



Learning stochastic processes with intrinsic noise from cross-sectional biological data

Suryanarayana Maddu^{a,1,2} , Victor Chardès^{a,1,2} , and Michael J. Shelley^{a,b}

Edited by Risto Nieminen, Aalto-yliopisto Teknillisen fysiikan laitos, Espoo, Finland; received October 10, 2024; accepted July 24, 2025

Inferring dynamical models from data continues to be a significant challenge in computational biology, especially given the stochastic nature of many biological processes. We explore a common scenario in omics, where statistically independent cross-sectional samples are available at a few time points, and the goal is to infer the underlying diffusion process that generated the data. Existing inference approaches often simplify or ignore noise intrinsic to the system, compromising accuracy for the sake of optimization ease. We circumvent this compromise by inferring the phase-space probability flow that shares the same time-dependent marginal distributions as the underlying stochastic process. Our approach, probability flow inference (PFI), disentangles force from intrinsic stochasticity while retaining the algorithmic ease of ordinary differential equation (ODE) inference. Analytically, we prove that for Ornstein–Uhlenbeck processes the regularized PFI formalism yields a unique solution in the limit of well-sampled distributions. In practical applications, we show that PFI enables accurate parameter and force estimation in high-dimensional stochastic reaction networks, and that it allows inference of cell differentiation dynamics with molecular noise, outperforming state-of-the-art approaches.

generative modeling | probability flow | intrinsic noise | single-cell omics | gene regulation

1. Introduction

From gene expression (1, 2), collective motion in animal groups (3, 4), to growth in ecological communities (5), the behavior of biological processes is driven by a dynamic interplay between deterministic mechanisms and intrinsic noise. In these systems, stochasticity plays a pivotal role, often leading to outcomes that diverge significantly from those predicted by deterministic frameworks (4, 6, 7). To elucidate the respective roles of deterministic forces and intrinsic noise, many inference approaches exploit time-correlations in high-resolution stochastic trajectories (8, 9). Comparatively, inferring biological processes at a single-cell level presents a unique challenge: Due to the destructive nature of the single-cell omics measurements, we can only observe statistically independent cross-sectional samples from the latent stochastic process (10, 11). This limitation imposes fundamental constraints on the identifiable dynamics and makes it difficult to disentangle deterministic forces from intrinsic noise (12, 13).

In this study, we investigate this issue for systems that can be modeled as diffusion processes (8), for which the continuous-time evolution of the degrees of freedom $\mathbf{x} \in \mathbb{R}^d$ follows a stochastic differential equation (SDE) (14). Under the assumption that there are no exogenous factors or unobserved variables driving the evolution of the system, we consider autonomous Itô processes of the form

$$d\mathbf{x} = \mathbf{f}(\mathbf{x})dt + \sqrt{2}\mathbf{G}(\mathbf{x})d\mathbf{W}, \quad [1]$$

where \mathbf{W} is a standard d -dimensional Wiener process, $\mathbf{f} : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is a deterministic force and $\mathbf{G} : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$ the intrinsic noise model. Many molecular processes in cells have been shown to follow negative binomial statistics (15, 16), revealing a strong correlation between the amplitude of regulation and intrinsic noise. As a result, a biologically plausible intrinsic noise model $\mathbf{G}(\mathbf{x})$ can be force and state-dependent.

The formulation via Eq. 1 in terms of stochastic trajectories $\{\mathbf{x}(t), t \geq 0\}$ is equivalent to a formulation in terms of the transition probability $p(\mathbf{x}, t|\mathbf{y}, s)$, which describes the probability to reach the state \mathbf{x} at time t , having started at \mathbf{y} at time s (14). Its evolution obeys the Kolmogorov forward equation,

$$\partial_t p(\mathbf{x}, t|\mathbf{y}, s) = -\nabla \cdot \left[\mathbf{f}(\mathbf{x})p(\mathbf{x}, t|\mathbf{y}, s) - \nabla \cdot (\mathbf{D}(\mathbf{x})p(\mathbf{x}, t|\mathbf{y}, s)) \right], \quad [2]$$

Significance

High-throughput omics datasets provide detailed snapshots of the cellular transcriptome, but their destructive nature prevents tracking the time evolution of a cell's molecular state. This limitation, combined with the stochastic nature of cellular processes, poses a significant challenge for interpreting and modeling the variability in omics data. To address these challenges, we propose Probability Flow Inference, a method that fits the deterministic flow of probability in phase space to time-resolved cross-sectional data. On model examples, we show that the regularized probability flow formulation yields a unique solution, and can account for stochasticity of arbitrary structure. This enables the inference of biophysically informed stochastic models of cellular differentiation directly from cross-sectional omics data.

Author affiliations: ^aCenter for Computational Biology, Flatiron Institute, New York, NY 10010; and ^bCourant Institute of Mathematical Sciences, New York University, New York, NY 10012

Author contributions: S.M., V.C., and M.J.S. designed research; performed research; contributed new reagents/analytic tools; analyzed data; and wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2025 the Author(s). Published by PNAS. This article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

¹S.M. and V.C. contributed equally to this work.

²To whom correspondence may be addressed. Email: smaddu@flatironinstitute.org or vchardes@flatironinstitute.org.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2420621122/-DCSupplemental>.

Published September 11, 2025.

for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, $t, s \geq 0$, with $\mathbf{D} = \mathbf{G}\mathbf{G}^T \in \mathbb{S}_+^d$. When time-resolved trajectories are available, both the force field and the diffusion can be inferred simultaneously by fitting either of these equations to the data (8). To simplify computations, most approaches rely on discretizing Eq. 1 rather than fitting transition probabilities with Eq. 2, with successful applications in fields such as soft matter and finance (8, 17–19).

However, with single-cell omics data, the lack of trajectory information makes it impossible to reconstruct the transition probabilities $p(\mathbf{x}, t|\mathbf{y}, s)$. Within this setting, it is more appropriate to model the evolution of marginal distributions over time with the Fokker–Planck equation

$$\partial_t p_t(\mathbf{x}) = -\nabla \cdot [\mathbf{f}(\mathbf{x})p_t(\mathbf{x}) - \nabla \cdot (\mathbf{D}(\mathbf{x})p_t(\mathbf{x}))], \quad [3]$$

which is obtained by marginalizing Eq. 2 over an initial condition $p_0(\mathbf{y})$. The inverse problem now reduces to learning how the probability mass is moved between empirical distributions at successive time points rather than how one trajectory evolves. Unlike trajectory-based methods, it is no longer possible to infer both the force and noise models simultaneously, requiring a prior on one to infer the other.

Approaches based on optimal transport have been used to tackle this question, first in static settings by learning pairwise couplings between successive empirical distributions, and subsequently in dynamical settings by learning a time-continuous model connecting distributions at all times. While static methods cannot model time-continuous and nonlinear dynamics (11, 20–23), their dynamical counterparts lift these constraints, but all methods remain limited to additive priors on the noise model (24–27). Other approaches that integrate trajectory data with mechanistic differential equation models also use deterministic frameworks (28) or assume additive noise (29, 30). Even popular methods that infer cell-fate directionality from messenger RNA splicing and spatial transcriptomics data also adopt additive noise models for parameter estimation (31). However, force and state-dependent noise models not only better capture biological variability, but they also have the capacity to shift, create, or eliminate fixed points in the energy landscape, which is of paramount importance to model processes like cell differentiation (1, 2, 6). This underscores the need for inference methods that accommodate molecular noise while retaining algorithmic simplicity.

We introduce Probability Flow Inference (PFI), a method that separates the force field from any intrinsic noise model while retaining the computational efficiency of ODE optimization. First, by analytically solving the PFI approach for Ornstein–Uhlenbeck (OU) processes, we prove that the inverse problem admits a unique solution when the force field is regularized, and that without regularization, it can lead to infinitely many solutions. By focusing on a subclass of isotropic OU processes, we show that all nonconservative force contributions can be recovered in the limit of vanishingly small regularization. We contrast this observation by demonstrating the existence of a bias–variance trade-off coming from finite sampling of the cross-sectional data, thereby advocating for a nonzero regularization. Finally, with the same analytical solution, we show that a good prior on intrinsic noise is necessary to accurately infer the underlying force field.

Building on these theoretical insights, we apply the PFI approach to stochastic reaction networks, widely used to model cellular processes. We show that PFI reliably infers the continuous diffusion approximation of these discrete-state stochastic processes, with a particular focus on estimating rate parameters

and inferring gene regulatory networks. In agreement with our theoretical results for the OU processes, we underscore the importance of an informed biophysical prior on the noise model to achieve accurate force inference. Finally, using a curated hematopoietic stem cell model, we demonstrate that PFI outperforms state-of-the-art generative models in inferring cell differentiation dynamics and predicting *in silico* gene knockdown perturbations.

2. Methodology

2.1. Problem Statement. We assume that the data is given as $K + 1$ statistically independent cross-sectional snapshots, each composed of n_k samples, taken from the true process at successive times $t_0 = 0 < \dots < t_k < \dots < t_K = T$, with uniform spacing Δt . We further assume that the n_k i.i.d. samples $\{\mathbf{x}_{i,t_k} | 1 \leq i \leq n_k\}$ are measured at each time t_k from the true marginals, giving access to an empirical estimator of p_{t_k} ,

$$p_{t_k}(\mathbf{x}) \approx \frac{1}{n_k} \sum_{i=1}^{n_k} \delta(\mathbf{x} - \mathbf{x}_{i,t_k}). \quad [4]$$

The objective is to infer the force field $\mathbf{f}(\mathbf{x})$ associated with a latent stochastic process interpolating the observed marginals. Since it is not possible to simultaneously infer both the force field and diffusion from statistically independent cross-sectional samples, we impose strong priors on the noise model. For example, in the case of molecular noise arising from chemical reactions, the diffusion tensor $\mathbf{D}(\mathbf{x})$ is force and state dependent and follows a known functional form (32). We leverage such intrinsic noise priors to accurately infer the force fields and reconstruct the underlying stochastic dynamics.

2.2. PFI. A common approach to inferring the force field $\mathbf{f}(\mathbf{x})$ is to fit the SDE by minimizing a distance metric between the observed empirical marginals and the generated samples, in a predict and correct manner (33, 34). However, this optimization is both memory- and compute-intensive and is limited to additive or diagonal noise models (34). As previously noted, we can improve on this by observing that the SDE formulation Eq. 1 contains redundant information for modeling the evolution of marginals, and that Eq. 3 is sufficient.

Rather than fitting the SDE, we instead fit the Fokker–Planck equation directly. This can be accomplished using its Lagrangian-frame formulation, known as the probability flow ODE (35), which reads:

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}) - \nabla \cdot \mathbf{D}(\mathbf{x}) - \mathbf{D}(\mathbf{x})\nabla \log p_t(\mathbf{x}). \quad [5]$$

The term $\nabla \log p_t(\mathbf{x})$, or the gradient of the log-probability of the marginals, is known as the score (36). Solving this ODE from the initial condition $p_0(\mathbf{x})$ generates samples from the same marginal distributions $p_t(\mathbf{x})$ as the underlying SDE. Notably, this property has been used to simulate the Fokker–Planck equation (37, 38), as well as to estimate entropy production in active matter systems (39). The PFI approach consists of two steps: i) estimating the score function, and ii) fitting the probability flow ODE Eq. 5 to the observed marginals. We now outline these two steps:

2.2.1. Score estimation. The first step of the PFI approach requires estimating the time-dependent score function from empirical samples at various time points. To efficiently solve this task, we leverage recent advancements in generative modeling

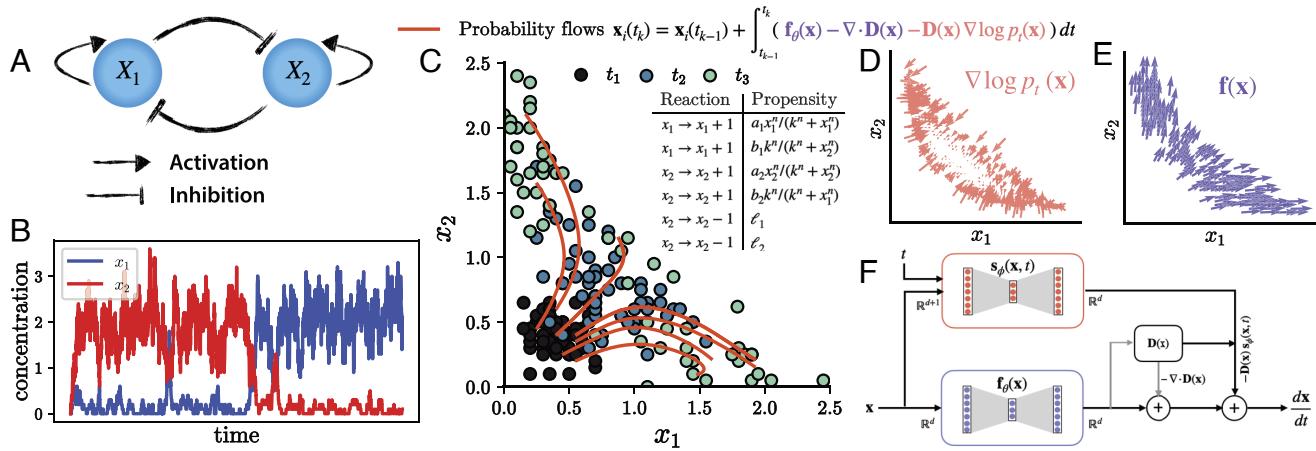


Fig. 1. Probability flow inference. (A) Dynamics of the canonical fate decision regulatory circuit of two mutually opposing transcription factors (x_1, x_2) that positively self-regulate themselves. (B) Time series generated from direct stochastic simulation (Gillespie algorithm) of the toggle-switch model showing the concentration levels of x_1 (blue) and x_2 (red), illustrating their dynamics over time. The Gillespie simulation was run with parameters $a_1 = a_2 = b_1 = b_2 = k = 1.0$, and $n = 4$. The degradation rates are $\ell_1 = \ell_2 = 1$. (C) Cross-sectional snapshots generated from the discrete Gillespie simulations at time points $\{t_1, t_2, t_3\}$ represent the marginals' evolution through the stochastic process. Characteristics lines (in red) show the probability flows deforming the initial state t_1 to a future state t_3 in the state space, illustrating the dynamic transition of the system's state. The *Inset* describes the reactions and associated propensities of the two-gene fate decision circuit shown in (A). Score (*D*) and force (*E*) approximation at time t_1 . (F) Network architecture of our probability flow inference (PFI) approach. Both the autonomous force field $f_\theta(\mathbf{x}) : \mathbb{R}^d \rightarrow \mathbb{R}^d$ and the time-dependent score model $s_\phi(\mathbf{x}, t) : \mathbb{R}^{d+1} \rightarrow \mathbb{R}^d$ are parameterized using feedforward networks, with parameters θ and ϕ , respectively.

that allow fast and accurate score estimation in high dimensions (35, 40). Specifically, we use sliced score matching (*Materials and Methods*) to train a score network $s_\phi(\mathbf{x}, t)$ that approximates $\nabla \log p_t(\mathbf{x})$.

2.2.2. Force inference. Once an accurate score model is available, we seek to fit the force via Eq. 5, following the *predict and correct* strategy outlined in Alg. 1. Using the estimated score function and an initial guess of the force field, we push the observed samples from time t_k to t_{k+1} using the probability flow ODE in Eq. 5, generating a predicted distribution $\hat{p}_{t_{k+1}}(\mathbf{x})$ as defined in Eq. 4. The force field is then optimized to minimize a distance \mathcal{D} between the predicted distribution and the observed distributions. In practice, we minimize the total discrepancy across all cross-sectional time points, i.e., $\hat{\mathbf{f}} = \arg \min_{\mathbf{f}} \sum_{i=1}^K \mathcal{D}(\hat{p}_{t_i}, p_{t_i})$.

In Fig. 1, we illustrate the two step PFI approach applied to a bistable genetic switch system (Fig. 1 A and B). The probability flows in transcription factor concentration space (Fig. 1C) have two components: one derived from the estimated score (Fig. 1D) and the other from the force field (Fig. 1E). Finally, as shown in Fig. 1F, both the score and the force field can be parameterized using feedforward neural networks.

In many cases, it is impossible to uniquely identify a force field matching the observed marginals. To see this, consider a force field $\hat{\mathbf{f}}(\mathbf{x}) = \mathbf{f}(\mathbf{x}) + \mathbf{h}(\mathbf{x})$. Using Eq. 3, we see that $\mathbf{f}(\mathbf{x})$ and $\hat{\mathbf{f}}(\mathbf{x})$ generate the same marginals $p_t(\mathbf{x})$ if $\mathbf{h}(\mathbf{x})$ satisfies

$$\nabla \cdot (\mathbf{h}(\mathbf{x}) p_t(\mathbf{x})) = 0, \forall t \geq 0, \forall \mathbf{x}. \quad [6]$$

This equation can have multiple nongradient solutions (*SI Appendix, Text 1*). For instance, if $p_t(\mathbf{x})$ has radial symmetry in \mathbf{x} , any force field $\mathbf{h}(\mathbf{x}) = \mathbf{K}\mathbf{x}$, with \mathbf{K} a skew-symmetric matrix, is a solution of Eq. 6. This lack of uniqueness, referred to as the identifiability issue, has been of long-standing concern in the analysis of single-cell RNA-seq data (6, 13, 33).

To circumvent this ill-posedness, we choose to introduce an ℓ_2 -regularization on the Jacobian matrix of the inferred force field to the loss function which reads

$$\mathcal{L}_{\Delta t, K} = \sum_{i=1}^K \left[\mathcal{W}_2^2(\hat{p}_{t_i}(\mathbf{x}), p_{t_i}(\mathbf{x})) + \lambda \Delta t \int_{t_{i-1}}^{t_i} \int \|\nabla \hat{\mathbf{f}}(\mathbf{x})\|_F^2 p_{t_i}(\mathbf{x}) d\mathbf{x} dt \right]. \quad [7]$$

Here, $\lambda \geq 0$ is a tunable parameter that controls the strength of regularization, and we use the Wasserstein distance \mathcal{W}_2 to quantify the discrepancy between distributions. Though the above regularization penalizes the gradient of the force field, a regularization minimizing the kinetic energy is also appropriate (24).

Computing the exact Wasserstein distance \mathcal{W}_2 requires $\mathcal{O}(n^3 \log n)$ operations, and its estimation in d dimensions has a sample complexity of $\mathcal{O}(n^{-1/d})$. Its entropy-regularized version, known as the Sinkhorn divergence, reduces the computational cost to $\mathcal{O}(n^2)$ with a dimension-independent sample complexity $\mathcal{O}(n^{-1/2})$ for large entropic regularization (41, 42). In all our numerical examples, we use the Sinkhorn divergences to approximate the Wasserstein distance. Overall, the PFI approach allows for the use of accurate forward solvers (43) and constant-memory gradient computations (44) to fit the force field. More broadly, and relevant to PFI, reverse-mode automatic differentiation tools now allow the fitting of ODEs with millions of parameters to data (44), enabling more flexible approaches to density estimation and time series modeling (45, 46). Before stepping into numerical examples, in the next section, we study analytically the identifiability issue for OU processes.

Analytical Case Study: Ornstein-Uhlenbeck Process. Linear models are a very popular choice for gene regulatory network inference (47), and reconstructing cellular dynamics based on RNA velocity (48, 49). Despite their frequent use for such inference tasks, the challenges related to the identifiability issue, the role of regularization, and various sources of error have not been addressed. In this section, we tackle this issue in a continuous-time limit, $\Delta t \rightarrow 0$, with which we establish a

uniqueness result for the inferred process in the presence of regularization. For this purpose, we assume that the underlying latent process to be inferred is a d -dimensional OU process with an interaction matrix Ω . That is,

$$d\mathbf{x} = \Omega \mathbf{x} dt + \sqrt{2\mathbf{D}} d\mathbf{W}, \text{ with } \mathbf{x}_0 \sim p_0(\mathbf{x}),$$

where Ω has eigenvalues with strictly negative real part. When $p_0 = \mathcal{N}(\mathbf{m}_0, \Sigma_0)$, the solution of the OU process is Gaussian at all times with $\mathbf{x}_t \sim \mathcal{N}(\mathbf{m}_t, \Sigma_t)$, with Σ_t and \mathbf{m}_t being, respectively, the covariance and mean of the process at time t (50). We assume that Σ_0 is full rank, so that Σ_t is positive definite at all later times (50). The covariance matrix Σ_t can be decomposed as $\Sigma_t = \sum_i \sigma_{i,t}^2 \mathbf{w}_{i,t} \mathbf{w}_{i,t}^T$, where $\sigma_{i,t}$ are its eigenvalues and $\mathbf{w}_{i,t}$ are the corresponding eigenvectors. In this problem, we restrict the inferred force model to be linear, $\hat{\mathbf{f}}(\mathbf{x}) = \hat{\Omega}\mathbf{x}$, and the diffusion tensor to be a given constant $\hat{\mathbf{D}}$, perhaps previously estimated. We present an analytical form of the loss function (Eq. 7) as a function of $\hat{\Omega}$ in the continuous-time limit ($\Delta t \rightarrow 0$) and large sample limit ($n \rightarrow \infty$).

2.2.3. Continuous-time loss function. With the following theorem, we prove in *SI Appendix, Text 2*, that the loss in Eq. 7 converges to a strongly convex loss for $\lambda > 0$.

Theorem 1. With $K = \lfloor T/\Delta t \rfloor$, when $n \rightarrow \infty$ and $\Delta t \rightarrow 0$, the loss function $\mathcal{L}_{\Delta t, K}/\Delta t \rightarrow \mathcal{L}$ with

$$\begin{aligned} \mathcal{L} = & \text{tr}((\hat{\Omega} - \Omega)\mathbf{P}(\hat{\Omega} - \Omega)^T + \lambda T \hat{\Omega} \hat{\Omega}^T) \\ & + \int_0^T \sum_{i,p} \frac{\sigma_{i,t}^2}{(\sigma_{i,t}^2 + \sigma_{p,t}^2)^2} \left(\mathbf{w}_{i,t}^T (\sigma_{p,t}^2 (\hat{\Omega} - \Omega)) \right. \\ & \left. + \sigma_{i,t}^2 (\hat{\Omega}^T - \Omega^T) + 2(\hat{\mathbf{D}} - \mathbf{D}) \right) \mathbf{w}_{p,t}^T dt, \end{aligned} \quad [8]$$

$$\text{where } \mathbf{P} = \int_0^T \mathbf{m}_t \mathbf{m}_t^T.$$

For $\lambda > 0$ this loss function is strongly convex, and thus has a unique minimum on \mathbb{R}^d .

In *SI Appendix, Text 2C*, we also provide a version of this theorem for the Sinkhorn divergence with entropic regularization ϵ , of which Eq. 8 is the particular case when $\epsilon \rightarrow 0$. In the discrete-time setting, the asymptotic loss function \mathcal{L} of Eq. 8 is achieved once Δt is sufficiently small to resolve all the relevant timescales. More specifically, when $\Omega_{\max} \Delta t, D \Delta t / \sigma_{\min} \ll 1$, where Ω_{\max} denotes the largest singular value of Ω and σ_{\min} is the smallest eigenvalue of Σ_0 . The λ regularization ensures that the objective function is strongly convex, as without regularization, we can face the identifiability issue.

2.2.4. An analytically solvable isotropic process. To gain more insight into the role of the regularization, we simplify our model by considering an isotropic process, where $\Sigma_0 = \sigma_0^2 \mathbf{I}$, $\mathbf{D} = D \mathbf{I}$, and $\Omega = \Omega_s \mathbf{I} + \Omega_a$, with Ω_a a skew-symmetric matrix. As a result, the covariance matrix is isotropic at all times, i.e. $\Sigma_t = \sigma_t^2 \mathbf{I}$ (50). For such an isotropic OU process the nonconservative and conservative forces correspond to the skew-symmetric and the symmetric parts of Ω , respectively. The nonconservative force $\Omega_a \mathbf{x}$ generates rotations around the origin $\mathbf{x} = 0$, while the conservative part $\Omega_s \mathbf{I} \mathbf{x}$ induces inward flows.

We illustrate the PFI approach with a numerical example of an isotropic OU process in $d = 10$ dimensions, using $K = 10$ snapshots, $n = 8,000$ samples and with $\lambda = 10^{-2}$. Fig. 2A shows the evolution of the true OU process projected along the two leading eigendirections of \mathbf{P} , with the true probability flow (dashed red lines) interpolating the successive Gaussian marginal distributions (shaded discs). The probability flow lines spiral toward the origin under the combined effect of the nonconservative and conservative forces, while diffusion is reflected in the spreading over time of the distributions. In

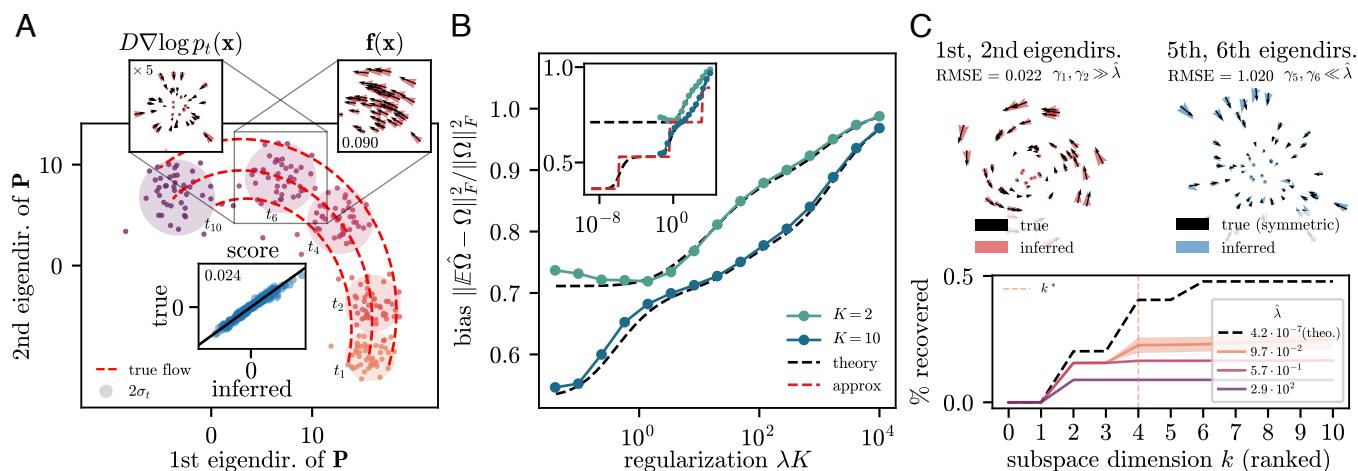


Fig. 2. Theoretical insight for the inference of an Ornstein–Uhlenbeck process. (A) Illustration of the inference process: Samples are pushed from at time t_i to time t_{i+1} using the PF ODE. The inferred score and inferred force field are shown (red arrows) overlaid on the true score and the true force field (dark arrows) at time t_6 . The inferred force has an error of 9% (computed over the whole state-space). As shown in *Inset* for t_6 , the score inference is also accurate with an error of 2.4%. (B) Bias of the inferred interaction matrix as a function $\lambda = \lambda K$. In dashed lines are shown the theory (dark) and the dots correspond to numerical minimization of Eq. 7 using a precomputed score model. In *Inset* the bias is shown for a wider range of λ , and the dashed red line shows the approximate solution. (C) In the two *Upper* panels are shown, at $\lambda \simeq 10^{-1}$, projections of the inferred force field (centered in $\mathbf{x} = 0$) in the first and second eigendirections of \mathbf{P} (red arrows), and in the fifth and sixth (blue arrows). In the leading eigendirections, the force field inferred matches the true force field, while as the weaker eigendirections, the inference recovers the symmetric part of the true force field. In the *Lower* panel is shown the % of true skew-symmetric interactions recovered in the k -leading eigendirections of \mathbf{P} , as measured by $\|\mathbf{Q}_k \hat{\Omega}_a \mathbf{Q}_k\|_F^2 / \|\Omega_a\|_F^2$, where \mathbf{Q}_k is the projector on subspace spanned by the k eigenvectors associated with the k largest eigenvalues of \mathbf{P} . For $\lambda \simeq 10^{-1}$, beyond the effective dimension $k^* = 4$ the inference only infers equilibrium (here symmetric) dynamics. The parameters used are $d = 10, n = 8,000, \|m_0\| = 20, \Sigma_0 = 1, D = 8, \Omega_s = 2, \Delta t = 0.05, \Omega_a = 3\Omega_s \mathbf{A}$, with \mathbf{A} a skew-symmetric matrix is chosen at random (*Materials and Methods*). The score is precomputed with the same samples (*SI Appendix, Text 4*).

this example, the relative strength of the nonconservative and conservative forces is 3 to 1, while the time scales of diffusion and nonconservative forces are comparable (as measured by the ratio $\tau_{\text{force}}/\tau_{\text{diff}} = D/(\Sigma_0 \Omega_{\max})$). The score estimated by sliced score-matching (*Materials and Methods*) is accurate (RMSE of $\sim 2.5\%$), as shown in the *Inset* by comparing it to its analytical prediction $\nabla \log p_t(\mathbf{x}) \sim -\Sigma_t^{-1}(\mathbf{x} - \mathbf{m}_t)$. Using this score model, we infer the force using the Gaussian Wasserstein estimator to avoid the curse of dimensionality associated with the empirical Wasserstein distance (41). At time t_6 , we overlay the inferred force field on the true force field, and find that we accurately infer the force with RMSE $\sim 9\%$.

Because the covariance matrix is diagonal at all times, the skew-symmetric part can only be inferred from the motion of the mean of the distribution \mathbf{m}_t , which is encoded in \mathbf{P} . Decomposing $\mathbf{P} = \sum_i \gamma_i \mathbf{u}_i \mathbf{u}_i^T$ in terms of its eigenvalues and eigenvectors, we derive in *SI Appendix, Text 2D*, an analytical formula for $\hat{\Omega}$, the minimum of the loss for $\lambda > 0$. This formula shows an excellent agreement with the PFI solution, as shown in Fig. 2B by plotting the bias $\|\mathbb{E}\hat{\Omega} - \Omega\|_F^2/\|\Omega\|_F^2$ as a function of $\tilde{\lambda} = \lambda K$ for $K \in \{2, 10\}$. We can gain real insight into the role of the regularization using an approximation to $\hat{\Omega}$:

$$\hat{\Omega} \approx \Omega - \mathbf{Q}_{\tilde{\lambda}} \Omega_a \mathbf{Q}_{\tilde{\lambda}},$$

where $\mathbf{Q}_{\tilde{\lambda}} = \sum_i \chi(\gamma_i < \tilde{\lambda}) \mathbf{u}_i \mathbf{u}_i^T$ is the projector onto the eigenspace of \mathbf{P} having eigenvalues smaller than the penalty $\tilde{\lambda}$, with χ the indicator function. This approximation, derived in *SI Appendix, Text 2D*, is shown in *Inset* of Fig. 2B (dashed red) in comparison with the analytical solution (dashed black). This solution suggests that the nonconservative force $\Omega_a \mathbf{x}$ is inferred only in the subspace spanned by $\{\mathbf{u}_i | \gamma_i \gg \tilde{\lambda}\}$ and is set to zero in its orthogonal complement. In these remaining directions, only the conservative force is correctly estimated. This idea is

further exemplified in Fig. 2C (*Upper panel*) by showing the projections of the inferred force field on eigendirections of \mathbf{P} sorted by decreasing eigenvalues γ_i . We can observe that the force field is fully estimated in the leading eigendirections, for which $\gamma_i \gg \tilde{\lambda}$, but that only the symmetric contribution is inferred when $\gamma_i \ll \tilde{\lambda}$. This observation is rationalized in Fig. 2C (*Lower panel*) where we plot the fraction of recovered skew-symmetric interactions in the subspace spanned by the k leading eigenvectors of \mathbf{P} . We see that beyond an effective dimension k^* the fraction plateaus, suggesting that the time-dependent deformations of the marginal distributions in the remaining eigendirections are indistinguishable from equilibrium dynamics.

These findings highlight the role of regularization as a recovery threshold for nonequilibrium dynamics, and taking $\lambda \rightarrow 0$ ensures exact recovery of all nonconservative forces available in the data. However, decreasing $\tilde{\lambda}$ also incurs an increased variance coming from the finite sample size n . As shown in Fig. 3A, we observe a bias–variance trade-off in the expected relative mean square error for the force. In *SI Appendix, Text 2E*, we derive the first-order, finite sample size correction to the continuous-time loss function, allowing us to estimate analytically the variance $\mathbb{E}\|\hat{\Omega} - \mathbb{E}\hat{\Omega}\|_F^2$. Given K , with all nondimensional quantities being fixed and for n large, we predict the variance to read as $C/(n\Delta t^2)$, with C a constant that quantifies the magnitude of deformations of the marginal distributions (*SI Appendix, Text 2E*). The scaling $(n\Delta t^2)^{-1}$ is in excellent agreement with the PFI solution, as shown by the collapse in Fig. 3B as Δt is varied.

Finally, our analytical solution allows us to study the error coming from the misestimation of the diffusion tensor $\hat{\mathbf{D}}$. Assuming that $\hat{\mathbf{D}} = \hat{\mathbf{D}}\mathbf{S}$, where \mathbf{S} is an arbitrary matrix in \mathbb{S}_+^d (*Materials and Methods*) with unit maximum eigenvalue, we plot in Fig. 3C the bias as a function of $\hat{\mathbf{D}}$. These results show that the bias is minimum when $\hat{\mathbf{D}} = D$, and that using a deterministic noise model, where $\hat{\mathbf{D}} = 0$, results in a larger error

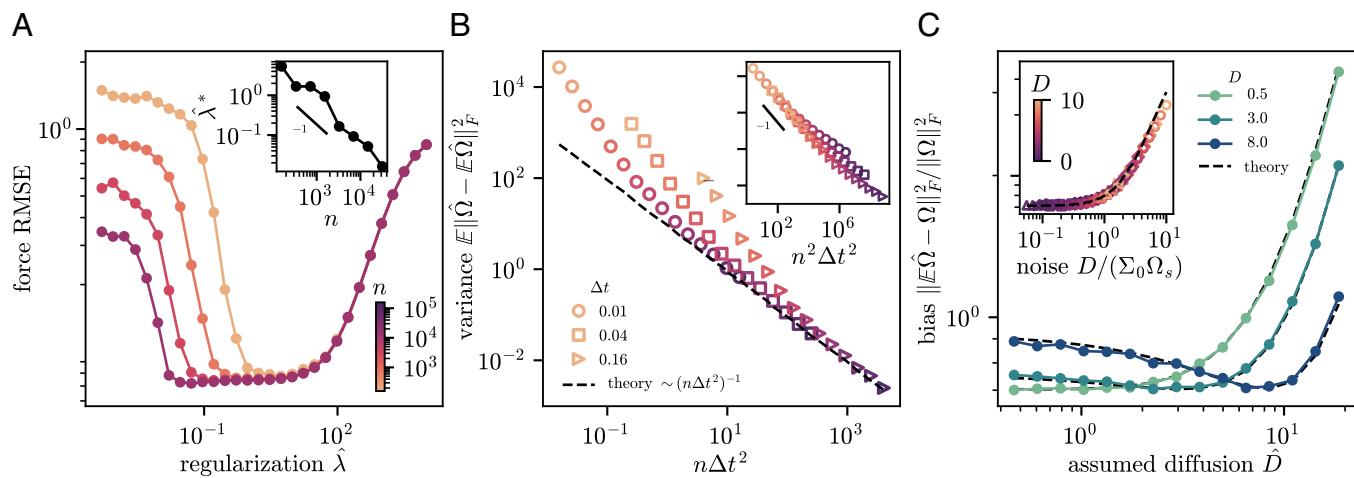


Fig. 3. Sources of error for the inference of an Ornstein-Uhlenbeck process. (A) Bias-variance trade-off as a function of the regularization $\tilde{\lambda}$ in the relative mean square error of the force field. The inflection point of the trade-off is reached at smaller $\tilde{\lambda}$ for larger sample sizes n . (B) Variance on the inferred interaction matrix as a function of $n\Delta t^2$, with $\Omega_0 \Delta t = 0.1$, $D/\Sigma_0 \Omega_0 = 4$. For various Δt the variance collapses at large n on the prediction (black dashed line). As shown in *Inset*, at smaller n the scaling is found to be $n^2 \Delta t^2$. (C) Bias on the inferred interaction matrix as a function of the assumed diffusion tensor $\hat{\mathbf{D}} = \hat{\mathbf{D}}\mathbf{S}$ with \mathbf{S} a semidefinite positive symmetric matrix with unit maximum eigenvalue (*Materials and Methods*). The bias shows a minimum when $\hat{\mathbf{D}} = D$. When $\hat{\mathbf{D}} = 0$ (which corresponds to fitting a deterministic model), the bias increases with D . This is shown in *Inset* as a function of the nondimensional strength of noise $D/(\Sigma_0 \Omega_0) \sim \frac{\tau_{\text{force}}}{\tau_{\text{diff}}}$ for different values of Σ_0 and Ω_0 . With respectively increasing values for $\Sigma_0 \in \{0.5, 2, 3\}$, square, circle, and Right triangle markers correspond to $\Omega_0 = 2$, while diamonds, Left and Upper triangle markers corresponds to $\Omega_0 = 3$. The parameter $\|m_0\|^2/\Sigma_0$ is fixed to 400 (s.t. $\|m_0\| = 20$ when $\Sigma_0 = 1$). The color map for n is shared between (A) and (B). In (B and C) the regularization is $\tilde{\lambda} = 0.2$. In the three panels $K = 5$, $\Omega_a = 3\Omega_0 \mathbf{A}$ and all the remaining free parameters (notably **A**) in each panel are the same as Fig. 2. We use the analytical prediction for the score.

as the intrinsic noise D increases. This observation is rationalized using the analytical solution (*SI Appendix*, Text 2D), which predicts that for λ small, with K and all other nondimensional quantities fixed, the bias at $\hat{D} = 0$ (which corresponds to fitting a deterministic model) is a known function of $\tau_{\text{force}}/\tau_{\text{diff}}$ the ratio of the timescales associated with the force and with the diffusion. This prediction agrees very well with the PFI solution, as shown in the *Inset* of Fig. 3C for various values of Σ_0 and Ω_s . This shows that estimating $\tau_{\text{force}}/\tau_{\text{diff}}$ is an efficient way to decide whether fitting a deterministic model is sufficient ($\tau_{\text{force}}/\tau_{\text{diff}} \ll 1$) or noise is necessary ($\tau_{\text{force}}/\tau_{\text{diff}} \gtrsim 1$).

In more general cases, it may not be possible to quantify the relative strengths of intrinsic noise and force. However, due to the Poisson nature of the chemical reactions underlying cellular processes, the intrinsic noise variance is often closely linked to the amplitude of the force (15, 16), resulting in comparable timescales for both intrinsic noise and force (i.e. $\tau_{\text{force}} \sim \tau_{\text{diff}}$). Therefore, accounting for intrinsic noise is crucial to accurately infer cellular processes from single-cell omics data. In the following, we perform numerical experiments to demonstrate how the PFI approach offers a flexible mechanism to incorporate stochasticities of arbitrary nature, and in particular molecular noise.

3. Numerical Results

In this section, we exploit the flexibility of the PFI approach to infer gene regulatory networks and model cell differentiation dynamics, incorporating molecular noise.

3.1. PFI Allows Accurate Parameter and Force Estimation for Stochastic Reaction Networks. Cellular processes are driven by an intricate array of chemical reactions (51, 52). While techniques like flow cytometry, microscopy, and high-throughput omics provide extensive data on cellular processes, interpreting these data, their variability, and estimating reaction rate constants from them requires mechanistic models. Under the assumption that the system is well mixed, the chemical master equation (CME) offers a detailed probabilistic representation of these stochastic reaction networks (SRNs) (53, 54). However, the use of CME is limited in both simulation and inference tasks due to the significant computational complexity involved in solving it (55–57). Diffusion approximations, such as the Chemical Langevin Equation (CLE), offer a computationally tractable alternative by approximating the discrete CME with a continuous diffusion process. These methods accurately capture stochastic effects at moderate molecule counts (54, 58), making them a practical and effective approximation to the CME.

However, when dealing with single-cell omics data, the stoichiometry of gene regulatory networks is unknown. In addition, simultaneous measurements of both protein and mRNA counts are generally unavailable, complicating efforts to develop detailed descriptions of the underlying stochastic regulatory networks. To address this challenge, simple coarse-grained models have been introduced that leverage the separation of timescales between transcription factor binding to regulatory DNA sites and the processes of transcription and translation (59, 60). One such model describes the stochastic evolution of mRNA counts, denoted by $\mathbf{x} \in \mathbf{R}^d$, with d the number of genes, assuming that the transcription rate of gene i is proportional to an activation function $g_i(\mathbf{x}, V)$, $0 < g_i(\mathbf{x}, V) < 1$ with V being the reaction volume, typically that of the nucleus. Under well-mixed assumptions, the stochastic mRNA dynamics are then approximated by the CLE

$$d\mathbf{x} = (mV\mathbf{g}(\mathbf{x}, V) - \ell\mathbf{x}) dt + \sqrt{mV\mathbf{g}(\mathbf{x}, V) + \ell\mathbf{x}} d\mathbf{W}, \quad [9]$$

where ℓ is the degradation rate of mRNA molecules, m is the transcription rate, and the square root is taken entrywise. This corresponds to the following deterministic force and diffusion tensor in the probability flow ODE:

$$\mathbf{f}(\mathbf{x}) = mV\mathbf{g}(\mathbf{x}, V) - \ell\mathbf{x},$$

$$\mathbf{D}(\mathbf{x}) = \frac{1}{2} \text{diag}(mVg_1(\mathbf{x}, V) + \ell x_1, \dots, mVg_d(\mathbf{x}, V) + \ell x_d).$$

Although the CLE in Eq. 9 is a simplified approximation to the underlying SRNs governing gene regulation (*Materials and Methods*), it has been shown to quantitatively reproduce experimental steady-state single-cell transcriptomics profiles for known gene regulatory networks (61). By leveraging the CLE formulation in Eq. 9, the PFI approach can be readily applied to infer parameters and forces in high-dimensional SRNs. To demonstrate this, we consider the Mammalian Cortical Area Development (mCAD) gene regulatory network (see Fig. 4, $d = 5$) and use the boolODE framework (62) to compute $\mathbf{g}(\mathbf{x})$. To generate the marginal data, we conduct a detailed simulation of the mCAD stochastic reaction network using the Gillespie algorithm (*Materials and Methods*), producing $K = 10$ snapshots, each containing $n = 6,000$ samples.

We begin the PFI procedure by training a score network $s_\phi(\mathbf{x}, t)$ to estimate the score function from the marginal data in concentration space \mathbf{x}/V (*SI Appendix*, Fig S2). Using the PFI approach, we fit the Eq. 9 to the cross-sectional data by minimizing the loss function of Eq. 7. We parameterize the force using a feed-forward neural network consisting of four fully connected layers, each with 50 nodes and smooth ELU activation (*Materials and Methods*). The explicit regularization parameter is set to $\lambda = 10^{-4}$. To compare the predicted and measured distributions, we apply the Sinkhorn divergence with $\epsilon = 0.1$. To assess the usefulness of the CLE description, we compare the results with those obtained by using common models for the intrinsic noise: i) additive, ii) simple state-dependent ($\sqrt{\mathbf{x}}$), and iii) deterministic (*Materials and Methods*). The explicit forms of the diffusion tensor $\mathbf{D}(\mathbf{x})$ for different noise models are discussed in *SI Appendix*, Table S1. Fig. 4A shows the errors in force fields under these noise models and for increasing reaction volumes V . As expected, because the system is well mixed, at large volumes V all the models become deterministic and perform equally well. However, across all reaction volumes, the CLE yields the most accurate force field estimates, outperforming both the state-dependent and additive noise models. Its performance degrades at small volumes when the CLE approximation breaks down.

These differences in inferred dynamics between noise models are yet better quantified by examining the estimated Jacobian matrix ($\nabla \hat{\mathbf{g}}(\mathbf{x})$), which can be used to directly infer the gene regulatory network. In Fig. 4D, we compare the average estimated Jacobian matrices for each diffusion model to the analytical Jacobian matrix (Fig. 4C). Only the CLE model accurately identifies regulatory interactions, as indicated by AUC values approaching unity, based on the precision–recall curves (Fig. 4E) (62). Conversely, non-CLE noise models yield an AUC close to 0.5, indicating incorrect identification of interactions in the network. This result shows that while the non-CLE noise models can attain moderate errors of ~ 10 to 20% at intermediate volumes, they do so by learning a completely inaccurate regulatory network.

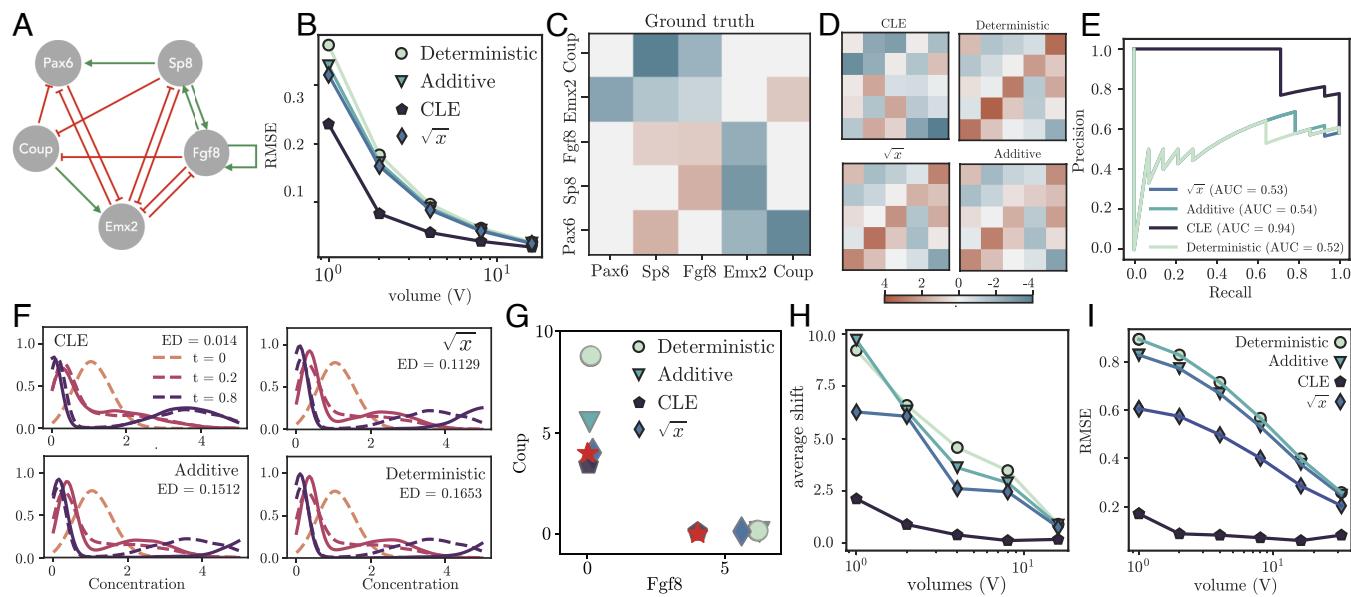


Fig. 4. Parameter and force estimation for stochastic reaction networks. (A) mCAD gene regulatory network. Red lines denote inhibition and green arrows activation. (B) The RMSE $\|\mathbf{g}(\mathbf{x}) - \hat{\mathbf{g}}(\mathbf{x})\|_2^2 / \|\mathbf{g}(\mathbf{x})\|_2^2$ of the inferred force shown for different compartment volumes V . The Jacobian corresponding to the true force field $\mathbf{f}(\mathbf{x})$ (C), with estimated Jacobians $\sqrt{v}\hat{\mathbf{g}}(\mathbf{x})$ obtained under different noise models shown in (D). The associated precision-recall (PR) curves are displayed in (E), with the area under the curve (AUC) values reported in the Inset. (F) Comparison of empirical marginals from the inferred diffusion process (solid) and Gillespie simulations (dashed) for $V = 4$, under various noise models (Inset). The Inset also reports the Energy Distance (ED) between predicted marginals and Gillespie simulations for each noise model. (G and H): Visualization of fixed points, which are solutions to the equation $\hat{\mathbf{g}}(\mathbf{x}) - \mathbf{x} = 0$, where $\hat{\mathbf{g}}(\mathbf{x})$ represents the force fields inferred under different noise models. Different markers indicate the fixed points derived from the corresponding noise models, plotted against the *Pax6* and *Coup* genes. Results are shown for two reaction volumes: $V = 4$ (G) and $V = 16$ (H). The red star represents the true fixed point (Ground Truth, \mathbf{x}^*), while the marginal densities at the initial state $p_0(\mathbf{x})$ (orange) and final state $p_T(\mathbf{x})$ (blue) are shown as contour plots. (I) The average shift $\|\mathbf{x} - \mathbf{x}^*\|_2$ as a function of reaction volume V , for different noise models.

This observation should reflect a poor generalization performance of the non-CLE models. To test this, we ran the inferred process using initial conditions different from those in the training data (*Materials and Methods*). To quantify generalization, we calculated the Energy Distance (Inset of Fig. 4F) between the predicted marginals from various diffusion models and the ground-truth stochastic simulations. These results show that the inferred CLE process accurately tracks the true marginals, again outperforming the other noise models. This underscores the importance of the inferred force model's accuracy in determining generalization performance.

Furthermore, we explored the dynamical behavior of these inferred force models by deterministically evolving and plotting the corresponding fixed points obtained for initial conditions sampled from the marginal distribution $\hat{p}_0(\mathbf{x})$. As molecular noise in the data increases, the fixed points for non-CLE force fields deviate from the ground truth, as shown in Fig. 4 G–I. These findings complement earlier one-dimensional studies (6), which showed that multiplicative noise can shift, create, or eliminate fixed points. Our results demonstrate this effect in a high-dimensional inverse setting, underscoring the importance of selecting appropriate noise priors to accurately capture the system's underlying dynamical properties.

Finally, in the scenario where the stoichiometry of the SRN is known, we can apply the PFI approach to estimate reaction rates in a high-dimensional setting. We consider a linear cyclic network consisting of $d = 30$ species and $R = 30$ reactions (*Materials and Methods*). In this constrained setting, the results are consistent with those observed for the mCAD network. The CLE approximation yields more accurate parameter estimates and better generalization performance compared to other noise models, as shown in *SI Appendix*, Fig S5.

3.2. PFI Enables Inference of Cell Differentiation Dynamics With Molecular Noise.

Cell differentiation dynamics arise from a complex interplay between deterministic mechanisms and stochastic fluctuations. However, many popular computational approaches for modeling differentiation are either purely deterministic or rely on oversimplified noise models. In this section, we examine how biological noise affects the predictive accuracy of generative models of cellular differentiation, and compare with the CLE model inferred using the PFI approach (PFI-CLE). While various methods exist, we focus on two prominent techniques: the TrajectoryNet framework, based on dynamical optimal transport (24), and the PRESCIENT technique (63), which models diffusion through a global potential function. TrajectoryNet fits a nonautonomous force field $\mathbf{f}(\mathbf{x}, t)$ to interpolate between the marginals data with minimal kinetic energy, while PRESCIENT models differentiation dynamics as a diffusion process, with the force field $\hat{\mathbf{f}}(\mathbf{x}) = -\nabla\hat{\Phi}(\mathbf{x})$ and additive noise $\hat{\mathbf{D}}(\mathbf{x}) = \sigma^2\mathbf{I}$.

We begin the comparison by generating longitudinal gene expression profiles using an expert-curated model of Myeloid Progenitor differentiation that incorporates realistic intrinsic stochasticity. This Hematopoietic Stem Cell (HSC) differentiation model includes 11 transcription factors and captures the differentiation of common myeloid progenitors into Erythrocytes, Megakaryocytes, Monocytes, and Granulocytes. We simulate the HSC regulatory network (Fig. 5A) with the Gillespie algorithm (*Materials and Methods*) to generate the marginal data consisting of $K = 8$ snapshots with $n = 5,000$ samples. The data are grouped into four clusters, with each cluster's average gene count profile shown in *SI Appendix*, Fig S6, for varying reaction volumes. Each profile represents one of the four HSC differentiation cell types and aligns with experimentally measured mRNA data (64, 65).

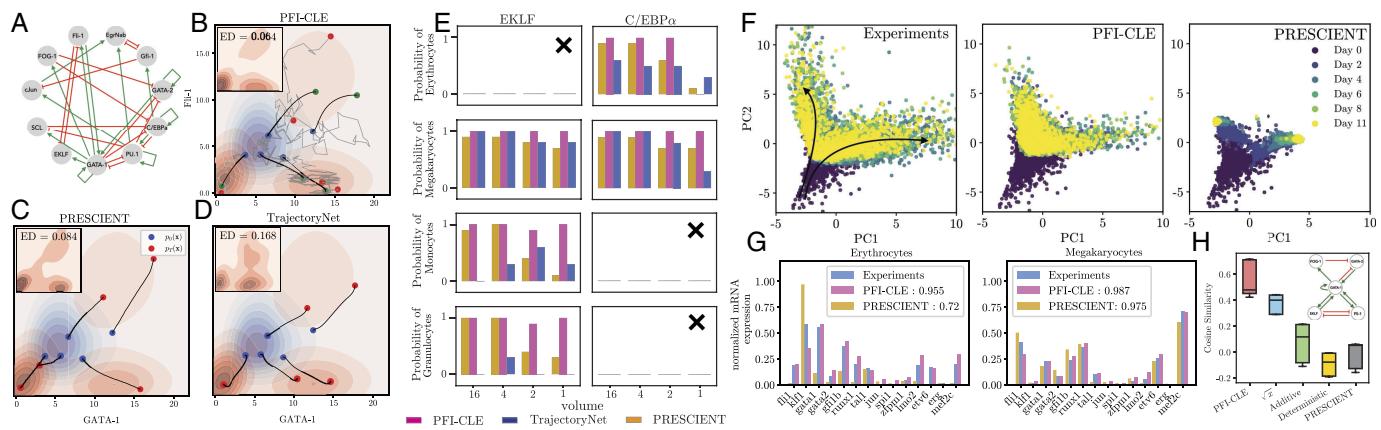


Fig. 5. Inferring HSC differentiation dynamics on simulated and experimental data. (A) HSC gene regulatory network. Red lines denote inhibition and green arrows activation. (B–D) Cell differentiation trajectories using different approaches (shown in title). The marginal densities at the initial state $p_0(\mathbf{x})$ (blue) and final state $p_T(\mathbf{x})$ (orange) are shown as contour plots. In the *Inset*, we show the reconstructed marginal at time T along with the reconstruction error (ED). (E) Probability of observing Erythrocytes, Megakaryocytes, Monocytes, and Granulocytes at steady state. Each panel displays the estimated probability of recovering the different cell types (rows) under specific perturbation conditions (columns). The averages are computed across varying initial conditions. The panels also show how these probabilities change with system size, V . Panels marked with a cross indicate the absence of the respective cell type under the corresponding perturbations. The results are color-coded by the approach used: TrajectoryNet (blue), PRESCIENT (gold), and Inferred Chemical Langevin model (purple). (F) (*Left* panel) PCA projection of mRNA counts of 14 key TFs associated with ex vivo hematopoiesis measured on days (0, 2, 4, 6, 8, 11). The two emergent branches (highlighted with black solid lines) lead to Erythrocytes and Megakaryocytes, the terminal cell types. Predicted differentiation trajectories following induction at day 0, using PFI-CLE (*Middle* panel) and PRESCIENT (*Right* panel). (G) Normalized expression levels of the key TFs from PFI-CLE and PRESCIENT compared with experimental observations on day 11 for each branch. The *Inset* shows the cosine-similarity between the predicted and the measured gene profiles. (H) Cosine similarity score between the inferred response matrix and known regulatory model (shown in *Inset*) for different noise models, and also PRESCIENT. Additional metrics and model comparisons are shown in *SI Appendix*, Fig S8.

We train all three models on the marginal data with varying levels of intrinsic noise, corresponding to a reaction volume $V \in \{1, 2, 4, 16\}$. The training procedure for all three methods is discussed in *Materials and Methods*. To compare the inferred dynamics of the three models, we examine the individual cell trajectories generated by simulating the inferred diffusion processes. Starting from the same initial position in gene space (marked in blue), each approach predicts a different final state (marked in red) as illustrated in Fig. 5 B–D. In other words, the predicted differentiated state of the same progenitor cell varies significantly between the three methods. However, despite the markedly distinct differentiation trajectories, all three methods accurately reconstruct the marginals, as shown by the low reconstruction error when measuring the distance between the predicted marginals and the data (*Inset* Fig. 5 B–D and *SI Appendix*, Table S3). For the PFI-CLE approach, we also illustrate the probability flow lines (black lines) to show how the deterministic phase velocity transports particles, and how it differs from the corresponding stochastic trajectories starting from the same initial condition. To further quantify the difference between the inferred process, we calculate the cosine similarity between the inferred probability flow lines and the CLE probability flow (PF) lines for each cell state. The probability flow of TrajectoryNet is taken to be its nonautonomous force field $\mathbf{f}(\mathbf{x}, t)$, while for PRESCIENT, it is estimated as $-\nabla \hat{\Psi}(\mathbf{x}) - \ell \mathbf{x} - \kappa^2 \nabla \log p_t(\mathbf{x})$. As intrinsic stochasticity increases, the flow lines corresponding to TrajectoryNet and PRESCIENT diverge significantly from the PFI-CLE flow lines in each cell state, as shown in *SI Appendix*, Fig. S7A. These results illustrate that while all methods effectively interpolate between the marginal distributions, increasing intrinsic stochasticity in the data leads them to learn markedly different stochastic dynamics, as evidenced by the analysis of their probability flows and the observation of individual cell trajectories.

Given the distinct dynamics learned by each approach, we next explored whether good interpolation accuracy correlates with

learning the correct dynamics, and how molecular noise affects this relationship. To evaluate this, we performed perturbation studies to certify the correctness of the learned dynamics under initial conditions different from the training data (*Materials and Methods*). Specifically, we conduct knockdown experiments by setting the expression levels of specific genes to zero. We focus on experimentally measured outcomes for knockdowns of the genes *C/EBP α* and *EKLF*. Experimental studies have reported that disruption of *C/EBP α* blocks the transition from the common myeloid to the granulocyte/monocyte progenitor (66), and knockdown of the *EKLF* gene leads to the absence of the Erythrocyte cell line (67). In Fig. 5E, we report the probability of observing each cell type under various perturbation conditions. To separate the effect of the force model from the influence of noise, we perform in silico perturbation experiments across a range of noise levels. At low noise levels, both PRESCIENT and PFI-CLE, despite using different force models, accurately predict the perturbation outcomes. However, as molecular noise increases, differences in their inferred dynamics become more pronounced, with PRESCIENT being the most affected. For instance, under high noise conditions, PRESCIENT fails to recover the Monocyte/Granulocyte lineage in multiple knockdown scenarios, whereas PFI-CLE, which accounts for molecular noise, consistently reconstructs this lineage across all noise levels. In contrast, TrajectoryNet struggles to recover the Monocyte/Granulocyte lineage even at low noise, and as stochasticity increases, its ability to reconstruct the Megakaryocyte/Erythrocyte lineage also diminishes monotonically. These findings underscore the critical role of intrinsic stochasticity and illustrate how it can impede state-of-the-art methods from accurately capturing gene regulatory interactions during differentiation.

Guided by insights from our simulation study of the HSC network, we apply the PFI framework to time-resolved mRNA count data collected during ex vivo hematopoiesis (68). This dataset comprises cells undergoing 12-d ex vivo induced differentiation from CD34 $^{+}$ hematopoietic stem and progenitor cells

toward erythrocytes and megakaryocytes. Single-cell RNA-seq measurements are collected on days 2, 4, 6, 8, and 11 following induction. In Fig. 5*F*, the PCA projection of mRNA counts for 14 key transcription factors (TFs) reveals two branching trajectories over time, corresponding to erythropoiesis and megakaryopoiesis. Notably, the key regulatory genes identified in this experimental study overlap with those previously considered in the HSC differentiation (Fig. 5*A*), reinforcing the observation that hematopoietic differentiation is driven by a small subset of transcription factors among the approximately 10,000 genes profiled.

Using our two-step PFI approach, we directly fit the CLE given in Eq. 9 to the measured mRNA count data. The average degradation rate is set to $\ell = 1 \text{ d}^{-1}$ consistent with typical measurements in mammalian cells (69). To assess the validity of the PFI-CLE formulation, we also infer force fields under other commonly used noise models. Fig. 5*F* shows that, starting from day 0 initial conditions, both the PFI-CLE (*Middle* panel) and the PRESCIENT method (*Right* panel) effectively recover the branching trajectories associated with erythropoiesis and megakaryopoiesis. To quantify agreement with experimental data, we calculate the cosine similarity between predicted and measured normalized gene expression on day 11 along both branches (Fig. 5*G*), showing that both methods accurately recover the expected terminal cell types.

To go further, we identify the causal relationships between the genes through an *in silico* perturbation of the inferred force field $\mathbf{f}(\mathbf{x})$ (*Materials and Methods*). As highlighted in Fig. 5*H*, the PFI-CLE produced the highest similarity score when evaluated against known regulatory interactions involving the key genes driving erythrocyte-megakaryocyte progenitors to their respective terminal states. The performance degrades successively from the state-dependent $\sqrt{\mathbf{x}}$ model, to the additive noise model, and finally the deterministic model. This trend strikingly mirrors our earlier findings on the simulated mCAD network, where CLE accurately identifies regulatory interactions (Fig. 4 *D* and *E*), and non-CLE models lead to the identification of spurious interactions between the genes. Interestingly, even though PRESCIENT captures the branching differentiation (Fig. 5 *F*, *Right* panel), it does so by learning incorrect regulations between the genes as quantified in Fig. 5*H*. To evaluate the robustness of these findings, we also inferred force fields across different gene sets (*SI Appendix, Table S4*) and observed that the same trend consistently held across settings with varying dimensionality d . Consistent with our analysis on the simulated HSC network, the PFI-CLE formulation, by accounting for intrinsic stochasticity, enables a clearer separation of noise and regulatory signals and accurately captures the regulatory interactions among key genes.

4. Discussion

In this paper, we propose PFI, an approach that transforms the problem of learning SDEs into inferring their corresponding phase-space probability flow. This is facilitated by recent advances in score-based generative modeling, which allow for efficient computation of the gradient log-probability, $\nabla \log p_t(\mathbf{x})$, from high-dimensional cross-sectional data of the time-evolving distribution $p_t(\mathbf{x})$. This reformulation dramatically simplifies optimization and, crucially, separates the impact of intrinsic noise from inferring the force field.

We first limited our analysis of the PFI problem to the analytically tractable class of Ornstein–Uhlenbeck processes. We proved that the regularization ensures a strongly convex loss

with a unique global minimum in the limit of well-sampled distributions, both in time and in state-space. For an isotropic Ornstein–Uhlenbeck process, we minimized analytically this loss function, showing that the relative magnitudes of the regularization and of the time-dependent deformations of the marginals select the learnable nonequilibrium contributions to the force. While large deformations are used to learn nonconservative forces, smaller deformations are washed out by the regularization. Although this observation suggests to use a smaller regularization, we showed that the bias reduction obtained thereby trades-off with an increasing variance stemming from the finite sampling of distributions, both in time Δt and in the number of samples n . This interplay between n , Δt , and the regularization calls for careful model selection when considering complex models.

Using the same analytical solution, we subsequently showed that an inaccurate estimation of the noise strength leads to a dramatic decrease in performance when the stochastic effects are of the same order of magnitude as the force. This is of paramount importance for the inference of gene regulatory networks for which intrinsic noise is strongly correlated with the amplitude of regulation. To further explore this role of the noise model, we numerically investigated more realistic models of stochastic reaction networks, using the PFI approach. Our study underscores the critical role of intrinsic noise in parameter estimation, regulatory network inference, and generalization to unseen data. An incorrect noise model can lead to spurious relationships between species, which affects the inferred probabilistic landscape (6). Consequently, force fields based on such models exhibit poor generalization performance when tested with initial conditions different from those used during training. The PFI approach, therefore, proves to be a valuable tool for analyzing single-cell omics data, even when the stoichiometry of the reaction network is unknown.

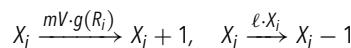
Finally, we applied our framework to learn data-driven models of cell differentiation. To account for molecular noise, we inferred a Chemical Langevin model using the PFI approach, and compared it with popular generative models in predicting the effects of gene knockdowns in the hematopoiesis system for increased noise strength. While PRESCIENT successfully predicts the effects of interventions when noise is negligible, it struggles when it becomes significant. On the other hand, TrajectoryNet fails to accurately capture regulatory interactions, particularly in high-dimensional settings, even with minimal molecular noise. In contrast, PFI-CLE consistently makes accurate predictions, emphasizing the importance of modeling intrinsic noise to infer accurate regulatory pathways. Applied to time-resolved mRNA count data collected during *ex vivo* human hematopoiesis, the PFI-CLE model yields stable temporal dynamics and captures the branching trajectories corresponding to erythropoiesis and megakaryopoiesis. In addition, the PFI CLE model interpolates the data by learning experimentally-verified regulatory interactions between genes. This is a demonstration on time-resolved single-cell RNAseq data of the necessity to account for intrinsic noise to predict accurate regulatory interactions. Indeed, our results show that accurately interpolating the data is not sufficient, and the interpolation needs to be guided by comprehensive biophysical priors of the latent stochastic processes, and special care should be given to modeling intrinsic noise. For this purpose, the PFI approach provides a very flexible solution to incorporate more realistic noise models in the inference of regulatory processes from single-cell omics data. In this direction, future work should aim at extending the PFI approach to account for the unobserved protein dynamics between protein production and its regulatory effects on transcription (70–74).

In this study, we primarily focus on intrinsic stochasticity in the form of molecular noise. However, extending the PFI approach to include other sources of stochasticity, like fluctuations in transcriptional rates (72, 75, 76), would be a natural step forward. More importantly, future studies should build upon the PFI framework by integrating cell death and proliferation. We believe this is a necessary step to take to successfully and reliably apply the PFI approach to real data. Given the prevalence of noise in cellular processes, our approach marks a significant step toward developing biophysically accurate, data-driven models that incorporate nontrivial stochastic effects.

5. Materials and Methods

5.1. Choice of Matrices for the Ornstein–Uhlenbeck Process. To generate a random skew-symmetric matrix, we first generate a matrix $\mathbf{U} \in \mathbb{R}^{d \times d}$ with i.i.d. entries drawn uniformly at random in $[0, 1]$, and $\mathbf{e} \in \mathbb{R}^d$ a vector with i.i.d. entries drawn uniformly at random in $[0.9, 1]$. Denoting $\mathbf{W} \in \mathbb{C}^{d \times d}$ the eigenbasis of $\mathbf{U} - \mathbf{U}^T$, the matrix \mathbf{A} is set to be $\mathbf{A} = \mathbf{W}\mathbf{e}\mathbf{W}^\dagger/\max(\mathbf{e})$, where i is the imaginary number and \dagger denotes the Hermitian conjugate. The qualitative behavior shown in Figs. 2 and 3 is unchanged by repeatedly drawing new instances of \mathbf{A} . In Figs. 2 and 3, we used the same instance of the matrix \mathbf{A} . To generate a random symmetric matrix \mathbf{S} we use a similar approach, with the matrix \mathbf{W} being the eigenbasis of $\mathbf{U} + \mathbf{U}^T$.

5.2. Modeling Gene Regulation. Under the assumption that the system is well mixed, the regulation of gene i by a set R_i of regulators obeys the following set of stochastic reactions



Here, the term $mVg(R_i)$ represents the propensity associated with the reaction that produces the mRNA for gene x_i , where m denotes the transcription rate, and R_i denotes the set of regulators of node i . The term ℓX_i corresponds to the propensity of the degradation reaction, with ℓ being the degradation rate. These propensities denote the number of these reactions happening per unit of time. The nonlinear function $g(R_i)$ encapsulates the regulatory interactions governing the expression of gene i and is mediated by proteins. Given the volume of the reaction compartment, the functional form of R_i as a function of the concentration of proteins is derived from equilibrium statistical mechanics considerations (59). In this work, we adopt its Boolean network implementation as introduced in ref. 62.

In the absence of joint measurements of mRNA molecules and proteins, we assume that protein and mRNA levels are strongly correlated, allowing us to replace the protein regulators R_i with their mRNA counterparts. Although this assumption suggests that proteins are in quasi-steady-state with mRNA, which contrasts with observations (77, 78) of faster mRNA turnover compared to proteins (72), it still enables the recovery of single-cell RNA-seq profiles that are quantitatively comparable at steady state (61). In practice, most gene regulatory network inference approaches ignore protein dynamics, implicitly making a similar assumption.

5.2.1. Simulation of linear stochastic reaction networks. For the linear cyclic network $X_i \xrightarrow{k_i} X_{i+1}$, rate constants k_i were generated using a logarithmic scale spanning from $10^{0.1}$ to $10^{1.5}$, distributed across $d = 30$ values, and scaled by a factor of 0.1. Gillespie simulations were then computed using the stoichiometric matrix and the rate constants.

5.3. Comparison With Existing Methods. In the following, we describe the existing methods used to infer cell differentiation dynamics from empirical marginal distributions. We consider that $K + 1$ distributions are observed at times $t_0 \leq \dots \leq t_K$, each with n samples. The PFI-CLE model is trained using the two-step PFI approach, with the force function $\mathbf{f}_\theta : \mathbb{R}^d \rightarrow \mathbb{R}^d$ represented by a feed-forward neural network. The network consists of four fully connected layers, each with 30 nodes and smooth ELU activations.

5.3.1. PRESCIENT. Following the original idea by Hashimoto (33), popular generative models model cellular differentiation as a diffusion process $\mathbf{x}(t)$ (63, 79), governed by the stochastic differential equation:

$$\frac{d\mathbf{x}}{dt} = (-\nabla \Psi(\mathbf{x}) - \ell \mathbf{x})dt + \sqrt{2\kappa^2} d\mathbf{W}(t), \quad [10]$$

where the drift is the gradient of a potential function $\Psi(\mathbf{x}) : \mathbb{R}^d \rightarrow \mathbb{R}$. PRESCIENT proceeds by finding the function Ψ that minimizes the loss function

$$\mathcal{L}_{\text{PRESCIENT}} = \sum_{i=1}^K \left[\mathcal{W}_2^2(\hat{p}_{t_i}(\mathbf{x}), p_{t_i}(\mathbf{x}))^2 + \tau \sum_{j=1}^n \frac{\Psi(\mathbf{x}_j)}{\kappa^2} \right].$$

The Wasserstein distance measures the difference between the observed distribution $p_{t_i}(\mathbf{x})$ and the predicted distribution $\hat{p}_{t_i}(\mathbf{x})$, and the parameter τ controls the entropic regularization. The probability distributions are fitted to the observed data by simulating the stochastic differential equation Eq. 10. The potential function $\Psi(\mathbf{x})$ is parameterized with a fully connected neural network with ELU activation function. In the original study, the noise scale is set to be $\kappa = 0.1$ and hyperparameter $\tau = 10^{-6}$, and we use the same parameters in the training of the PRESCIENT model. The potential function $\Psi(\mathbf{x})$ is parameterized with a feed-forward neural network consisting of four fully connected layers, each with 30 nodes and smooth ELU activation.

5.3.2. TrajectoryNet. Another popular approach for modeling cellular differentiation involves parameterizing the force field as a nonautonomous Neural ODE (24, 80), without any explicit noise model:

$$\frac{d\mathbf{x}}{dt} = \hat{\mathbf{f}}(\mathbf{x}, t).$$

Using this framework, TrajectoryNet fits a continuous normalizing flow connecting the successive distributions, enforcing an analytically tractable reference distribution $p_{t_{-1}} \sim \mathcal{N}(0, 1)$ at time t_{-1} . The loss function reads

$$\mathcal{L}_{\text{TrajectoryNet}} = - \sum_{i=0}^K \mathbb{E}_{p_{t_i}} \log \hat{p}_{t_i}(\mathbf{x}) + \text{regularization}.$$

The first term corresponds to the log-likelihood of the predicted distributions $\hat{p}_{t_i}(\mathbf{x})$ evaluated on the observed data $p_{t_i}(\mathbf{x})$. TrajectoryNet includes different choices for the regularization on the force field, among which are the penalization of the curvature of the force field (used in PFI) and of its kinetic energy (24). To train TrajectoryNet, we use the default parameters with the OT-inspired regularization on the kinetic energy suggested in the original study. The nonautonomous force field is modeled using a neural network with three fully connected layers, each containing 64 nodes and employing leaky ReLU activations.

5.4. In Silico Perturbations. We conduct in silico perturbations by simulating the inferred models from a given initial condition until a steady state is reached, with the concentration of the knocked-down gene set to zero. The models are tested with varying initial conditions derived from the training data, given by $p = (1 - c)p^* + cU[0.25, 0.5]$, to evaluate generalization. In this setup, $c = 0$ corresponds to the first marginal of the training data $p^* = \hat{p}_0(\mathbf{x})$, while $c = 1$ represents a uniform initialization of mRNA expression within the hypercube $[0.25, 0.5]^d$. We generate 10 distinct initial conditions by selecting c from a uniformly spaced grid between 0 and 1, with increments of 0.1. Cells are assigned to a specific cell state if the cosine similarity between the cell state's expression profile and the predicted expression vector is 0.95 or higher.

Alternatively, to directly assess causal relationships between genes, we perform perturbation analysis by measuring the change in f_j resulting from a fold change in x_i (81):

$$\Delta_{jj} \equiv f_j(\dots, x_i) - f_j(\dots, \xi x_i), \quad 0 < \xi < 1.$$

This approach mimics biological knockdown experiments, with the parameter ξ controlling the perturbation strength. For example, $\xi = 0$ corresponds to a full knockout of the input, while $\xi \approx 1$ approximates a knockdown derivative. To quantify the regulatory influence of x_i on x_j , we compute the average response matrix $\langle \Delta_{ij} \rangle$ across all time points and samples, sweeping over perturbation strengths in the range $0.7 < \xi < 0.95$. We quantify the agreement between the inferred response matrix Δ and the known regulatory matrix M using the cosine similarity metric. The entries of M take values in $\{1, -1, 0\}$, corresponding to activation, inhibition, and no interaction, respectively. The similarity score is defined as

1. M. Thattai, A. van Oudenaarden, Intrinsic noise in gene regulatory networks. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 8614–8619 (2001).
2. R. Losick, C. Desplan, Stochasticity and cell fate. *Science* **320**, 65–68 (2008).
3. W. Bialek *et al.*, Statistical mechanics for natural flocks of birds. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 4786–4791 (2012).
4. J. Jhawar *et al.*, Noise-induced schooling of fish. *Nat. Phys.* **16**, 488–493 (2020).
5. J. Grilli, Macroecological laws describe variation and diversity in microbial communities. *Nat. Commun.* **11**, 4743 (2020).
6. M. A. Coomer, L. Ham, M. P. H. Stumpf, Noise distorts the epigenetic landscape and shapes cell-fate decisions. *Cell Syst.* **13**, 83–102.e6 (2022).
7. N. Maheshri, E. K. O’Shea, Living with noisy genes: How cells function reliably with inherent variability in gene expression. *Annu. Rev. Biophys. Biomol. Struct.* **36**, 413–434 (2007).
8. A. Frishman, P. Ronceray, Learning force fields from stochastic trajectories. *Phys. Rev. X* **10**, 021009 (2020).
9. D. B. Brückner, P. Ronceray, C. P. Broedersz, Inferring the dynamics of underdamped stochastic systems. *Phys. Rev. Lett.* **125**, 058103 (2020).
10. W. Saelens, R. Cannoodt, H. Todorov, Y. Saeyns, A comparison of single-cell trajectory inference methods. *Nat. Biotechnol.* **37**, 547–554 (2019).
11. G. Schiebinger *et al.*, Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell* **176**, 928–943.e22 (2019).
12. R. D. Brackston, E. Lakatos, M. P. Stumpf, Transition state characteristics during cell differentiation. *PLoS Comput. Biol.* **14**, e1006405 (2018).
13. C. Weinreb, S. Wolock, B. K. Tusi, M. Socolovsky, A. M. Klein, Fundamental limits on dynamic inference from single-cell snapshots. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E2467–E2476 (2018).
14. C. Gardiner, *Stochastic Methods*, 0172-7389 (Springer-Verlag, Berlin Heidelberg, 2009), vol. 13.
15. C. Ahlmann-Eltze, W. Huber, Comparison of transformations for single-cell RNA-seq data. *Nat. Methods* **20**, 665–672 (2023).
16. S. Choudhary, R. Satija, Comparison and evaluation of statistical error models for scRNA-seq. *Genome Biol.* **23**, 27 (2022).
17. L. Pérez García, J. Donlucas Pérez, G. Volpe, A. V. Arzola, G. Volpe, High-performance reconstruction of microscopic force fields from Brownian trajectories. *Nat. Commun.* **9**, 5166 (2018).
18. F. Ferretti, V. Chardès, T. Mora, A. M. Walczak, I. Giardina, Building general Langevin models from discrete datasets. *Phys. Rev. X* **10**, 031018 (2020).
19. Y. A. Kutoyants, *Statistical Inference for Ergodic Diffusion Processes*, Springer Series in Statistics (Springer, London, 2004).
20. K. D. Yang, C. Uhler, “Scalable unbalanced optimal transport using generative adversarial networks” in *International Conference on Learning Representations* (OpenReview.net, 2019).
21. G. Schiebinger, Reconstructing developmental landscapes and trajectories from single-cell data. *Curr. Opin. Syst. Biol.* **27**, 100351 (2021).
22. S. Zhang, A. Afanassiev, L. Greenstreet, T. Matsumoto, G. Schiebinger, Optimal transport analysis reveals trajectories in steady-state systems. *PLoS Comput. Biol.* **17**, e1009466 (2021).
23. C. Bunne *et al.*, Learning single-cell perturbation responses using neural optimal transport. *Nat. Methods* **20**, 1759–1768 (2023).
24. A. Tong *et al.*, “A dynamic optimal transport network for modeling cellular dynamics” in *International Conference on Machine Learning*, H. Daumé, A. Singh, Eds. (PMLR, 2020), pp. 9526–9536.
25. L. Chizat, S. Zhang, M. Heitz, G. Schiebinger, Trajectory inference via mean-field Langevin in path space. *Adv. Neural Inf. Process. Syst.* **35**, 16731–16742 (2022).
26. H. Lavenant, S. Zhang, Y.-H. Kim, G. Schiebinger, Toward a mathematical theory of trajectory inference. *Ann. Appl. Probab.* **34**, 428–500 (2024).
27. C. Bunne, Y. P. Hsieh, M. Cuturi, A. Krause, “The schrödinger bridge between Gaussian measures has a closed form” in *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics*, F. Ruiz, J. Dy, J.-W. van de Meent, Eds. (PMLR, 2023), pp. 5802–5833.
28. A. Oceno, L. Haghverdi, N. S. Mueller, F. J. Theis, Reconstructing gene regulatory dynamics from high-dimensional single-cell snapshot data. *Bioinformatics* **31**, i89–i96 (2015).
29. H. Matsumoto *et al.*, Scode: An efficient regulatory network inference algorithm from single-cell RNA-seq during differentiation. *Bioinformatics* **33**, 2314–2321 (2017).
30. M. Sanchez-Castillo, D. Blanco, I. M. Tienda-Luna, M. Carrión, Y. Huang, A Bayesian framework for the inference of gene regulatory networks from time and pseudo-time series data. *Bioinformatics* **34**, 964–970 (2018).
31. P. Zhou, F. Bocci, T. Li, Q. Nie, Spatial transition tensor of single cells. *Nat. Methods* **21**, 1053–1062 (2024).
32. D. T. Gillespie, The chemical Langevin equation. *J. Chem. Phys.* **113**, 297–306 (2000).
33. T. Hashimoto, D. Gifford, T. Jaakkola, “Learning population-level diffusions with generative RNNs” in *Proceedings of The 33rd International Conference on Machine Learning*, M. F. Balcan, K. Q. Weinberger, Eds. (PMLR, 2016), pp. 2417–2426.
34. X. Li, T. K. L. Wong, R. T. Chen, D. Duvenaud, “Scalable gradients for stochastic differential equations” in *International Conference on Artificial Intelligence and Statistics*, S. Chiappa, R. Calandri, Eds. (PMLR, 2020), pp. 3870–3882.
35. Y. Song *et al.*, “Score-based generative modeling through stochastic differential equations” in *International Conference on Learning Representations* (OpenReview.net, 2020).
36. A. Hyvärinen, Estimation of non-normalized statistical models by score matching. *J. Mach. Learn. Res.* **6**, 695–709 (2005).
37. D. Maoutsa, S. Reich, M. Opper, Interacting particle solutions of Fokker-Planck equations through gradient-log-density estimation. *Entropy* **22**, 802 (2020).
38. N. M. Boffi, E. Vanden-Eijnden, Probability flow solution of the Fokker-Planck equation. *Mach. Learn. Sci. Technol.* **4**, 035012 (2023).
39. N. M. Boffi, E. Vanden-Eijnden, Deep learning probability flows and entropy production rates in active matter. *Proc. Natl. Acad. Sci. U.S.A.* **121**, e2318106121 (2024).
40. Y. Song, S. Garg, J. Shi, S. Ermon, “Sliced score matching: A scalable approach to density and score estimation” in *Proceedings of The 35th Uncertainty in Artificial Intelligence Conference*, R. P. Adams, V. Gogate, Eds. (PMLR, 2020), pp. 574–584.
41. A. Genevay, G. Peyré, M. Cuturi, “Learning generative models with Sinkhorn divergences” in *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*, A. Storkey, F. Perez-Cruz, Eds. (PMLR, 2018), pp. 1608–1617.
42. A. Genevay, L. Chizat, F. Bach, M. Cuturi, G. Peyré, “Sample complexity of Sinkhorn divergences” in *The 22nd international conference on artificial intelligence and statistics*, K. Chaudhuri, M. Sugiyama, Eds. (PMLR, 2019), pp. 1574–1583.
43. K. Atkinson, *An Introduction to Numerical Analysis* (John Wiley & Sons, 1991).
44. R. T. Chen, Y. Rubanova, J. Bettencourt, D. K. Duvenaud, Neural ordinary differential equations. *Adv. Neural Inf. Process. Syst.* **31**, 07366V2 (2018).
45. W. Grathwohl, R. T. Chen, J. Bettencourt, I. Sutskever, D. Duvenaud, “Ffjord: Free-form continuous dynamics for scalable reversible generative models” in *International Conference on Learning Representations* (OpenReview.net, 2018).
46. Y. Rubanova, R. T. Chen, D. K. Duvenaud, Latent ordinary differential equations for irregularly-sampled time series. *Adv. Neural Inf. Process. Syst.* **32**, 5320–5330 (2019).
47. B. Yuan *et al.*, Cellbox: Interpretable machine learning for perturbation biology with application to the design of cancer combination therapy. *Cell Syst.* **12**, 128–140 (2021).
48. G. La Manno *et al.*, RNA velocity of single cells. *Nature* **560**, 494–498 (2018).
49. V. Bergen, M. Lange, S. Peidli, F. A. Wolf, F. J. Theis, Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat. Biotechnol.* **38**, 1408–1414 (2020).
50. S. Särkkä, A. Solin, *Applied Stochastic Differential Equations* (Cambridge University Press, 2019), vol. 10.
51. H. H. McAdams, A. Arkin, Stochastic mechanisms in gene expression. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 814–819 (1997).
52. J. M. Raser, E. K. O’Shea, Noise in gene expression: Origins, consequences, and control. *Science* **309**, 2010–2013 (2005).
53. D. T. Gillespie, A rigorous derivation of the chemical master equation. *Phys. A Stat. Mech. Appl.* **188**, 404–425 (1992).
54. D. Schnoer, G. Sanguineti, R. Grima, Approximation and inference methods for stochastic biochemical kinetics—a tutorial review. *J. Phys. A Math. Theor.* **50**, 093001 (2017).
55. F. Fröhlich, B. Kaltenbacher, F. J. Theis, Scalable parameter estimation for genome-scale biochemical reaction networks. *PLoS Comput. Biol.* **13**, e1005331 (2017).
56. C. Zechner, M. Unger, S. Pelet, M. Peter, F. Koeppl, Scalable inference of heterogeneous reaction kinetics from pooled single-cell recordings. *Nat. Methods* **11**, 197–202 (2014).
57. C. Drovandi, A. N. Pettitt, M. J. Faddy, Approximate Bayesian computation using indirect inference. *J. R. Stat. Soc. Ser. C Appl. Stat.* **60**, 317–337 (2011).
58. P. Mozgunov *et al.*, A review of the deterministic and diffusion approximations for stochastic chemical reaction networks. *React. Kinet. Mech. Catal.* **123**, 289–312 (2018).
59. D. Marbach *et al.*, Revealing strengths and weaknesses of methods for gene network inference. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 6286–6291 (2010).
60. J. M. Bower, H. Bolouri, *Computational Modeling of Genetic and Biochemical Networks* (The MIT Press, 2001).
61. P. Dibaeinia, S. Sinha, Sergio: A single-cell expression simulator guided by gene regulatory networks. *Cell Syst.* **11**, 252–271 (2020).
62. A. Pratapa, A. P. Jalihal, J. N. Law, A. Bharadwaj, T. M. Murali, Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. *Nat. Methods* **17**, 147–154 (2020).
63. G. H. T. Yeo, S. D. Saksena, D. K. Gifford, Generative modeling of single-cell time series with prescient enables prediction of cell trajectories with interventions. *Nat. Commun.* **12**, 3222 (2021).
64. C. J. Pronk *et al.*, Elucidation of the phenotypic, functional, and molecular topography of a myeloerythroid progenitor cell hierarchy. *Cell Stem Cell* **1**, 428–442 (2007).
65. J. Krumrieck, C. Marr, T. Schroeder, F. J. Theis, Hierarchical differentiation of myeloid progenitors is encoded in the transcription factor network. *PLoS One* **6**, e22649 (2011).

$$\text{Cosine similarity} = \frac{\sum_{i,j} \Delta_{ij} M_{ij}}{\sqrt{\sum_{i,j} \Delta_{ij}^2} \sqrt{\sum_{i,j} M_{ij}^2}}.$$

For hematopoiesis, the matrix M is constructed based on the boolean network model discussed in ref. 65.

Data, Materials, and Software Availability. Simulation code to reproduce the figures, along with the simulated data used for the figures, has been deposited in <https://github.com/SuryanarayanaMK/PFI> (82). We also used publicly available data from the following Nature paper: <https://www.nature.com/articles/s41467-021-27159-x> (83).

66. P. Zhang *et al.*, Enhancement of hematopoietic stem cell repopulating capacity and self-renewal in the absence of the transcription factor C/EBP α . *Immunity* **21**, 853–863 (2004).
67. B. Nuez, D. Michalovich, A. Bygrave, R. Ploemacher, F. Grosveld, Defective haematopoiesis in fetal liver resulting from inactivation of the EKLF gene. *Nat. Int. Wkly. J. Sci.* **375**, 316–318 (1995).
68. G. Georgopoulos *et al.*, Discrete regulatory modules instruct hematopoietic lineage commitment and differentiation. *Nat. Commun.* **12**, 6790 (2021).
69. M. Shamir, Y. Bar-On, R. Phillips, R. Milo, Snapshot: Timescales in cell biology. *Cell* **164**, 1302–1302 (2016).
70. K. Josić, J. M. López, W. Ott, L. Shiau, M. R. Bennett, Stochastic delay accelerates signaling in gene networks. *PLoS Comput. Biol.* **7**, e1002264 (2011).
71. D. Bratsun, D. Volfson, L. S. Tsimring, J. Hasty, Delay-induced stochastic oscillations in gene regulation. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 14593–14598 (2005).
72. U. Herbach, A. Bonnaffoux, T. Espinasse, O. Gandrillon, Inferring gene regulatory networks from single-cell data: A mechanistic approach. *BMC Syst. Biol.* **11**, 105 (2017).
73. U. Herbach, Gene regulatory network inference from single-cell data using a self-consistent proteomic field. arXiv [Preprint] (2021). <http://arxiv.org/abs/2109.14888> (Accessed 2 August 2024).
74. E. Ventre, U. Herbach, T. Espinasse, G. Benoit, O. Gandrillon, One model fits all: Combining inference and simulation of gene regulatory networks. *PLoS Comput. Biol.* **19**, e1010962 (2023).
75. L. Ham, R. D. Brackston, M. P. Stumpf, Extrinsic noise and heavy-tailed laws in gene expression. *Phys. Rev. Lett.* **124**, 108101 (2020).
76. G. Gorin, J. J. Vastola, M. Fang, L. Pachter, Interpretable and tractable models of transcriptional noise for the rational design of single-molecule quantification experiments. *Nat. Commun.* **13**, 7620 (2022).
77. R. d. S. Abreu, L.O. Penalva, E. M. Marcotte, C. Vogel, Global signatures of protein and mRNA expression levels. *Mol. Biosyst.* **5**, 1512–1526 (2009).
78. V. Shahrezaei, P. S. Swain, Analytical distributions for stochastic gene expression. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 17256–17261 (2008).
79. D. R. Sisan, M. Halter, J. B. Hubbard, A. L. Plant, Predicting rates of cell state change caused by stochastic fluctuations using a data-driven landscape model. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 19262–19267 (2012).
80. Y. Sha, Y. Qiu, P. Zhou, Q. Nie, Reconstructing growth and dynamic trajectories from single-cell transcriptomics data. *Nat. Mach. Intell.* **6**, 25–39 (2024).
81. J. Shen, F. Liu, Y. Tu, C. Tang, Finding gene network topologies for given biological function with recurrent neural network. *Nat. Commun.* **12**, 3125 (2021).
82. S. Maddu, V. Chardès, PFI. GitHub. <https://github.com/SuryanarayanaMK/PFI>. Deposited 6 July 2025.
83. G. Georgopoulos *et al.*, Discrete regulatory modules instruct hematopoietic lineage commitment and differentiation. Zenodo. <https://doi.org/10.5281/zenodo.5291737>. Deposited 24 June 2021.