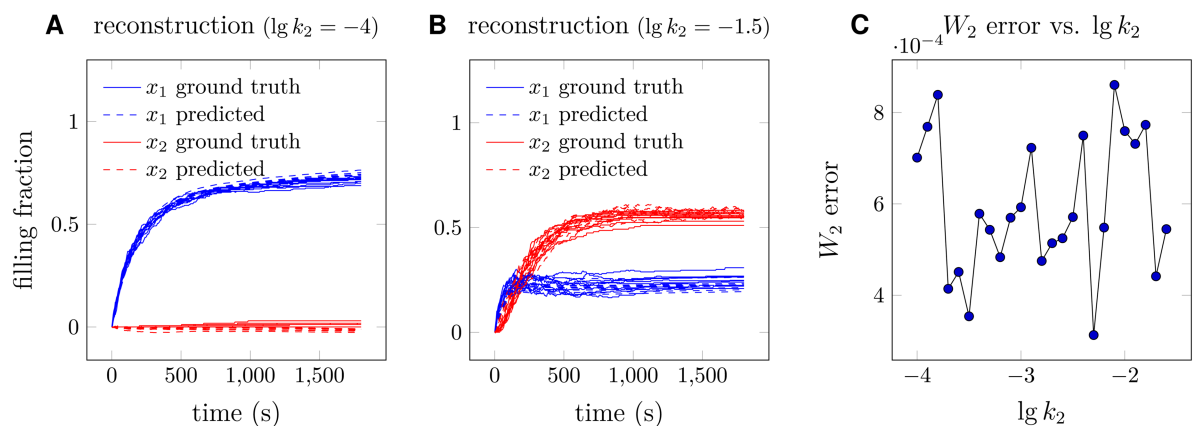
The extrinsic-noise-driven temporally decoupled squared  $W_2$  distance loss Eq (8) between the distribution of the ground truth trajectories and the distribution of the predicted trajectories generated by our END-nSDE reconstructed SDE model is the smallest among all methods (shown in Table 1). The underlying reason is an SDE well approximates the genuine Markov counting process underlying the continuum-limit RPA-DNA binding process [49]. The RNN and LSTM models do not capture the intrinsic fluctuations in the counting process. The neural ODE model is a deterministic model and cannot capture the stochasticity in the RPA-DNA binding dynamics. Additionally, the Gaussian process can only accurately approximate linear SDEs, which is not an appropriate form for an SDE describing the RPA-DNA binding process.

In Fig 6A and 6B, we plot the predicted trajectories obtained by the trained neural SDE model for two different values  $\lg k_2 = -4$  and  $\lg k_2 = -1.5$ . Actually, for all different values of  $k_2$ , trajectories generated by our END-nSDE method match well with the ground truth trajectories on the testing set, as the temporally decoupled squared  $W_2$  loss is maintained small for all  $k_2$  (shown in Fig 6C). This demonstrates the ability of our method to capture the dependence of the stochastic dynamics on biochemical kinetic parameters.

**Table 1.** The extrinsic-noise-driven time-decoupled squared  $W_2$  distance Eq (8) between the ground truth and predicted trajectories generated by different models on the testing set.

Model	Loss
END-nSDE	<b>0.0006</b>
LSTM	0.062
RNN	0.087
nODE	0.0012
Gaussian Process	0.0010

<https://doi.org/10.1371/journal.pcbi.1013462.t001>



**Fig 6. Reconstructed trajectories of the RPA-DNA binding model.** A. Sample ground truth and reconstructed trajectories evaluated at  $\lg k_2 = -4$ , where we use the convention that  $\lg = \log_{10}$ . B. Sample ground truth and reconstructed parameters evaluated at  $\lg k_2 = -1.5$ . C. Temporally decoupled squared  $W_2$  distances (see Eq (8)) between the ground truth and reconstructed trajectories evaluated at different  $\lg k_2$  values. In A and B, blue and red trajectories represent the filling fractions of DNA by 20nt-mode and 30nt-mode RPA, respectively. The dashed lines represent the predicted trajectories, and the solid lines represent the ground truth. Throughout the figure, the data are generated by a single neural SDE model that accepts the conversion rate  $k_2$  as a parameter and outputs the trajectories.

<https://doi.org/10.1371/journal.pcbi.1013462.g006>

### 3.3. Reconstructing high-dimensional NF $\kappa$ B signaling dynamics from simulated and experimental data

Finally, we evaluate the effectiveness of the END-nSDE framework in reconstructing high-dimensional NF $\kappa$ B signaling dynamics under varying noise intensities and investigate the performance of the neural SDE method in reconstructing experimentally measured noisy NF $\kappa$ B dynamics. The procedure is divided into two parts. First, we trained and tested our END-nSDE method on synthetic data generated by the NF $\kappa$ B SDE model Eq (10) under different noise intensities ( $\sigma_1, \sigma_2$ ). Second, we test whether the trained END-nSDE can reproduce the experimental dynamic trajectories.

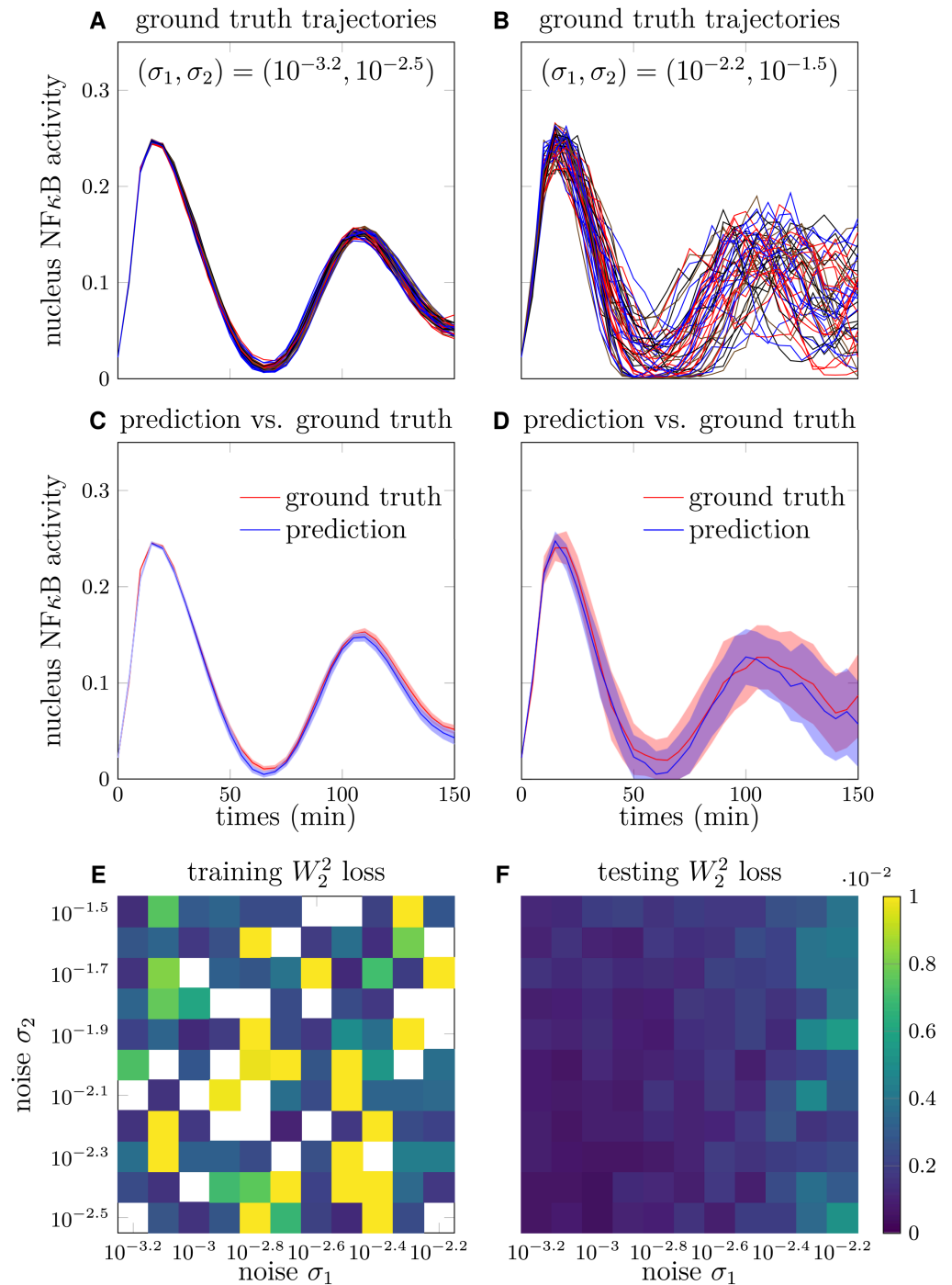
**3.3.1. Reconstructing a 52-dimensional stochastic model for NF $\kappa$ B dynamics.** For training END-nSDE models, we first generated synthetic data from the 52-dimensional SDE model of NF $\kappa$ B signaling dynamics Eqs (10) and established models [51,56]. The synthetic trajectories are generated under 121 combinations of noise intensity ( $\sigma_1, \sigma_2$ ) in Eqs (10) (see Appendix D of S1 Text). The resulting NF $\kappa$ B trajectories vary depending on noise intensity, with low-intensity noise producing more consistent dynamics across cells (see Fig 7A) and higher-intensity noise yielding more heterogeneous dynamics (see Fig 7B). The simulated ground truth trajectories are split into training and testing datasets (see Appendix E in S1 Text for details). Specifically, we excluded 25 combinations of noise intensities ( $\sigma_1, \sigma_2$ ) from the training set in order to test the generalizability of the trained neural SDE model on noisy intensities.

Next, as detailed in Appendix E of S1 Text, we trained a 52-dimensional neural SDE model using our END-nSDE method on synthetic trajectories. The loss function is based on the  $W_2$  distance between the distributions of the neural SDE predictions in Eqs (10) and the simulated nuclear I $\kappa$ B $\alpha$ -NF $\kappa$ B complex and nuclear NF $\kappa$ B activities ( $u_5(t)$  and  $u_{10}(t)$ , respectively) and the corresponding END-nSDE predictions. The remaining 50 variables of the NF $\kappa$ B system were treated as latent variables, as they are not directly included in the loss function calculation.

Although the NF $\kappa$ B dynamics vary under different noise intensities ( $\sigma_1, \sigma_2$ ), the trajectories generated by our trained neural SDE closely align with the ground truth synthetic NF $\kappa$ B dynamics under different noise intensities ( $\sigma_1, \sigma_2$ ) (see Fig 7C and 7D). The neural SDE model demonstrates greater accuracy in reconstructing NF $\kappa$ B dynamics when the noise in I $\kappa$ B $\alpha$  transcription ( $\sigma_1$ ) is smaller, as evidenced by the reduced squared  $W_2$  distance between the predicted and ground-truth trajectories on both the training and validation sets (see Fig 7E and 7F). The temporally decoupled squared  $W_2$  loss Eq (8) on the validation set is close to that on the training set for different values of noise intensities ( $\sigma_1, \sigma_2$ ). The mean squared  $W_2$  distance across all combinations of noise intensities ( $\sigma_1, \sigma_2$ ) is 0.0013 for the training set, and the validation set shows a mean squared  $W_2$  distance of 0.0017.

Since the loss function for this application involves only two variables out of 52, we also tested whether the “full” 52-dimensional NF $\kappa$ B system can be effectively modeled by a two-dimensional neural SDE. After training, we found that the reduced model was insufficient for reconstructing the full 52-dimensional dynamics, as it disregarded the 50 latent variables not included in the loss function (see Fig D in Appendix F of S1 Text). This result underscores the importance of incorporating latent variables from the system, even when they are not explicitly included in the loss function.

**3.3.2. Reproducing NF $\kappa$ B data with a trained END-nSDE.** We assessed whether our proposed END-nSDE can accurately reconstruct the experimentally measured NF $\kappa$ B dynamic

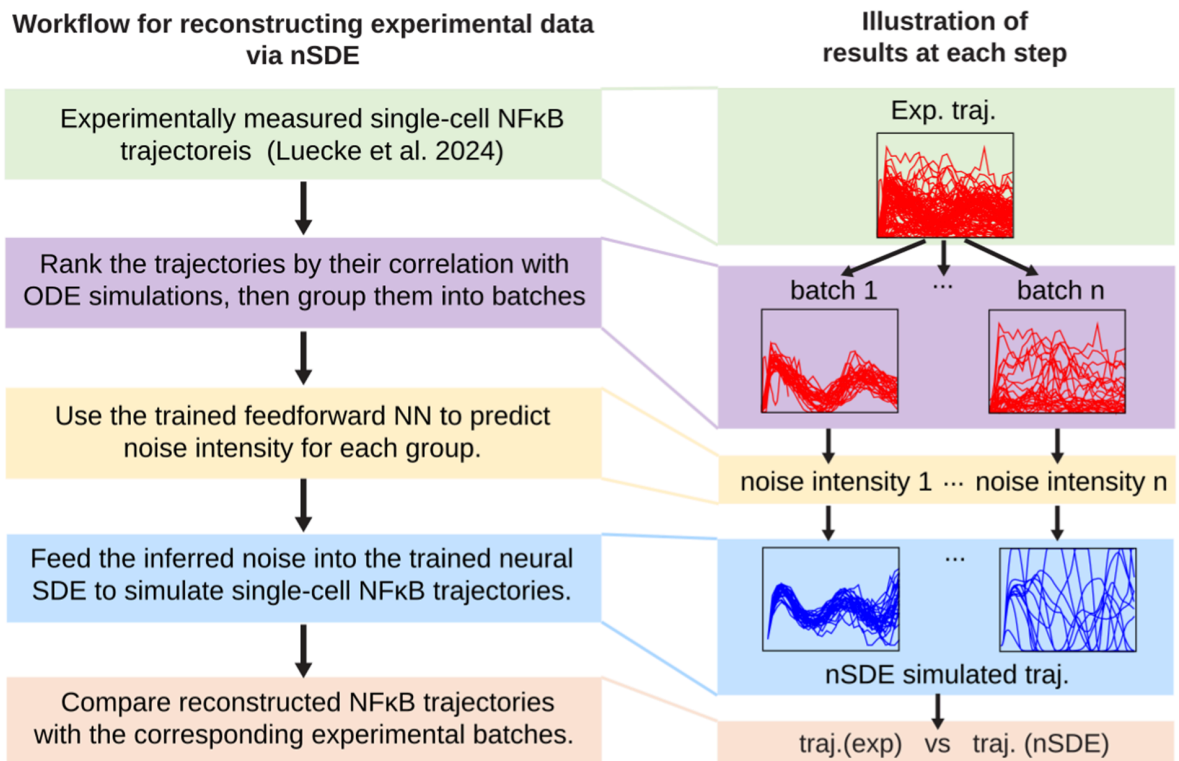


**Fig 7. Reconstruction of NFκB signaling dynamics.** A. Sample trajectories of nuclear NFκB concentration as a function of time with  $\sigma_1 = 10^{-3.2}, \sigma_2 = 10^{-2.5}$ . B. Sample trajectories of nuclear NFκB concentration as a function of time with  $\sigma_1 = 10^{-2.2}, \sigma_2 = 10^{-1.5}$ . C. Reconstructed nuclear NFκB trajectories generated by the trained neural SDE versus the ground truth nuclear NFκB trajectories under noise intensities  $\sigma_1 = 10^{-3.2}, \sigma_2 = 10^{-2.5}$  in Eq (10). D. Reconstructed nuclear NFκB trajectories generated by the trained neural SDE versus the ground truth nuclear NFκB trajectories under noise intensities  $\sigma_1 = 10^{-2.2}, \sigma_2 = 10^{-1.5}$ . E. The squared  $W_2$  distance between the distributions of the predicted trajectories and ground truth trajectories on the training set under different noise strengths ( $\sigma_1, \sigma_2$ ). For training, we randomly selected 50% sample trajectories in 80 combinations of noise strengths ( $\sigma_1, \sigma_2$ ) as the training dataset. Blank cells indicate that the corresponding parameter set is not included in the training set. F. Validation of the trained model by evaluating the squared  $W_2$  distance between the distributions of predicted trajectories and ground truth trajectories on the validation set.

<https://doi.org/10.1371/journal.pcbi.1013462.g007>

trajectories. For simplicity and feasibility, we tested the END-nSDE under the assumption that: (1) all cells share the same drift function, and (2) cells with trajectories deviating similarly from their ODE predictions have the same noise intensities. Based on these assumptions, we developed the following workflow (see Fig 8):

1. We used experimentally measured single-cell trajectories of NF $\kappa$ B concentration, obtained through live-cell image tracking of macrophages from mVenus-tagged RelA mouse with a frame frequency of five minutes [63], yielding a total of 31 consecutive time points. These trajectories correspond to the sum of nuclear I $\kappa$ B $\alpha$ -NF $\kappa$ B and NF $\kappa$ B concentration in the 52D SDE model ( $u_5(t)$  and  $u_{10}(t)$  in Eq (10)).
2. The experimental dataset was divided into subgroups. Cosine similarity was calculated between the ODE-generated trajectory (representative-cell NF $\kappa$ B dynamics) and experimental trajectories. The trajectories are then ranked and divided into different groups based on their cosine similarity with the trajectory generated from the ODE model [64]. Experimental trajectories with higher similarity to the ODE trajectory are expected to exhibit smaller intrinsic fluctuations, corresponding to lower noise intensities (see Appendix G in S1 Text for details).
3. Each group of experimental trajectories was input into the trained neural network (see the next paragraph for more details) to infer the corresponding noise intensities ( $\sigma_1, \sigma_2$ ). For simplicity, we assume that trajectories within each group share the same noise intensities.



**Fig 8. Workflow of reconstructing experimental data via END-nSDE.** Workflow for reconstructing experimental data using the trained parameterized nSDE and the parameter-inference neural network (NN). The boxes on the left outline the steps of the experimental data reconstruction process, while the boxes on the right illustrate the corresponding results at each step.

<https://doi.org/10.1371/journal.pcbi.1013462.g008>

4. The inferred noise is then used as inputs for the trained END-nSDE to simulate NF $\kappa$ B trajectories.
5. The simulated trajectories were compared with the corresponding experimental data to evaluate the model's performance.

To estimate noise intensities from different groups of experimentally measured single-cell nuclear NF $\kappa$ B trajectories (step (3) in the proposed workflow), we trained another neural network to predict the corresponding I $\kappa$ B $\alpha$  transcription and NF $\kappa$ B translocation noise intensities from the groups of NF $\kappa$ B trajectories in the synthetic training data, similar to the approach taken in [65]. The trained neural network can then be used for predicting noise intensities in the validation set (see Appendix H in [S1 Text](#) for technical details).

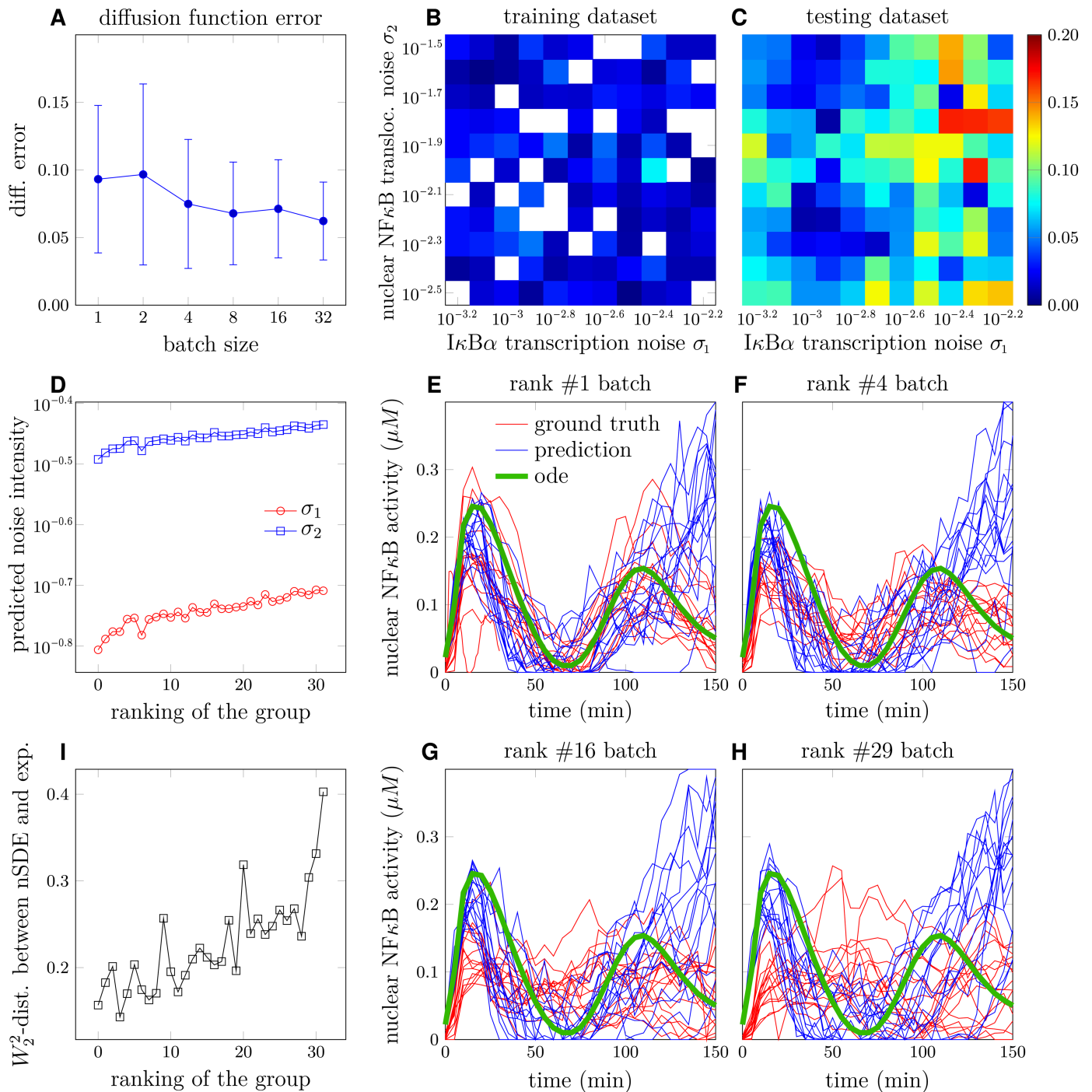
Assessing the impact of group size (number of trajectories) on noise intensity prediction performance, we found that taking a group size of at least two leads to a relative error of around 0.1 (see [Fig 9A](#)). Given the high heterogeneity present in experimental data, we took a group size of 32 as the input into the neural network. Under this group size, the relative errors in the predicted noise intensities were 0.021 on the training set and 0.062 on the testing set (see [Fig 9B](#) and [9C](#)).

Using the trained neural network, we inferred noise intensities for the experimental data, which were grouped based on their cosine similarities with the representative-cell trajectory (deterministic ODE) with a group size of 32. The predicted noise intensities on the experimental dataset are larger than the noise intensities on the training set, possibly because unmodeled extrinsic noise complicates the inference of noise intensity. The transcription noise of I $\kappa$ B $\alpha$  is predicted to be within the range of  $[10^{-0.81}, 10^{-0.71}]$  (see [Fig 9D](#)). In addition, the inferred noise for NF $\kappa$ B translocation fell within  $[10^{-0.49}, 10^{-0.43}]$  (see [Fig 9D](#)). These inferred noise intensities were then used as inputs to the END-nSDE to simulate NF $\kappa$ B trajectories.

We compare the reconstructed NF $\kappa$ B trajectories generated by the trained neural SDE model with the experimentally measured NF $\kappa$ B trajectories (see [Fig 9E–9I](#)). The trajectories generated using our END-nSDE method successfully reproduce the experimental dynamics for the majority of time points for the top 50% of cell subgroups most correlated with the representative-cell ODE model (see [Fig 9E–9G](#), [Fig 9I](#)).

For the top-ranked subgroups (#1 to #16), the heterogeneous nSDE-reconstructed dynamics align well with the experimental data for the first 100 minutes. The predicted trajectories deviate more from ground truth trajectories observed in experiments after 100 minutes possibly due to error accumulation and errors in the predicted noise intensity. For experimental subgroups that significantly deviate from the representative-cell ODE model, the END-nSDE struggles to fully capture the heterogeneous trajectories. This limitation likely arises from the assumption that all cells in a group share the same underlying dynamics, whereas in reality, substantial cellular differences in underlying dynamics exist due to heterogeneity in the drift term, an aspect not accounted for in END-nSDE due to the high computational cost.

With sufficient data and computational resources, our proposed workflow is able to incorporate extrinsic noise in the drift terms, allowing for further discrimination of experimental trajectories. Our END-nSDE method can partially reconstruct experimental datasets and has the potential to fully capture experimental dynamics. Furthermore, trajectories generated from the trained END-nSDE model can reproduce the intrinsic fluctuations in the observed NF $\kappa$ B signaling dynamics which are inaccessible to the representative-cell ODE model.



**Fig 9. Inferring intrinsic noise intensities and reconstructing experimental data via END-nSDE.** A. Plots showing the mean (solid circles) and variance (error bars) of the relative error in the reconstructed noise intensities ( $\sigma_1, \sigma_2$ ) predicted by the parameter-inference NN for the testing dataset, as a function of the group size of input trajectories. B. Heatmaps showing the relative error in the reconstructed noise intensities for the training dataset. Colored cells represent results from the parameter-inference NN for the training dataset, while blank cells indicate noise strength values not included in the training set. C. Heatmaps showing the relative error in the diffusion function for the testing dataset. D. The inferred intensity of  $\text{I}\kappa\text{B}\alpha$  transcription noise ( $\sigma_1$ ) and NF $\kappa$ B translocation noise ( $\sigma_2$ ) in different groups of experimental trajectories, plotted against the group's ranking in decreasing similarity with the representative ODE trajectory. E-H. Groups of experimental and nSDE-reconstructed trajectories ranked by decreasing cosine similarity: #1 (E), #4 (F), #16 (G), #29 (H). The squared  $W_2$ -distance between experimental and SDE-generated trajectories are 0.157 (E), 0.143 (F), 0.212 (G), 0.236 (H). The inferred noises are  $(10^{-0.49}, 10^{-0.81})$  (E),  $(10^{-0.47}, 10^{-0.78})$  (F),  $(10^{-0.46}, 10^{-0.74})$  (G),  $(10^{-0.44}, 10^{-0.71})$  (H). I. The temporally decoupled squared  $W_2$  distance between reconstructed trajectories generated by the trained END-nSDE and groups of experimental trajectories, ordered according to decreasing similarity with the representative ODE trajectory.

<https://doi.org/10.1371/journal.pcbi.1013462.g009>

## 4. Discussion

In this work, we used the  $W_2$ -distance to develop an END-nSDE reconstruction method that takes into account extrinsic noise in gene expression dynamics as observed across various biophysical and biochemical processes such as circadian rhythms, RPA-DNA binding, and NF $\kappa$ B translocation. We first demonstrated that our END-nSDE method can successfully reconstruct a minimal noise-driven fluctuating SDE characterizing the circadian rhythm, showcasing its effectiveness in reconstructing SDE models that contain both intrinsic and extrinsic noise. Next, we used our END-nSDE method to learn a surrogate extrinsic-noise-driven neural SDE, which approximates the RPA-DNA binding process. Molecular binding processes are usually modeled by a Markov counting process and simulated using Monte-Carlo-type stochastic simulation algorithms (SSAs) [48]. Our END-nSDE reconstruction approach can effectively reconstruct the stochastic dynamics of the RPA-ssDNA binding process while also taking into account extrinsic noise (heterogeneity in biological parameters among different cells). Our END-nSDE method outperforms several benchmark methods such as LSTMs, RNNs, neural ODEs, and Gaussian processes.

Finally, we applied our methodology to analyze NF $\kappa$ B trajectories collected from over a thousand cells. Not only did the neural SDE model trained on the synthetic dataset perform well on the validation set, but it also partially recapitulated experimental trajectories of NF $\kappa$ B abundances, particularly for subgroups with dynamics similar to those of the representative cell. These results underscore the potential of neural SDEs in modeling and understanding the role of intrinsic noise in complex cellular signaling systems [66–68].

When the experimental trajectories were divided into subgroups, we assumed that all cells across different groups shared the same drift function (as in the representative ODE) and cells within each group shared the same diffusion term. We found that subgroups with dynamics more closely aligned with the deterministic ODE model resulted in better reconstructions. In contrast, for experimental trajectories that deviated significantly from the representative ODE model, their underlying dynamics may differ from those defined by the representative cell's ODE. Therefore, the assumption that a group shares the same drift function as the representative cell ODE holds only when the trajectories closely resemble the ODE. Incorporating noise into the drift term for training the neural SDE could potentially address this issue. We did not consider this approach due to the high computational cost required for training.

Applying our method to high-dimensional synthetic NF $\kappa$ B datasets, we showed the importance of incorporating latent variables. This necessity arises because the ground-truth dynamics of the measured quantities (nuclear NF $\kappa$ B) are not self-closed and inherently depend on additional variables. Consequently, the 52-dimensional SDE reconstruction requires more variables than just the “observed” dynamics of nuclear NF $\kappa$ B. In this example, the remaining 50 variables in the nSDE were treated as latent variables, even though they were not explicitly included in the loss function.

For high-dimensional systems (e.g., 52 dimensions as in our NF $\kappa$ B example), analyzing stochastic dynamics remains challenging. Even though regulated processes do not follow gradient dynamics in general, imposing a self-consistent energy landscape and adopting lower dimensional projections can provide a valuable framework for studying stochastic dynamics of high-dimensional biological systems [69–71]. Once an effective energy landscape is identified, prior knowledge about the system structure can be incorporated into the neural SDE framework through the following formulation:

$$d\mathbf{X} = (F(\mathbf{X}) + \hat{f}(\mathbf{X}; \omega))dt + \hat{\sigma}(\mathbf{X}; \omega)d\mathbf{B}_t, \quad (17)$$

where  $F(\mathbf{X}) = \nabla E(\mathbf{X})$  represents the prior knowledge of the energy landscape. The neural networks  $\hat{f}$  and  $\hat{\sigma}$  then learn deviations from the prior knowledge and the unknown intrinsic noise. Such prior information on the energy landscape could facilitate training and improve accuracy of the learned model [37]. How imposition of a high-dimensional landscape as a constraint affects our  $W_2$ -distance-based inference and how this potential should be interpreted should be explored in more depth. If meaningful and informative, prior results on how landscapes can be used to characterize neural networks across various tasks can be leveraged [72, 73].

Finally, neural SDEs can serve as surrogate models for complex biomedical dynamics [74, 75]. Combining such surrogate models with neural control functions [72,76,77] can be useful for tackling complex biomedical control problems. As shown in preceding work [32,37], a larger number of training trajectories led to a more accurate reconstructed neural SDE. However, in biological experiments, obtaining more training trajectories could be more expensive. Therefore, it is of biological significance to find out the number of training trajectories that can be practically obtained in real experiments and that are necessary for an accurate reconstruction of the intrinsic-noise-aware SDE using our END-nSDE approach. Finally, it is worth further investigation to find out the biophysical molecular processes in which taking into account intrinsic fluctuations is necessary. In such problems, using our END-nSDE framework to reconstruct the noisy molecular dynamics could yield a more biologically reasonable, noise-aware model than first-principle-based mass-action ODE models.

While our work focuses on gene regulation dynamics, it is important to emphasize that the END-nSDE reconstruction method is general and can potentially be applied to biological systems beyond gene regulation. The method's ability to capture both intrinsic and extrinsic noise makes it suitable for modeling various stochastic biological processes, including, but not limited to, signal transduction networks, metabolic pathways, population dynamics, and developmental processes. The examples we chose—circadian rhythms, RPA-DNA binding dynamics, and NF $\kappa$ B signaling—were selected to demonstrate the method's capabilities and bring the neural SDE approach to the attention of the molecular and cell biology community. Future applications could extend to other domains such as epidemiology, ecology, and systems biology, where stochastic dynamics that could be described by SDEs with heterogeneity among different cells or individuals are prevalent.

Besides better understanding effective energy landscape constraints, there are several promising directions for future research. First, techniques to extract an explicit form of the learned neural network SDEs can be developed. For example, one could employ a polynomial model as the reconstructed drift and diffusion functions in the SDE [78]. Such an explicit functional form of the approximate SDE may facilitate biological interpretation of the underlying model. Recent research has also shed light on directly interpreting trained neural networks using simple functions such as polynomials [79]. Therefore, one can apply such methods to extract the approximate forms from the learned drift and diffusion functions in the neural SDE and interpret their biophysical meaning.

Another promising avenue of investigation is to combine discrete and continuous modeling approaches to account for both mRNA and protein dynamics. Such a hybrid approach would use discrete Markov jump processes for low-abundance species (such as mRNA) while employing SDEs for high-abundance species (such as proteins), thereby addressing the limitations of pure SDE approaches when molecular counts approach zero.

Finally, the presence of unobserved variables in cellular systems poses a significant challenge for accurate SDE modeling. Many cellular processes involve hidden regulatory mechanisms, unmeasured metabolites, or latent cellular states that influence the observed dynamics but are not directly captured in experimental measurements. This limitation can lead to

model misspecification, where the inferred drift and diffusion functions compensate for missing variables, potentially resulting in biased parameter estimates and poor predictive performance. A more realistic scenario occurs when we already know what molecules can have an effect on the dynamics, but experiments can only report a few molecular species. In such cases, we can model the full system dynamics with the full dimension with a parameterized model for sampling the initial values of those unobserved variables. The rest of the training procedure would be the same as in the main text.

## Supporting information

**S1 Text. Technical appendices.**  
(PDF)

## Acknowledgments

We acknowledge Stefanie Luecke for providing the experimental datasets for NF $\kappa$ B dynamics.

## Author contributions

**Formal analysis:** Xiangting Li, Mingtao Xia.

**Funding acquisition:** Lucas Böttcher, Tom Chou.

**Investigation:** Jiancheng Zhang, Xiangting Li, Xiaolu Guo, Zhaoyi You, Lucas Böttcher, Alex Mogilner, Alexander Hoffmann, Tom Chou, Mingtao Xia.

**Methodology:** Jiancheng Zhang, Xiangting Li, Mingtao Xia, Xiaolu Guo.

**Project administration:** Mingtao Xia.

**Resources:** Xiangting Li, Xiaolu Guo, Lucas Böttcher, Mingtao Xia.

**Software:** Jiancheng Zhang, Xiangting Li, Tom Chou, Mingtao Xia, Xiaolu Guo.

**Supervision:** Xiaolu Guo, Mingtao Xia.

**Validation:** Jiancheng Zhang, Xiangting Li, Mingtao Xia, Xiaolu Guo.

**Visualization:** Jiancheng Zhang, Xiangting Li, Xiaolu Guo, Tom Chou, Mingtao Xia.

**Writing – original draft:** Jiancheng Zhang, Xiangting Li, Xiaolu Guo, Zhaoyi You, Mingtao Xia.

**Writing – review & editing:** Jiancheng Zhang, Xiangting Li, Xiaolu Guo, Zhaoyi You, Lucas Böttcher, Alex Mogilner, Alexander Hoffmann, Tom Chou, Mingtao Xia.

## References

1. Swain PS, Elowitz MB, Siggia ED. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc Natl Acad Sci U S A*. 2002;99(20):12795–800. <https://doi.org/10.1073/pnas.162041399> PMID: 12237400
2. Elowitz MB, Levine AJ, Siggia ED, Swain PS. Stochastic gene expression in a single cell. *Science*. 2002;297(5584):1183–6. <https://doi.org/10.1126/science.1070919> PMID: 12183631
3. Sanchez A, Choubey S, Kondev J. Regulation of noise in gene expression. *Annu Rev Biophys*. 2013;42:469–91. <https://doi.org/10.1146/annurev-biophys-083012-130401> PMID: 23527780
4. Foreman R, Wollman R. Mammalian gene expression variability is explained by underlying cell state. *Mol Syst Biol*. 2020;16(2):e9146. <https://doi.org/10.15252/msb.20199146> PMID: 32043799

5. Mitchell S, Hoffmann A. Identifying Noise Sources governing cell-to-cell variability. *Curr Opin Syst Biol.* 2018;8:39–45. <https://doi.org/10.1016/j.coisb.2017.11.013> PMID: 29623300
6. Thattai M, van Oudenaarden A. Intrinsic noise in gene regulatory networks. *Proc Natl Acad Sci U S A.* 2001;98(15):8614–9. <https://doi.org/10.1073/pnas.151588598> PMID: 11438714
7. Tsimring LS. Noise in biology. *Rep Prog Phys.* 2014;77(2):026601. <https://doi.org/10.1088/0034-4885/77/2/026601> PMID: 24444693
8. Fu AQ, Pachter L. Estimating intrinsic and extrinsic noise from single-cell gene expression measurements. *Stat Appl Genet Mol Biol.* 2016;15(6):447–71. <https://doi.org/10.1515/sagmb-2016-0002> PMID: 27875323
9. Llamasi A, Gonzalez-Vargas AM, Versari C, Cinquemani E, Ferrari-Trecate G, Hersen P, et al. What population reveals about individual cell identity: single-cell parameter estimation of models of gene expression in yeast. *PLoS Comput Biol.* 2016;12(2):e1004706. <https://doi.org/10.1371/journal.pcbi.1004706> PMID: 26859137
10. Dharmarajan L, Kaltenbach H-M, Rudolf F, Stelling J. A simple and flexible computational framework for inferring sources of heterogeneity from single-cell dynamics. *Cell Syst.* 2019;8(1):15–26.e11. <https://doi.org/10.1016/j.cels.2018.12.007> PMID: 30638813
11. Finkenstädt B, Woodcock DJ, Komorowski M, Harper CV, Davis JRE, White MRH. Quantifying intrinsic and extrinsic noise in gene transcription using the linear noise approximation: an application to single cell data. *The Annals of Applied Statistics.* 2013; p. 1960–82.
12. Dixit PD. Quantifying extrinsic noise in gene expression using the maximum entropy framework. *Biophys J.* 2013;104(12):2743–50. <https://doi.org/10.1016/j.bpj.2013.05.010> PMID: 23790383
13. Fang Z, Gupta A, Kumar S, Khammash M. Advanced methods for gene network identification and noise decomposition from single-cell data. *Nat Commun.* 2024;15(1):4911. <https://doi.org/10.1038/s41467-024-49177-1> PMID: 38851792
14. Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S. Stochastic mRNA synthesis in mammalian cells. *PLoS Biol.* 2006;4(10):e309. <https://doi.org/10.1371/journal.pbio.0040309> PMID: 17048983
15. Raser JM, O'Shea EK. Control of stochasticity in eukaryotic gene expression. *Science.* 2004;304(5678):1811–4. <https://doi.org/10.1126/science.1098641> PMID: 15166317
16. Sigal A, Milo R, Cohen A, Geva-Zatorsky N, Klein Y, Liron Y, et al. Variability and memory of protein levels in human cells. *Nature.* 2006;444(7119):643–6. <https://doi.org/10.1038/nature05316> PMID: 17122776
17. Singh A, Razooky BS, Dar RD, Weinberger LS. Dynamics of protein noise can distinguish between alternate sources of gene-expression variability. *Mol Syst Biol.* 2012;8:607. <https://doi.org/10.1038/msb.2012.38> PMID: 22929617
18. Paulsson J. Models of stochastic gene expression. *Physics of Life Reviews.* 2005;2(2):157–75.
19. Singh A, Soltani M. Quantifying intrinsic and extrinsic variability in stochastic gene expression models. *PLoS One.* 2013;8(12):e84301. <https://doi.org/10.1371/journal.pone.0084301> PMID: 24391934
20. Wang D-G, Wang S, Huang B, Liu F. Roles of cellular heterogeneity, intrinsic and extrinsic noise in variability of p53 oscillation. *Sci Rep.* 2019;9(1):5883. <https://doi.org/10.1038/s41598-019-41904-9> PMID: 30971810
21. Voit EO, Martens HA, Omholt SW. 150 years of the mass action law. *PLoS Comput Biol.* 2015;11(1):e1004012. <https://doi.org/10.1371/journal.pcbi.1004012> PMID: 25569257
22. Ferner RE, Aronson JK. Cato Guldberg and Peter Waage, the history of the Law of Mass Action, and its relevance to clinical pharmacology. *Br J Clin Pharmacol.* 2016;81(1):52–5. <https://doi.org/10.1111/bcp.12721> PMID: 26174880
23. Bressloff PC, Newby JM. Metastability in a stochastic neural network modeled as a velocity jump Markov process. *SIAM Journal on Applied Dynamical Systems.* 2013;12(3):1394–435.
24. Kurtz TG. Limit theorems and diffusion approximations for density dependent Markov chains. In: Wets RJB, editor. Berlin, Heidelberg: Springer; 1976. p. 67–78.
25. Tian T, Burrage K, Burrage PM, Carletti M. Stochastic delay differential equations for genetic regulatory networks. *Journal of Computational and Applied Mathematics.* 2007;205(2):696–707.
26. Chen K-C, Wang T-Y, Tseng H-H, Huang C-YF, Kao C-Y. A stochastic differential equation model for quantifying transcriptional regulatory network in *Saccharomyces cerevisiae*. *Bioinformatics.* 2005;21(12):2883–90. <https://doi.org/10.1093/bioinformatics/bti415> PMID: 15802287
27. Xia M, Chou T. Kinetic theories of state- and generation-dependent cell populations. *Phys Rev E.* 2024;110(6–1):064146. <https://doi.org/10.1103/PhysRevE.110.064146> PMID: 39916132
28. Zechner C, Unger M, Pelet S, Peter M, Koepl H. Scalable inference of heterogeneous reaction kinetics from pooled single-cell recordings. *Nat Methods.* 2014;11(2):197–202. <https://doi.org/10.1038/nmeth.2794> PMID: 24412977

29. Sukys A, Öcal K, Grima R. Approximating solutions of the Chemical Master equation using neural networks. *iScience*. 2022;25(9):105010. <https://doi.org/10.1016/j.isci.2022.105010> PMID: 36117994
30. Öcal K, Gutmann MU, Sanguinetti G, Grima R. Inference and uncertainty quantification of stochastic gene expression via synthetic models. *Journal of The Royal Society Interface*. 2022;19(192):20220153.
31. Cao Z, Chen R, Xu L, Zhou X, Fu X, Zhong W, et al. Efficient and scalable prediction of stochastic reaction-diffusion processes using graph neural networks. *Math Biosci*. 2024;375:109248. <https://doi.org/10.1016/j.mbs.2024.109248> PMID: 38986837
32. Xia M, Li X, Shen Q, Chou T. Squared Wasserstein-2 Distance for Efficient Reconstruction of Stochastic Differential Equations; 2024. <https://arxiv.org/abs/2401.11354>.
33. Li X, Wong TKL, Chen RTQ, Duvenaud D. Scalable gradients for stochastic differential equations. In: *International Conference on Artificial Intelligence and Statistics*, 2020.
34. Kidger P, Foster J, Li X, Lyons TJ. Neural SDEs as Infinite-Dimensional GANs. In: *Proceedings of the 38th International Conference on Machine Learning*. 2021. p. 5453–63.
35. Kidger P, Foster J, Li X, Lyons TJ. Neural SDEs as infinite-dimensional GANs. In: 2021. p. 5453–63.
36. Arjovsky M, Chintala S, Bottou L. Wasserstein GAN. 2017. <http://arxiv.org/abs/1701.07875v3>
37. Xia M, Li X, Shen Q, Chou T. An efficient Wasserstein-distance approach for reconstructing jump-diffusion processes using parameterized neural networks. *Machine Learning: Science and Technology*. 2024;5:045052.
38. Xia M, Shen Q. A local squared Wasserstein-2 method for efficient reconstruction of models with uncertainty. 2024. <https://arxiv.org/abs/2406.06825>
39. Xia M, Shen Q, Maini P, Gaffney E, Mogilner A. A new local time-decoupled squared Wasserstein-2 method for training stochastic neural networks to reconstruct uncertain parameters in dynamical systems. *Neural Networks*. 2025.
40. Yu R, Wang R. Learning dynamical systems from data: an introduction to physics-guided deep learning. *Proc Natl Acad Sci U S A*. 2024;121(27):e2311808121. <https://doi.org/10.1073/pnas.2311808121> PMID: 38913886
41. Gonze D. Modeling circadian clocks: from equations to oscillations. *Open Life Sciences*. 2011;6(5):699–711.
42. Westermark PO, Welsh DK, Okamura H, Herzog H. Quantification of circadian rhythms in single cells. *PLoS Comput Biol*. 2009;5(11):e1000580. <https://doi.org/10.1371/journal.pcbi.1000580> PMID: 19956762
43. Dueva R, Iliakis G. Replication protein A: a multifunctional protein with roles in DNA replication, repair and beyond. *NAR Cancer*. 2020;2(3):zcaa022. <https://doi.org/10.1093/narcan/zcaa022> PMID: 34316690
44. Caldwell CC, Spies M. Dynamic elements of replication protein A at the crossroads of DNA replication, recombination, and repair. *Crit Rev Biochem Mol Biol*. 2020;55(5):482–507. <https://doi.org/10.1080/10409238.2020.1813070> PMID: 32856505
45. Nguyen HD, Yadav T, Giri S, Saez B, Graubert TA, Zou L. Functions of Replication Protein A as a Sensor of R Loops and a Regulator of RNaseH1. *Mol Cell*. 2017;65(5):832–847.e4. <https://doi.org/10.1016/j.molcel.2017.01.029> PMID: 28257700
46. Wold MS. Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. *Annu Rev Biochem*. 1997;66:61–92. <https://doi.org/10.1146/annurev.biochem.66.1.61> PMID: 9242902
47. Ding J, Li X, Shen J, Zhao Y, Zhong S, Lai L, et al. ssDNA accessibility of Rad51 is regulated by orchestrating multiple RPA dynamics. *Nature Communications*. 2023;14:3864.
48. Gillespie DT. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*. 1977;81(25):2340–61.
49. Gillespie DT. The chemical Langevin equation. *The Journal of Chemical Physics*. 2000;113(1):297–306.
50. Sheu K, Luecke S, Hoffmann A. Stimulus-specificity in the responses of immune sentinel cells. *Curr Opin Syst Biol*. 2019;18:53–61. <https://doi.org/10.1016/j.coisb.2019.10.011> PMID: 32864512
51. Adelaja A, Taylor B, Sheu KM, Liu Y, Luecke S, Hoffmann A. Six distinct NF- $\kappa$ B signaling codons convey discrete information to distinguish stimuli and enable appropriate macrophage responses. *Immunity*. 2021;54(5):916–930.e7. <https://doi.org/10.1016/j.immuni.2021.04.011> PMID: 33979588
52. Hoffmann A, Levchenko A, Scott ML, Baltimore D. The IkappaB-NF-kappaB signaling module: temporal control and selective gene activation. *Science*. 2002;298(5596):1241–5. <https://doi.org/10.1126/science.1071914> PMID: 12424381

53. Cheng QJ, Ohta S, Sheu KM, Spreafico R, Adelaja A, Taylor B, et al. NF- $\kappa$ B dynamics determine the stimulus specificity of epigenomic reprogramming in macrophages. *Science*. 2021;372(6548):1349–53. <https://doi.org/10.1126/science.abc0269> PMID: 34140389
54. Sen S, Cheng Z, Sheu KM, Chen YH, Hoffmann A. Gene regulatory strategies that decode the duration of NF $\kappa$ B dynamics contribute to LPS- versus TNF-specific gene expression. *Cell Syst*. 2020;10(2):169–182.e5. <https://doi.org/10.1016/j.cels.2019.12.004> PMID: 31972132
55. Adelaja A, Taylor B, Sheu KM, Liu Y, Luecke S, Hoffmann A. Six distinct NF $\kappa$ B signaling codons convey discrete information to distinguish stimuli and enable appropriate macrophage responses. *Immunity*. 2021;54(5):916–930.e7. <https://doi.org/10.1016/j.immuni.2021.04.011> PMID: 33979588
56. Guo X, Adelaja A, Singh A, Wollman R, Hoffmann A. Modeling heterogeneous signaling dynamics of macrophages reveals principles of information transmission in stimulus responses. *Nat Commun*. 2025;16(1):5986.
57. Bressloff PC. *Stochastic processes in cell biology*. Springer; 2014.
58. Spagnolo B, Spezia S, Curcio L, Pizzolato N, Fiasconaro A, Valenti D. Noise effects in two different biological systems. *The European Physical Journal B*. 2009;69:133–46.
59. Pahle J, Challenger JD, Mendes P, McKane AJ. Biochemical fluctuations, optimisation and the linear noise approximation. *BMC Syst Biol*. 2012;6:86. <https://doi.org/10.1186/1752-0509-6-86> PMID: 22805626
60. Swain PS, Elowitz MB, Siggia ED. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc Natl Acad Sci U S A*. 2002;99(20):12795–800. <https://doi.org/10.1073/pnas.162041399> PMID: 12237400
61. Kelleher JD. *Deep learning*. MIT Press. 2019.
62. Saptadi NTS, Kristiawan H, Nugroho AY, Rahayu N, Waseso B, Intan I. *Deep Learning: Teori, Algoritma, dan Aplikasi*. Sada Kurnia Pustaka; 2025.
63. Luecke S, Guo X, Sheu KM, Singh A, Lowe SC, Han M, et al. Dynamical and combinatorial coding by MAPK p38 and NF $\kappa$ B in the inflammatory response of macrophages. *Mol Syst Biol*. 2024;20(8):898–932. <https://doi.org/10.1038/s44320-024-00047-4> PMID: 38872050
64. Nakamura T, Taki K, Nomiyama H, Seki K, Uehara K. A shape-based similarity measure for time series data with ensemble learning. *Pattern Analysis and Applications*. 2013;16(4):535–48.
65. Frishman A, Ronceray P. Learning force fields from stochastic trajectories. *Physical Review X*. 2020;10(2):021009.
66. Rao CV, Wolf DM, Arkin AP. Control, exploitation and tolerance of intracellular noise. *Nature*. 2002;420(6912):231–7. <https://doi.org/10.1038/nature01258> PMID: 12432408
67. Arias AM, Hayward P. Filtering transcriptional noise during development: concepts and mechanisms. *Nat Rev Genet*. 2006;7(1):34–44. <https://doi.org/10.1038/nrg1750> PMID: 16369570
68. Eling N, Morgan MD, Marioni JC. Challenges in measuring and understanding biological noise. *Nat Rev Genet*. 2019;20(9):536–48. <https://doi.org/10.1038/s41576-019-0130-6> PMID: 31114032
69. Kang X, Li C. A dimension reduction approach for energy landscape: identifying intermediate states in metabolism-EMT network. *Adv Sci (Weinh)*. 2021;8(10):2003133. <https://doi.org/10.1002/advs.202003133> PMID: 34026435
70. Lang J, Nie Q, Li C. Landscape and kinetic path quantify critical transitions in epithelial-mesenchymal transition. *Biophys J*. 2021;120(20):4484–500. <https://doi.org/10.1016/j.bpj.2021.08.043> PMID: 34480928
71. Li C, Wang J. Landscape and flux reveal a new global view and physical quantification of mammalian cell cycle. *Proc Natl Acad Sci U S A*. 2014;111(39):14130–5. <https://doi.org/10.1073/pnas.1408628111> PMID: 25228772
72. Böttcher L, Asikis T. Near-optimal control of dynamical systems with neural ordinary differential equations. *Machine Learning: Science and Technology*. 2022;3(4):045004.
73. Böttcher L, Wheeler G. Visualizing high-dimensional loss landscapes with Hessian directions. *Journal of Statistical Mechanics: Theory and Experiment*. 2024;2024(2):023401.
74. Fonseca LL, Böttcher L, Mehrad B, Laubenbacher RC. Optimal control of agent-based models via surrogate modeling. *PLoS Comput Biol*. 2025;21(1):e1012138. <https://doi.org/10.1371/journal.pcbi.1012138> PMID: 39808665
75. Böttcher L, Fonseca LL, Laubenbacher RC. Control of medical digital twins with artificial neural networks. *Philosophical Transactions of the Royal Society A*. 2025;383:20240228.
76. Asikis T, Böttcher L, Antulov-Fantulin N. Neural ordinary differential equation control of dynamics on graphs. *Physical Review Research*. 2022;4(1):013221.
77. Böttcher L, Antulov-Fantulin N, Asikis T. AI Pontryagin or how artificial neural networks learn to control dynamical systems. *Nat Commun*. 2022;13(1):333. <https://doi.org/10.1038/s41467-021-27590-0> PMID: 35039488

78. Fronk C, Petzold L. Interpretable polynomial neural ordinary differential equations. *Chaos*. 2023;33(4):043101. <https://doi.org/10.1063/5.0130803> PMID: 37097945
79. Morala P, Cifuentes JA, Lillo RE, Ucar I. NN2Poly: a polynomial representation for deep feed-forward artificial neural networks. *IEEE Trans Neural Netw Learn Syst*. 2025;36(1):781–95. <https://doi.org/10.1109/TNNLS.2023.3330328> PMID: 37962996

# S1 Text: Technical Appendices

## Appendix A: Hyperparameters and conditions used in the neural SDE model

In this section, we provide the hyperparameters in the neural SDE models as well as the training details in Table A in S1 Text.

Table A: The hyperparameters for training the neural SDE model of each example.

	Example 1	Example 2	Example 3
Gradient descent method	Adam	Adam	Adam
Learning rate	0.002	0.002	0.002
Weight decay	0.005	0.005	0.005
Activation function	ReLU	ReLU	ReLU
# of epochs	500	2000	2000
# of hidden layers in $\hat{f}$	3	2	2
# of neurons in hidden layer in $\hat{f}$	100	200	400
# of hidden layers in $\hat{\sigma}$	3	2	2
# of neurons in each hidden layer in $\hat{\sigma}$	100	200	400
$\Delta t$	0.1 (s)	$\frac{1}{720}$ (hr)	$\frac{1}{12}$ (hr)
$T$	11	361	31

## Appendix B: RPA dynamic binding to long ssDNA can be simulated by a generalized random sequential adsorption (RSA) model

To elucidate the biophysical mechanism of replication protein A (RPA) binding dynamics to long single-stranded DNA (ssDNA), we developed a continuous-time discrete Markov chain model. The Random Sequential Adsorption (RSA) model in a one-dimensional (1D) finite-length context effectively represents the process of protein binding to DNA, capturing the key property that each nucleotide (nt) of ssDNA cannot be occupied by more than one protein molecule. This unique characteristic leads to incomplete occupation even with protein oversaturation, distinguishing DNA-relevant reactions from those described by the mass action law. To reveal finer structures such as gap distribution, we implemented an exact stochastic sampling approach.

We adapted the model based on current knowledge of RPA binding modes, incorporating multiple binding modes and volume exclusion effects. RPA has two binding modes: a 20-nt mode (partial binding mode, PBM) and a 30-nt mode (full-length binding mode, FLBM). RPA initially binds to a 20-nt ssDNA with a rate of  $k_1$  and dissociates at a rate of  $k_{-1}$ . This 20-nt mode assumes constant  $k_1$ . One scenario involves DBD-A, DBD-B, and DBD-C binding to ssDNA, with subsequent DBD-D binding leading to the 30-nt mode. In the 20-nt mode, DBD-D binds an extra 10-nt ssDNA with a rate of  $k_2$  and dissociates at  $k_{-2}$ , forming the 30-nt mode (FLBM). RPA always aligns in the same direction along DNA.

The multiple binding modes result in interesting kinetic features like facilitated exchange and desorption. In this model, the ssDNA fragment state is represented by a vector of length  $L$  with each component

taking values in  $\{0, 1\}$ , where 0 indicates an unoccupied nucleotide and 1 indicates an occupied one. Each RPA initially occupies  $\ell = 20$  nts and can further occupy  $\Delta\ell = 10$  nts if the local state allows, moving in the 3' direction. Each DNA site can be occupied by only one RPA molecule, altering the RSA model's available reactions. Each consecutive unoccupied segment of length  $\ell$  recruits one RPA at the rate of  $k_1$ , with the total binding rate given by:

$$v_1 = k_1 \sum_{j=1}^{L-\ell+1} \delta_{\mathbf{0}^{\times\ell}}(\text{state}[j, j+1, \dots, j+\ell-1]),$$

where  $\delta_a(b)$  is the Kronecker delta function,  $\mathbf{0}^{\times\ell}$  is a zero vector of length  $\ell$ , and  $\text{state}[j, \dots, j+\ell-1]$  is the ssDNA state vector.

To occupy another 10 nts, we assigned a rate parameter  $k_2$  and calculate the overall rate by:

$$v_2 = k_2 \sum_{q_j} \delta_{\mathbf{0}^{\times\Delta\ell}}(\text{state}[q_j + \ell, \dots, q_j + \ell + \Delta\ell - 1]),$$

where  $q_j$  represents the leftmost position of each bound RPA in the 20-nt mode. For unbinding, the 30-nt mode reopens to the 20-nt mode at rate  $k_{-2}$ , and the 20-nt mode desorbs at rate  $k_{-1}$ . The overall rates are:

$$\begin{aligned} v_{-2} &= k_{-2} \#\{\text{FLBM RPA}\}, \\ v_{-1} &= k_{-1} \#\{\text{PBM RPA}\}. \end{aligned}$$

The total possible reaction rate is:

$$v_{\text{tot}} = v_1 + v_{-1} + v_2 + v_{-2}.$$

Reactions occur stochastically according to exponentially distributed waiting times with parameter  $v_{\text{tot}}$ . The waiting time  $\delta t$  follows the exponential distribution with a probability density function:

$$\text{pdf}(\delta t = t) = v_{\text{tot}} e^{-v_{\text{tot}} t}.$$

After each reaction, the DNA state updates, and the possible reactions are re-evaluated. The Gillespie algorithm was used to sample the trajectories of this stochastic model. We used Julia to perform exact stochastic simulations of all known RPA-ssDNA interactions, and codes are available at ([https://github.com/hsianktin/RPA\\_model](https://github.com/hsianktin/RPA_model)).

## Appendix C: Implementation of benchmarks

In this section, we introduce the benchmark methods; hyperparameters and settings for training of all methods used in this paper are shown in Table B in S1 Text.

### RNN networks

Recurrent neural networks (RNN) are often used for language processing, but they can also be used to analyze time series data with temporal correlations. For reconstructing the RPA-DNA binding dynamics, we used the RNN model. A neural network that contains two layers of RNN and two linear layers was used to model RPA's dynamic binding with single-stranded DNA. All layers of RNN are built using `torch.nn.Module` package. Hyperparameters in the neural network are initialized by Pytorch default settings. The parameters of this neural network are initialized by default. At each time step,  $\mathbf{X}(t; \omega) \in \mathbb{R}^d$  representing the RPA dynamics at this time point is inputted in Example 2.2.2, and the dynamics at the next time point:  $\mathbf{X}(t + \Delta t; \omega)$  is outputted as the prediction. The RNN is trained by optimizing the loss function Eq. (8). The pipeline of the model could be seen in Fig A in S1 Text.

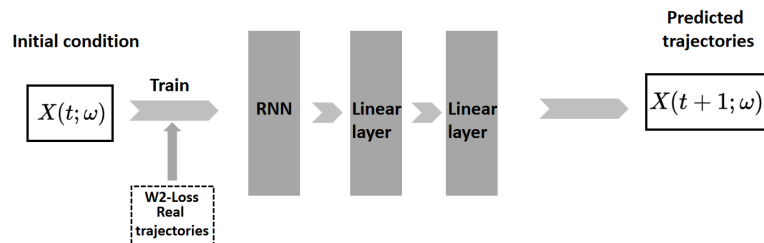


Figure A: The structure of the RNN model used. The RNN layer in this figure can be built using `torch.nn.RNN`.

## LSTM networks

Long Short-Term Memory (LSTM) networks, a class of variants of Recurrent Neural Networks (RNNs), have been widely used in modeling sequential data such as time series and natural language. We used the LSTM network model implemented through the `torch.nn.Module` package for reconstructing the RPA-DNA dynamics. Hyperparameters in the neural network are initialized by default. At each time step, the  $\mathbf{X}(t; \omega) \in \mathbb{R}^d$  discussed above is inputted in Example 2.2.2, and the state at the next time point:  $\mathbf{X}(t + \Delta t; \omega)$  is outputted as the prediction. Then, the gradients of the loss function Eq. (8) are calculated to update the parameters in the LSTM model. The input size of the two layers of LSTM is 2 and the hidden size is 4. The input size and output size of the two linear layers are 8,4 and 4,2 respectively. The LSTM is trained by optimizing the loss function Eq. (8). The pipeline of the model is shown in Fig B in S1 Text.

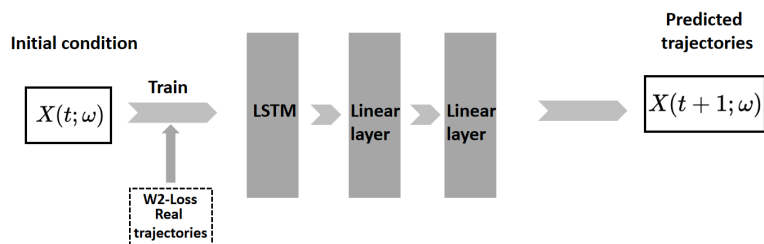


Figure B: The structure of the LSTM model used. The LSTM layer in this figure can be built using `torch.nn.LSTM`.

## Gaussian Process

The Gaussian Process (GP) is also widely used in modeling time-series data; it is implemented using the `gaussianprocess.GaussianProcessRegressor` package in Python. We used radial basis functions as kernel functions (using `sklearn.gaussianprocess.kernels` package), and the hyperparameters of  $\alpha$  and `n-restarts-optimizer` are 0.3 and 5, respectively. When training the GP model, we inputted the trajectories of all time points in the training set to the GP for fitting. Then, we inputted the trajectory of the current time point in the testing set, and the model predicts the trajectory of the next time point based on the kernel.

## Appendix D: Generating the simulated NF $\kappa$ B dataset

To evaluate the performance of our proposed END-nSDE method in reconstructing dynamics of NF $\kappa$ B signaling, we generated training trajectories by simulating a previously developed 52-dimensional model

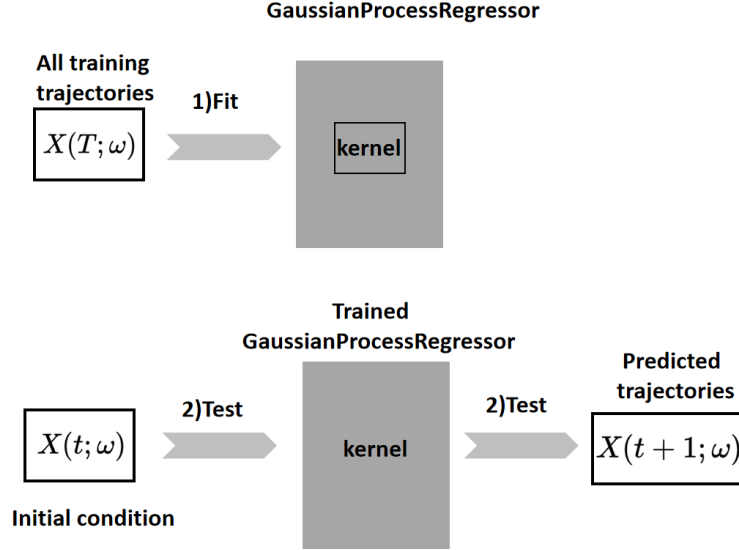


Figure C: The structure of the GP model used in this work.

Table B: Hyperparameters of benchmark methods

Model	Implementation
RNN	one RNN layer plus four MLP layers width of RNN layer: 256 learning rate: $5 \times 10^{-4}$ learning epoch: 2000 Optimization method: Adam
LSTM	one LSTM layer plus four MLP layers width of LSTM layer: 128 learning rate: $5 \times 10^{-4}$ learning epoch: 3000 Optimization method: Adam

for the NF $\kappa$ B signaling network under 100ng/mL TNF stimulation (see Eqs. (10) for the SDEs). The synthetic trajectories were generated using 100 sets of noise intensities. We set  $\sigma_1 \in [10^{-3.2}, 10^{-2.2}]$ ,  $\sigma_2 \in [10^{-2.5}, 10^{-1.5}]$  in Eqs. (10) and use 100 combinations of  $(\sigma_1, \sigma_2)$ :  $\sigma_1 \in \{10^{-3.2+i\delta_1}, i = 0, \dots, 9\}$ ,  $\delta_1 = 0.1$  and  $\sigma_2 \in \{10^{-2.5+j\delta_2}, j = 0, \dots, 9\}$ ,  $\delta_2 = 0.1$  respectively. Other parameters were fixed constants. The parameter values in Eqs. (10) are listed in Table C in S1 Text.

The corresponding SDEs were simulated using the 'SDEProblem' function from the 'DifferentialEquations' package in Julia. Simulations were conducted from 0 minutes (stimulus application time) to 150 minutes, and the state was recorded at every 5 minute intervals. Initial values were set to the steady-state solutions of the ordinary differential equations (ODEs), which were obtained using the ode15s function in MATLAB.

Table C: Parameter values for the NF $\kappa$ B model.

Parameter	Value	description
$k_{\text{basal}}$	$5 \times 10^{-7}$	basal I $\kappa$ B $\alpha$ mRNA synthesis
$k_{\text{max}}$	$6 \times 10^{-5}$	maximal rate of I $\kappa$ B $\alpha$ mRNA synthesis induced by NF $\kappa$ B
$n_{\text{NF}\kappa\text{B}}$	2.938	Hill coefficient for mRNA syn
$K_{\text{NF}\kappa\text{B}}$	0.1775	EC50 for mRNA syn
$k_{\text{deg}}$	0.33	degradation rate of I $\kappa$ B $\alpha$ mRNA
$k_{\text{imp}}$	0.6	import rate of NF $\kappa$ B
$k_{\text{a-I}\kappa\text{B-NF}\kappa\text{B}}$	200	association rate of NF $\kappa$ B and I $\kappa$ B $\alpha$
$k_{\text{deg-NF}\kappa\text{B}}$	0	degradation rate of NF $\kappa$ B
$k_{\text{exp}}$	0.042	export rate of NF $\kappa$ B
$v$	3.5	Volume ratio: cytoplasmic volume/nuclear volume
$k_{\text{d-I}\kappa\text{B-NF}\kappa\text{B}}$	0.008	dissociation rate of NF $\kappa$ B-I $\kappa$ B $\alpha$
$k_{\text{phos}}$	2	Phosphorylation/degradation of complexed I $\kappa$ B $\alpha$

## Appendix E: Training a neural SDE using simulated NF $\kappa$ B trajectories

We partitioned the simulated ground-truth trajectories into training and validation sets as follows: 50% of each of 96 sets of trajectories (out of a total of 121) associated with different noise intensities were used for training, while the remaining simulated ground-truth trajectories were used as the validation dataset. Since we could only observe NF $\kappa$ B activity, we defined our loss function (see Eq. (8)) to focus solely on the nuclear NF $\kappa$ B and the I $\kappa$ B $\alpha$ -NF $\kappa$ B nuclear complex. In the loss function Eq. (8),  $\omega \equiv (\sigma_1, \sigma_2)$  denoted two noise intensities (see Eqs. (10)).  $\mu(\omega)$  and  $\hat{\mu}(\omega)$  are the distributions of  $\mathbf{X}(t; \omega)$  and  $\hat{\mathbf{X}}(t; \omega)$ , respectively.  $\mathbf{X}(t; \omega)$  represented the values of  $(u_5, u_{10})$  in Eqs. (10) at time  $t$ . The other 50 variables were not included in the calculation of the loss function.

## Appendix F: Reconstructing I $\kappa$ B $\alpha$ -NF $\kappa$ Bn and NF $\kappa$ Bn in NF $\kappa$ B signaling as a 2D SDE model

To investigate the importance of latent variables that are not included in the loss function, instead of reconstructing the 52D surrogate SDE model for the NF $\kappa$ B signaling dynamics, we attempted to directly reconstruct the dynamics of the nuclear complex I $\kappa$ B $\alpha$ -NF $\kappa$ B and nuclear NF $\kappa$ B using the 2D SDE

$$\begin{aligned} d\hat{u}_5 &= f_5(\hat{u}_5, \hat{u}_{10}, t)dt + \sigma_1(\hat{u}_5, \hat{u}_{10}, t)dB_{1,t}, \\ d\hat{u}_{10} &= f_{10}(\hat{u}_5, \hat{u}_{10}, t)dt + \sigma_2(\hat{u}_5, \hat{u}_{10}, t)dB_{2,t} \end{aligned} \quad (1)$$

to approximate the two SDEs in NF $\kappa$ B signaling model that describes nuclear I $\kappa$ B $\alpha$ -NF $\kappa$ B complex and NF $\kappa$ B.

Using a 2D SDE model to reconstruct the NF $\kappa$ B signaling dynamics could not accurately reconstruct the noisy dynamics of nuclear I $\kappa$ B $\alpha$ -NF $\kappa$ B and NF $\kappa$ B. As shown in Fig D G-H, for certain noise strengths  $(\sigma_0, \sigma_1)$ , both the training and testing losses (the temporally decoupled squared  $W_2$  distance Eq. (3)) are large compared to those obtained from the direct reconstruction of the full 52-dimensional model. Furthermore, when providing the inferred noise strengths from experimentally observed trajectories (the same as in Fig 7A and 7B), the reconstructed 2D SDE fails to generate predicted trajectories  $(\hat{u}_5, \hat{u}_{10})$  that align well with experimental data. Thus, it is necessary to retain the remaining 50 variables in the model, although they are not directly used in the calculation of the loss function.

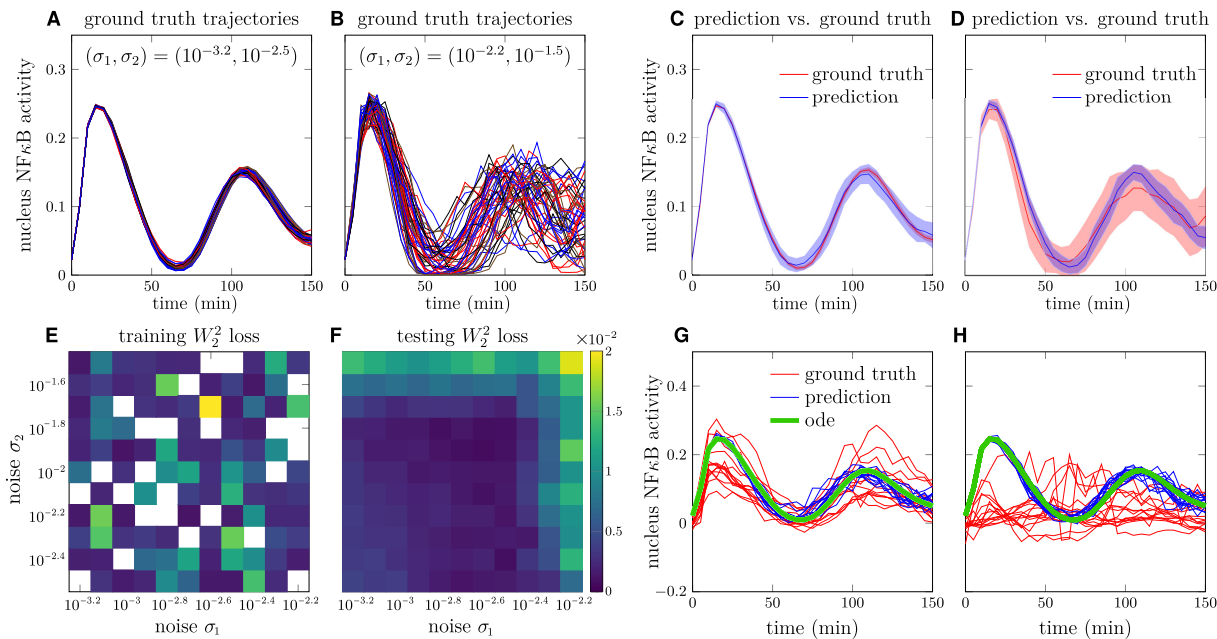


Figure D: **Reconstructed NF $\kappa$ B dynamics using a two-dimensional nSDE.** A. Trajectories of nuclear NF $\kappa$ B concentration over time in the synthetic dataset with noise intensities  $\sigma_1 = 10^{-3.2}$ ,  $\sigma_2 = 10^{-2.5}$ . B. trajectories of nuclear NF $\kappa$ B concentration over time in the synthetic dataset with noise intensities  $\sigma_1 = 10^{-2.2}$ ,  $\sigma_2 = 10^{-1.5}$ . C. Comparison of NF $\kappa$ Bn trajectories predicted by the neural SDE with the ground truth trajectories under noise intensities  $\sigma_1 = 10^{-3.2}$ ,  $\sigma_2 = 10^{-2.5}$ . D. Comparison of NF $\kappa$ Bn predicted by the neural SDE with the ground truth trajectories under noise intensities  $\sigma_1 = 10^{-2.2}$ ,  $\sigma_2 = 10^{-1.5}$ . E. The squared  $W_2$  distance between distributions of predicted trajectories and ground truth trajectories in the training set under different noise strengths  $\sigma_0, \sigma_1$ . Empty cells indicate that the corresponding parameter set is not included in the training set. F. The squared  $W_2$  distance between distributions of the predicted trajectories and the ground truth trajectories on the testing set. G.-H. After grouping the experimental trajectories by cosine similarity to the ODE reference trajectory (shown in green), the trained neural network estimates the noise  $(\sigma_1, \sigma_2)$  of each group of experimental trajectories (shown in red). Then, these estimates were fed into the nSDE to generate a group of (reconstructed) trajectories (shown in blue). The highest- and lowest-ranked similarity groups (#1 and #31, see Fig 9D, 9E, and 9I) are shown in G. and H., respectively.

## Appendix G: Dividing experimental NF $\kappa$ B trajectories into subgroups

We divided the experimentally observed trajectories into 32 groups, each consisting of 32 trajectories. The experimental data were divided into different groups based on each trajectory’s correlation with a ODE-model-based deterministic trajectory

$$\text{corr}(\mathbf{v}_j, \mathbf{v}_{\text{ODE}}) = \frac{\sum_{i=1}^n v_j(t_i)v_{\text{ODE}}(t_i)}{\sqrt{\sum_{i=1}^n v_j^2(t_i)} \cdot \sqrt{\sum_{i=1}^n v_{\text{ODE}}^2(t_i)}} \quad (2)$$

where  $\mathbf{v}_j(t_i), \mathbf{v}_{\text{ODE}}(t_i)$  denote the  $j^{\text{th}}$  observed trajectory in experimental data and the ODE trajectory. The closer a trajectory is to the first-principle-based ODE trajectories, the higher probability that the fluctuations result from intrinsic noise (*i.e.* the Brownian-type noise).

## Appendix H: Training the neural network to infer noise intensities in NF $\kappa$ B dynamics

We adopted a neural network that takes a group of trajectories ( $u_5(t) + u_{10}(t)$ ) under the same noise intensity  $(\sigma_1, \sigma_2)$  as the input and outputs the inferred noise intensity  $(\hat{\sigma}_1, \hat{\sigma}_2)$ . Weights and biases in this neural network were optimized by minimizing the mean squared error (MSE) loss:

$$\text{MSE}(\Lambda) = \sum_{i=1}^m \|\sigma_i - \hat{\sigma}_i\|^2, \quad (3)$$

where  $\sigma_i$  are the ground truth noise intensity parameters underlying the  $i^{\text{th}}$  group of observed trajectories and  $\hat{\sigma}_i$  are the corresponding predicted parameters. Despite the assumption of all cells sharing the same drift function (the same underlying dynamics), different trajectories naturally arise from intrinsic noise.

Detailed steps on inferring noise intensities from a group of trajectories are provided in Fig EA, EB, and EC. The training and testing sets are the same as those in Appendices and of the Supporting Information. The workflow for splitting NF $\kappa$ B SDE-simulated trajectories into training and testing datasets is illustrated in Fig EA and EB. Specifically, out of the 121 combinations of noise intensities (each containing 100 simulation trajectories), 20% were designated for the testing dataset pool. For the remaining 80% noise intensities, 50% of the trajectories under each parameter were randomly selected and added to the training dataset pool (indicated by the olive box in Fig EB). The remaining 50% of trajectories from these parameter sets were used for the testing dataset pool (red box in Fig EB). From both the training and testing dataset pools, for each combination of noise intensity, a group of trajectories was randomly sampled using a permutation sampling approach to construct the training and testing datasets (olive and red solid box in Fig EB).

The neural network used is equipped with an attention structure followed by a feed-forward structure of 2 hidden layers with 64 and 128 neurons in each layer, respectively, where the attention mechanism is designed for assigning weights to different trajectories in a group (sequenced by their similarities to the deterministic ODE trajectory) (Fig EC). The hidden dimension of the attention structure in the query layer is 30, and the hidden dimension of the attention structure in the key and value layers is both 32.

To assess the impact of group size on the performance of the trained neural network in predicting noise intensities, we tested group sizes corresponding to 2, 4, 8, 16, and 32 trajectories per group. The accuracy of the predictions was evaluated using the relative error metric:

$$\frac{|\lg(\hat{\sigma}_1) - \lg(\sigma_1)| + |\lg(\hat{\sigma}_2) - \lg(\sigma_2)|}{|\lg(\sigma_1)| + |\lg(\sigma_2)|}.$$

To explore how the proposed noise inference depends on the frequency of trajectory sampling, we trained the neural network (Fig EC.) using different numbers of sampled time points. Fig F plots the errors with

downsampled time points, plotted in terms of the time interval  $\Delta t = 150min/(\# \text{ time points} - 1)$ , and shows that the error in the predicted noise intensity is fairly insensitive to reducing the number of time points, down to about 6 time points (30 minute intervals). By using a group of 32 trajectories, inference of the intrinsic noise intensity, allows a smaller number of observed time points. In our example, errors increase appreciably when fewer than three time points are used.

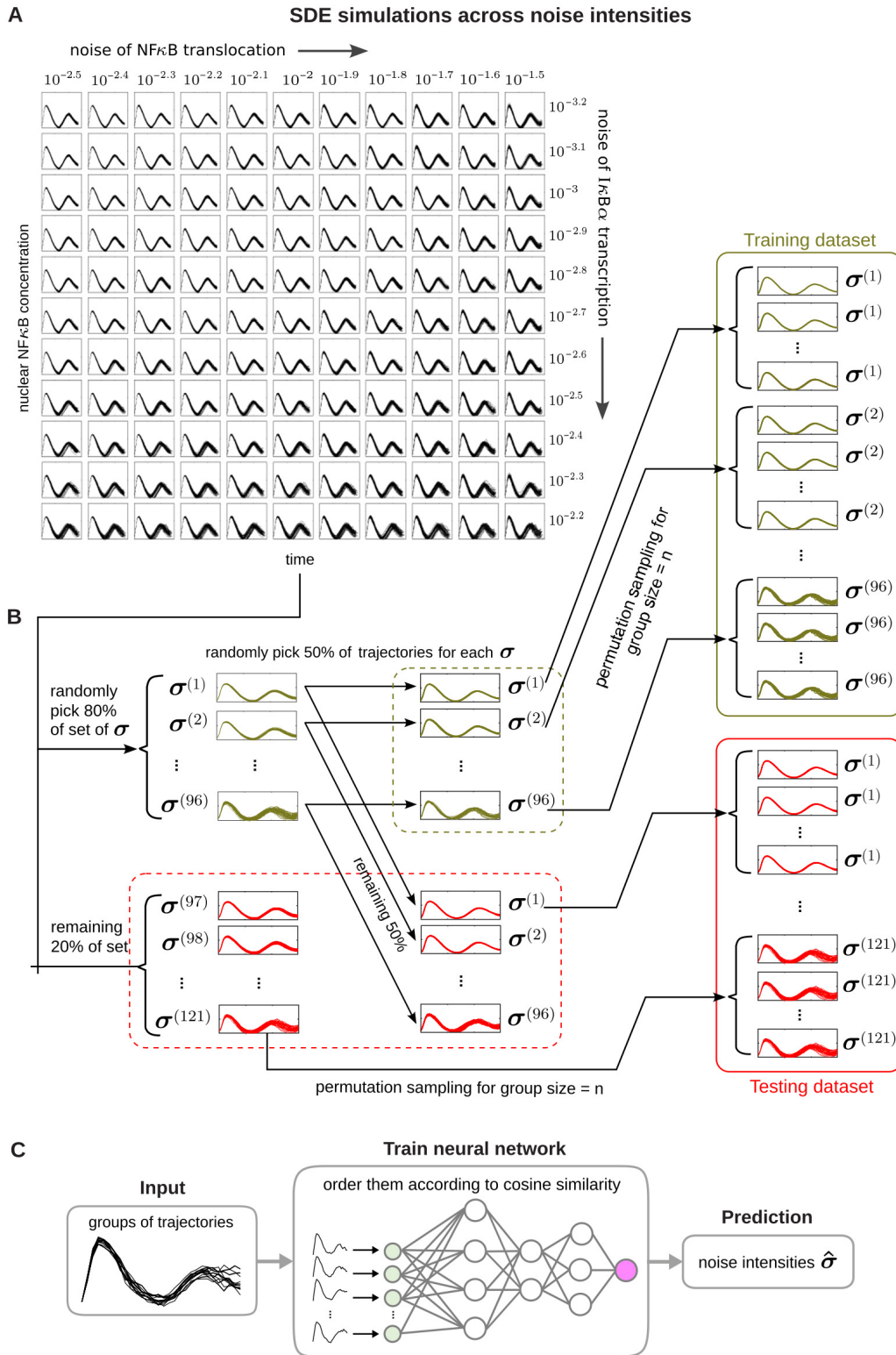


Figure E: **Workflow of using neural network to infer noise intensities from a group of trajectories.** A.  $\text{NF}\kappa\text{B}$  SDE simulations under 121 different noise intensity settings. B. Schematic workflow for splitting the dataset into training and testing sets, where each group of trajectories is used to train and test the neural network for noise prediction. C. Schematic of the training of a neural network to predict noise intensities for each group of trajectories.

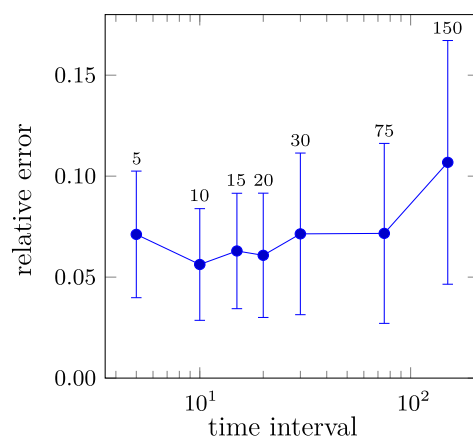


Figure F: **Neural network performance in inferring noise intensities from downsampled trajectories.** Relative error and uncertainty of inferred noise as a function of the trajectory sampling interval. Experimental trajectories were originally measured every 5 minutes over a 150-minute period; these measurements were then divided into trajectory groups for noise-intensity inference (Fig 8, Step 2). To assess the effect of longer intervals, the same dataset was downsampled to 10, 15, 20, 30, 75, and 150 minute sampling intervals.