

# Sparse-to-Field Reconstruction via Stochastic Neural Dynamic Mode Decomposition

**Yujin Kim**  
**Sarah Dean**

*Department of Computer Science, Cornell University, Ithaca, NY, USA.*

YK826@CORNELL.EDU  
SDEAN@CORNELL.EDU

**Editors:** G. Sukhatme, L. Lindemann, S. Tu, A. Wierman, N. Atanasov

## Abstract

Many consequential real-world systems, like wind fields and ocean currents, are dynamic and hard to model. Learning their governing dynamics remains a central challenge in scientific machine learning. Dynamic Mode Decomposition (DMD) provides a simple, data-driven approximation, but practical use is limited by sparse/noisy observations from continuous fields, reliance on linear approximations, and the lack of principled uncertainty quantification. To address these issues, we introduce Stochastic NODE–DMD, a probabilistic extension of DMD that models continuous-time, nonlinear dynamics while remaining interpretable. Our approach enables continuous spatiotemporal reconstruction at arbitrary coordinates and quantifies predictive uncertainty. Across four benchmarks, a synthetic setting and three physics-based flows, it surpasses a baseline in reconstruction accuracy when trained from only 10% observation density. It further recovers the dynamical structure by aligning learned modes and continuous-time eigenvalues with ground truth. Finally, on datasets with multiple realizations, our method learns a calibrated distribution over latent dynamics that preserves ensemble variability rather than averaging across regimes. Our code is available [here](#).

**Keywords:** system identification, dynamic mode decomposition, neural ODEs, uncertainty quantification, stochastic dynamics learning, sparse observations

## 1. Introduction

Data-driven identification of high-dimensional systems—inferring dynamics from sequential data—is a core challenge in scientific machine learning (Brunton et al., 2016; Raissi et al., 2019; Lusch et al., 2018). Applications span large-scale phenomena such as weather and climate (Shi et al., 2015; Rasp et al., 2020; Kurth et al., 2023) and finer-scale neuroscience problems like brain imaging reconstruction (Lee et al., 2021; Mechelli et al., 2005; Misra and Pessoa, 2025). Beyond scientific value, system identification underpins planning and control in open environments: wind-field models aid balloon navigation and UAV control (Bellemare et al., 2020; Sydney et al., 2013), ocean-current models guide marine robots (Vasiljević et al., 2017; Wiggert et al., 2022), and traffic-flow models inform autonomous driving (Campbell et al., 2010). In these applications, probabilistic modeling improves robustness and long-horizon stability.

Existing approaches fall into three groups: physics-based, operator-theoretic (Koopman-based), and black-box. Physics-based models (Brunton et al., 2016; Raissi et al., 2019) leverage governing equations and are interpretable, but can be noise-sensitive, expensive at scale, and rarely probabilistic. Operator methods grounded in spectral/Koopman theory (Mezić, 2005)—including DMD (Rowley et al., 2009; Schmid, 2010a; Tu, 2013; Lusch et al., 2018) and neural operators (Lu et al., 2019; Li et al., 2020b; Kovachki et al.,

2023)—yield interpretable linear surrogates (Kutz et al., 2016a) but are constrained by linear evolution and grid-based operations, limiting performance with sparse observations and strong nonlinearities. Hybrid efforts that pair neural implicit representations (Mildenhall et al., 2021; Niemeyer et al., 2022; Müller et al., 2022) with optimized DMD (Askham and Kutz, 2018; Sashidhar and Kutz, 2022) (e.g., (SaraerToosi et al., 2025)) expand expressivity yet remain fundamentally linear and deterministic. Black-box sequential models (Chen et al., 2018; Pathak et al., 2018; Kidger et al., 2020) scale well and are resolution-agnostic (Brunton et al., 2020), but often lack uncertainty quantification, struggle to generalize, destabilize over long horizons, and may violate physics (Chen et al., 2018; Kidger et al., 2020; Brunton et al., 2020). Probabilistic variants (Krishnan et al., 2015; Doerr et al., 2018; Alvarez et al., 2013) add uncertainty quantification but demand more data and computation and reduce interpretability.

In many dynamical systems, stochasticity reflects both imperfect observation and unmodeled dynamics (e.g., measurement noise, environmental perturbations). Moreover, practical constraints—high experimental costs, physical inaccessibility, and sensor limitations—lead to sparse observations. Consequently, a method that is (i) physically interpretable, provides (ii) uncertainty quantification, and learns effectively from (iii) sparse, noisy datasets is needed. This paper introduces Stochastic Neural Ordinary Differential Equation Dynamic Mode Decomposition (Stochastic NODE-DMD), a deep probabilistic model for system identification using DMD. We focus on stochastic system identification from sparse and nonlinear datasets.

Prior work has developed probabilistic frameworks for DMD (Takeishi et al., 2017) and neural implicit representations for spatial bases (SaraerToosi et al., 2025). However, neither addresses the joint challenges of continuous spatiotemporal reconstruction with temporally coherent uncertainty quantification from sparse observations. Our method advances beyond these approaches by (1) replacing fixed spatial bases with an implicit neural encoder that enables continuous reconstruction from arbitrary sparse measurements, and (2) introducing a stochastic Neural ODE (Chen et al., 2018; Tzen and Raginsky, 2019; Li et al., 2020a) for latent dynamics that maintains long-term temporal coherence. This unified framework provides uncertainty quantification for both measurement noise and dynamics variance across multiple realizations.

We evaluate Stochastic NODE-DMD on four datasets: a synthetic dataset with linear mode evolution and added noise, and three physics-based benchmarks including the Gray-Scott reaction-diffusion model for chemical dynamics (Pearson, 1993), Navier-Stokes equations for modeling vorticity flow, and flow past a cylinder (Brunton et al., 2020). Using only 10% of fixed datapoints across sequences to simulate sparsity, we assess reconstruction quality and dynamics retrieval. Comparisons of learned modes and coefficients against ground-truth parameters demonstrate the model’s reliability, while tests on datasets with multiple dynamics highlight effective uncertainty quantification.

## 2. Background and Problem Formulation

### 2.1. Problem setting

Consider a continuous-time nonlinear dynamical system with state  $\mathbf{x}(t)$  governed by dynamics  $f$ :

$$\dot{\mathbf{x}}(t) = f(\mathbf{x}(t)), \quad \mathbf{x}(t) \in \mathcal{X}, \quad (1)$$

In many applications, such as fluid dynamics or wave propagation, the overall state  $\mathbf{x}(t)$  is as a continuous function over a spatial domain. With spatial coordinate  $\mathbf{s} \in \Omega \subset \mathbb{R}^d$ ,  $\mathbf{x}(\mathbf{s}, t) \in \mathbb{C}$  is the value of the state

at coordinate  $\mathbf{s}$  and time  $t$ , and  $\mathcal{X} = L^2(\Omega)$  is the function space of square-integrable fields over the spatial domain  $\Omega$ . As a result, Eq. (1) typically manifests as a partial differential equation (PDE), though it is commonly approximated by discretization for numerical simulation. In particular, a grid of  $\Omega$  is defined, denoted  $S_\Omega \subset \Omega$  with  $|S_\Omega| = n$ . Then the system is approximated with a finite dimensional state in  $\mathbb{C}^n$  and corresponding approximate dynamics. For an  $\epsilon$  grid, we typically require  $n$  to scale like  $\text{vol}(\Omega)/\epsilon^d$ , so even in this finite dimensional approximation, the state is high dimensional.

We instead consider observations that are spatially sparse, such as measurements from a few fixed meteorological or oceanographic stations. Define a finite set of spatial coordinates  $S = \{\mathbf{s}_1, \dots, \mathbf{s}_m\} \subset \Omega$  and let  $\mathbf{x}(S, t) \in \mathbb{C}^m$  denote the values of the state at those coordinates at time  $t$ . We observe noisy measurements at discrete times  $t_k$  ( $k = 0, \dots, p-1$ ) as  $\mathbf{y}_k = \mathbf{x}(S, t_k) + \eta_k$ , where  $\eta_k \in \mathbb{C}^m$  represents additive measurement noise. The resulting dataset consists of the observation sequence  $\{\mathbf{y}_k\}_{k=0}^{p-1}$ , such as time-indexed sensor readings. The goal is to learn a low-dimensional and interpretable model that captures the underlying spatiotemporal dynamics, enables prediction of fields  $\mathbf{x}(\mathbf{s}, t)$  at arbitrary coordinates  $\mathbf{s}$  (including unseen full-resolution grids), and quantifies uncertainty due to factors like noise, subsampling, and model mismatch. To this end, we formulate the problem as inferring a stochastic dynamic mode decomposition over continuous space by modeling the mode coefficient as a latent state.

## 2.2. Linear Evolution and Dynamic Mode Decomposition

While the governing dynamics in Eq. (1) are nonlinear and infinite-dimensional, many phenomena of interest can be well represented in a low dimensional manner. This is achieved by projecting the spatiotemporal data onto a set of spatial modes  $W_i$ , yielding continuous-time mode coefficients  $\phi_i(t)$  and their discrete-time samples  $\phi_{i,k} = \phi_i(t_k)$ :

$$\mathbf{x}(\mathbf{s}, t) \approx \sum_{i=1}^r \phi_i(t) W_i(\mathbf{s}), \quad \mathbf{y}_k \approx \sum_{i=1}^r \phi_{i,k} \mathbf{w}_i + \eta_k, \quad (2)$$

where  $\mathbf{w}_i = W_i(S) \in \mathbb{C}^m$  are the sampled modes and  $r$  is the number<sup>1</sup> of modes. Under an additional assumption of linear mode evolution, the coefficients evolve as:

$$\dot{\phi}_i(t) = \lambda_i \phi_i(t), \quad \phi_{i,k+1} = e^{\lambda_i \Delta t} \phi_{i,k}, \quad (3)$$

with complex eigenvalue  $\lambda_i = \alpha_i + j\beta_i$  encoding growth rate  $\alpha_i$  and frequency  $\beta_i$ . This linear structure, while an approximation, enables predictive modeling and spectral analysis of complex systems (Rowley et al., 2009; Schmid, 2010a). This approximation is theoretically grounded in Koopman operator theory (Mezić, 2005), which lifts nonlinear dynamics into a linear (infinite-dimensional) space of observables.

Dynamic Mode Decomposition (DMD) is a tractable approach to linear dynamics on high or infinite dimensional states. Given an observation  $\mathbf{y}_k \in \mathbb{C}^m$  at times  $t_k$ , collection of sequential observation data pair  $Y$  and  $Y'$  are defined as  $Y = [\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_{p-1}] \in \mathbb{C}^{m \times p}$  and  $Y' = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_p] \in \mathbb{C}^{m \times p}$ .

1. In standard DMD, the maximum number of computable modes is limited by the rank of the data matrix. However, it is common to truncate to a smaller  $r$  by retaining only the dominant singular values from the SVD step, as this reduced-rank approximation is sufficient to capture the essential dynamics in many systems of interest (Tu, 2013; Schmid, 2010b).

DMD computes a matrix  $A$  such that  $Y' \approx AY$ , from which the eigenvalues  $\mu_i$  and mode matrix  $W \in \mathbb{C}^{m \times r}$  (with  $r$  effective modes) are obtained. The discrete eigenvalues satisfy  $\mu_i \approx e^{\lambda_i \Delta t}$ , linking DMD to the continuous linear evolution in Eq. (3). In practice, operating on vectorized, fixed-grid snapshots provides computational simplicity but reduces flexibility on irregular or continuous spatial domains; the linear-evolution assumption yields clear mode structure yet can leave nonlinear residuals unmodeled in complex systems; and the standard deterministic formulation emphasizes point estimates over calibrated uncertainty, which can be challenging under sparse sensors, noise, or missing data.

### 2.3. Probabilistic DMD as a Generative Model

To incorporate uncertainty and enable principled probabilistic inference, DMD can be reformulated as a latent-variable generative model (Takeishi et al., 2017). In this framework, the mode coefficients from the linear evolution (Section 2.2) are treated as latent variables  $\phi_{i,k} \in \mathbb{C}$ . For each *pair* of consecutive snapshots  $(\mathbf{y}_k, \mathbf{y}_{k+1})$ , the conditional likelihoods are defined as:

$$\mathbf{y}_k \mid \{\phi_{i,k}\}_{i=1}^r \sim \mathcal{CN}\left(\sum_{i=1}^r \phi_{i,k} \mathbf{w}_i, \sigma^2 I_m\right), \quad \mathbf{y}_{k+1} \mid \{\phi_{i,k}\}_{i=1}^r \sim \mathcal{CN}\left(\sum_{i=1}^r e^{\lambda_i \Delta t} \phi_{i,k} \mathbf{w}_i, \sigma^2 I_m\right), \quad (4)$$

with standard complex Gaussian priors  $\phi_{i,k} \sim \mathcal{CN}(0, 1)$ . Maximizing the marginal likelihood recovers classical DMD in the noise-free limit. However, because the generative model is defined *pairwise*—with independent latent states  $\phi_k$  for each transition—the global linear evolution in Eq. (3) is not enforced across time steps. This breaks long-term dynamical consistency, potentially leading to accumulating errors in multi-step forecasting and unstable mode amplitude trajectories under noise or missing data.

### 2.4. Nonlinear State-Space Formulation for Stochastic NODE–DMD

We extend the probabilistic generative model of Bayesian DMD (Takeishi et al., 2017) into a nonlinear state-space framework supporting continuous-time evolution and uncertainty quantification. Let  $\phi_k \in \mathbb{C}^r$  be the low-dimensional latent state vector encoding  $r$  modes (with  $r \ll m$ ),  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_r) \in \mathbb{C}^{r \times r}$  the diagonal matrix of eigenvalues, and  $W \in \mathbb{C}^{m \times r}$  the mode matrix. Then:

$$\phi_{k+1} = e^{\Lambda \Delta t} \phi_k + f_\theta(\phi_k, t_k) \Delta t + \zeta_k, \quad \zeta_k \sim \mathcal{CN}(\mathbf{0}, \tau^2 \Delta t \cdot I_r), \quad (5)$$

$$\mathbf{y}_k = W \phi_k + \eta_k, \quad \eta_k \sim \mathcal{CN}(\mathbf{0}, \sigma^2 I_m), \quad (6)$$

where  $f_\theta$  is a neural network capturing nonlinear residual dynamics (parameterized by  $\theta$ ),  $\zeta_k$  represents process noise, and  $\eta_k$  denotes observation noise. To derive a differentiable continuous-time model, set  $t_k = k \Delta t$  and take the limit  $\Delta t \rightarrow 0$ . The discrete increment  $\zeta_k = \tau \cdot \Delta B_k$  with  $\Delta B_k \sim \mathcal{CN}(\mathbf{0}, \Delta t \cdot I_r)$  converges in distribution to a Brownian increment, yielding the stochastic differential equation:

$$d\phi_t = (\Lambda \phi_t + f_\theta(\phi_t, t)) dt + \tau dB_t. \quad (7)$$

Equation (7) thus defines a neural stochastic differential equation combining linear evolution drift, nonlinear correction, and process noise.  $\Lambda \phi_t$  corresponds to the classical linear evolution, while the neural residual  $f_\theta$  provides adaptive corrections for nonlinearity and model mismatch. In the limit  $f_\theta \rightarrow 0$  and  $\tau^2 \rightarrow 0$ , the formulation reduces to deterministic DMD.

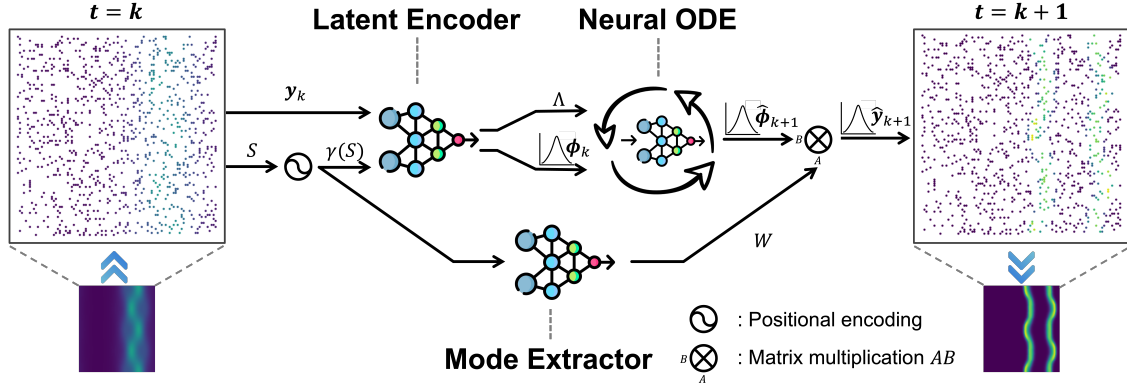


Figure 1: **Overview of the Stochastic NODE-DMD architecture.** At time step  $k$ , subsampled measurements  $\mathbf{y}_k$  and set of spatial coordinates  $S$  are fed into the `Latent Encoder`, which outputs eigenvalues  $\Lambda$  and the latent distribution  $p(\phi_k)$ . The `Neural ODE` evolves this distribution forward to time  $k+1$ . The predicted latent state distribution  $p(\hat{\phi}_{k+1})$  is then combined with spatial modes from the `Mode Extractor`  $W$  to reconstruct the distribution of subsampled field at time  $k+1$ , enabling uncertainty-aware, grid-free forecasting.

### 3. Stochastic Neural Ordinary Differential Equation DMD (NODE-DMD)

We extend the probabilistic DMD framework proposed by [Takeishi et al. \(2017\)](#), by incorporating continuous-time nonlinear latent evolution via Neural ODEs with stochastic diffusion ([Chen et al., 2018](#); [Tzen and Raginsky, 2019](#)). We also build upon NeuralDMD (NDMD) ([SaraerToosi et al., 2025](#)) which integrates neural implicit representation for sparse reconstruction. Our proposed method, termed Stochastic Neural ODE DMD (NODE-DMD), learns uncertainty-aware dynamics within a generative modeling framework. It combines the linear spectral structure of DMD with data-driven nonlinear corrections, and supports sparse, noisy observations. Figure 1 illustrates the overall architecture of Stochastic NODE-DMD. Given subsampled measurements  $\mathbf{y}_k$  at fixed spatial coordinates  $S$ , the model predicts the state distribution at time  $k+1$ . Inputs may originate from the dataset (“teacher forcing” with  $\mathbf{y}_k$ ) or prior model predictions (autoregressive forecasting from  $\hat{\mathbf{y}}_k$ ).

There are three main components: the `Mode Extractor` outputs spatial mode functions, and the `Latent Encoder` outputs eigenvalues  $\Lambda$  and the initial latent distribution  $\mathcal{CN}(\boldsymbol{\mu}_{\phi_k}, \Sigma_{\phi_k})$ . Both depend on a positional encoding  $\gamma(\mathbf{s})$ . The stochastic `Neural ODE` block evolves this distribution forward in continuous time. The predicted latent state distribution  $p(\hat{\phi}_{k+1}) = \mathcal{CN}(\boldsymbol{\mu}_{\phi_{k+1}}, \Sigma_{\phi_{k+1}})$  is combined with the mode function  $W_\psi(\gamma(\mathbf{s}))$  to generate the field prediction at arbitrary coordinates:  $\hat{\mathbf{y}}(\mathbf{s}, t_{k+1}) = W_\psi(\gamma(\mathbf{s}))^\top \hat{\phi}_{k+1}$ . By evaluating  $\hat{\mathbf{y}}_{k+1} = \hat{\mathbf{y}}(S, t_{k+1})$ , this architecture enables uncertainty-aware, multi-step forecasting from sparse and noisy observations.

### 3.1. Mode Extractor and Latent Encoder with Positional Encoding

To enable reconstruction at arbitrary coordinates  $\mathbf{s} \in \Omega$ , we parameterize spatial modes using a neural implicit representation  $W_\psi$  (Mildenhall et al., 2021; Müller et al., 2022), parameterized by  $\psi$ . To effectively encode raw spatial coordinates and enable the network to capture high-frequency details in continuous domains, we apply a positional encoding  $\gamma : \mathbb{R}^d \rightarrow \mathbb{R}^{2Ld+1}$  (where  $d$  is the coordinate dimensionality and  $L$  is the number of frequency bands) to individual coordinates before feeding them into the relevant modules (Tancik et al., 2020; Mildenhall et al., 2021). This encoding transforms each coordinate dimension  $s_j$  as  $\gamma(s_j) = [s_j, \sin(2^0\pi s_j), \cos(2^0\pi s_j), \dots, \sin(2^{L-1}\pi s_j), \cos(2^{L-1}\pi s_j)]$ . We compute the encoded set  $\gamma(S) = \{\gamma(\mathbf{s}_i)\}_{i=1}^m$ , which is used as input to the `Mode Extractor`. The pair  $(\mathbf{y}_k, \gamma(S))$  is used as input to the `Latent Encoder` which determines the mode coefficient  $\phi$  as probabilistic latent state using variational inference encoding (Kingma and Welling, 2013).

### 3.2. Stochastic Neural ODE for Latent Evolution

The `stochastic Neural ODE` enforces linear evolution across time steps while allowing nonlinear corrections and intrinsic uncertainty propagation. It evolves the latent mode state  $\phi(t) \in \mathbb{C}^r$  continuously via the SDE in Eq. (7), where the linear drift is encoded by  $\Lambda$ , the nonlinear residual is modeled by  $f_\theta$ , and the diffusion term  $\tau d\mathbf{B}_t$  captures intrinsic uncertainty. To overcome the independent pairwise-transition limitation of the generative model in Bayesian DMD (Takeishi et al., 2017), we propagate the distribution of  $\phi$  with a `stochastic Neural ODE`, allowing data-driven corrections and uncertainty transport. Furthermore, we reinforce temporal continuity using the consistency loss in Sec. 3.3. We discretize the SDE over substeps of  $\delta t$  with an uncertainty-aware Euler–Maruyama scheme (Kloeden and Pearson, 1977). Implementation details, including multi-substep updates and covariance propagation, are given in Appendix A.

### 3.3. Training Objective and Rollout Strategy

The model is trained end-to-end using a composite loss function that balances reconstruction accuracy, latent regularization, and temporal consistency. (Full mathematical details are provided in Appendix B). For a transition from time  $t_k$  to  $t_{k+1}$ , let  $\mathbf{y}_{k+1} \in \mathbb{C}^m$  be the target observation,  $(\hat{\boldsymbol{\mu}}_y, \hat{\boldsymbol{\sigma}}_y^2)$  the predicted mean and variance of the reconstructed field, and  $(\boldsymbol{\mu}_\phi, \boldsymbol{\sigma}_\phi^2), (\hat{\boldsymbol{\mu}}_\phi, \hat{\boldsymbol{\sigma}}_\phi^2)$  the latent distributions from the `Latent Encoder` and `Stochastic Neural ODE`, respectively.

- **Reconstruction Loss:** A Gaussian negative log-likelihood from  $p(\hat{\mathbf{y}}_{k+1}) = (\hat{\boldsymbol{\mu}}_y, \hat{\boldsymbol{\sigma}}_y^2)$  to  $\mathbf{y}_{k+1}$  encourages accurate field prediction while adapting to data uncertainty.
- **Latent KL Regularization:** A KL divergence term regularizes the latent mode coefficients  $\hat{\phi}_i$  toward a standard complex Gaussian prior.
- **Consistency Loss:** The latent distribution matching loss aligns the encoder-estimated distribution  $p(\phi_{k+1} | \mathbf{y}_{k+1}, \gamma(S))$  with the SDE-propagated distribution  $p(\hat{\phi}_{k+1} | \phi_k)$  using MSE and KL-divergence.

The total loss is a weighted sum of these three loss terms. We train the model end-to-end with a curriculum that gradually transitions inputs from ground truth  $(\mathbf{y}_k)$  to model predictions  $(\hat{\mathbf{y}}_k)$ , stabilizing training and enabling robust long-horizon autoregressive forecasts; details are provided in Appendix C.

## 4. Experiments

**Models.** We evaluated Stochastic NODE–DMD on four datasets and benchmarked it against Neural DMD (NDMD) (SaraerToosi et al., 2025), which learns DMD components via a neural implicit representation and makes predictions through open-loop spectral rollout. We set the mode rank to  $r = 4$  for Synthetic Sequence task and  $r = 8$  for all other tasks, and aligned key training hyperparameters such as the number of epochs and batch size across methods.

**Simulated dynamical systems.** We evaluate Stochastic NODE–DMD on four simulated systems, ranging from a linear synthetic sequence with known ground truth to nonlinear PDE-governed fluid flows. In Fig. 2, the second column shows the ground-truth field for each task. All systems are discretized on structured grids and evolved for  $T = 50$ – $150$  time steps to capture salient dynamics and to enable both short- and long-horizon evaluation. In the sparse-observation experiments (Sec. 4.1 and 4.2), we train with 10% fixed spatial sensors to emulate sparse measurements and evaluate reconstructions on the full grid.

- **Synthetic Sequence:** Four predefined spatial modes with exponential temporal evolution (Eq. (2)) and additive observation noise. This mirrors common synthetic setups used to validate DMD/Koopman methods (Dawson et al., 2016; Askham and Kutz, 2018; Zhang et al., 2017) preserving the ground-truth modes and eigenvalues. The dataset uses a  $32 \times 32$  grid and sequence length  $T = 50$  (Fig. 2(a)).
- **Gray-Scott Reaction-Diffusion:** A coupled PDE system modeling two reacting and diffusing chemical species (Pearson, 1993). The dataset uses a  $100 \times 100$  grid and sequence length  $T = 100$  (Fig. 2(b)).
- **2D Navier-Stokes Spectral Vorticity:** Incompressible 2D flow in vorticity using `torch-cfd` (Cao et al., 2024). The dataset uses a  $128 \times 128$  grid and sequence length  $T = 50$  (Fig. 2(c)).
- **2D Flow Past a Cylinder:** Viscous incompressible flow around a circular cylinder, governed by the Navier-Stokes equations simulated using `torch-cfd` (Cao et al., 2024) with finite volume method and pressure projection. The dataset uses a  $100 \times 100$  grid and sequence length  $T = 150$ . (Fig. 2(d)).

Details of parameters, initial conditions, numerical schemes, and implementation notes are provided in Appendix D.

### 4.1. Reconstructing continuous spatiotemporal dynamics from sparse observations.

We evaluate whether NODE–DMD can infer missing spatiotemporal information from sparse observations while remaining reliable to the underlying dynamics. To this end, only 10% of spatial locations are used as fixed sensors and observed over the entire sequence.

**Qualitative Results.** Figure 2 visualizes the final rollout segment for each task. We report NODE–DMD results in two evaluation modes: teacher-forced one-step prediction and autoregressive multi-step rollout. With only sparse measurements, NODE–DMD preserves coherent structures and sharp phase alignment, whereas NDMD produces blurred reconstructions that lose fine-scale features. The Cylinder Flow task highlights characteristic behaviors: in teacher-forced one-step prediction, NODE–DMD captures sharp geometric detail and phase alignment, whereas in autoregressive multi-step rollout, it maintains plausible geometry but exhibits phase lag due to accumulated frequency misalignment. By contrast, Neural DMD tends to average out the vortex shedding. Learning spatial modes as continuous functions of coordinates

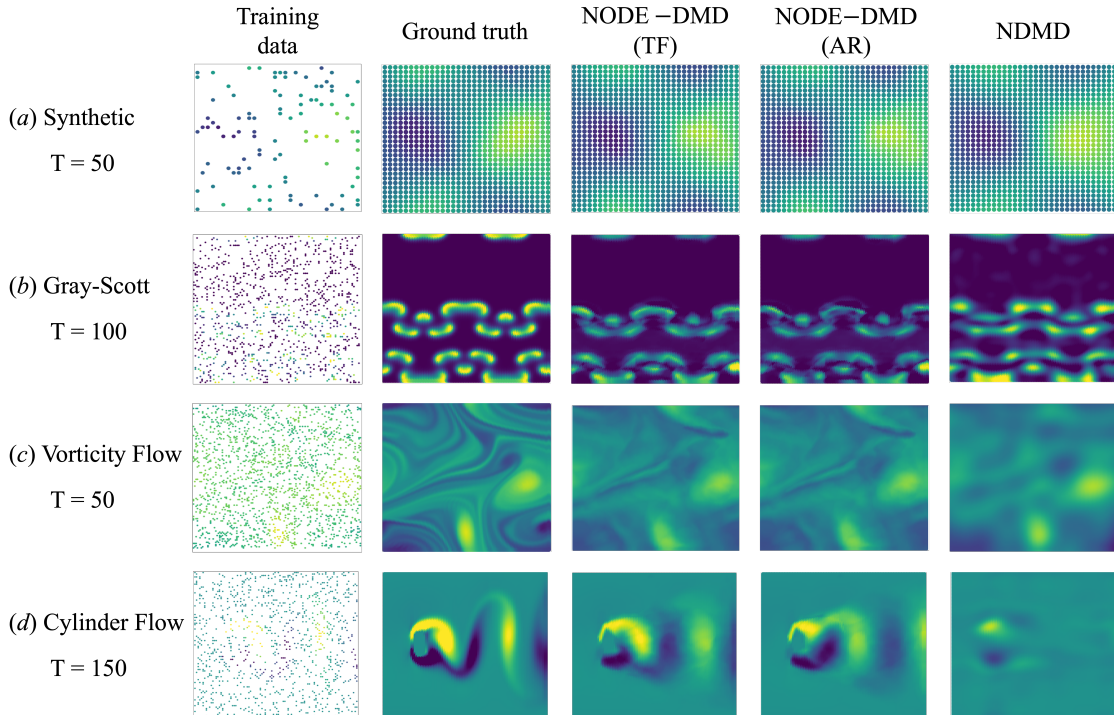


Figure 2: **Qualitative reconstruction results from sparse observations.** Row labels list the task and roll-out length  $T$  (final time step). Columns: (1) training data—10% of fixed spatial points sampled from the ground truth, (2) ground-truth full field, (3) NODE-DMD with teacher-forced one-step prediction, (4) NODE-DMD with autoregressive multi-step rollout, and (5) NDMD.

enables interpolation at unobserved locations and effectively super-resolves missing regions, mitigating the ambiguity induced by fixed sparse sensors.

**Quantitative results.** Table 1 reports the time-averaged per-pixel MSE. On the near-linear, mode-separable Synthetic DMD Sequence task, NODE-DMD is comparable to NDMD. On the more nonlinear systems—Gray-Scott, Spectral Vorticity, and Cylinder Flow—NODE-DMD achieves consistently lower single-step prediction error than NDMD, and shows comparable error in longer multi-step forecasts.

As expected, multi-step rollouts generally underperform one-step prediction because autoregressive recovery proceeds open-loop from the first frame, so small modeling errors accumulate over time. On Cylinder, periodicity makes pointwise MSE highly phase-sensitive: small frequency bias causes phase drift and is penalized without time-shift alignment, increasing the multi-step error. Note that the smaller error achieved by NDMD comes at the cost of “averaging out” uncertainty in the phase as shown in Fig. 2. We further analyze the role of stochastic variance in Section 4.3.

System	DMD SEQUENCE	GRAY-SCOTT	VORTICITY FLOW	CYLINDER FLOW
NDMD	<b>0.0442</b>	0.0049	0.0047	0.0467
NODE-DMD(TF)	0.0466	<b>0.0046</b>	<b>0.0024</b>	<b>0.0067</b>
NODE-DMD(AR)	0.0455	0.0076	<b>0.0024</b>	0.0473

Table 1: MSE of full snapshot reconstruction from sparse (10%) observations

## 4.2. Recovering Ground-Truth Dynamics from Noisy, Sparse Observations

We test whether NODE-DMD recovers continuous-time dynamics from noisy sequences observed at fixed 10% sensors. Additional details are presented in Appendix E; here we report qualitative results.

Fig. 3 visualizes mode portraits at a random time step ( $t=14$ ). For each mode  $k$ , we draw isocontours of the instantaneous spatial contribution  $|W_k(x)\phi_k(t)|$ . To ensure comparability across time and between GT/predictions, we fix absolute contour levels per mode using the GT sequence: pool  $|W_k(x)\phi_k(t)|$  over all  $(x, t)$  and set levels to the (30, 60, 90)-th percentiles of that pooled distribution.<sup>2</sup> Portraits in GT mode (left side in Fig. 3) and portraits predicted by the model (right side in Fig. 3) exhibit matched high-contribution regions, indicating that the model organizes spatial energy consistently with the ground-truth dynamics.

## 4.3. Learning a distribution over dynamics across realizations

We test whether NODE-DMD learns a posterior over spatiotemporal fields that captures variability across realizations of the same system (identical initial condition, different dynamics parameters). We train two models on the Vorticity Flow benchmark at a lower resolution than in Section 4.1. To isolate dynamics uncertainty from sparse-observation effects, training uses fully observed spatiotemporal fields (no sensor subsampling). For evaluation, we visualize how a massless object would move in each flow field. To track its trajectory, we recover velocities from vorticity using the classical streamfunction-vorticity formulation for 2D incompressible flow on a periodic domain, implemented via FFT-based Poisson inversion; see Hussaini and Zang (1986); Boyd (2001); Kundu et al. (2024) for details.

Figure 4 compares two training regimes. On the left, from a single velocity field realization, the posterior concentrates and the  $N=10$  sampled trajectories cluster tightly around the mean drift, indicating limited epistemic variability when trained on one realization only. On the right, training on a dataset of 10 realizations (same initial state but different dynamics parameters) yields a posterior that captures inter-realization variability: the ensemble exhibits a wider spread while preserving the mean drift. In other words, NODE-DMD does not collapse to an averaged, time-invariant pattern but learns a distribution that covers the family of dynamics induced by the training realizations.

## 5. Conclusion

We introduced Stochastic NODE-DMD, a probabilistic and interpretable framework for system identification that (i) reformulates DMD as a generative model, (ii) employs a neural implicit representation to enable

2. A spatially constant GT mode (e.g., mode 1) yields nearly flat  $|W_k|$  and degenerate isocontours; such cases may be omitted as in Fig. 3(a).

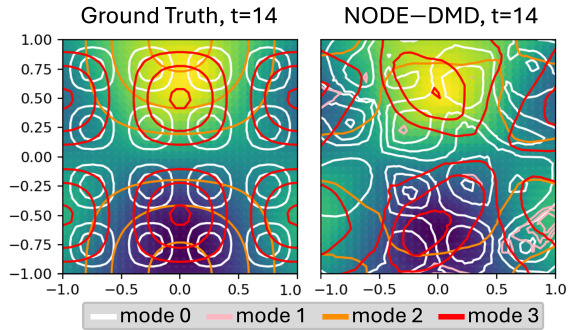


Figure 3: **Mode-portrait overlays at absolute levels.** Colored contours show per-mode isocontours of  $|W_k(x) \phi_k(t)|$  at thresholds fixed once per mode from the GT sequence (30/60/90-th percentiles). Backgrounds show the corresponding field (Left: ground-truth mode portrait; Right: model-predicted mode portrait).

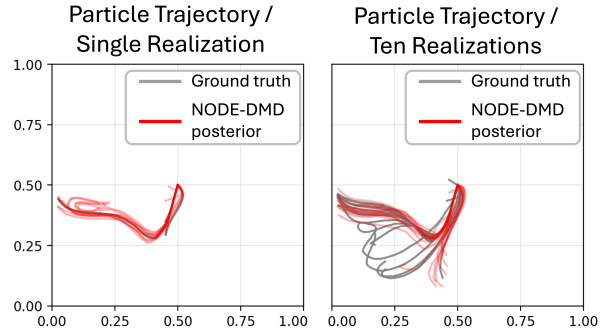


Figure 4: **Particle trajectories.** Trajectories obtained by transporting particles in velocity fields sampled from NODE-DMD (red), compared with ground-truth trajectories (gray). Left: Model trained on a single vorticity-field realization. Right: Model trained on 10 realizations.

grid-free, continuous spatial reconstruction from sparse observations, and (iii) augments the linear DMD drift with a stochastic Neural ODE to capture nonlinear residuals while propagating uncertainty. In doing so, the method preserves the spectral interpretability of DMD yet extends its applicability to sparse, noisy, and strongly nonlinear settings.

Across four benchmarks (one synthetic; reaction–diffusion, vorticity flow, cylinder flow), the method reconstructs from 10% fixed sensors, maintains coherent geometry and phase in long-horizon rollouts, and recovers principal dynamical factors (modes and eigenvalues). Training over multiple realizations learns a calibrated distribution over dynamics rather than collapsing to time-averaged patterns. Our proposed method is competitive on near-linear cases and stronger on more nonlinear systems, indicating that combining linear spectral structure with learned nonlinear corrections improves robustness under sparsity and noise while retaining interpretability.

**Limitations and future work.** Scaling to large, real-world settings will benefit from faster latent-SDE solvers and memory-/parallel-aware training, as well as temporal regularization and physics-informed priors to further improve long-horizon stability. An important potential application of this framework is modeling dynamic open environment for planning and control, such as balloons and UAVs operating in wind-fields. Such settings often include irregular, sparse sensor networks which may be augmented through online data assimilation or active sensing. We hope that our method for tractable, uncertainty-aware reconstruction and forecasting can interface directly with downstream planning and control in autonomous systems.

## References

- Mauricio A. Alvarez, David Luengo, and Neil D. Lawrence. Linear latent force models using gaussian processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(11):2693–2705, 2013.
- Travis Askham and J Nathan Kutz. Variable projection methods for an optimized dynamic mode decomposition. *SIAM Journal on Applied Dynamical Systems*, 17(1):380–416, 2018.
- Marc G Bellemare, Salvatore Candido, Pablo Samuel Castro, Jun Gong, Marlos C Machado, Subhdeep Moitra, Sameera S Ponda, and Ziyu Wang. Autonomous navigation of stratospheric balloons using reinforcement learning. *Nature*, 588(7836):77–82, 2020.
- John P Boyd. *Chebyshev and Fourier spectral methods*. Courier Corporation, 2001.
- Steven L Brunton, Joshua L Proctor, and J Nathan Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113(15):3932–3937, 2016.
- Steven L Brunton, Bernd R Noack, and Petros Koumoutsakos. Machine learning for fluid mechanics. *Annual Review of Fluid Mechanics*, 52:477–508, 2020.
- Mark Campbell, Magnus Egerstedt, Jonathan P How, and Richard M Murray. Autonomous driving in urban environments: approaches, lessons and challenges. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 368(1928):4649–4672, 2010.
- Shuhao Cao, Francesco Brarda, Ruipeng Li, and Yuanzhe Xi. Spectral-refiner: Accurate fine-tuning of spatiotemporal fourier neural operator for turbulent flows. *arXiv preprint arXiv:2405.17211*, 2024.
- Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. *Advances in neural information processing systems*, 31, 2018.
- Scott TM Dawson, Maziar S Hemati, Matthew O Williams, and Clarence W Rowley. Characterizing and correcting for the effect of sensor noise in the dynamic mode decomposition. *Experiments in Fluids*, 57(3):42, 2016.
- Andreas Doerr, Christian Daniel, Martin Schiegg, Duy Nguyen-Tuong, Stefan Schaal, Marc Toussaint, and Sebastian Trimpe. Probabilistic recurrent state-space models. *arXiv preprint arXiv:1801.10395*, 2018.
- M Yousuff Hussaini and Thomas A Zang. Spectral methods in fluid dynamics. Technical report, 1986.
- Patrick Kidger, James Morrill, James Foster, and Terry Lyons. Neural controlled differential equations for irregular time series. *Advances in neural information processing systems*, 33:6696–6707, 2020.
- Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

- Peter E Kloeden and RA Pearson. The numerical solution of stochastic differential equations. *The ANZIAM Journal*, 20(1):8–12, 1977.
- Nikola Kovachki, Zongyi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural operator: Learning maps between function spaces with applications to pdes. *Journal of Machine Learning Research*, 24(89):1–97, 2023.
- Rahul Krishnan, Uri Shalit, and David Sontag. Structured inference networks for nonlinear state space models. In *Proceedings of the AAAI conference on artificial intelligence*, volume 31, 2017.
- Rahul G. Krishnan, Uri Shalit, and David Sontag. Deep kalman filters. *arXiv preprint arXiv:1511.05121*, 2015.
- Pijush K Kundu, Ira M Cohen, David R Dowling, and Jesse Capecelatro. *Fluid mechanics*. Elsevier, 2024.
- Thorsten Kurth, Shashank Subramanian, Peter Harrington, Jaideep Pathak, Morteza Mardani, David Hall, Andrea Miele, Karthik Kashinath, and Anima Anandkumar. Fourcastnet: Accelerating global high-resolution weather forecasting using adaptive fourier neural operators. In *Proceedings of the platform for advanced scientific computing conference*, pages 1–11, 2023.
- J Nathan Kutz, Steven L Brunton, Bingni W Brunton, and Joshua L Proctor. *Dynamic mode decomposition: data-driven modeling of complex systems*. SIAM, 2016a.
- J Nathan Kutz, Steven L Brunton, Bingni W Brunton, and Joshua L Proctor. *Dynamic mode decomposition: data-driven modeling of complex systems*. SIAM, 2016b.
- Mi Hyun Lee, Nambeom Kim, Jaeun Yoo, Hang-Keun Kim, Young-Don Son, Young-Bo Kim, Seong Min Oh, Soohyun Kim, Hayoung Lee, Jeong Eun Jeon, et al. Multitask fmri and machine learning approach improve prediction of differential brain activity pattern in patients with insomnia disorder. *Scientific reports*, 11(1):9402, 2021.
- Xuechen Li, Ting-Kam Leonard Wong, Ricky T Q Chen, and David Duvenaud. Scalable gradients for stochastic differential equations. *International Conference on Artificial Intelligence and Statistics*, pages 3870–3882, 2020a.
- Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895*, 2020b.
- Lu Lu, Pengzhan Jin, and George E Karniadakis. Deeponet: Learning nonlinear operators for identifying differential equations based on the universal approximation theorem of operators. *arXiv preprint arXiv:1910.03193*, 2019.
- Bethany Lusch, J Nathan Kutz, and Steven L Brunton. Deep learning for universal linear embeddings of nonlinear dynamics. *Nature Communications*, 9(1):4950, 2018.

- Andrea Mechelli, Cathy J Price, Karl J Friston, and John Ashburner. Voxel-based morphometry of the human brain: methods and applications. *Current Medical Imaging*, 1(2):105–113, 2005.
- Igor Mezić. Spectral properties of dynamical systems, model reduction and decompositions. *Nonlinear Dynamics*, 41(1):309–325, 2005.
- Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021.
- Joyneel Misra and Luiz Pessoa. Brain dynamics and spatiotemporal trajectories during threat processing. *bioRxiv*, pages 2024–04, 2025.
- Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM transactions on graphics (TOG)*, 41(4):1–15, 2022.
- Michael Niemeyer, Jonathan T Barron, Ben Mildenhall, Mehdi SM Sajjadi, Andreas Geiger, and Noha Radwan. Regnerf: Regularizing neural radiance fields for view synthesis from sparse inputs. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5480–5490, 2022.
- Jaideep Pathak, Brian Hunt, Michelle Girvan, Zhixin Lu, and Edward Ott. Model-free prediction of large spatiotemporally chaotic systems from data: A reservoir computing approach. *Physical Review Letters*, 120(2):024102, 2018.
- John E Pearson. Complex patterns in a simple system. *Science*, 261(5118):189–192, 1993.
- Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019.
- Stephan Rasp, Peter D Dueben, Sebastian Scher, Jonathan A Weyn, Soukayna Mouatadid, and Nils Thuerey. Weatherbench: A benchmark data set for data-driven weather forecasting. *Journal of Advances in Modeling Earth Systems*, 12(11):e2020MS002203, 2020.
- Clarence W Rowley, Igor Mezić, Shervin Bagheri, Philipp Schlatter, and Dan S Henningson. Spectral analysis of nonlinear flows. *Journal of fluid mechanics*, 641:115–127, 2009.
- Yulia Rubanova, Ricky TQ Chen, and David K Duvenaud. Latent ordinary differential equations for irregularly-sampled time series. *Advances in neural information processing systems*, 32, 2019.
- Ali SaraerToosi, Renbo Tu, Kamyar Azizzadenesheli, and Aviad Levis. Neural dynamic modes: Computational imaging of dynamical systems from sparse observations. *arXiv preprint arXiv:2507.03094*, 2025.

- Diya Sashidhar and J Nathan Kutz. Bagging, optimized dynamic mode decomposition for robust, stable forecasting with spatial and temporal uncertainty quantification. *Philosophical Transactions of the Royal Society A*, 380(2229):20210199, 2022.
- Peter J Schmid. Dynamic mode decomposition of numerical and experimental data. *Journal of fluid mechanics*, 656:5–28, 2010a.
- Peter J. Schmid. Dynamic mode decomposition of numerical and experimental data. *J. Fluid Mech.*, 656: 5–28, 2010b.
- Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-chun Woo. Convolutional lstm network: A machine learning approach for precipitation nowcasting. In *Advances in Neural Information Processing Systems*, volume 28, 2015.
- Nitin Sydney, Brandon Smyth, and Derek A Paley. Dynamic control of autonomous quadrotor flight in an estimated wind field. *52nd IEEE Conference on Decision and Control*, pages 3609–3616, 2013.
- Naoya Takeishi, Yoshinobu Kawahara, Yasuo Tabei, and Takehisa Yairi. Bayesian dynamic mode decomposition. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17)*, pages 2814–2821, 2017.
- Matthew Tancik, Pratul Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. *Advances in neural information processing systems*, 33:7537–7547, 2020.
- Jonathan H Tu. *Dynamic mode decomposition: Theory and applications*. PhD thesis, Princeton University, 2013.
- Belinda Tzen and Maxim Raginsky. Neural stochastic differential equations: Deep latent gaussian models in the diffusion limit. *arXiv preprint arXiv:1905.09883*, 2019.
- Antonio Vasilijević, Đula Nađ, Filip Mandić, Nikola Mišković, and Zoran Vukić. Coordinated navigation of surface and underwater marine robotic vehicles for ocean sampling and environmental monitoring. *IEEE/ASME transactions on mechatronics*, 22(3):1174–1184, 2017.
- Marius Wiggert, Manan Doshi, Pierre FJ Lermusiaux, and Claire J Tomlin. Navigating underactuated agents by hitchhiking forecast flows. In *2022 IEEE 61st Conference on Decision and Control (CDC)*, pages 2417–2424. IEEE, 2022.
- Hao Zhang, Scott Dawson, Clarence W Rowley, Eric A Deem, and Louis N Cattafesta. Evaluating the accuracy of the dynamic mode decomposition. *arXiv preprint arXiv:1710.00745*, 2017.

## Appendix A. Stochastic Neural ODE procedure

We evolve the latent state by the SDE in Eq. (7). For implementation we work in  $\mathbb{R}^{2r}$  via real–imag stacking and use the real lift  $\Lambda_{\mathbb{R}}$  of  $\Lambda$ . Over  $\Delta t$ , we apply an uncertainty-aware Euler–Maruyama with  $p$  substeps  $\delta t$ . See Algorithm 1 for the full procedure.

**Algorithm 1:** Stochastic Neural ODE Integration with Uncertainty Propagation

**Input:** Initial latent mean  $\boldsymbol{\mu}_{\phi_t} \in \mathbb{C}^r$ , covariance  $\Sigma_{\phi_t} \in \mathbb{C}^{r \times r}$ , eigenvalues  $\Lambda$ , neural drift  $f_{\theta}$ , diffusion  $\tau$ , time step  $\delta t$ , horizon  $\Delta t$

**Output:** Final latent distribution  $\mathcal{CN}(\boldsymbol{\mu}_{\phi_{t+\Delta t}}, \Sigma_{\phi_{t+\Delta t}})$

```

for  $p = 0, 1, \dots, P - 1$  ; // where  $t_p = t + p\delta t$ ,  $\Delta t = P\delta t$ 
do
     $\boldsymbol{\delta} \leftarrow \Lambda \boldsymbol{\mu}_{\phi_{t_p}} + f_{\theta}(\boldsymbol{\mu}_{\phi_{t_p}}, t_p)$  // Total drift
     $\boldsymbol{\mu}_{\phi_{t_{p+1}}} \leftarrow \boldsymbol{\mu}_{\phi_{t_p}} + \Delta t \cdot \boldsymbol{\delta}$  // Mean update
     $J \leftarrow \nabla_{\phi}(\Lambda \phi + f_{\theta}(\phi, t_p))|_{\phi=\boldsymbol{\mu}_{\phi_{t_p}}}$ ; // Jacobian of drift
     $A \leftarrow I + \Delta t \cdot J$ 
     $\Sigma_{\phi_{t_{p+1}}} \leftarrow A \Sigma_{\phi_{t_p}} A^H + \Delta t \cdot \tau^2 I$  // Covariance propagation
end
return  $\mathcal{CN}(\boldsymbol{\mu}_{\phi_{t+\Delta t}}, \Sigma_{\phi_{t+\Delta t}})$ 
    
```

## Appendix B. Training Loss Function

For a transition from time  $t_k$  to  $t_{k+1}$ , let  $\mathbf{y}_{k+1} \in \mathbb{C}^m$  be the target observation,  $(\hat{\boldsymbol{\mu}}_y, \hat{\boldsymbol{\sigma}}_y^2)$  the predicted mean and variance of the reconstructed field, and  $(\boldsymbol{\mu}_{\phi}, \boldsymbol{\sigma}_{\phi}^2)$ ,  $(\hat{\boldsymbol{\mu}}_{\phi}, \hat{\boldsymbol{\sigma}}_{\phi}^2)$  the latent distributions from the Latent Encoder and Stochastic Neural ODE, respectively.

The total loss is

$$\mathcal{L} = w_{\text{recon}} \cdot \mathcal{L}_{\text{recon}} + w_{\text{kl}} \cdot \mathcal{L}_{\text{kl}} + w_{\text{cons}} \cdot \mathcal{L}_{\text{cons}}, \quad (8)$$

with weights  $w_{\text{recon}} = 3$ ,  $w_{\text{kl}} = 10^{-3}$ ,  $w_{\text{cons}} = 0.15$ .

- **Reconstruction Loss:** A Gaussian negative log-likelihood encourages accurate field prediction while adapting to data uncertainty:

$$\mathcal{L}_{\text{recon}} = \frac{1}{2} \left[ \log \hat{\boldsymbol{\sigma}}_y^2 + \frac{(\mathbf{y}_{k+1} - \hat{\boldsymbol{\mu}}_y)^2}{\exp(\log \hat{\boldsymbol{\sigma}}_y^2)} + \log(2\pi) \right].$$

- **Latent KL Regularization:** A KL divergence term regularizes the latent mode coefficients  $\hat{\phi}_i$  toward a standard complex Gaussian prior, preventing overfitting and promoting structured representations:

$$\mathcal{L}_{\text{kl}} = -\frac{1}{2} \sum_i (1 + \log \hat{\boldsymbol{\sigma}}_{\phi}^2 - \hat{\boldsymbol{\mu}}_{\phi}^2 - \exp(\log \hat{\boldsymbol{\sigma}}_{\phi}^2)).$$

- **Consistency Loss:** Latent distribution matching loss aligns the encoder-estimated distribution  $p(\phi_{k+1} | y_{k+1}, \gamma(S))$  with the SDE-propagated distribution  $p(\hat{\phi}_{k+1} | \phi_k)$  to enforce temporal consistency:

$$\mathcal{L}_{\text{cons}} = \text{MSE}(\mu_\phi, \hat{\mu}_\phi) + \kappa \cdot \text{KL}(\mathcal{CN}(\mu_\phi, \sigma_\phi^2) \| \mathcal{CN}(\hat{\mu}_\phi, \hat{\sigma}_\phi^2)),$$

where  $\kappa = 0.001$  scales the variance term.

## Appendix C. Training Detail

The model is trained end-to-end using gradient descent on the training dataset. Due to its recursive structure, we can select either ground-truth observations  $y_k$  or model predictions  $\hat{y}_k$  as inputs during training. To stabilize the process and enable robust multi-step forecasting, we apply a curriculum learning schedule that transitions from teacher forcing (using  $y_k$ ) to autoregressive mode (using  $\hat{y}_k$ ). For each training batch, we uniformly select teacher forcing mode with probability  $\epsilon$  (applying it to all sequences and timesteps in the batch) or autoregressive mode with probability  $1 - \epsilon$ , where  $\epsilon$  decays linearly from 1 to 0 over the course of training. In fully autoregressive mode, the model reconstructs the entire sequence of duration  $T$  given only the initial observation.

## Appendix D. Simulation Dataset Details

### D.1. Synthetic DMD Sequence

We construct a synthetic dataset based on the principles of DMD, where the underlying dynamics are generated from a known set of modes, eigenvalues, and initial coefficients. This allows direct comparison between learned Koopman modes and ground-truth DMD components. The state at each time step is defined as a linear combination of spatial modes  $W(x)$  modulated by complex exponentials. Let

$$\begin{aligned} m_0(x, y) &= \sin\left(\frac{\pi}{2}(x+1)\right) \cos\left(\frac{\pi}{2}(y+1)\right), \\ m_1(x, y) &= \cos(\pi(x+1)) \sin(\pi(y+1)), \\ m_2(x, y) &= \sin(2\pi x) \sin(2\pi y), \\ m_3(x, y) &= 0.5, \end{aligned}$$

and define  $W(x) = [m_0(x), m_1(x), m_2(x), m_3(x)]^\top$ . The field is a linear combination

$$I(x, t) = W(x)^\top \phi(t), \quad \phi_k(t) = b_k \exp((\alpha_k + j\omega_k) t \Delta t_{\text{eff}}),$$

with  $\Delta t_{\text{eff}} = 0.1$  (implemented via a fixed time-scaling so that sequences share identical temporal factors across experiments).

We use

$$\begin{aligned} \alpha &= [-0.01, -0.05, -0.20, -0.01], \quad \omega = [2.00, 4.00, 1.00, 0.30], \\ b &= [1.0+0.5j, 0.8-0.3j, 0.7+0.2j, 0.2+0.0j]. \end{aligned}$$

The domain is a uniform  $32 \times 32$  grid over  $[-1, 1] \times [-1, 1]$  ( $n=1024$ ). At each time  $t$ , only 10% of grid points (a single, time-invariant random index set) are observed to emulate sparse sensing. Observations are corrupted by i.i.d. complex Gaussian noise  $\eta \sim \mathcal{N}_{\mathbb{C}}(0, \sigma^2)$  with  $\sigma = 0.1$ :

$$y_t(x_i) = I(x_i, t) + \eta_{i,t}.$$

We roll out  $T=100$  steps with timestamps  $t = 0, \dots, T-1$ . For evaluation, we also retain the full noiseless fields  $I(x, t)$ . The ground-truth discrete-time eigenvalues used for comparison are

$$\mu_k = \exp((\alpha_k + j\omega_k) \Delta t_{\text{eff}}).$$

## D.2. Gray-Scott Reaction-Diffusion Model

The Gray-Scott model simulates the interaction of two chemical species  $u$  and  $v$  through reaction, diffusion, and feed/kill processes, exhibiting rich pattern formation such as spots, waves, and mazes. The evolution is governed by the coupled PDEs:

$$\frac{\partial u}{\partial t} = D_u \nabla^2 u - uv^2 + F(1 - u), \quad (9)$$

$$\frac{\partial v}{\partial t} = D_v \nabla^2 v + uv^2 - (F + k)v, \quad (10)$$

with periodic boundary conditions. The Laplacian is approximated using finite differences with nearest-neighbor rolls. We use parameters  $D_u = 2 \times 10^{-4}$ ,  $D_v = 1 \times 10^{-5}$ ,  $F = 0.035$ ,  $k = 0.065$ , known to produce **wave-like and spotted patterns** (Pearson, 1993). A small perturbation is added to  $F$  at initialization ( $\sigma = 10^{-3}$ ) to break symmetry.

The system is simulated on a  $100 \times 100$  grid over  $[-1, 1] \times [-1, 1]$ , with spatial step  $\Delta x = 0.01$ . Initial conditions are wave-modulated fields:

$$u(x, y, 0) = 0.9 + 0.1 \sin(4\pi x) \cos(2\pi y), \quad v(x, y, 0) = 0.1 + 0.05 \sin(\pi x). \quad (11)$$

Only 10% of spatial points are observed. The concentration of species  $v$  is used as the observable.

### D.2.1. 2D NAVIER-STOKES VORTICITY FLOW (SPECTRAL METHOD)

We simulate incompressible 2D fluid flow using the Navier-Stokes equations in vorticity-stream function form, using (Cao et al., 2024). The vorticity  $\omega = \nabla \times \mathbf{u}$  evolves as:

$$\frac{\partial \omega}{\partial t} + (\mathbf{u} \cdot \nabla) \omega = \nu \nabla^2 \omega, \quad (12)$$

with  $\mathbf{u} = \nabla^\perp \psi$  and  $\nabla^2 \psi = -\omega$ . The initial vorticity field is a filtered random field with peak wavenumber  $k = 2$ , ensuring smooth, coherent structures. The simulation uses the `torch-cfd` library with Crank-Nicolson + RK4 time stepping, viscosity  $\nu = 10^{-3}$ , and domain  $[0, 2\pi] \times [0, 2\pi]$ .

The grid resolution is  $128 \times 128$  (coordinates normalized to  $[-1, 1] \times [-1, 1]$ ). Time step is  $\Delta t = 10^{-3}$ , with snapshots saved every 100 steps ( $\sim \Delta t_{\text{save}} = 0.1$ ), yielding 100 snapshots over  $T = 10.0$ . The flow develops vortex streets and turbulent-like structures.

### D.2.2. 2D NAVIER-STOKES FLOW PAST A CYLINDER

We simulate viscous incompressible flow around a circular cylinder using (Cao et al., 2024), modeling the von Kármán vortex street. The Navier-Stokes equations are:

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{u} = -\nabla p + \nu \nabla^2 \mathbf{u}, \quad (13)$$

$$\nabla \cdot \mathbf{u} = 0, \quad (14)$$

The grid is  $100 \times 100$  in  $[-1, 1] \times [-1, 1]$ . Viscosity is  $\nu = 1/500$ , time step  $\Delta t = 10^{-3}$ , and snapshots are saved every 150 steps over  $T = 2.0$ . The vorticity field exhibits periodic shedding behind the cylinder. Spatial observations are densely available (full cropped grid), but only vorticity is used as the observable.

## Appendix E. Recovering Ground-Truth Dynamics from Noisy, Sparse Observations

We evaluate the model trained in Sec. 4.2 and report (i) spatial mode similarity and (ii) continuous-time eigenvalue accuracy.

### E.1. Metric details

**Mode similarity** ( $\text{sim}_{\cos}(\widehat{W}, W_{\text{GT}})$ ). Learned modes  $\widehat{W} \in \mathbb{C}^{N \times r}$  are matched to ground-truth modes  $W \in \mathbb{C}^{N \times r}$  by solving a linear assignment (Hungarian) problem over permutations  $\pi$  that maximizes per-mode cosine similarity after a phase correction  $e^{i\theta_k}$ :

$$\cos(k) = \frac{\langle \widehat{W}_{:,k}, e^{i\theta_k} W_{:,\pi(k)} \rangle}{\|\widehat{W}_{:,k}\| \|W_{:,\pi(k)}\|} \in [0, 1].$$

We report the mean per-pair cosine similarity.

**Eigenvalue accuracy** ( $|\widehat{\lambda} - \lambda_{\text{GT}}|$ ). Pairs  $(\widehat{\lambda}_i, \lambda_j)$  follow the mode assignment. Because  $\phi_k(t)$  evolves under a Neural ODE, we estimate continuous-time eigenvalues using a log-ratio estimator (Tu, 2013; Schmid, 2010b; Kutz et al., 2016b):

$$\widehat{\lambda}_k(t) = \frac{1}{\Delta t} \text{Log} \left( \frac{\phi_k(t + \Delta t)}{\phi_k(t)} \right), \quad \widehat{\lambda}_k = \text{median}_t \widehat{\lambda}_k(t),$$

with a time-consistent branch choice. We then report the mean absolute complex difference.

### E.2. Quantitative results.

Table 2 compares the proposed NODE-DMD and NDMD on mode agreement and spectral accuracy.

**Mode similarity.** Our average cosine is 0.708 (per-mode 0.589, 0.811, 0.628, 0.807), while NDMD records 0.767 on average, reflecting very high alignment on three modes (0.999, 0.999, 0.996) and a lower value on one mode (0.074). Overall, NDMD attains a slightly higher per-mode cosine, whereas our method exhibits a more even alignment across modes. Since the cosine is computed per mode and is sensitive to

Metric	mode 0	mode 1	mode 2	mode 3	Avg.
$\text{sim}_{\cos}(\hat{W}, W_{\text{GT}})_{\text{NODE-DMD}}$	0.589	0.811	0.628	0.807	<b>0.708</b>
$\text{sim}_{\cos}(\hat{W}, W_{\text{GT}})_{\text{NDMD}}$	0.0741	0.999	0.999	0.996	<b>0.7674</b>
$ \hat{\lambda} - \lambda_{\text{GT}} _{\text{NODE-DMD}}$	0.060	0.985	1.248	0.135	<b>0.607</b>
$ \hat{\lambda} - \lambda_{\text{GT}} _{\text{NDMD}}$	0.266	1.0198	1.937	3.8754	<b>1.7748</b>
$\lambda_{\text{GT}}$	$-0.01+2.0j$	$-0.05+4.0j$	$-0.20+1.0j$	$-0.01+0.3j$	
$\hat{\lambda}_{\text{NODE-DMD}}$	$-0.22+1.03j$	$-0.03+2.75j$	$-0.25+0.88j$	$0.05+0.32j$	
$\hat{\lambda}_{\text{NDMD}}$	$-0.002+0.033j$	$0+0j$	$-0+0.06j$	$0.002+0.125j$	

Table 2: **Mode agreement and continuous-time eigenvalue errors.** Cosine similarity is computed after permutation and per-mode phase alignment. Eigenvalue errors are absolute complex differences after optimal assignment.

basis rotations within a shared subspace, these differences do not necessarily imply a gap in reconstruction quality.

**Eigenvalue accuracy.** Our average absolute complex error is **0.607**, compared to **1.775** for NDMD. NDMD’s estimates are closer to low/near-zero frequencies, yielding larger differences from the ground-truth spectrum. By contrast, our estimates preserve the overall ordering and scales of damping and frequency. Given that reconstructions follow  $y(t) = W \phi(t)$  and long-horizon behavior is largely governed by spectral factors in  $\phi(t)$  (phase and amplitude growth), these results suggest that, despite a slightly lower per-mode cosine, the improved spectral estimates of our method are particularly beneficial for stable, long-horizon evolution.

**Side note (fair comparison with NDMD).** Neural DMD (NDMD) parameterizes one DC mode with eigenvalue  $0+0j$  and  $r/2$  complex-conjugate pairs. For real-field reconstruction it uses only the positive member with a factor  $2 \text{Re}(\cdot)$  (since  $z + z^* = 2 \text{Re}(z)$ ). For a like-for-like comparison, we evaluate NDMD using the single DC mode and the three *positive* members of the conjugate pairs (the  $2 \text{Re}(\cdot)$  form is equivalent to summing both members). Eigenvalue metrics are computed on the positive-branch eigenvalues accordingly. We also note that the ground-truth decomposition includes a spatial DC component, which aligns well with NDMD’s dedicated DC mode; thus this setting is, if anything, slightly favorable to NDMD. Despite these architectural differences, this protocol is intended to keep the comparison as fair as possible.

## Appendix F. Resolution flexibility Experiment

A key advantage of NODE–DMD is its grid-free formulation, which allows the model to reconstruct the full spatial field at arbitrary position without requiring retraining. To evaluate this, we compare NODE–DMD with NDMD (SaraerToosi et al., 2025) under low- and high-resolution reconstruction settings. For each task, we downsample the field to a  $(50 \times 50)$  grid for the low-resolution experiment and upsample it to a  $(200 \times 200)$  grid for the high-resolution experiment. All models are trained only on the original resolution data with 10% spatial sampling.

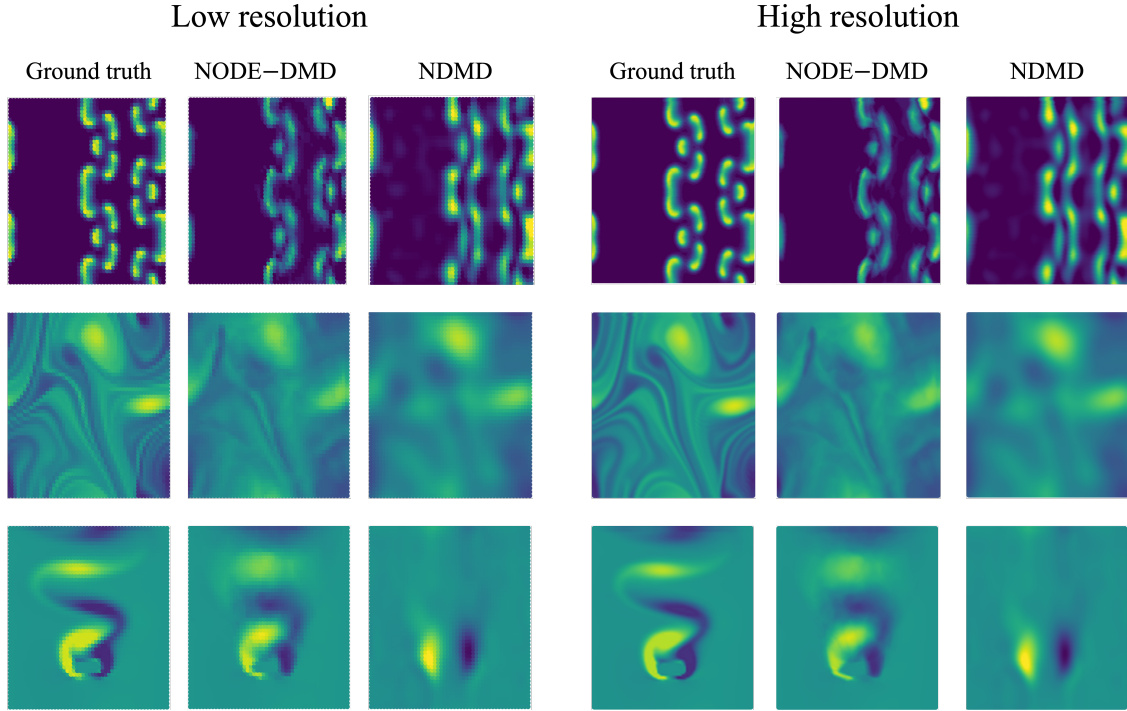


Figure 5: Reconstruction results at the final time step for low-resolution ( $50 \times 50$ ) and high-resolution ( $200 \times 200$ ) grids for the Gray–Scott system (first row), Vorticity flow (second row), and Cylinder flow (third row). Visualizations are generated using NDMD and the 1-step NODE–DMD model.

Method	LOW RESOLUTION			HIGH RESOLUTION		
	NDMD	NODE DMD(TF)	NODE DMD(AR)	NDMD	NODE DMD(TF)	NODE DMD(AR)
GRAY–SCOTT	0.0048(−2.0%)	0.0047(+2.2%)	0.0078(+2.6%)	0.0047(−4.1%)	0.0045(−2.2%)	0.0075(−1.3%)
VORTICITY FLOW	0.0047(0%)	0.0024(0%)	0.0024(0%)	0.0047(0%)	0.0024(0%)	0.0024(0%)
CYLINDER FLOW	0.0507(+8.6%)	0.0064(−4.5%)	0.0462(−2.3%)	0.0530(+13.5%)	0.0067(0%)	0.0477(+0.8%)

Table 3:  $L_1$  reconstruction error and its relative change under different spatial resolutions. NODE–DMD(TF) and NODE–DMD(AR) denote teacher-forced one-step prediction and autoregressive multi-step rollout, respectively. Percentage change is computed with respect to the original-resolution full-field reconstruction error reported in Table 1, where negative values indicate error reduction and positive values indicate error increase.

Figure 5 presents qualitative reconstruction results at the final time step for the three benchmark systems. Table 3 reports the average per-pixel  $L_1$  error over the full temporal rollout for low and high resolutions, along with the percentage change relative to the original-resolution reconstruction error reported in Table 1. Negative percentages indicate error reduction, and positive values indicate error increase.

Across all tasks, NODE–DMD exhibits consistently small positive changes in reconstruction error when the target resolution differs from the training resolution compared to NDMD. These results highlight the robustness of NODE–DMD to changes in grid density and demonstrate that its continuous implicit representation enables stable full-field reconstruction, independent of the resolution at which the field is queried.

### Appendix G. Ablation on DMD spectral evolution: stochastic latent-dynamics baseline

To isolate the role of the DMD-specific spectral inductive bias in Stochastic NODE–DMD, we additionally compare against a probabilistic baseline that removes the DMD-style spectral evolution while retaining the same coordinate-conditioned spatial decoder. This comparison is intended to distinguish the effect of stochastic latent dynamics alone from the benefit of combining them with the linear spectral structure of DMD.

The baseline follows a standard probabilistic latent-dynamics design: an encoder infers a Gaussian latent posterior from observations, the latent state is evolved by a learned continuous-time stochastic flow, and the field is reconstructed by a decoder (Krishnan et al., 2017; Rubanova et al., 2019; Li et al., 2020a). Concretely, we retain the coordinate-conditioned mode extractor  $W_\psi(\mathbf{s})$ , which allows reconstruction at arbitrary spatial coordinates, but replace the DMD-style latent evolution with a generic stochastic latent dynamics model. Relative to Stochastic NODE–DMD, the only removed ingredient is the DMD-style spectral evolution bias. Let  $\mathbf{z}_k \in \mathbb{C}^r$  denote the latent state at time  $t_k$ . The baseline infers a Gaussian latent posterior

$$q(\mathbf{z}_k \mid \mathbf{y}_k, S) = \mathcal{CN}(\boldsymbol{\mu}_{z_k}, \Sigma_{z_k}), \tag{15}$$

using a latent encoder that takes the sparse observation  $\mathbf{y}_k$  and sampled coordinates  $S$  as input. Unlike Stochastic NODE–DMD, this baseline does not predict eigenvalues and does not impose the linear evolution term. Instead, the latent state is propagated by a generic learned stochastic flow,

$$d\mathbf{z}_t = f_\theta(\mathbf{z}_t, t) dt + \tau d\mathbf{B}_t, \tag{16}$$

where  $f_\theta$  is a neural drift function and  $\tau d\mathbf{B}_t$  is the diffusion term. The reconstructed field is then obtained through the same coordinate-conditioned decoder,

$$\hat{\mathbf{y}}(\mathbf{s}, t) = W_\psi(\gamma(\mathbf{s}))^\top \mathbf{z}_t. \tag{17}$$

We compare Stochastic NODE–DMD and this stochastic latent-dynamics baseline using four probabilistic forecasting metrics: full-field mean squared error (MSE), continuous ranked probability score (CRPS), and empirical coverage at the nominal 68% and 95% levels. Table 4 summarizes the results.

The results show that removing the spectral evolution bias substantially weakens both predictive accuracy and uncertainty quality on the nonlinear flow benchmarks. On Gray–Scott, Vorticity Flow, and Cylinder Flow, NODE–DMD achieves lower MSE and lower CRPS than the stochastic baseline, indicating that the

System	Method	MSE ↓	CRPS ↓	Cov.@68	Cov.@95
Synthetic	NODE–DMD	0.0466	0.1191	0.7823	0.9639
	SDE baseline	0.0308	0.0460	0.3325	0.5826
Gray–Scott	NODE–DMD	0.0046	0.0187	0.8450	0.9082
	SDE baseline	0.0113	0.0231	0.5629	0.9349
Vorticity Flow	NODE–DMD	0.0024	0.0249	0.8241	0.9610
	SDE baseline	0.0051	0.0279	0.6250	0.9163
Cylinder Flow	NODE–DMD	0.0067	0.0277	0.7917	0.9394
	SDE baseline	0.0241	0.0409	0.8771	0.9860

Table 4: Comparison with a stochastic latent-dynamics baseline without DMD spectral evolution. The baseline retains the same coordinate-conditioned mode extractor and decoder, but replaces the DMD-style latent evolution with a generic stochastic latent flow. Coverage is the empirical fraction of ground-truth values contained in the nominal 68% and 95% predictive intervals.

DMD-structured latent dynamics improve both deterministic reconstruction and probabilistic forecast quality.

Coverage further clarifies the role of the spectral prior. For the Synthetic task, the SDE baseline attains lower MSE and CRPS, but its empirical coverage collapses far below the nominal levels, indicating that its uncertainty estimates are severely under-dispersed and poorly calibrated. In contrast, NODE–DMD yields coverage much closer to the nominal 68% and 95% targets while remaining competitive in reconstruction error. On Gray–Scott and Vorticity Flow, NODE–DMD again provides substantially better-calibrated uncertainty, with coverage consistently closer to the target intervals. On Cylinder Flow, the SDE baseline over-covers at both levels, indicating conservative but weakly informative uncertainty, whereas NODE–DMD remains closer to the nominal range while preserving much better reconstruction fidelity.

Overall, these results support the claim that the benefit of Stochastic NODE–DMD is not simply due to adding stochastic latent dynamics. Rather, the linear spectral structure of DMD provides a strong inductive bias for organizing long-horizon latent evolution, while the stochastic Neural ODE supplies nonlinear correction and uncertainty transport around that structure. This combination yields more accurate forecasts and more informative uncertainty estimates than a generic stochastic latent model without spectral evolution.

### G.1. Sensitivity analysis with respect to observation density

All main experiments in the paper use a fixed observation density of 10%. To assess robustness to data sparsity, we additionally evaluate our method across sampling ratios of 1%, 5%, 10%, 20%, 30%, and 100%.

For each task and sampling ratio, we run the reconstruction experiment with 5 random seeds and report the resulting distribution of reconstruction errors. To compare tasks with different absolute error scales, we normalize each task using the worst error observed at 1% sampling and the best error observed at 100%

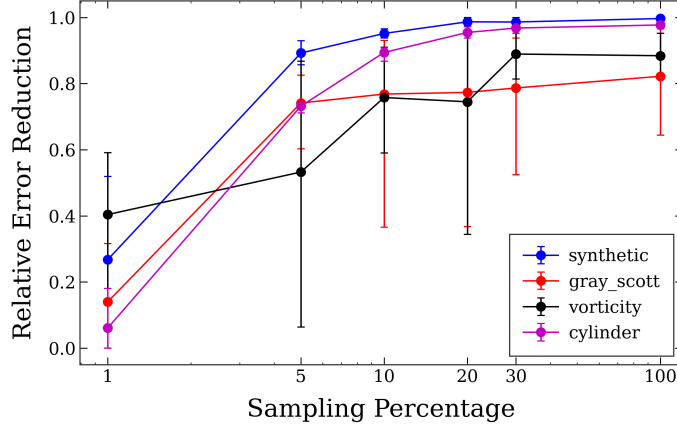


Figure 6: Sensitivity of reconstruction performance to observation density. For each task, the normalized score is defined using the worst reconstruction error at 1% sampling and the best reconstruction error at 100% sampling across seeds. Solid lines and markers denote the normalized mean across seeds, and vertical bars indicate the seed-wise best–worst range at each sampling ratio. The results show rapid improvement from extremely sparse observations, followed by saturation as the sampling density increases.

sampling across seeds. Specifically, for task-wise visualization we define

$$\tilde{P}_s(\rho) = \frac{E_{\max}(1\%) - E_s(\rho)}{E_{\max}(1\%) - E_{\min}(100\%)}, \quad (18)$$

where  $E_s(\rho)$  denotes the reconstruction error for seed  $s$  at sampling ratio  $\rho$ ,  $E_{\max}(1\%)$  is the worst error among seeds at 1%, and  $E_{\min}(100\%)$  is the best error among seeds at 100%. Under this definition, larger  $\tilde{P}$  indicates better performance. Figure 6 reports the normalized mean across seeds together with the seed-wise best–worst range at each sampling ratio. As shown in Fig. 6, performance improves substantially as the sampling ratio increases from 1% to 10%, and then tends to level off beyond approximately 10%~30% for most tasks. In particular, the Synthetic, Gray–Scott, and Cylinder tasks recover most of their full-observation performance by around 10% observations, indicating that the sparse-to-field reconstruction quality saturates relatively early.

The Vorticity task shows noticeably larger seed-to-seed variability and a non-monotone trend at intermediate sampling ratios. In our runs, this behavior is driven by a small number of seeds with substantially worse reconstruction quality, rather than by a consistent degradation shared across all seeds. We therefore interpret the irregularity in Fig. 6 as seed-sensitive optimization variability for this task. Even with this variability, the typical performance at 10% already approaches that at denser observation settings, so the main conclusion remains unchanged: the method is effective well before the full-observation regime, while additional observations beyond roughly 10%~30% yield diminishing returns.