Stochastic Process Inference Without Trajectories: A Probabilistic Approach

David Hathcock, Mark S. Squillante, Yuhai Tu
Mathematical Sciences Department, IBM Research
Thomas J. Watson Research Center, Yorktown Heights, NY 10598, USA
david.hathcock@ibm.com, mss@us.ibm.com, yuhai@us.ibm.com

1. INTRODUCTION

A fundamental problem in computer system performance, as well as in the natural sciences, concerns inferring from observations an understanding of the behavior of stochastic processes of interacting system components whose dynamics are driven by an unknown underlying stochastic differential equation (SDE). The objective in solving this problem is to infer the underlying equations of the dynamics of the system from sets of system measurements, indexed over time. Given the stochastic nature of such systems, together with a lack of information on stochastic trajectories in many cases [1, 3], this represents a very challenging problem in general.

In this paper we consider the above inference problem within the context of complex computer systems modeled as stochastic processes of interacting system components whose continuous-time evolution follows SDEs of the general form

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x})\boldsymbol{\eta}(t), \tag{1}$$

where $\mathbf{x} \in \mathbb{R}^n$ is the n-dimensional state vector, $\mathbf{f}(\mathbf{x})$ the drift term, $\mathbf{D}(\mathbf{x}) = \frac{1}{2}\mathbf{g}(\mathbf{x})\mathbf{g}(\mathbf{x})^\top$ the diffusion term, and $\boldsymbol{\eta}(t)$ a Gaussian noise vector. The dynamics can be equivalently formulated in terms of a Fokker-Planck (FP) equation characterizing the time evolution of the probability density $P(\mathbf{x}, t)$ for an ensemble of particles with dynamics given by (1). Using the Itô interpretation for multiplicative noise, we have

$$\frac{dP(\mathbf{x},t)}{dt} = \nabla\left[-\mathbf{f}(\mathbf{x})P(\mathbf{x},t) + \nabla(\mathbf{D}(\mathbf{x})P(\mathbf{x},t))\right]. \tag{2}$$

This problem has been extensively studied in cases where complete stochastic trajectories $\{\mathbf{x}(t), t \geq 0\}$ are available. Under the assumption of a lack of information on the stochastic trajectories, which is consistent with computer system measurement processes as well as the recent work in [1, 3], we seek to infer the drift and diffusion terms without such trajectory information. More specifically, the available data is limited to K sets of N cross-sectional measurements collected at a discrete selection of K time epochs; namely, the data consists of a series of empirical distributions $\nu^1, \nu^2, \ldots, \nu^K$ measured at times $t_1 < t_2 < \cdots < t_K$, respectively, where

$$\nu^k(\mathbf{x}) = \frac{1}{N}\sum_{i=1}^{N} \delta(\mathbf{x} - \mathbf{x}_i(t_k)), \quad k \in [K] := \{1, \ldots, K\}, \tag{3}$$

$\mathbf{x}_i(t_k)$ is the ith value measured at time t_k, $i \in [N]$, $k \in [K]$, and $\delta(\mathbf{x})$ is the Dirac delta-function. The correspondence between measured points at different times that lie on the same trajectory is assumed to be unknown, in contrast to problems where the available data includes trajectory information.

Our goal is to simultaneously infer both the underlying dynamics (i.e., drift $\mathbf{f}$ and diffusion $\mathbf{D}$) and the assignment of measured points to trajectories from the empirical measurement distributions $\nu^1, \nu^2, \ldots, \nu^K$. In contrast to related work, we focus directly on the corresponding mathematics without resorting to any simplifying approximations, and then we efficiently solve the resulting optimization problem.

Related work. The recent work in [1, 3] considers this inference problem under the additional assumption that the diffusion term $\mathbf{D}$ is known. Their approach consists of two key elements: first, they approximate the SDE in (1) by a *deterministic probability flow* (DPF) *ordinary differential equation* (ODE) $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) - \nabla\mathbf{D}(\mathbf{x}) - \mathbf{D}(\mathbf{x})\nabla \log P(\mathbf{x}, t)$ whose trajectories preserve the evolution of the probability density under (2); and second, they apply score-function estimation to approximate the score term $\nabla \log P(\mathbf{x}, t)$ of the DPF ODE. The problem is then solved through a combination of score-based generative modeling, a neural network parameterization of the drift, and optimal transport distance measures on the empirical distributions ν^k.

More specifically, the first step consists of using the evolution of the empirical distribution to infer the score $s(\mathbf{x}, t) = \nabla \log P(\mathbf{x}, t)$ by solving the following optimization problem:

$$\hat{s}(\mathbf{x}, t_k) = \arg\min_{s} \sum_{k=1}^{K} \lambda(t_k)\mathbb{E}_{\nu^k}\left[\mathrm{tr}(\nabla s(\mathbf{x}, t_k)) + \frac{1}{2}||s(\mathbf{x}, t_k)||_2^2\right],$$

where $\lambda(t_k)$ is a weighting function for the measurements at time t_k, $k \in [K]$, and $s(\mathbf{x}, t)$ is parameterized by a fully connected neural network whose weights are tuned for the optimization problem. Once an accurate score model is obtained, the second step consists of solving for the drift $\mathbf{f}(\mathbf{x})$ associated with the DPF ODE by minimizing the weighted distance between measured distributions and those predicted using the ODE as follows:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{k=1}^{K} \gamma(t_k)\mathcal{L}(\hat{\mu}_{\boldsymbol{\theta}}^k, \nu^k), \tag{4}$$

$$\hat{\mu}_{\boldsymbol{\theta}}^k(\mathbf{x}) = \frac{1}{N}\sum_{i=1}^{N} \delta(\mathbf{x} - \hat{\mathbf{x}}_i(t_k)), \qquad k \in [K], \tag{5}$$

$$\hat{\mathbf{x}}_i(t_k) = \mathbf{x}_i(t_{k-1}) + \int_{t_{k-1}}^{t_k} (\mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x}_i) - \nabla\mathbf{D}(\mathbf{x}_i) - \mathbf{D}(\mathbf{x}_i)\hat{s}(\mathbf{x}_i, \tau))\, d\tau, \quad i \in [N], \tag{6}$$

where $\gamma(t_k)$ is a weighting function for the measurements at

MAMA 2025 Stony Brook, NY, USA.

time t_k, $\mathcal{L}(\cdot,\cdot)$ a loss function providing a distance measure between empirical distributions, and the drift $\mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x})$ parameterized by a fully connected neural network whose weights $\boldsymbol{\theta}$ are tuned as part of the optimization. Here $\mathcal{L}(\cdot,\cdot)$ provides an optimal transport distance measure between the predicted density $\hat{\mu}^k_{\boldsymbol{\theta}}(\mathbf{x})$ at time t_k, given the previous measurement, and the empirical distribution $\nu^k(\mathbf{x})$ at time t_k, $k \in [K]$. Given its high computational and sample complexity, Sinkhorn divergences are used to estimate such optimal transport measures together with related algorithms to more efficiently solve the minimization problem in (4)–(6) [1, 3].

2. OUR APPROACH

We derive two approaches to address the general inference problem of interest, both by studying and dealing directly with the stochasticity of the process without resorting to any simplifying approximations, and then efficiently solving the resulting optimization problem. Specifically, in our first approach, we exploit the full stochastic dynamics of the interacting components of the system, each formulated in terms of the FP equation (2), to generate the predicted density $\hat{\mu}^k_{\boldsymbol{\theta}}(\mathbf{x})$ with respect to the unknown drift $\mathbf{f}(\mathbf{x})$ and diffusion $\mathbf{D}$. This then leads to the optimization problem:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{k=1}^{K} \mathcal{L}(\hat{\mu}^k_{\boldsymbol{\theta}}, \nu^k), \tag{7}$$

$$\hat{\mu}^k_{\boldsymbol{\theta}}(\mathbf{x}) = \frac{1}{N}\sum_{i=1}^{N} \hat{P}_i(\mathbf{x}, t_k), \quad \hat{P}_i(\mathbf{x}, t_{k-1}) = \delta(\mathbf{x} - \mathbf{x}_i(t_{k-1})),$$

$$\frac{d\hat{P}_i(\mathbf{x},t)}{dt} = \nabla\left[-\mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x})\hat{P}_i(\mathbf{x},t) + \nabla(\mathbf{D}_{\boldsymbol{\theta}}(\mathbf{x})\hat{P}_i(\mathbf{x},t))\right],$$

where $\mathcal{L}(\cdot,\cdot)$ is a loss function providing an optimal transport distance measure between the predicted density $\hat{\mu}^k_{\boldsymbol{\theta}}(\mathbf{x})$ at time t_k, given the previous measurement, and the empirical distribution $\nu^k(\mathbf{x})$ at time t_k, $k \in [K]$. In contrast with (4)–(6), the predicted densities $\hat{\mu}^k_{\boldsymbol{\theta}}(\mathbf{x})$ are expressed as the arithmetic mean of the densities $\hat{P}_i(\mathbf{x}, t_k)$ obtained by evolving the FP equation from t_{k-1} to t_k, assuming a Dirac delta function for the initial distribution on the previous measurements $\mathbf{x}_i(t_{k-1})$. Both Wasserstein metrics and Sinkhorn divergences are considered for $\mathcal{L}(\cdot,\cdot)$. We note that the weights $\gamma(t_k)$ can be easily accommodated in (7) of our formulation as in (4) of the DPF approach, but omit these details for ease of exposition.

The predicted densities $\hat{\mu}^k_{\boldsymbol{\theta}}(\mathbf{x})$ can be equivalently expressed in terms of a transition probability, or a Green's function, associated with the FP equation as

$$\hat{P}_i(\mathbf{x}, t_k) = \hat{P}_{\boldsymbol{\theta}}(\mathbf{x}, t_k | \mathbf{x}_i(t_{k-1}), t_{k-1}). \tag{8}$$

In practice, the transition probability $\hat{P}_{\boldsymbol{\theta}}(\mathbf{x}, t_k|\mathbf{y}, t_{k-1})$ can be computed by various approaches, including: (i) evolving numerically the FP equation forward in time, which is exact but inefficient; (ii) assuming a Gaussian transition distribution, which is efficient but primarily accurate for linear systems and small time intervals $\delta t = t_k - t_{k-1}$; and (iii) employing refined approximations of the Green's function, which is efficient and more broadly accurate.

In our second approach, we evaluate the transition probability corresponding to state $\mathbf{x}_i$ at time t_k, $i \in [N]$, $k \in [K]$, by expressing the original optimization problem in terms of the following maximum likelihood optimization

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}, \varpi_1, \ldots, \varpi_K} \prod_{k=1}^{K}\prod_{i=1}^{N} \hat{P}_{\boldsymbol{\theta}}(\mathbf{x}_{\varpi_k(i)}(t_k), t_k | \mathbf{x}_i(t_{k-1}), t_{k-1}), \tag{9}$$

where the conditional densities $\hat{P}_{\boldsymbol{\theta}}(\cdot,\cdot|\cdot,\cdot)$ depend on the parameterization $\boldsymbol{\theta}$ of the drift $\mathbf{f}_{\boldsymbol{\theta}}$ and diffusion $\mathbf{D}_{\boldsymbol{\theta}}$ by means of (8) associated with the FP equation. The optimization is then performed over permutations $\varpi_k(i)$, $k \in [K]$, $i \in [N]$, that determine the most likely pairing between measurements at adjacent times, thus attempting to reconstruct the desired trajectories. We can equivalently formulate the optimization problem (9) in terms of minimizing the negative summations of the log-likelihoods as

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}, \varpi_1, \ldots, \varpi_K} \left(-\sum_{k=1}^{K}\sum_{i=1}^{N} \log \hat{P}_{\boldsymbol{\theta}}(\mathbf{x}_{\varpi_k(i)}(t_k), t_k | \mathbf{x}_i(t_{k-1}), t_{k-1}) \right), \tag{10}$$

and hence (10) is equivalent to a linear assignment optimization problem that can be solved in polynomial time.

When the time intervals $\delta t = t_k - t_{k-1}$, $k \in \{2, \ldots, N\}$, of the general FP equation are small, then the transition probability can be approximated by a Gaussian distribution [4]:

$$\hat{P}_{\boldsymbol{\theta}}(\mathbf{x}_k, t_k | \mathbf{x}_{k-1}, t_{k-1}) = \frac{1}{\sqrt{(2\pi\delta t)^n \det \mathbf{D}_{\boldsymbol{\theta}}(\mathbf{x}_{k-1})}} \exp\Bigg(-\frac{1}{4\delta t}(\mathbf{x}_k^\top - \mathbf{x}_{k-1}^\top - \delta t \mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x}_{k-1})^\top)\mathbf{D}_{\boldsymbol{\theta}}(\mathbf{x}_{k-1})^{-1}(\mathbf{x}_k - \mathbf{x}_{k-1} - \mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x}_{k-1})\delta t)\Bigg).$$

Upon substituting the above equation into (10), we derive

$$\begin{aligned}\hat{\boldsymbol{\theta}} &= \arg\min_{\boldsymbol{\theta}, \varpi_1, \ldots, \varpi_K} \sum_{k=1}^{K}\sum_{i=1}^{N} \Bigg(\frac{1}{4\delta t}\Big[\mathbf{x}_{\varpi_k(i)}(t_k)^\top - \mathbf{x}_i(t_{k-1})^\top \\ &\quad - \delta t \mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x}_i(t_{k-1}))^\top\Big] \cdot \mathbf{D}_{\boldsymbol{\theta}}(\mathbf{x}_i(t_{k-1}))^{-1} \cdot \Big[\mathbf{x}_{\varpi_k(i)}(t_k) \\ &\quad - \mathbf{x}_i(t_{k-1}) - \delta t \mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x}_i(t_{k-1}))\Big] + \frac{1}{2}\log\det \mathbf{D}_{\boldsymbol{\theta}}(\mathbf{x}_i(t_{k-1}))\Bigg), \\ &= \arg\min_{\boldsymbol{\theta}} \sum_{k=1}^{K} \Bigg(\frac{1}{4\delta t} W_{\mathbf{D}_{\boldsymbol{\theta}}(\mathbf{x}_i(t_{k-1}))^{-1}}(\mu^k_{\boldsymbol{\theta}}, \nu^k) \\ &\quad + \frac{1}{2}\langle \log\det \mathbf{D}_{\boldsymbol{\theta}}(\mathbf{x})\rangle_{\mathbf{x}_i(t_{k-1})}\Bigg), \end{aligned} \tag{11}$$

where the inner product above is interpreted as a distance between a measurement $\mathbf{x}_{\varpi_k(i)}(t_k)$ at time t_k and the mean $\hat{m}_{i,k} = \mathbf{x}_i(t_{k-1}) + \delta t \mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x}_i(t_{k-1}))$ of the Gaussian likelihood based on a measurement $\mathbf{x}_i(t_{k-1})$ at time t_{k-1} with respect to the spatially dependent metric $\mathbf{D}_{\boldsymbol{\theta}}(\mathbf{x}_i(t_{k-1}))^{-1}$. The optimization over the permutations $\varpi_k(i)$, $k \in [K]$, $i \in [N]$, implies that this distance is exactly the Wasserstein distance with metric $\mathbf{D}_{\boldsymbol{\theta}}(\mathbf{x}_i(t_{k-1}))^{-1}$ between the two empirical distributions: ν^k, with masses at the measured data (3), and $\mu^k_{\boldsymbol{\theta}}$ given by $\mu^k_{\boldsymbol{\theta}}(\mathbf{x}) = \frac{1}{N}\sum_{i=1}^{N} \delta(\mathbf{x} - \mathbf{x}_i(t_{k-1}) - \delta t \mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x}_i(t_{k-1}))$, with masses at the predicted Gaussian likelihood means. The last term in (11) depends only on $\mathbf{D}_{\boldsymbol{\theta}}(\mathbf{x})$ and thus, when the diffusion matrix is unknown as assumed herein, this last term plays the role of a regularization term (otherwise, when $\mathbf{D}_{\boldsymbol{\theta}}(\mathbf{x})$ is known, the last term is a constant and plays no role). The minimization of the Wasserstein distance with metric $\mathbf{D}_{\boldsymbol{\theta}}(\mathbf{x}_i(t_{k-1}))^{-1}$ in (11) will tend to increase the spectra of the diffusion matrix $\mathbf{D}_{\boldsymbol{\theta}}(\mathbf{x})$, while the corresponding regularization term will tend to favor small eigenvalues.

3. PERFORMANCE EVALUATION

We first note that the computational complexity of each of our approaches is no greater than that of the drift optimization step of the DPF approach in (4)–(6) from [1, 3], while the the score function inference required for the DPF approach introduces additional computational complexity. Next, we compare the prediction accuracy of our second approach based on the maximum likelihood formulation in (11) with that of the DPF approach (4)–(6), both of which can be performed explicitly in the case of Ornstein-Uhlenbeck (OU) dynamics where the drift is linear and fixed with $\mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x}) = -\mathbf{A}\mathbf{x}$ and the diffusion is constant noise with $\mathbf{D}_{\boldsymbol{\theta}}(\mathbf{x}) = \mathbf{D}$. We assume the steady-state score $\hat{s}(\mathbf{x})$ in (6) is inferred perfectly, i.e., the stationary distribution is given by $P_{ss}(\mathbf{x}) = ((2\pi)^2 \det(\boldsymbol{\Sigma}))^{1/2} \exp\Big(-\frac{1}{2}\mathbf{x}^\top \boldsymbol{\Sigma}^{-1}\mathbf{x}\Big)$, where the covariance $\boldsymbol{\Sigma}$ is the solution to the Lyapunov equation $\mathbf{A}\boldsymbol{\Sigma} + \boldsymbol{\Sigma}\mathbf{A}^\top = 2\mathbf{D}$, and therefore the steady-state score is given by $\hat{s}(\mathbf{x}) = \nabla \log P_{ss}(\mathbf{x}) = -\boldsymbol{\Sigma}^{-1}\mathbf{x}$.

Analytical comparison. Under the above assumptions, the DPF in (6) becomes $d\mathbf{x} = (-\mathbf{A}\mathbf{x} + \mathbf{D}\boldsymbol{\Sigma}^{-1}\mathbf{x})d\tau$, which from the solution of (4)–(6) yields the desired prediction for time t_k, $k \in [K]$, as follows

$$\begin{aligned}\hat{\mathbf{x}}(t_k) &= \exp\big((-\mathbf{A} + \mathbf{D}\boldsymbol{\Sigma}^{-1})\delta t\big)\mathbf{x}(t_{k-1}) \\ &\approx (\mathbf{I} - \delta t(\mathbf{A} - \mathbf{D}\boldsymbol{\Sigma}^{-1}))\mathbf{x}(t_{k-1}). \end{aligned} \tag{12}$$

The corresponding maximum likelihood mean prediction from the solution of our formulation in (11) renders

$$\hat{m}(t_k) = \exp(-\mathbf{A}\delta t)\mathbf{x}(t_{k-1}) \approx (\mathbf{I} - \delta t\mathbf{A})\mathbf{x}(t_{k-1}). \tag{13}$$

The superior prediction accuracy of (13) over (12) can be observed analytically by considering the difference between the term $\delta t(\mathbf{A} - \mathbf{D}\boldsymbol{\Sigma}^{-1})$ in (12) and the term $\delta t\mathbf{A}$ in (13). In the DPF ODE, the additional $\mathbf{D}\boldsymbol{\Sigma}^{-1}$ term introduces a systematic bias in the predicted position of the measurement at the next time epoch. Such bias may help to explain the limitation that this previous DPF approach cannot accurately infer the drift from stationary data as demonstrated in [1]. Interestingly, they found that using stationary data led to inference of the equilibrium solution corresponding to $\mathbf{A}\boldsymbol{\Sigma} = \boldsymbol{\Sigma}\mathbf{A}^\top$ (and therefore $\boldsymbol{\Sigma} = \mathbf{A}^{-1}\mathbf{D}$), even for systems that are not in equilibrium. We conjecture that their optimization scheme may implicitly prefer $\hat{\mathbf{x}}(t_k) = \mathbf{x}(t_{k-1})$ in the stationary case, thus leading to this equilibrium solution. In contrast, our approach yields an unbiased prediction for the mean location of the next measurement. As our numerical example below further demonstrates, this allows us to accurately infer the drift even from stationary data.

Numerical experiment. As a representative example, we consider the six-dimensional OU process previously studied in [1, 2]. The process has non-isotropic diffusion and an asymmetric drift matrix; see equation (H5) in [2] for the drift and diffusion matrix elements. We generate trajectories of this process using the Euler-Maruyama discretization with step size $dt = 0.001$, discarding the initial transient portions of the trajectories so that the data are samples of the stationary distribution of the OU process.

Cross-sectional density measurements are collected at a time interval $\delta t = 0.1$ across K time epochs with N measurements at each time epoch. As emphasized above, the correspondence between measurements at different time epochs (i.e., the trajectory information) is unknown. For the representative example in Fig. 1, we use $K = 20$ and $N = 200$.

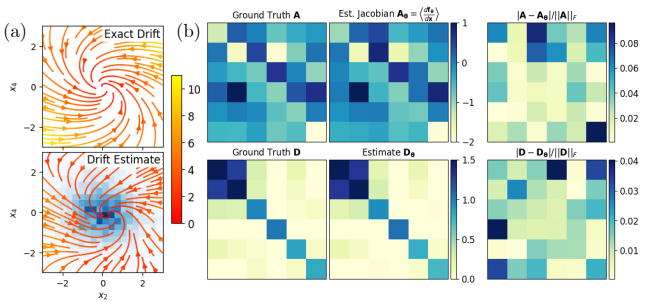

Figure 1: Inference of OU process without trajectories. (a) Exact (top) and inferred (bottom) drift in a 2D slice of the 6D state space. A histogram of the measured data (across all times) is shown in blue in the background of the lower plot. (b) Comparison of inferred and exact drift and diffusion matrices. Since our drift model is nonlinear we compare against the Jacobian $d\mathbf{f}_{\boldsymbol{\theta}}/d\mathbf{x}$ averaged over measured datapoints. The final column shows the difference between exact and inferred matrices relative to the matrix norms.

The drift $\mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x})$ is parameterized with a fully connected feed-forward neural network with two hidden layers and 10 nodes per layer. We assume the diffusion is constant $\mathbf{D}_{\boldsymbol{\theta}}(\mathbf{x}) = \mathbf{D}_{\boldsymbol{\theta}}$, but the entries of the diffusion matrix are unknown. The drift and diffusion terms are learned using the loss function given in (11), i.e., using the Wasserstein distance small-δt limit of the maximum likelihood optimization problem.

Fig. 1(a) respectively shows the exact and inferred drift fields $\mathbf{f}(\mathbf{x})$ and $\mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x})$ for a two-dimensional slice (along the (x_2, x_4)-axis) of the six-dimensional state space. Our inference approach accurately reproduces quantitatively the drift of the OU process, performing best in the region near the origin where the data is concentrated (as shown by the histogram in the background of the estimated drift field).

To further quantify the performance of our inference scheme, Fig. 1(b) shows comparisons of the exact and inferred OU drift and diffusion matrices. For the drift comparison (where our neural network model is nonlinear), we compute the Jacobian and average over measured data points. Relative to the matrix norms, our approach accurately captures elements of these matrices with less than 10% and 5% error, respectively.

The above analytical/numerical comparisons demonstrate two major advances in our approach to stochastic process inference without trajectories: (1) we infer the dynamics from stationary data; (2) we simultaneously infer the drift and diffusion. It is important to note that neither (1) nor (2) was possible in previous work [1, 3]. Ongoing work is studying performance scaling with both data quantity and regularization schemes to prevent overfitting, particularly in areas of the state space with limited data.

4. REFERENCES

[1] V. Chardès, et al. Stochastic force inference via density estimation. In *NeurIPS 2023 AI for Science Workshop*, 2023.

[2] A. Frishman and P. Ronceray. Learning force fields from stochastic trajectories. *Phys. Rev. X*, 10:021009, Apr 2020.

[3] S. Maddu, et al. Inferring biological processes with intrinsic noise from cross-sectional data. *arXiv:2410.07501v1*, 2024.

[4] H. Risken. *The Fokker-Planck Equation: Methods of Solution and Applications*. Springer Berlin Heidelberg, 1996.