NAGO: Noise-Aware Generative Operator via Flow Matching

Guanyu Chen[1]* **Shengze Xu**[2]* **Pengwei Liu**[1] **Xingyu Ren**[1]
Pengkai Wang[3] **Antonios Armaou**[4] **Tieyong Zeng**[2] **Dong Ni**[1]†
[1]Zhejiang University [2]The Chinese University of Hong Kong
[3]The Hong Kong Polytechnic University [4]The Pennsylvania State University

Abstract

We introduce NAGO, a noise-aware generative operator for efficient sampling from invariant measures of stochastic partial differential equations (SPDEs) in function space. Our central idea is to treat the source measure as a design variable rather than a fixed Gaussian default. We establish admissibility conditions for well-defined infinite-dimensional flows and show that the finite-band source–target covariance ratio controls the spectral stiffness of the marginal velocity field. This analysis yields an admissibility-constrained spectral-matching rule: choose a source that is sufficiently rough for well-posedness, then minimize covariance mismatch over the resolved Fourier modes. On three representative stochastic-field benchmarks, this source design improves few-step spectral accuracy and several complementary distributional and physical statistics. In a controlled Matérn-field ablation, it reduces the accurate RK4 budget from about 30 steps for Gaussian functional flow matching to 3–5 steps, a 6–10$\times$ reduction in neural ODE evaluations. These gains are strongest for stationary multiscale fields and depend on the metric, resolution, and guidance protocol.

1 Introduction

Stochastic partial differential equations (SPDEs) constitute a fundamental mathematical framework for modeling the evolution of infinite-dimensional random systems under uncertainty and provide a rigorous description of diverse spatiotemporal phenomena, including statistical physics [Huang, 2009] and composite materials [Hornung, 1992]. However, in many practical applications, such as geophysical modeling [Lindgren et al., 2011], turbulent flow [Kida and Murakami, 1987, Hairer and Mattingly, 2006, Carmona, 1999], and thin film growth [Hairer, 2013], the exact evolution of a single realization is often less important than the long-term statistical behavior of the system [Eckmann and Hairer, 2001].

A central problem in this context is the characterization of invariant measures [Da Prato and Zabczyk, 1996, 2014], which represent the most compressed yet complete description of long-term stochastic behavior, as discussed in Appendix B. For many SPDEs of physical and statistical interest, these invariant measures take the form of Gaussian or non-Gaussian probability measures on suitable function spaces, with representative examples including the Ornstein–Uhlenbeck equations [Särkkä and Solin, 2019] and nonlinear models such as the stochastic Allen–Cahn equation [Xu et al., 2025] and the stochastic Navier–Stokes equation [Bensoussan, 1995], which may admit unique ergodic invariant measures under appropriate conditions. Despite their importance, sampling invariant measures of SPDEs remains challenging due to the infinite-dimensional nature of the dynamics, the absence of closed-form solutions in nonlinear regimes, and the intricate interaction between noise,

*These authors contributed equally to this work.
†Corresponding author: dni@zju.edu.cn.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

dissipation, and spatial operators. Classical numerical approaches, including long-time integration of discretized SPDEs [Lord et al., 2014, Da Prato and Zabczyk, 2014] and Markov chain Monte Carlo (MCMC) sampling in function space [Chen, 2025], may approximate invariant measures but are often computationally prohibitive, particularly in high-dimensional or stiff systems exhibiting pronounced multiscale structure.

These challenges have motivated growing interest in data-driven and generative modeling approaches that aim to learn invariant measures directly, bypassing explicit long-time simulation. Recent advances in generative modeling [Ho et al., 2020, Song et al., 2021a,b, Lipman et al., 2023] and operator-learning frameworks [Kovachki et al., 2023, Li et al., 2021] have shown considerable promise: generative models provide expressive distributional approximations, while neural operators enable mappings between functional distributions. Recent progress has been made toward generative modeling directly in Hilbert or Banach spaces. Rahman et al. [2022], Seidman et al. [2023] proposed GANO and VANO, while Pidstrigach et al. [2024], Lim et al. [2025], Kerrigan et al. [2023, 2024], Franzese et al. [2023] develop the theoretical foundations for diffusion models and flow matching in infinite-dimensional settings. More details about related work are reviewed in Appendix A.

Nevertheless, an important gap remains between the theory of functional generative modeling and its practical use for multiscale SPDE data. Most implementations still adopt standard Gaussian noise as the source distribution by default. For invariant measures with nontrivial spatial regularity, this choice can be poorly conditioned: the discretized Gaussian noise injects excessive high-frequency energy, forcing the learned velocity field to suppress unresolved modes and thereby making training and inference resolution-dependent. Existing remedies, such as interpolation schedules [Chen and Vanden-Eijnden, 2026], Schrödinger bridge formulations [Shi et al., 2023], and adaptive numerical solvers [Chen et al., 2024], can improve the transport dynamics or numerical integration, but they do not directly modify the source measure that induces the stiffness. We address this issue by treating the source measure as a design variable in functional flow matching. This perspective leads to an admissibility-constrained spectral-matching principle: the source must first be rough enough for function-space well-posedness, and should then be chosen to reduce covariance mismatch over the frequencies resolved by the model.

Based on this principle, we propose `NAGO`, a noise-designed generative operator for sampling from multiscale SPDE invariant measures. By selecting an admissible Matérn source with a finite-band spectrum aligned to the target, `NAGO` reduces high-frequency mismatch and yields better-conditioned marginal velocity fields. We support this approach with an infinite-dimensional analysis, a quantitative link between source–target covariance ratios and spectral stiffness, and an operational source-selection rule. Our main contributions are summarized as follows:

- We identify a function-space mechanism behind source design: source regularity determines admissibility, while the resolved source–target covariance ratio governs spectral stiffness.
- We develop an admissibility-constrained spectral-matching rule for selecting Matérn sources from target samples, including both smoothness and correlation-length selection.
- Across three stochastic-field benchmarks, we demonstrate improved few-step generation and complementary distributional statistics. A controlled ablation reduces the accurate RK4 budget by 6–$10\times$ relative to a Gaussian-source model; a separate post-training benchmark shows an $89.1\times$ raw-throughput ratio to the SNS solver.

Code availability. The code release is available at `https://github.com/cgymmy/NAGO`.

2 Preliminary

In this section, the fundamental concepts of random fields and SPDEs are first introduced, followed by a discussion of invariant measures of SPDEs, and concluded with a brief overview of the basic theory underlying functional flow-matching generative models.

2.1 Random Field and Stochastic PDEs

To formalize invariant measures and noise distributions in function space, we briefly recall the notion of random fields. Usually, a random field is regarded as a random variable taking values in a function space; we formalize this as follows:

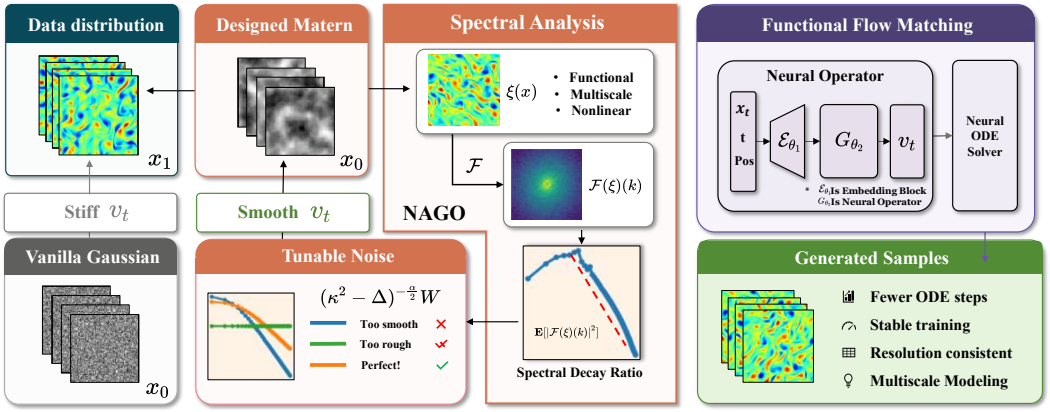


Figure 1: The structure of the NAGO model.

Definition 2.1 (Random Field)**.** *For a set $D \subset \mathbb{R}^d$, a (real-valued) random field $\{u(x) : x \in D\}$ is a collection of real-valued random variables defined on a probability space $(\Omega, \mathcal{F}, P)$, where Ω denotes the sample functional space, $\mathcal{F}$ is a σ-algebra of subsets of Ω, and P is a probability measure on $(\Omega, \mathcal{F})$.*

The class of random fields is broad and contains many important special cases with distinct structural properties as discussed in Appendix C.2.4. Among them, the Matérn field [Lindgren et al., 2011] plays a particularly prominent role, as it defines a Gaussian random field with tunable smoothness and correlation length. These properties make it both mathematically tractable and practically useful, especially for studying the effect of regularity in function-space probabilistic modeling. We therefore introduce the Matérn field as a representative example.

Definition 2.2 (Matérn Field)**.** *A Gaussian random field $u(x)$ with Matérn covariance is called a Matérn field, which is a solution to the linear fractional stochastic partial differential equation:*

$$\mathcal{L}^{\alpha/2} u(x) = (\kappa^2 - \Delta)^{\alpha/2} u(x) = W(x), \tag{1}$$

where $x \in D \subset \mathbb{R}^d$, $\nu = \alpha - d/2 > 0$ describes the smoothness, $\rho = \frac{\sqrt{2\nu}}{\kappa} > 0$ for the correlation length, Δ is the Laplacian operator, and $W(x)$ is a spatial Gaussian white noise.

The Matérn field provides a canonical example of Gaussian measures. On $\mathbb{R}^d$, or under periodic boundary conditions, the spectral solution of u defined by Eq. (1) is given by

$$u(x) = \mathcal{F}^{-1}\left[G(k)\widehat{W}(k)\right](x), \quad G(k) = (\kappa^2 + \|k\|^2)^{-\alpha/2}, \tag{2}$$

where $\mathcal{F}$ is the Fourier transform defined in Appendix C.12, and $G(k)$ is the Fourier multiplier of the solution operator. The corresponding spectral density of u is $S_u(k) \propto |G(k)|^2 = (\kappa^2 + \|k\|^2)^{-\alpha}$. The covariance function of u is then given by:

$$c(x) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} (\kappa\|x\|)^\nu K_\nu(\kappa\|x\|), \tag{3}$$

where $\nu = \alpha - d/2$, $\rho = \frac{\sqrt{2\nu}}{\kappa}, \sigma^2 = \frac{\Gamma(\nu)}{(4\pi)^{d/2}\kappa^{2\nu}\Gamma(\alpha)}$ and K_ν is the modified Bessel function of the second kind. Detailed derivations are provided in Appendix C.2.3. In addition, Gaussian noise, white noise, and general Matérn fields are closely related, as summarized in Table 1. Here, "Normalized" indicates whether the corresponding noise has been rescaled to have unit marginal variance in the discretized setting. In the remainder of this paper, we use the term **Matérn noise** to denote noise generated from a Matérn field.

Table 1: Comparison of different random fields.

	Matérn	Gaussian	White
Smoothness α	$> d/2$	0	0
Range parameter κ	> 0	–	–
Normalized	–	Yes	No

Having introduced random fields and the Matérn field, we now turn to the SPDEs of interest in this work. In general, SPDEs can be

divided into two categories [Lord et al., 2014]: stationary SPDEs, which describe mappings between probability measures as introduced in Appendix C.3, and semilinear evolution equations, which describe the temporal evolution of systems perturbed by noise. Analogous to ordinary time-dependent PDEs, semilinear evolution SPDEs are treated as semilinear stochastic ordinary differential equations (SODEs) on a Hilbert space $\mathcal{H}_U$, taking the form:

$$\begin{cases} du = [\mathcal{L}u + F(u)]dt + G(u)dW(t), \\ u(0,x) = u_0, \quad x \in D. \end{cases} \tag{4}$$

Here, D is a bounded spatial domain in $\mathbb{R}^d$, $\mathcal{L}$ denotes a linear operator that generates the semigroup $P_t = e^{t\mathcal{L}}$; $W(t)$ is an $\mathcal{H}_\xi$-valued Q-Wiener process, where $\mathcal{H}_\xi$ is a separable Hilbert space; $F : \mathcal{H}_U \to \mathcal{H}_U$ denotes a nonlinear operator; and $G : \mathcal{H}_U \to \mathcal{L}(\mathcal{H}_{0,\xi}, \mathcal{H}_U)$ is a Hilbert–Schmidt operator associated with the Cameron–Martin space $\mathcal{H}_{0,\xi}$ of $\mathcal{H}_\xi$. Under appropriate dissipativity and ergodicity conditions (see Thm. C.19), the SPDE admits a unique invariant measure μ, which is defined as follows:

Definition 2.3 (Invariant Measure)**.** *A Borel probability measure μ on $(H, \mathcal{B}(H))$ is called an* invariant measure *for Equ.*(4) *if it is invariant under the semigroup $(P_t)_{t\geq 0}$, i.e. $\forall\, t \geq 0$, $\forall\, \varphi \in B_b(H)$,*

$$\int_H (P_t\varphi)(x)\, \mu(dx) \;=\; \int_H \varphi(x)\, \mu(dx),$$

Equivalently, if $u(0) \sim \mu$, then $u(t) \sim \mu$ for every $t \geq 0$.

Moreover, we say the SPDE is *ergodic* (or *mixing*) with invariant measure μ if for every initial condition $u_0 \in H$, $Law(u^x(t)) \to \mu$ as $t \to \infty$, where $\to$ denotes weak convergence of probability measures on H. Further details on the existence and properties of invariant measures for SPDEs are provided in Appendix C.3.

2.2 Infinite-dimensional Flow Matching

Compared with finite-dimensional generative models, the marginal velocity field in infinite-dimensional generative modeling is well-defined only under rather stringent theoretical conditions [Pidstrigach et al., 2024, Kerrigan et al., 2024]. Here, we briefly review the theoretical framework for infinite-dimensional generative models and the conditions required for well-posedness.

Assume that the initial noise $g \in \mathcal{U}$ follows a Gaussian measure μ_0 with covariance operator $Q : \mathcal{U} \to \mathcal{U}$, where Q is self-adjoint, nonnegative, and trace-class, and $(\mathcal{U}, \langle\cdot,\cdot\rangle)$ is a separable Hilbert space. Data samples are denoted by f, distributed according to the target measure μ_1. Our objective is to construct an infinite-dimensional flow $\psi_t : [0,1] \times \mathcal{U} \to \mathcal{U}$ that pushes μ_0 to μ_1. Then the marginal probability measure at time t is defined as $\mu_t := [\psi_t]_{\#}\mu_0$, which admits the representation $\mu_t(A) = \int_{\mathcal{U}} \mu_t(A|f)\, d\mu_1(f), \quad \forall A \in \mathcal{B}(\mathcal{U})$, where $\mu_t(\cdot|f)$ denotes the conditional measure and $\mathcal{B}(\mathcal{U})$ is the Borel σ-algebra on $\mathcal{U}$. Then we define the map ψ_t which is associated with a time-dependent velocity field $v_t : [0,1] \times \mathcal{U} \to \mathcal{U}$ through

$$\frac{d}{dt}\psi_t(g) = v_t(\psi_t(g)), \qquad \psi_0(g) = g \sim \mu_0. \tag{5}$$

When the target measure μ_1 is fixed, the central object of interest is the *marginal velocity field* v_t, whose regularity and well-posedness directly determine the stability and learnability of the generative flow. The corresponding infinite-dimensional flow matching objective is $\mathcal{L}_{\mathrm{FFM}} = \mathbb{E}_{t,\, h\sim\mu_t}\left[\|v_t(h) - v_\theta(h)\|^2\right]$. As in finite-dimensional settings, the marginal velocity field is generally intractable. We therefore consider conditional velocity fields $v_t(\cdot \mid f)$ associated with conditional measures $\mu_t(\cdot \mid f)$. We recall two standard results from functional flow matching that express the marginal field through conditional dynamics and characterize the required absolute continuity; our new source-design analysis begins in Sec. 3.2.

Theorem 2.4 (Recalled marginalization result)**.** *Given $\int_0^1 \iint \|v_t(h|f)\|^2\, d\mu_t(h|f)\, d\mu_1(f)\, dt$ is finite, and that for a.e. $f \sim \mu_1$ and $t \in [0,1]$, the conditional measure $\mu_t(\cdot|f)$ is absolutely continuous with respect to μ_t ($\mu_t(\cdot|f) \ll \mu_t$). Then the marginal velocity field admits the representation*

$$v_t(h) = \int_{\mathcal{U}} v_t(h|f)\, \frac{d\mu_t(h|f)}{d\mu_t(h)}\, d\mu_1(f). \tag{6}$$

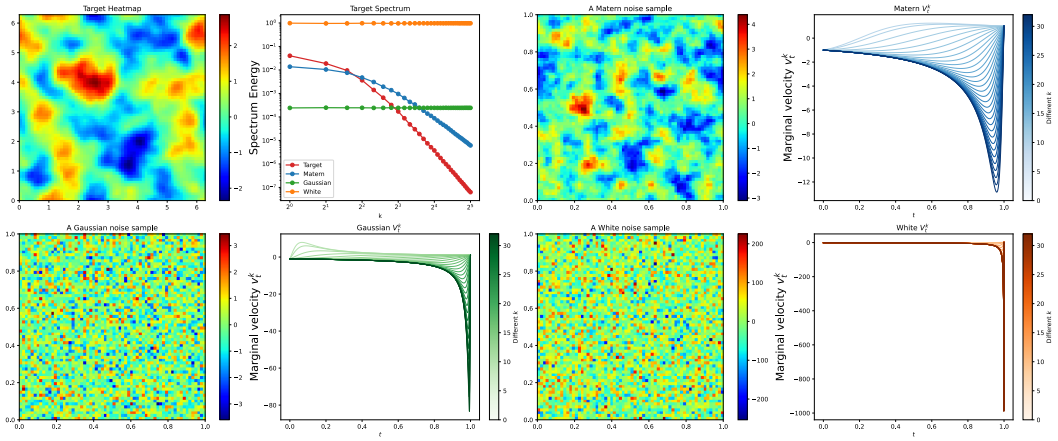


Figure 2: Samples under various noise distributions and the target distribution, their spectra, and the marginal velocity fields at different frequencies for each noise type.

The absolute continuity condition $\mu_t(\cdot|f) \ll \mu_t$ is therefore a key requirement for the existence of a well-defined marginal velocity field. We adopt affine conditional flows of the form

$$f_t = \psi_t(g|f) = \alpha_t g + \beta_t f, \tag{7}$$

which push $g \sim \mu_0 = \mathcal{N}(0, Q)$ toward a given f. The corresponding conditional measure is $\mu_t(\cdot|f) = \mathcal{N}(\beta_t f, \alpha_t^2 Q)$.

Theorem 2.5 (Recalled absolute-continuity criterion)**.** *Consider a given measure μ on a Hilbert space $\mathcal{U}$ and the corresponding conditional probability measure $\mu_t(\cdot|f), f \sim \mu$. If $a.e. f, g \sim \mu$, $\mu_t(\cdot|f)$ is equivalant to $\mu_t(\cdot|g)$, then the marginal probability measure μ_t satisfies $\mu_t(\cdot|f) \ll \mu_t$ for $a.e. f \sim \mu$. Proof is shown in Appendix D.2.*

By the Feldman–Hájek theorem, this equivalence holds if and only if $f - g \in \alpha_t \mathcal{Q}^{1/2}(\mathcal{U})$, implying that the data distribution must be supported on the Cameron–Martin space associated with the noise covariance. In other words, the noise distribution must be rougher than the data distribution, an observation that underlies the conclusions of most existing theoretical works.

3 Methodology

The framework of the proposed Noise-Aware Generative Operator (`NAGO`) is presented in this section. We first introduce the problem setting and establish the notation used throughout the paper. We then extend the theoretical framework for infinite-dimensional generative models, focusing on how different noise measures affect the marginal velocity field in flow matching. Finally, we describe the complete `NAGO` framework together with the associated algorithmic strategy.

3.1 Problem Setting

Our goal is to generate independent functional samples from the invariant distribution μ defined in Def. (2.3), without explicitly simulating the underlying SPDE over long time horizons. In addition to unconditional generation, a learned generative model should support reconstruction under partial observations and enable principled uncertainty quantification through guided sampling.

3.2 Noise and Marginal Velocity Fields

Here, we focus on how different noise choices affect infinite-dimensional generative models, and analyze their impact primarily through a spectral perspective. In practice, the i.i.d. Gaussian array used after discretization is extremely rough. It is a finite-dimensional baseline rather than a trace-class Gaussian measure on the continuum Hilbert space assumed in Sec. 2.1; this distinction is important when interpreting cross-resolution behavior. We therefore generalize the noise family to Matérn measures mentioned above, which are analytically tractable and practically flexible.

Theorem 3.1. *Let u be Matérn noise with smoothness parameter ν in dimension d. Then $u \in H^s$ almost surely for all $s < \nu$, while the associated Cameron–Martin space is $\mathcal{H}_{\mathrm{CM}} = H^{\nu+d/2}$. The proof is provided in Appendix D.3.*

This theorem identifies the source regularity for which typical target samples lie in the Matérn source's Cameron–Martin space.

However, we find that satisfying Thm. 3.1 remains a rather loose criterion for Matérn noise. We therefore study the consequences of such regularity mismatch and quantify how noise regularity shapes the geometry and spectral stiffness of the marginal velocity field. To make the analysis explicit, we first derive a closed-form expression for the marginal velocity field in the Gaussian setting.

Theorem 3.2. *Given $\alpha_t, \beta_t \in C^1([0,1])$ with $\alpha_t^2 + \beta_t^2 > 0$ for every t, consider the interpolant process defined in Equ. (7), where $g \sim \mathcal{N}(0, Q_0)$ and $f \sim \mathcal{N}(0, Q_1)$ are independent samples from Gaussian measures supported on a Hilbert space $\mathcal{H}$. Assume that Q_0 and Q_1 have a common eigenbasis with strictly positive eigenvalues. Then the marginal velocity field admits the closed-form expression*

$$\begin{aligned} v_t(x) &= \mathbb{E}[\dot{f}_t \mid f_t = x] = V(t)x, \\ V(t) &= \left(\dot{\alpha}_t \alpha_t I + \dot{\beta}_t \beta_t Q_1 Q_0^{-1}\right) \left(\alpha_t^2 I + \beta_t^2 Q_1 Q_0^{-1}\right)^{-1}. \end{aligned} \tag{8}$$

The proof is provided in Appendix D.4.

To further examine the impact of noise regularity, we consider two Matérn fields with covariance operators Q_i, $i = 0, 1$ as in Fig. 2. In this case, the marginal velocity field can be analyzed in the Fourier domain, where it admits the representation

$$\tilde{V}_k(t) = \frac{(t-1) + t\mu_k}{(t-1)^2 + t^2\mu_k}, \qquad \mu_k = \frac{\mathcal{F}_k(Q_1)}{\mathcal{F}_k(Q_0)}. \tag{9}$$

Here, $\tilde{V}_k(t)$ denotes the Fourier coefficient of the velocity field at frequency k. For Matérn covariances of the form $Q_i = c_i^2(\kappa_i^2 - \Delta)^{-(\nu_i + d/2)}$, we obtain

$$\mu_k = \frac{c_1^2}{c_0^2} \frac{(\kappa_0^2 + 4\pi^2\|k\|^2)^{\nu_0 + d/2}}{(\kappa_1^2 + 4\pi^2\|k\|^2)^{\nu_1 + d/2}}. \tag{10}$$

Observation. When the extreme case $Q_0 = Q_1$ is considered, we have $\mu_k = 1$ for all k, and the marginal velocity field in the Fourier domain reduces to $\tilde{V}_k(t) = \frac{2t-1}{2t^2-2t+1}$, which is independent of frequency. In this regime, the marginal velocity field is spectrally smooth, making it comparatively easier for a model to learn. In contrast, when the noise and data distributions are mismatched in regularity, particularly when $\nu_0 < \nu_1$, it follows from Chen and Vanden-Eijnden [2026] that $\mu_k \to 0$ as $\|k\| \to \infty$. Consequently, $\tilde{V}_k(t)$ develops increasingly sharp variations. For high-frequency modes with $0 < \mu_k \leq 1$, its minimum on $t \in [0,1]$ is $\tilde{V}_k^{\min} = -\frac{1}{2}\left(\sqrt{\mu_k} + \frac{1}{\sqrt{\mu_k}}\right)$, diverging to $-\infty$ as $\mu_k \to 0$. This behavior indicates severe spectral stiffness in the marginal velocity field, which poses significant challenges for both learning and numerical integration as in Fig. 2.

In practice, functional data are represented on discretized meshes $k < \infty$. As a result, the spectral-stiffness effect is partially mitigated in experiments. As illustrated in Fig. 2, when the regularity mismatch between the noise and target measures is pronounced, as in the case of white noise, the high-frequency components of the marginal velocity field become extremely sharp as $t \to 1$. Although discretized Gaussian noise partially mitigates this effect, an admissible Matérn source can reduce the mismatch over the modes represented by the model. This analysis motivates selecting the source through both continuum admissibility and finite-band spectral alignment.

3.3 `NAGO`: Noise-Aware Generative Operator

Here, we propose a noise-aware generative operator framework (NAGO) that uses target spectra to choose the source measure before training. The overall pipeline is illustrated in Fig. 1. At a high level, NAGO consists of three steps: (i) estimating the spectral decay of the target data measure; (ii) selecting a Matérn noise model whose Cameron–Martin space contains typical target samples, while controlling spectral stiffness; and (iii) training a functional flow-matching model with a neural-operator backbone.

Spectral characterization. Motivated by Equ. (2), we characterize the regularity of the target measure μ via its spectral energy. Let $\tilde{f}(k)$ denote the Fourier coefficient of $f \sim \mu$ at frequency $k \in \mathbb{Z}^d$. We define the spectral energy function as

$$S_d(k) := \mathbb{E}_{f\sim\mu}\big[\|\tilde{f}(k)\|^2\big], \qquad k \in \mathbb{Z}^d. \tag{11}$$

A classical characterization states that a function belongs to the Sobolev space H^s if and only if $\sum_{k\in\mathbb{Z}^d}(1+\|k\|^2)^s\,\|\tilde{f}(k)\|^2 < \infty$. Accordingly, the decay behavior of $S_d(k)$ provides an empirical proxy for the Sobolev regularity of the data measure. In practice, we estimate the decay exponent β by fitting a line to the log-log plot of the radially averaged spectrum over a high-frequency band; further details are provided in Appendix D.6.

Admissible noise selection. The following theorem gives a sufficient smoothness condition for typical target samples to lie in the source Cameron–Martin space.

Theorem 3.3. *If the spectrum of the data measure exhibits polynomial decay of the form $S_d(k) \sim \|k\|^{-\beta}$ as $\|k\| \to \infty$, then choosing Matérn source smoothness $\nu_0 < \frac{\beta}{2} - d$ ensures this Cameron–Martin inclusion. Proof is shown in Appendix D.5. For example, if the target Matérn field is characterized by $\nu = 2.0$ (corresponding to $\beta = 6$ in $d = 2$), then an admissible Matérn noise must satisfy $\nu_0 < 1$.*

The admissible set from Thm. 3.3 can be refined by controlling the spectral velocity: excessively rough noise induces stiffness through the covariance ratio μ_k in Equ. (10), leading to the spiky behavior discussed in Observation 3.2. Therefore, subject to the admissibility constraint in Thm. 3.3, we give the following theorem:

Theorem 3.4. *Assume the source and target are Matérn measures with the same κ and positive Fourier spectra. Let the target smoothness be ν_1, and let $\nu_{\mathrm{adm}} = \nu_1 - d/2$ be the Cameron–Martin admissibility threshold from Thm. 3.3. Define $r_0 = \mathcal{F}_0(Q_1)/\mathcal{F}_0(Q_0)$, the target-to-source covariance ratio at the zero mode. For a prescribed uniform bound $L > 1$ on the magnitude of the Fourier marginal velocity coefficient, choose ν_0 so that*

$$0 < \nu_0 < \nu_{\mathrm{adm}} - \varepsilon, \quad r_0 M_N^{\nu_1-\nu_0} > C_L, \quad r_0 < C_L^{-1}, \tag{12}$$

where $M_N = \frac{\kappa^2}{\kappa^2+4\pi^2N^2} < 1$, $C_L = (L-\sqrt{L^2-1})^2 < 1$, N is the maximum represented frequency, and $\varepsilon > 0$ enforces strict admissibility. Then $|\tilde{V}_k(t)| < L$ for every $t \in [0,1]$ and $\|k\| \le N$. The proof is provided in Appendix D.6.

The shared-κ result gives a spectral design guide for Matérn sources. Operationally, we (i) estimate the tail exponent $\widehat{\beta}$, (ii) restrict $\nu_0 \le \widehat{\beta}/2 - d - \varepsilon$, and (iii) choose (ν_0, κ_0) from this admissible set to minimize the empirical mismatch of the covariance ratio $\widehat{\mu}_k = \widehat{S}_1(k)/S_0(k)$ over the resolved band. We use

$$(\nu_0, \kappa_0) = \arg\min_{\mathcal{A}_\varepsilon} \sum_{k\in\mathcal{K}_N} w_k\big(\log\widehat{\mu}_k - \overline{\log\widehat{\mu}}\big)^2, \qquad \mathcal{A}_\varepsilon = \{\nu_0 \le \widehat{\beta}/2 - d - \varepsilon\}, \tag{13}$$

where $\mathcal{K}_N$ contains the represented Fourier modes and w_k can downweight unreliable tail estimates. This criterion seeks a nearly frequency-independent ratio, as suggested by Observation 3.2; the length-scale sweep in Appendix F.6 shows why smoothness matching alone is insufficient. **Implication for common noise choices.** Discretized Gaussian noise can be regarded as spectral noise with constant high-frequency energy. Although it is always available on a fixed grid, its high-frequency energy tends to induce sharper variations in the learned marginal velocity field than an admissible Matérn source selected by Eq. (13). This explains why appropriately regularized spectral noise can substantially improve both stability and efficiency.

Training and guided generation. The subsequent training follows the standard flow matching framework, with Fourier Neural Operator (FNO) used as the backbone architecture. To support guided generation under partial observations, we employ the infinite-dimensional Tweedie formula. Assume that $\mu_t \ll \mu_0$, $\dot{\alpha}_t \neq 0$, and $\beta_t - \dot{\beta}_t\alpha_t/\dot{\alpha}_t \neq 0$. Then the Tweedie formula is

$$\mathbb{E}[f_1|f_t] = \left(\beta_t - \dot{\beta}_t\frac{\alpha_t}{\dot{\alpha}_t}\right)^{-1}\left(f_t - \frac{\alpha_t}{\dot{\alpha}_t}v_t(f_t)\right). \tag{14}$$

The proof is provided in Appendix D.7. The complete training and sampling algorithms are summarized in Appendix E.

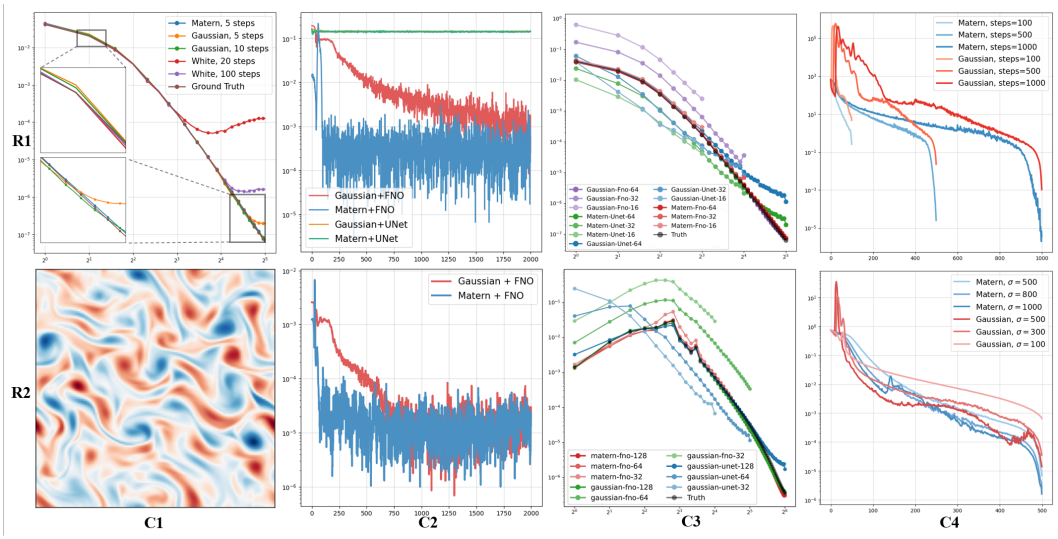


Figure 3: Noise-source comparison on the Matérn-field (top) and SNS (bottom) benchmarks. Columns show representative samples or spectra, error versus solver budget, cross-resolution spectra, and guided reconstruction loss, respectively.

4 Experiments

In this section, we empirically evaluate the proposed `NAGO` with the explicit goal of validating the hypotheses raised in the Introduction regarding noise-induced spectral stiffness and its impact on training stability, sampling efficiency, resolution consistency, and guided generation. Specifically, we focus on three aspects: (i) generation accuracy and efficiency at a fixed resolution, (ii) consistency across varying spatial resolutions, and (iii) numerical stability and accuracy under guided generation. Additional experimental details and ablation studies are provided in Appendix F.

Benchmarks. Three datasets are considered in this study, including linear Matérn fields, the stochastic Allen–Cahn (SAC) equation, and the stochastic Navier–Stokes (SNS) equation. Detailed descriptions of the underlying SPDEs and data generation procedures are provided in Appendix F.2.

Experimental Setup. We evaluate three complementary aspects corresponding to the core claims of this work: generation quality and efficiency, resolution consistency, and stability under guided generation. The spectral accuracy $E(k)$ is defined as $E(k) = \sum_{k \le \|m\| \le k+1} |\tilde{u}(m)|^2$, and the log-MSE (lMSE), $\frac{1}{N}\sum_k (\log \tilde{E}_k - \log E_k)^2$, is used as the primary spectral metric. For SNS, Appendix F.8 additionally reports multiscale MMD, marginal Wasserstein and Jensen–Shannon distances, structure functions, spatial correlations, and physical observables. All experiments employ RK4 as the neural ODE solver to ensure that performance differences arise from noise design rather than solver choice.

Baselines. We compare combinations of source distributions and backbones, summarized in Table 4. Controlled source ablations keep the data, architecture, optimizer, training budget, RK4 solver, and evaluation samples fixed; only the source distribution changes.

4.1 Generation at Fixed Resolution

We first evaluate unconditional generation at the same spatial resolution as the training data. For the Matérn field benchmark, Fig. 3 (top) shows that the source choice strongly affects the learned spectral energy distribution. White noise concentrates excessive energy in high-frequency modes, causing the marginal velocity field to become ill-conditioned as $t \to 1$, and even 100 integration steps are insufficient for accurate sampling. Gaussian noise partially alleviates this issue but still requires substantially finer temporal discretization. This behavior is consistent with the noise-induced spectral stiffness discussed in the Introduction. In contrast, the selected Matérn source attains low spectral error with 3–5 RK4 steps. Similar few-step trends occur for SAC and SNS.

Table 2: Log-MSE for different test cases. Values in parentheses indicate the number of ODE integration steps used during inference. Best results are highlighted in bold.

Noise	Model/Steps	Matérn Field ($\times 10^{-3}$)			Stochastic Allen–Cahn ($\times 10^{-3}$)			Stochastic NS ($\times 10^{-2}$)	
		C1(5)	C2(10)	C3(50)	C1(5)	C2(10)	C3(50)	C1(5)	C2(10)
Matérn	FNO	**1.70**	**0.638**	**1.44**	**0.553**	**0.778**	**0.612**	**4.44**	**4.40**
	UNet	312	342	275	30.7	32.1	32.7	7.27	7.55
Gaussian	FNO	113	5.66	1.52	7.16	2.31	0.899	5.11	9.03
	UNet	2860	2700	2760	7.99	6.03	4.84	44.7	28.7
White	FNO	1.77×10^4 (20)	8930 (50)	1780 (100)	204	5.16	1.20	5220	2670

Table 2 reports the benchmark-level comparison. A denser controlled sweep on the Matérn target (Appendix F.7) shows that Gaussian FFM reaches its stable spectral-error regime at about 30 RK4 steps, whereas NAGO attains lower lMSE at 3–5 steps. Thus, source design reduces the accurate step and neural-function-evaluation budget by 6–10$\times$ in this setting. The U-Net results indicate that the source can also help a grid-based backbone, but they do not establish architecture-independent gains.

Generation Time Cost. On one RTX 4090 at batch size 16, the SNS pseudo-spectral solver requires 20.3814 s per recorded batch, while five-step NAGO requires a median 0.228848 s, a raw post-training throughput ratio of $89.1\times$. This comparison excludes dataset construction and model training and is not an equal-fidelity or end-to-end break-even claim. The Matérn-specific efficiency evidence is the separate 6–10$\times$ reduction in RK4 steps above; the timing protocol and autocorrelation-adjusted accounting are given in Appendix F.10.

4.2 Generation at Varying Resolutions

We next separate same-resolution generation at 64^2 from transfer to 32^2 and 128^2. As shown in Fig. 3 (C3), the Matérn source better preserves the reference spectrum across these grids. At 128^2, it retains 93.99% of energy in the frequency band shared with training and places 2.39% in newly available modes, compared with 23.64% and 76.92% for the Gaussian source (Appendix F.9).

These results show preservation of the learned shared-frequency distribution and suppression of spurious new-band energy, not reconstruction of an unseen continuum tail. Indeed, the 128^2 Matérn palinstrophy ratio remains 6.72, so fine-scale extrapolation is still imperfect.

4.3 Generation with Guidance

Finally, we consider guided generation, where low-resolution observations are provided and the goal is to reconstruct the full field while quantifying uncertainty.

Previous studies [Lu et al., 2025] have reported numerical oscillations or divergence during guided generation, particularly under strong guidance, which is closely linked to steep or ill-conditioned marginal velocity fields. Several works [Chen and Vanden-Eijnden, 2026, Albergo and Vanden-Eijnden, 2023, Albergo et al., 2025] have proposed interpolation or guidance strength schedules to mitigate these effects.

Figure 3 (C4) and Table 3 compare guided SNS reconstruction with 25% observations under the original per-step guidance parameterization. Matérn runs show smoother early loss trajectories and lower error for several matched settings. However, when total guidance intervention is held fixed as the solver budget changes, Gaussian sources can be better on observation MSE while Matérn sources can retain better spectral fidelity. We therefore interpret guidance as a metric- and hyperparameter-dependent secondary result rather than direct causal confirmation of the stiffness mechanism; the fixed-total controls are reported in Appendix F.12.

Remark. Guidance strength and its accumulation across steps materially affect the comparison. Appendix F.12 reports this sensitivity and separates ODE refinement from cumulative guidance intervention.

Table 3: Sampling with guidance for the SNS benchmark.

Steps	Matérn + FNO		Gaussian + FNO		Matérn + U-Net		Gaussian + U-Net	
	L2	MSE	L2	MSE	L2	MSE	L2	MSE
100	$\mathbf{8.42\times10^{-2}}$	$\mathbf{7.27\times10^{-3}}$	1.16×10^{-1}	2.83×10^{-2}	6.26×10^{-2}	4.01×10^{-3}	3.10×10^{-1}	9.85×10^{-2}
500	$\mathbf{1.03\times10^{-1}}$	$\mathbf{1.08\times10^{-2}}$	1.59×10^{-1}	2.62×10^{-2}	2.76×10^{-1}	7.85×10^{-2}	1.52	2.37
1000	$\mathbf{8.70\times10^{-2}}$	$\mathbf{7.75\times10^{-3}}$	1.53×10^{-1}	2.39×10^{-2}	4.66×10^{-1}	2.27×10^{-1}	1.98	4.04

5 Conclusion and Limitations

In this work, we introduced NAGO, a noise-aware generative operator for sampling invariant measures of SPDEs. The key idea is to impose function-space admissibility first and then align the source and target covariance over the modes represented by the model. This separates the theoretical roughness requirement from the practical objective of reducing spectral stiffness. Across three stationary stochastic-field benchmarks, the selected Matérn source improves few-step spectral accuracy and several complementary distributional statistics; on the controlled Matérn target, it reduces the accurate RK4 budget from about 30 to 3–5 steps. The scope is nevertheless limited to stationary fields on periodic regular grids, where spectra can be estimated reliably and Matérn sampling is FFT-diagonalizable. Cross-resolution results preserve shared frequencies but do not recover an unseen spectral tail, and guidance comparisons remain metric- and hyperparameter-dependent. Extending the admissibility-constrained criterion to irregular geometries, nonstationary spectra, and learned anisotropic sources is an important direction for future work.

Acknowledgments and Disclosure of Funding

This work was supported by the National Key Research and Development Project of China under Grant No. 2025YFE0204600 and the Key Program of the National Natural Science Foundation of China under Grant No. 62533019. The authors declare no competing financial interests.

References

Michael Albergo, Nicholas M Boffi, and Eric Vanden-Eijnden. Stochastic interpolants: A unifying framework for flows and diffusions. *Journal of Machine Learning Research*, 26(209):1–80, 2025.

Michael Samuel Albergo and Eric Vanden-Eijnden. Building normalizing flows with stochastic interpolants. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://arxiv.org/abs/2209.15571.

Alain Bensoussan. Stochastic navier-stokes equations. *Acta Applicandae Mathematica*, 38(3): 267–304, 1995.

Nicholas Matthew Boffi, Michael Samuel Albergo, and Eric Vanden-Eijnden. Flow map matching with stochastic interpolants: A mathematical framework for consistency models. *Trans. Mach. Learn. Res.*, 2025, 2024. URL https://api.semanticscholar.org/CorpusID:270380005.

René Carmona. *Stochastic Partial Differential Equations: Six Perspectives: Six Perspectives*. Number 64. American Mathematical Soc., 1999.

Guanyu Chen, Shengze Xu, Dong Ni, and Tieyong Zeng. Dgnn: A neural pde solver induced by discontinuous galerkin methods, 2025. URL https://arxiv.org/abs/2503.10021.

Ting Chen. On the importance of noise scheduling for diffusion models, 2023. URL https://arxiv.org/abs/2301.10972.

Yifan Chen. New affine invariant ensemble samplers and their dimensional scaling, 2025. URL https://arxiv.org/abs/2505.02987.

Yifan Chen and Eric Vanden-Eijnden. Scale-adaptive generative flows for multiscale scientific data, 2026. URL https://arxiv.org/abs/2509.02971.

Yifan Chen, Eric Vanden-Eijnden, and Jiawei Xu. Lipschitz-guided design of interpolation schedules in generative models, 2026. URL https://arxiv.org/abs/2509.01629.

Yuzhu Chen, Fengxiang He, Shi Fu, Xinmei Tian, and Dacheng Tao. Adaptive time-stepping schedules for diffusion models. In *The 40th Conference on Uncertainty in Artificial Intelligence*, 2024. URL https://openreview.net/forum?id=lZzriJH2DC.

Giuseppe Da Prato and Jerzy Zabczyk. *Ergodicity for infinite dimensional systems*, volume 229. Cambridge university press, 1996.

Giuseppe Da Prato and Jerzy Zabczyk. *Stochastic equations in infinite dimensions*. Cambridge university press, 2014.

Jean-Pierre Eckmann and Martin Hairer. Invariant measures for stochastic partial differential equations in unbounded domains. *Nonlinearity*, 14(1):133–151, 2001.

Jiajie Fan, Laure Vuaille, Thomas Bäck, and Hao Wang. On the noise scheduling for generating plausible designs with diffusion models, 2023. URL https://arxiv.org/abs/2311.11207.

Giulio Franzese, Giulio Corallo, Simone Rossi, Markus Heinonen, Maurizio Filippone, and Pietro Michiardi. Continuous-time functional diffusion processes. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. URL https://openreview.net/forum?id=VPrir0p5b6.

Shiqi Gong, Peiyan Hu, Qi Meng, Yue Wang, Rongchan Zhu, Bingguang Chen, Zhiming Ma, Hao Ni, and Tie-Yan Liu. Deep latent regularity network for modeling stochastic partial differential equations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 7740–7747, 2023.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020.

Qiushan Guo, Sifei Liu, Yizhou Yu, and Ping Luo. Rethinking the noise schedule of diffusion-based generative models, 2024. URL https://openreview.net/forum?id=ylHLVq0psd.

Martin Hairer. Solving the kpz equation. *Annals of mathematics*, pages 559–664, 2013.

Martin Hairer and Jonathan C Mattingly. Ergodicity of the 2d navier-stokes equations with degenerate stochastic forcing. *Annals of Mathematics*, pages 993–1032, 2006.

Tiankai Hang, Shuyang Gu, Jianmin Bao, Fangyun Wei, Dong Chen, Xin Geng, and Baining Guo. Improved noise schedule for diffusion training. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 4796–4806. IEEE, 2025.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Ulrich Hornung. Applications of the homogenization method to flow and transport in porous media. *Summer School on Flow and Transport in Porous Media*, pages 167–222, 1992.

Peiyan Hu, Qi Meng, Bingguang Chen, Shiqi Gong, Yue Wang, Wei Chen, Rongchan Zhu, Zhi-Ming Ma, and Tie-Yan Liu. Neural operator with regularity structure for modeling dynamics driven by spdes, 2022. URL https://arxiv.org/abs/2204.06255.

Peiyan Hu, Haodong Feng, Yue Wang, and Zhiming Ma. Model-based closed-loop control algorithm for stochastic partial differential equation control. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, IJCAI '25, 2025. ISBN 978-1-956792-06-5. doi: 10.24963/ijcai.2025/599. URL https://doi.org/10.24963/ijcai.2025/599.

Jiahe Huang, Guandao Yang, Zichen Wang, and Jeong Joon Park. DiffusionPDE: Generative PDE-solving under partial observation. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. URL https://openreview.net/forum?id=z0I2SbjN0R.

Kerson Huang. *Introduction to statistical physics*. Chapman and Hall/CRC, 2009.

Gavin Kerrigan, Justin Ley, and Padhraic Smyth. Diffusion generative models in infinite dimensions. In *International Conference on Artificial Intelligence and Statistics*, 2023.

Gavin Kerrigan, Giosue Migliorini, and Padhraic Smyth. Functional flow matching. In *The 27th International Conference on AI and Statistics (AISTATS)*, 2024.

S Kida and Y Murakami. Kolmogorov similarity in freely decaying turbulence. *The Physics of fluids*, 30(7):2030–2039, 1987.

Junwan Kim, Jiho Park, Seonghu Jeon, and Seungryong Kim. Better source, better flow: Learning condition-dependent source distribution for flow matching, 2026. URL https://arxiv.org/abs/2602.05951.

Diederik P Kingma and Max Welling. Auto-encoding variational bayes, 2022. URL https://arxiv.org/abs/1312.6114.

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.

Zhiqi Li, Yuchen Sun, Greg Turk, and Bo Zhu. Functional mean flow in hilbert space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1928–1938, 2026.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations, 2021. URL https://arxiv.org/abs/2010.08895.

Jae Hyun Lim, Nikola B Kovachki, Ricardo Baptista, Christopher Beckham, Kamyar Azizzadenesheli, Jean Kossaifi, Vikram Voleti, Jiaming Song, Karsten Kreis, Jan Kautz, et al. Score-based diffusion models in function space. *Journal of Machine Learning Research*, 26(158):1–62, 2025.

Finn Lindgren, Håvard Rue, and Johan Lindström. An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach. *Journal of the royal statistical society: Series b (statistical methodology)*, 73(4):423–498, 2011.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=PqvMRDCJT9t.

Pengwei Liu, Zhongkai Hao, Xingyu Ren, Hangjie Yuan, Jiayang Ren, and Dong Ni. Papm: a physics-aware proxy model for process systems. In *Proceedings of the 41st International Conference on Machine Learning*, ICML'24. JMLR.org, 2024a.

Pengwei Liu, Qinxin Wu, Xingyu Ren, Yian Wang, and Dong Ni. A deep-learning-based surrogate modeling method with application to plasma processing. *Chemical Engineering Research and Design*, 211:299–317, 2024b.

Pengwei Liu, Pengkai Wang, Xingyu Ren, Hangjie Yuan, Zhongkai Hao, Chao Xu, Shengze Cai, and Dong Ni. Aerogto: An efficient graph-transformer operator for learning large-scale aerodynamics of 3d vehicle geometries. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 18924–18932, 2025.

Pengwei Liu, Xingyu Ren, Pengkai Wang, Hangjie Yuan, Zhongkai Hao, Guanyu Chen, Chao Xu, Dong Ni, and Shengze Cai. An efficient graph-transformer operator for learning physical dynamics with manifolds embedding. *Communications AI & Computing*, 1(1):9, 2026.

Gabriel J Lord, Catherine E Powell, and Tony Shardlow. An introduction to computational stochastic pdes. *Cambridge Texts Appl. Math*, 50, 2014.

Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. Dpm-solver++: Fast solver for guided sampling of diffusion probabilistic models. *Machine Intelligence Research*, 22 (4):730–751, 2025.

Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via deeponet based on the universal approximation theorem of operators. *Nature machine intelligence*, 3(3):218–229, 2021.

Shimon Malnick, Matan Rusanovsky, Ohad Fried, and Shai Avidan. Optimal transport flow matching by design, 2026. URL https://arxiv.org/abs/2606.04092.

Kunwoo Na, Junghyun Lee, Se-Young Yun, and Sungbin Lim. Probability-flow ode in infinite-dimensional function spaces, 2025. URL https://arxiv.org/abs/2503.10219.

Ariel Neufeld and Philipp Schmocker. Solving stochastic partial differential equations using neural networks in the wiener chaos expansion, 2026. URL https://arxiv.org/abs/2411.03384.

Jakiw Pidstrigach, Youssef Marzouk, Sebastian Reich, and Sven Wang. Infinite-dimensional diffusion models. *Journal of Machine Learning Research*, 25(414):1–52, 2024.

Md Ashiqur Rahman, Manuel A Florez, Anima Anandkumar, Zachary E Ross, and Kamyar Azizzadenesheli. Generative adversarial neural operators. *Transactions on Machine Learning Research*, 2022. ISSN 2835-8856. URL https://openreview.net/forum?id=X1VzbBU6xZ.

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.

Cristopher Salvi, Maud Lemercier, and Andris Gerasimovics. Neural stochastic pdes: Resolution-invariant learning of continuous spatiotemporal dynamics. *Advances in Neural Information Processing Systems*, 35:1333–1344, 2022.

Simo Särkkä and Arno Solin. Applied stochastic differential equations. *Institute of mathematical statistics textbooks*, 2019.

Jacob H. Seidman, Georgios Kissas, George J. Pappas, and Paris Perdikaris. Variational autoencoding neural operators. In *Proceedings of the 40th International Conference on Machine Learning*, ICML'23. JMLR.org, 2023.

Yuyang Shi, Valentin De Bortoli, Andrew Campbell, and Arnaud Doucet. Diffusion schrödinger bridge matching. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. URL https://openreview.net/forum?id=qy07OHsJT5.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *International Conference on Learning Representations*, 2021a. URL https://openreview.net/forum?id=St1giarCHLP.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021b. URL https://openreview.net/forum?id=PxTIG12RRHS.

Alexander Tong, Kilian FATRAS, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856. URL https://openreview.net/forum?id=CD9Snc73AW. Expert Certification.

Weijun Xu, Wenhao Zhao, and Shuhan Zhou. Interface fluctuations for 1d stochastic allen-cahn equation revisited, 2025. URL https://arxiv.org/abs/2402.19070.

Jiachen Yao, Abbas Mammadov, Julius Berner, Gavin Kerrigan, Jong Chul Ye, Kamyar Azizzadenesheli, and Anima Anandkumar. Guided diffusion sampling on function spaces with applications to PDEs. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. URL https://openreview.net/forum?id=oAgwvZay2U.

Dongkun Zhang, Ling Guo, and George Em Karniadakis. Learning in modal space: Solving time-dependent stochastic pdes using physics-informed neural networks. *SIAM Journal on Scientific Computing*, 42(2):A639–A665, 2020. doi: 10.1137/19M1260141. URL https://doi.org/10.1137/19M1260141.

Pascal Zwick, Nils Friederich, Maximilian Beichter, Lennart Hilbert, Ralf Mikut, and Oliver Bringmann. Lediflow: Learned distribution-guided flow matching to accelerate image generation, 2025. URL https://arxiv.org/abs/2505.20723.

Appendix Contents

A Related Work **16**

A.1 Neural Solvers 16

A.2 Generative Models 16

A.3 Difference between finite and infinite dimensional generative models. 17

B Invariant measures **18**

C Preliminaries and notation **19**

C.1 Random Fields 19

C.2 Noise Measure 21

C.3 Solutions of SPDEs 26

D Proofs of Main Theoretical Results **27**

D.1 Proof of Theorem 2.4 27

D.2 Proof of Theorem 2.5 28

D.3 Proof of Theorem 3.1 29

D.4 Proof of Theorem 3.2 30

D.5 Proof of Theorem 3.3 30

D.6 Proof of Theorem 3.4 31

D.7 Proof of Eq. 14 33

E Algorithm of `NAGO` **33**

F Experiments Details **35**

F.1 Baselines 35

F.2 Introduction of Datasets 36

F.3 Experiment 1: Generation at Fixed Resolution 37

F.4 Experiment 2: Generation at Varying Resolution 38

F.5 Experiment 3: Generation with Guidance 40

F.6 Ablation Experiments 41

F.7 Controlled RK4 Step Sweep 43

F.8 Complementary SNS Distribution Metrics 43

F.9 Resolution Diagnostics 43

F.10 Sampling Throughput Protocol 43

F.11 FFT Matérn Sampling Cost 43

F.12 Stationary Anisotropic Target 44

G Limitations of This Study **46**

A Related Work

In recent years, both operator-learning models and generative modeling techniques have advanced rapidly. We begin by briefly reviewing existing research on neural operators and their applications to modeling SPDEs. We then discuss recent progress in generative modeling within scientific computing and highlight how our approach differs from prior work in both areas.

A.1 Neural Solvers

Numerous neural solvers have been proposed to approximate the solutions or operators of PDEs and SPDEs using neural networks. Raissi et al. [2019] introduced Physics-Informed Neural Networks (PINNs), in which a neural network is optimized using PDE constraints as self-supervised losses to obtain PDE solutions. These methods have also been extended to solve SPDEs [Zhang et al., 2020, Neufeld and Schmocker, 2026], although they are difficult to generalize across different types of SPDEs. DGNN [Chen et al., 2025] combines discontinuous Galerkin formulations with piecewise neural trial spaces to construct a neural PDE solver. Among neural operators, the Fourier Neural Operator (FNO) [Kovachki et al., 2023, Li et al., 2021] and DeepONet [Lu et al., 2021] are among the most widely used due to their representative architectures, which were designed with universal approximation properties. The primary objective of neural operators is to learn mappings between distributions in function spaces. Many other related operator-based approaches have been proposed to address specific challenges, such as irregular domains and data [Liu et al., 2025, 2026], as well as long-term forecasting [Liu et al., 2024a,b]. These methods have demonstrated strong performance across a variety of tasks; however, their applicability remains largely restricted to PDE-related problems.

By incorporating time-dependent randomness, Neural SPDE [Salvi et al., 2022] and DLRNet [Gong et al., 2023, Hu et al., 2022, 2025] have been proposed for modeling SPDE operators. The distinction between modeling SPDE operators and PDE operators lies in the fact that the former accounts for infinite-dimensional, time-dependent randomness, whereas the latter does not. In the work of Salvi et al. [2022], kernel-based methods were employed to handle time evolution by learning the semigroup operator generated by the differential operator. In the work of Gong et al. [2023], the regularity features of SPDEs were combined with neural networks to improve generalization; however, the computational cost becomes prohibitively high for high-order signatures.

However, in real-world SPDE dynamics, the driving noise is rarely observable; only system states at selected spatial and temporal points are typically available. Moreover, numerical solvers require extremely fine temporal discretization, making full-trajectory prediction computationally costly. In this work, we therefore focus not on forecasting the SPDE evolution, but on sampling from its stationary distribution once equilibrium is reached. For example, in thin-film growth described by the EW/KPZ equations under fixed operating conditions, our interest lies in the stabilized surface statistics rather than the transient dynamics. Importantly, all objects of interest reside in function spaces, making operator-learning models essential for accurately capturing the underlying infinite-dimensional structure.

A.2 Generative Models

In recent years, rapid progress has been made in generative modeling, including DDPM [Ho et al., 2020], DDIM [Song et al., 2021a], score-based models [Song et al., 2021b], and flow matching [Lipman et al., 2023]. Beyond finite-dimensional settings, these ideas have been extended to infinite-dimensional generative models that aim to directly represent distributions over function spaces. For example, Pidstrigach et al. [2024], Lim et al. [2025], Kerrigan et al. [2023], Franzese et al. [2023] propose score- or diffusion-based approaches that are invariant to discretization and operate directly in Hilbert or Banach spaces, thereby enabling a variety of applications in scientific machine learning. Although FFM [Kerrigan et al., 2024] establishes functional flow matching, it uses Gaussian sources and does not empirically isolate source design. Other functional models adapt GAN or VAE architectures [Na et al., 2025, Rahman et al., 2022, Seidman et al., 2023], while recent work designs interpolants or conditional velocity fields [Albergo et al., 2025, Chen and Vanden-Eijnden, 2026, Chen et al., 2026, Boffi et al., 2024]. Functional Mean Flow targets one-step generation in Hilbert space [Li et al., 2026], and FunDPS develops guided diffusion sampling and an infinite-dimensional Tweedie formula for PDE inverse problems [Yao et al., 2025]. NAGO is complementary: it isolates

the source measure and links its continuum admissibility and finite-band covariance ratio to spectral stiffness.

Source design has also emerged in finite-dimensional flow matching. LeDiFlow learns a data-dependent image prior to shorten probability paths [Zwick et al., 2025]; Better Source, Better Flow learns condition-dependent text-to-image sources with variance and directional regularization [Kim et al., 2026]; and Optimal Transport Flow Matching by Design constructs a low-frequency image prior that admits an approximately optimal identity coupling [Malnick et al., 2026]. These works establish that source distributions matter beyond our setting. Our contribution is narrower and different: an analytic, admissibility-constrained source rule for probability measures on function spaces, with a spectral-stiffness analysis and cross-resolution SPDE evaluation.

A crucial design choice in diffusion-based generative models is the noise schedule, i.e., how noise levels are varied during the forward and reverse processes. Prior work has shown that the noise schedule significantly affects model behavior: for example, different noise schedules can shift optimal performance and robustness across tasks and resolutions, and the choice of schedule often needs to be tailored rather than fixed [Chen, 2023]. Subsequent research has undertaken both theoretical and empirical analysis of noise schedules, proposing new metrics and schedule designs to improve generation quality and training dynamics [Hang et al., 2025, Guo et al., 2024]. Moreover, studies focusing on specific application domains (e.g., design plausibility in conditional diffusion) have explored parametric noise scheduling to balance generative fidelity and structural constraints [Fan et al., 2023]. Together, these works highlight that noise choice and scheduling are fundamental for stable, high-quality generation in diffusion models.

A.3 Difference between finite and infinite dimensional generative models.

Generative models differ fundamentally depending on whether they are formulated in finite-dimensional Euclidean spaces or in infinite-dimensional function spaces. In finite-dimensional settings, the target distribution is defined on a fixed space $\mathbb{R}^d$, where the dimension d is prescribed by the chosen representation. For example, an image model may operate on a fixed pixel grid, while a latent-variable model may operate on a fixed-dimensional feature vector. In this regime, probability distributions can often be described through densities with respect to the Lebesgue measure, and generative dynamics such as diffusion processes, score-based models, normalizing flows, or flow matching can be formulated as stochastic or deterministic differential equations on $\mathbb{R}^d$. The associated analysis typically relies on finite-dimensional calculus, including gradients, divergences, Jacobians, change-of-variables formulas, and numerical discretizations of ordinary or stochastic differential equations.

Function-space generative modeling has a different mathematical structure. Here the target object is a probability measure on a space of functions or fields, such as a Hilbert space, a Banach space, or a Sobolev-type space. Samples are not merely finite vectors, but discretized observations of continuum objects. Typical examples include Gaussian random fields, solutions of stochastic partial differential equations, invariant measures of stochastic dynamics, and continuous physical fields. This change of state space affects the basic theoretical language of generative modeling. In infinite-dimensional spaces, there is no translation-invariant Lebesgue reference measure. Consequently, densities, likelihoods, and scores cannot in general be interpreted in the same way as in $\mathbb{R}^d$. Instead, the construction of generative dynamics must be formulated at the level of probability measures, operators, covariance structures, and infinite-dimensional evolution equations. Questions such as existence, uniqueness, regularity, and stability become central, since the flow must be well defined as a continuum object rather than only after a finite discretization.

The second major difference concerns the role of spatial resolution. In a finite-dimensional model, the resolution is usually part of the definition of the data representation. Changing the resolution changes the ambient dimension and therefore typically defines a different finite-dimensional modeling problem. For example, an image model trained on 64×64 images and another model trained on 256×256 images are usually treated as models on different Euclidean spaces. By contrast, in function-space generative modeling, different resolutions should be understood as different discretizations of the same underlying continuum measure. A model that is genuinely defined in function space should therefore be discretization-consistent: as the grid is refined, the learned distribution and the induced generative dynamics should converge to a well-defined limiting object. Equivalently, the model should not rely on artifacts of a particular grid size.

This distinction is especially important for multiscale SPDE data. The target invariant measures often have nontrivial spatial regularity and strongly nonuniform spectral structure. If the source noise is chosen without regard to this regularity, its discretized high-frequency content may dominate the induced transport dynamics. For instance, standard discretized Gaussian noise can inject substantial energy into modes that are not representative of the target measure. The learned velocity field must then spend significant capacity suppressing these unresolved high-frequency components. This can lead to spectrally stiff marginal velocity fields, unstable training, and generation behavior that changes with the spatial resolution. Thus, in function-space generative modeling, the source noise is not merely an implementation detail. It directly affects whether the learned flow is well conditioned, resolution robust, and compatible with the target continuum measure.

This perspective motivates the regularity-aware noise design considered in this work. Rather than fixing the source distribution to be standard Gaussian noise at every discretization level, we choose a noise measure whose regularity is matched to that of the target invariant measure. The goal is to reduce high-frequency mismatch at the level of the continuum formulation, thereby producing better-conditioned marginal velocity fields and more stable discretization-consistent generation.

B Invariant measures

In many engineering and scientific applications, the quantities of interest are not deterministic functions but uncertain spatial fields. For example, the permeability of a subsurface medium may vary sharply across space and cannot be fully determined from limited observations. Representing such a parameter by a single fixed function therefore ignores the uncertainty that is essential for prediction and decision-making. A more appropriate formulation is to model the parameter as a random field and characterize its distribution through a probability measure on a function space. In SPDE-based modeling, the invariant measure of a stochastic dynamics provides such a statistical description: it encodes the long-term spatial structure, variability, and correlations of the field. Sampling from this invariant measure then produces statistically plausible realizations of the unknown medium, which can be used for uncertainty propagation, inverse problems, risk assessment, and design optimization.

Why are multiscale invariant measures challenging for functional generative models? Multiscale SPDE invariant measures often contain spatial structures across a wide range of frequencies. A functional generative model must therefore learn not only the distribution of samples, but also a resolution-consistent transport mechanism between probability measures on function spaces. In practice, many implementations use standard Gaussian noise as the source distribution. This choice is convenient, but it can be poorly matched to the target invariant measure. When the source noise is much rougher than the data, it injects excessive high-frequency energy into the generative path. The induced marginal velocity field must then transport rough noise into smoother target fields, which creates strong spectral imbalance and leads to stiff dynamics. This stiffness can make training unstable and sampling expensive, especially when the spatial resolution changes.

Why are interpolation schedules or adaptive solvers not enough? Existing methods often try to improve stability through interpolation schedules [Tong et al., 2024, Chen et al., 2024], guidance strategies, or specialized ODE solvers. These techniques can be useful, but they mainly act on the generative path or on the numerical integration after the source and target measures have already determined the geometry of the flow. If the initial noise measure is strongly mismatched with the target regularity, the resulting marginal velocity field is already poorly conditioned. Downstream heuristics may reduce some symptoms of stiffness, but they do not remove its source. As a result, they may require additional tuning, increase computational cost, or fail to remain robust across discretization resolutions.

Our perspective: design the noise measure in function space. We instead treat the source noise measure as a design variable. The central idea is that the geometry of the functional flow should be controlled at the level of probability measures, rather than corrected only through downstream mechanisms. By choosing a noise distribution whose regularity is compatible with that of the target invariant measure, we can reduce high-frequency mismatch before it is inherited by the marginal velocity field. This leads to better-conditioned flows, more stable training, and improved cross-resolution behavior.

C Preliminaries and notation

We briefly review several standard facts on random fields, covariance operators, Karhunen–Loève expansions, and Gaussian measures on Hilbert spaces. These notions provide the functional-analytic foundation for the noise measures considered in this work.

C.1 Random Fields

C.1.1 Random Fields as Hilbert-Valued Random Variables

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $D \subset \mathbb{R}^d$ be a bounded domain. We denote by H a separable Hilbert space of functions on D, typically $H = L^2(D)$ or a Sobolev space. A random field is viewed as an H-valued random variable

$$u : \Omega \to H. \tag{15}$$

For each $\omega \in \Omega$, $u(\omega)$ is a function on D. When emphasizing the spatial variable, we write $u(x, \omega)$.

For $1 \le p < \infty$, the Bochner space $L^p(\Omega; H)$ consists of all strongly measurable random variables $u : \Omega \to H$ satisfying

$$\mathbb{E}\|u\|_H^p < \infty, \tag{16}$$

with norm

$$\|u\|_{L^p(\Omega;H)} := \left(\mathbb{E}\|u\|_H^p\right)^{1/p}. \tag{17}$$

When $p = 2$, $L^2(\Omega; H)$ is a Hilbert space with inner product

$$\langle u, v\rangle_{L^2(\Omega;H)} := \mathbb{E}\langle u, v\rangle_H. \tag{18}$$

The law of u is the probability measure

$$m := u_{\#}\mathbb{P} \tag{19}$$

on H. Thus, a random field can equivalently be regarded as a probability measure on a function space.

C.1.2 Mean and Covariance Operator

Assume that $u \in L^2(\Omega; H)$. Its mean is the element $\mu \in H$ defined by

$$\mu := \mathbb{E}u. \tag{20}$$

The covariance operator $\mathcal{C} : H \to H$ is defined weakly by

$$\langle \mathcal{C}\phi, \psi\rangle_H = \mathbb{E}\left[\langle u - \mu, \phi\rangle_H \langle u - \mu, \psi\rangle_H\right], \qquad \forall \phi, \psi \in H. \tag{21}$$

Equivalently,

$$\mathcal{C} = \mathbb{E}\big[(u - \mu) \otimes (u - \mu)\big], \tag{22}$$

where $(a \otimes b)\phi := a\langle b, \phi\rangle_H$.

The operator $\mathcal{C}$ is self-adjoint, positive semidefinite, and trace class. Moreover,

$$\mathrm{Tr}(\mathcal{C}) = \mathbb{E}\|u - \mu\|_H^2. \tag{23}$$

When $H = L^2(D)$ and $\mathcal{C}$ admits an integral kernel $C(x, y)$, we have

$$(\mathcal{C}\phi)(x) = \int_D C(x, y)\phi(y)\, dy, \qquad C(x, y) = \mathrm{Cov}(u(x, \cdot), u(y, \cdot)). \tag{24}$$

C.1.3 Karhunen–Loève Expansion

Assume that $H = L^2(D)$ and that the covariance operator $\mathcal{C}$ is compact, self-adjoint, positive semidefinite, and trace class. Let $\{(\lambda_i, \phi_i)\}_{i \ge 1}$ be its eigenpairs:

$$\mathcal{C}\phi_i = \lambda_i \phi_i, \qquad \lambda_i \ge 0, \tag{25}$$

where $\{\phi_i\}_{i\geq 1}$ is an orthonormal basis of H. Then the Karhunen–Loève expansion of u is

$$u(x,\omega) = \mu(x) + \sum_{i=1}^{\infty} \sqrt{\lambda_i}\, \xi_i(\omega)\phi_i(x), \tag{26}$$

where, for $\lambda_i > 0$,

$$\xi_i(\omega) = \frac{1}{\sqrt{\lambda_i}} \langle u(\omega) - \mu, \phi_i \rangle_H. \tag{27}$$

The random variables $\{\xi_i\}_{i\geq 1}$ satisfy

$$\mathbb{E}[\xi_i] = 0, \qquad \mathbb{E}[\xi_i \xi_j] = \delta_{ij}. \tag{28}$$

If u is Gaussian, then $\{\xi_i\}_{i\geq 1}$ are independent standard normal random variables.

The expansion converges in $L^2(\Omega; H)$, and

$$\mathbb{E}\|u - \mu\|_H^2 = \sum_{i=1}^{\infty} \lambda_i = \mathrm{Tr}(\mathcal{C}). \tag{29}$$

If $C(x,y)$ is continuous, then Mercer's theorem gives

$$C(x,y) = \sum_{i=1}^{\infty} \lambda_i \phi_i(x)\phi_i(y), \tag{30}$$

with convergence under the usual Mercer assumptions.

C.1.4 Gaussian Measures on Hilbert Spaces

An H-valued random variable u is Gaussian if

$$\langle u, \phi \rangle_H \tag{31}$$

is a real-valued Gaussian random variable for every $\phi \in H$. If u has mean μ and covariance operator $\mathcal{C}$, we write

$$u \sim \mathcal{N}(\mu, \mathcal{C}). \tag{32}$$

Equivalently,

$$\langle u, \phi \rangle_H \sim \mathcal{N}\left(\langle \mu, \phi \rangle_H, \langle \mathcal{C}\phi, \phi \rangle_H\right), \qquad \forall \phi \in H. \tag{33}$$

In infinite-dimensional Hilbert spaces, the covariance operator of a Gaussian measure must be trace class. Conversely, if $\mathcal{C} : H \to H$ is self-adjoint, positive semidefinite, and trace class, then there exists a Gaussian measure $\mathcal{N}(\mu, \mathcal{C})$ on H.

If $\mathcal{C}$ has eigenpairs $\{(\lambda_i, \phi_i)\}_{i\geq 1}$, then a centered Gaussian random variable with covariance $\mathcal{C}$ admits the representation

$$u = \sum_{i=1}^{\infty} \sqrt{\lambda_i}\, \xi_i \phi_i, \qquad \xi_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1). \tag{34}$$

The series converges in $L^2(\Omega; H)$ if and only if $\sum_{i=1}^{\infty} \lambda_i < \infty$.

C.1.5 Spectral Decay and Regularity

The covariance spectrum determines the regularity of Gaussian random fields. Assume that $H = L^2(D)$ and that $\{\phi_k\}$ is a Fourier-type basis indexed by frequency k. Let

$$u = \sum_{k} \sqrt{\lambda_k}\, \xi_k \phi_k, \qquad \xi_k \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1). \tag{35}$$

Then, formally,

$$\mathbb{E}\|u\|_{H^s(D)}^2 \asymp \sum_{k} (1 + |k|^2)^s \lambda_k. \tag{36}$$

Hence $u \in H^s(D)$ in mean square whenever

$$\sum_k (1 + |k|^2)^s \lambda_k < \infty. \tag{37}$$

Thus, faster decay of λ_k corresponds to smoother samples, while slower decay corresponds to rougher samples with stronger high-frequency content.

A Matérn-type Gaussian measure provides a convenient family of covariance spectra with tunable regularity. In a Fourier basis, a typical Matérn covariance has eigenvalues of the form

$$\lambda_k \propto \left(\kappa^2 + |k|^2\right)^{-\alpha}, \tag{38}$$

where $\kappa > 0$ controls the correlation length and $\alpha > 0$ controls spectral decay. Increasing α suppresses high-frequency modes and produces smoother random fields.

C.1.6 Implication for Source Noise Design

The preceding facts show that a Gaussian noise measure on a function space is determined by its covariance operator. In particular, choosing the source noise distribution is equivalent to choosing a covariance spectrum $\{\lambda_k\}$, which fixes the regularity and high-frequency energy of the initial random field.

Therefore, in function-space generative modeling, the source noise is not a neutral default. If the source noise is much rougher than the target measure, the induced generative dynamics must remove excessive high-frequency energy. This produces spectral imbalance in the marginal velocity field and may lead to stiff dynamics. Regularity-aware noise design aims to reduce this mismatch by choosing a source covariance spectrum whose decay is compatible with the target invariant measure.

C.2 Noise Measure

C.2.1 Gaussian Measure

Since we consider the infinite-dimensional case, unlike finite dimensions, not all translations preserve the measure. We need to consider those directions in H along which translating a Gaussian measure does not change its essential nature (i.e., keeps it equivalent). Including the SPDE, these problems are also dealt with in Hilbert Space. First, we have some definitions:

Definition C.1. *Let m_1, m_2 be two measures on H.*

- *$m_1 \ll m_2$ means measure m_1 is absolutely continuous respect to m_2: if $m_1(A) = 0$ for all A s.t. $m_2(A) = 0$. This means that the support of m_1 is a subset of the support of m_2.*
- *If $m_1 \ll m_2, m_2 \ll m_1$, m_1, m_2 are said to be equivalent $m_1 \sim m_2$.*
- *If there exists a measurable set A s.t. $m_1(A) = 0$ and $m_2(A^c) = 1$, then m_1, m_2 are singular $m_1 \perp m_2$.*

Theorem C.2 (Radon-Nikodym Theorem)**.** *Let $S = (H, \mathcal{B}(H))$ be a measurable space. m_1, m_2 are two σ-finite measures on S. If $m_1 \ll m_2$, then there exists a measurable function $f : H \to [0, \infty)$ s.t. $m_1(A) = \int_A f m_2(du), \forall A \in \mathcal{B}(H)$.*

The function f is called the Radon-Nikodym derivative of m_1 with respect to m_2 and is denoted by $f = \frac{dm_1}{dm_2}$.

Remark. *In the finite-dimensional case, the Radon-Nikodym derivative is the probability density function. Like m_2 is the Lebesgue measure, $m_1(dx) = p(x)dx$, then the Radon-Nikodym derivative is the probability density function $p(x)$.*

Definition C.3 (KL-divergence)**.** *The KL-divergence is actually related to the Radon-Nikodym derivative: if $m_1 \ll m_2$, then:*

$$\mathrm{KL}(m_1||m_2) = \int_H \log\left(\frac{dm_1}{dm_2}(x)\right) dm_1(x) = \int_H \log\left(f(x)\right) f(x) dm_2(x) \tag{39}$$

This is the KL divergence between measures. We define the Radon-Nikodym derivative to be infinite if $m_1 \not\ll m_2$. So for $H = \mathbb{R}^d, dm_1 = p_1(x)dx, dm_2 = p_2(x)dx$, the KL-divergence can be written as:

$$\mathrm{KL}(m_1||m_2) = \int_{\mathbb{R}^d} \log\left(\frac{p_1(x)}{p_2(x)}\right) p_1(x)dx \tag{40}$$

The Radon-Nikodym derivative captures local density ratios, and the KL divergence is their global average.

Then we need to consider the characteristic functional of Gaussian measure, cause one can show that characteristic functions of two measures are identical, then the measures are identical as well:

Definition C.4 (Characteristic Functional)**.** *The characteristic functional of a measure m is defined as:*

$$\hat{m}(\lambda) = \int_H e^{i\langle\lambda,u\rangle_H} m(du) \tag{41}$$

For Gaussian measure $m = N(\mu, \mathcal{C})$, we have the characteristic functional:

$$\hat{m}(\lambda) = \int_H e^{i\langle\lambda,u\rangle_H} m(du) = e^{i\langle\lambda,\mu\rangle_H - \frac{1}{2}\langle\lambda,\mathcal{C}\lambda\rangle_H}, \qquad \lambda \in H \tag{42}$$

So conversely, let $\mu \in H, \mathcal{C}$ is a trace class operator on H. Then there exists a unique probability measure m on H s.t. $\hat{m}(\lambda) = e^{i\langle\lambda,\mu\rangle_H - \frac{1}{2}\langle\lambda,\mathcal{C}\lambda\rangle_H}, \forall\lambda \in H$. Here $m = N(\mu, \mathcal{C})$. Meanwhile, we have $\int_H \|x\|_H^2 m(dx) = \mathrm{Tr}(\mathcal{C}) + \|\mu\|_H^2$.

Definition C.5 (Reproducing Kernel Hilbert Space)**.** *The subspace $\mathcal{C}^{1/2}H$ is called the reproducing kernel of measure $N_\mathcal{C}$. If $\mathrm{Ker}(\mathcal{C}) = 0$, then $\mathcal{C}^{1/2}H$ is dense in H.*

Then we have an isomorphism between H and $L^2(H, N_\mathcal{C})$:

$$f \in \mathcal{C}^{1/2}H \to W_f \in L^2(H, N_\mathcal{C}), \qquad W_f(x) = \langle\mathcal{C}^{-1/2}f, x\rangle_H, \quad x \in H \tag{43}$$

for $\int_H W_f(x)W_g(x)N_\mathcal{C}(dx) = \langle f, g\rangle_\mathcal{C}, \quad \forall f, g \in H$.

C.2.2 Camaron Martin Space

Unlike the finite dimensional case, there is no natural Brownian Motion process in infinite dimensions. Here is why: For example, the space time white noise W_t is defined on $H = L^2([0, 1])$, but it doesn't take values in H a.s. Since the white noise has covariance operator I, it is not trace class in H. So for Hilbert space H, we need to define the Brownian Motion process on H by using the Cameron-Martin space U. In other words, if we want to define the Diffusion Process on infinite dimensional space, we need to consider the direction in H along which translating a measure preserves the absolutely continuity.

Definition C.6 (Cameron-Martin Space)**.** *The Cameron-Martin space U is defined as:*

$$U := \{h \in H | m_h \ll m\}, m_h(A) = T_h^\# m(A) := m(A - h), \forall A \in \mathcal{B}(H) \tag{44}$$

where T_h is the translation operator: $T_h(u) = u + h$, $T_h^\# m$ is the push-forward measure of m by T_h,

So the Cameron-Martin space U is more regular than the original space H.

Theorem C.7 (Cameron-Martin Theorem)**.** *For $m = N(0, \mathcal{C})$, we have:*

$$m_h \sim m \text{ if and only if } h \in \mathcal{C}^{\frac{1}{2}}H := U \tag{45}$$

Since $\mathcal{C}^{-1}$ is unbounded, U is a proper subspace of $U = \mathcal{C}^{1/2}H$. The inner product is defined as:

$$\langle\phi, \psi\rangle_U = \langle\mathcal{C}^{-1/2}\phi, \mathcal{C}^{-1/2}\psi\rangle_H \tag{46}$$

This can be generalized to general Gaussian measure cases:

Theorem C.8 (Feldman-Hajek Theorem [Da Prato and Zabczyk, 2014])**.** *The following statements hold:*
1. $m_1 = N(\mu_1, \mathcal{C}_1), m_2 = N(\mu_2, \mathcal{C}_2)$ are either singular or equivalent.
2. They are equivalent if and only if

- $\mathcal{C}_1^{1/2}(H) = \mathcal{C}_2^{1/2}(H) = H_0$, $\mu_1 - \mu_2 \in H_0$
- $(\mathcal{C}_1^{-1/2}\mathcal{C}_2^{1/2})(\mathcal{C}_1^{-1/2}\mathcal{C}_2^{1/2})^* - I$ *is a Hilbert-Schmidt operator.*

Specifically, if $\mathcal{C}_1 = \mathcal{C}_2 = \mathcal{C}$*, we have the Radon Nikodym derivative:*

$$\frac{dm_1}{dm_2}(u) = \exp\left(\langle \mu_1 - \mu_2, \mathcal{C}^{-1}(u - \mu_2)\rangle_H - \frac{1}{2}\|\mathcal{C}^{-1/2}(\mu_1 - \mu_2)\|_H^2\right) \tag{47}$$

Therefore, the KL divergence between m_1 *and* m_2 *is:*

$$\mathrm{KL}[m_1|m_2] = \frac{1}{2}\left(\langle \Delta\mu, C^{-1}\Delta\mu\rangle_H\right) \tag{48}$$

First, we assume that the Cylindrical Wiener Process $\hat{W}_t$ on H, but it does not take values in H cause I is not a trace class. Here we suppose $\hat{W}_t$ is H_1-valued:

$$\hat{W}_t = \sum_{i=1}^{\infty} \beta_i(t)\phi_i \tag{49}$$

where $\beta_i(t)$ are independent standard Brownian motions and ϕ_i are the eigenbasis of H. Normally, we can define the $\mathcal{C}$-Wiener Process on H as:

Definition C.9 ($\mathcal{C}$-Wiener Process)**.** *A H-valued process* W_t *is called* $\mathcal{C}$*-Wiener process if: 1.* $W_0 = 0$ *2.* W_t *has continuous trajectories:* $\mathbb{R}^+ \to H$ *3.* W_t *has independent Gaussian increments:* $W_t - W_s \sim N(0, (t-s)\mathcal{C})$

Therefore by applying the covariance operator $\mathcal{C}$ to cylindrical Wiener Process $\hat{W}_t$, we can define the H-valued $\mathcal{C}$-Wiener process W_t as:

$$W_t = \sum_{i=1}^{\infty} \sqrt{\lambda_i}\beta_i(t)\phi_i = \mathcal{C}^{1/2}\hat{W}_t \tag{50}$$

where λ_i are the eigenvalues of $\mathcal{C}$. So H is actually the Cameron-Martin space of H_1.

To sum up, in infinite dimensional diffusion model, suppose we have $u \sim m_0, \eta \sim m_1$, where m_0 is the initial measure and m_1 is the target measure, all defined on H. In the diffusion process, we add noise process η to u, getting $v = u + \eta$. Then $v \sim \nu = m_0 * m_1$. Here we need to do some assumptions on measure m_0 and m_1: m_0 needs to be supported on the Cameron-Martin space U of m_1, that is $m_0 \ll m_1$, m_0 should be more regular than m_1. Only in this way, we can define the Radon-Nikodym derivative of ν with respect to m_1, and the perturbed measure ν is equivalent to m_1.

Normally, to guarantee the diffusion progress, we have two choices: 1. The support of m_{data} is contained in the Cameron-Martin space of $N(0, \mathcal{C}_{data})$ 2. Choose a large space K, s.t. the support of m_{data} and $N(0, \mathcal{C})$ are supported on K.

Definition C.10. *If A is a linear operator from* $\mathcal{D}(A) \subset H$ *to Hilbert space* H*, with an orthonormal basis of eigenfunctions* $\{\phi_j\}$ *and corresponding increasing eigenvalues* $\{\lambda_j\}$*, then* A^α *is defined as:*

$$A^\alpha u = \sum_{j=1}^{\infty} \lambda_j^\alpha \langle u, \phi_j\rangle \phi_j \tag{51}$$

and the domain $\mathcal{D}(A^\alpha)$ *is the set of all* $u \in H$ *such that* $A^\alpha u \in H$*.*

Definition C.11 (Absolute continuity of measures)**.** *Let* $(X, \mathcal{F})$ *be a measurable space, and let* μ *and* ν *be two measures on* $(X, \mathcal{F})$*. We say that* ν *is* absolutely continuous *with respect to* μ*, and write*

$$\nu \ll \mu,$$

if for every measurable set $A \in \mathcal{F}$*,* $\mu(A) = 0 \implies \nu(A) = 0$*.*

Definition C.12. *The Fourier transform of* $u(x)$ *in this paper is defined like this:*

$$\begin{cases} \mathcal{F}u(k) = \displaystyle\int_{\mathbb{R}^d} u(x)e^{-ikx}dx, \\ \mathcal{F}^{-1}u(x) = \dfrac{1}{(2\pi)^d}\displaystyle\int_{\mathbb{R}^d} u(k)e^{ikx}dk. \end{cases} \tag{52}$$

C.2.3 Random Fields on $\mathbb{R}^d$

In this section, background knowledge on random fields is first reviewed, including their formulation, representation, and numerical treatment, with an emphasis on stationary random fields. Subsequently, more general classes of random fields beyond the Matérn family are discussed, including non-stationary random fields that exhibit anisotropy or temporal dependence.

C.2.4 Stationary Random Fields

First we define stationary random field on $\mathbb{R}^d$ as:

Definition C.13 (Stationary Random Field)**.** *A second-order random field $u(x) : x \in D$ is called stationary if the mean is constant and covariance function depends only on the difference $x - y$, i.e. $\mu(x) = \mu,\ C(x, y) = C(x - y)$.*

Then we can define the covariance operator $\mathcal{C}$ as:

$$\mathcal{C}\phi = \int_{\mathbb{R}^d} C(x - y)\phi(y)dy \tag{53}$$

We find that it is actually the convolution operator of $C(x)$ with $\phi(x)$. Stationary random fields have some beautiful properties.

Theorem C.14 (Wiener-Khinchin Theorem)**.** *There exists a stationary random field $u(x) : x \in D$ with mean μ and covariance function $c(x)$ that is mean square continuous if and only if the function $c(x) : \mathbb{R}^d \to \mathbb{R}$ is such that*

$$c(x) = \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} e^{ik\cdot x}dF(k) = \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} e^{ik\cdot x}S(k)dk = \left(\mathcal{F}^{-1}S\right)(x) \tag{54}$$

where $F(k)$ is some measure on $\mathbb{R}^d$ called spectral distribution and $\hat{S}(x)$ is the Fourier transform of $S(k)$, called spectral density. Reversely, $S(k) = (\mathcal{F}c)(k) = \hat{c}(k)$. If $S(k)$ is non-negative and integrable, then $c(x)$ is a valid covariance function.

Theorem C.15 (Spectral Density of Random Field)**.** *Assume $u(x)$ has zero mean, then*

$$S_u(k) = \frac{1}{(2\pi)^d}\mathbb{E}[|\hat{u}(k)|^2] \tag{55}$$

By defining the pseudo-differential operators, the class of SPDEs is defined by:

$$\mathcal{L}_g u = W, \mathcal{L}_g = \mathcal{F}^{-1}g\mathcal{F} \tag{56}$$

where $g : \mathbb{R}^d \to \mathbb{C}$ must be a sufficiently regular and Hermitian-symmetric function, that is it must satisfy: $g(k) = \overline{g(-k)}$, $\bar{\cdot}$ denotes the complex conjugate. So if we have $\mathcal{L}_g u = W$, then:

$$u = \mathcal{L}_{\frac{1}{g}} W \tag{57}$$

Theorem C.16. *The spectral density of $\mathcal{L}_g u$ and of u are related by:*

$$S_{\mathcal{L}_g u}(k) = |g(k)|^2 S_u(k) \tag{58}$$

Generally, if

$$\mathcal{L}_g u = w \tag{59}$$

where w is a GeRF source term, then $S_w(k) = |g(k)|^2 S_u(k)$. Therefore, when $w = W$, $S_u = \frac{1}{(2\pi)^d|g(k)|^2}\mathbb{E}[|\hat{W}(k)|^2] = \frac{1}{|g(k)|^2}$. Then,

$$u(x) = \mathcal{L}_{\frac{1}{g}} w(x) = \mathcal{L}_{\sqrt{\frac{S_u}{S_w}}} w(x) \tag{60}$$

Then consider the exitence of the function.

Theorem C.17. *Let $w(x)$ be a real stationary GeRF over $\mathbb{R}^d$, and let $g : \mathbb{R}^d \to \mathbb{C}$ be a symbol function. Then for Equ.* (59)*, there exists a unique stationary solution $u(x)$ if and only if: there exists $N \in \mathbb{N}$ s.t.*

$$\int_{\mathbb{R}^d} \frac{dS_w(k)}{|g(k)|^2 (1 + \|k\|^2)^N} < \infty \tag{61}$$

and

$$S_u(k) = |g(k)|^{-2} S_w(k) \tag{62}$$

Moreover, $S_u(k)$ is unique if and only if $|g| > 0$.

Hence the key is the symbol function $g(k)$. The following theorem shows that solutions of SPDEs with White Noise source term is the starting point of more general solutions, when the source term can be any stationary GeRF.

Theorem C.18. *Let $w(x)$ be a real stationary GeRF over $\mathbb{R}^d$ with covariance distribution $C_w(x)$. Let g be a symbol function over $\mathbb{R}^d$ such that $\frac{1}{g}$ is smooth with polynomially bounded derivatives of all orders. Then, there exists a unique stationary solution to (59) and its covariance distribution is given by*

$$C_u(x) = C_u^W * C_w(x) \tag{63}$$

where C_u^W is the covariance function of the solution to the SPDE with White Noise source term.

Proof. The proof is straightforward by using Wiener-Khinchin theorem. □

For any precision operator which is a polynomial in the Laplacian, $Q = p(-\Delta)$, such as the Matern operator with $\nu \in \mathbb{N}$, this results in a polynomial $F(Q) = p(\|k\|^2)$.

Then we introduce the Matérn field.

Proof of Equ. (3). Do Fourier transform on Equ. (1):

$$\left\{\mathcal{F}(\kappa^2 - \Delta)^{\alpha/2} u\right\}(k) = (\kappa^2 + \|k\|^2)^{\alpha/2} (\mathcal{F}u)(k) \tag{64}$$

then we have

$$(\mathcal{F}u)(k) = \hat{u}(k) = \frac{\hat{W}(k)}{(\kappa^2 + \|k\|^2)^{\alpha/2}} \tag{65}$$

Therefore, u can be written as:

$$u(x) = \mathcal{F}^{-1}\left[\frac{\hat{W}(k)}{(\kappa^2 + \|k\|^2)^{\alpha/2}}\right] \tag{66}$$

Then the stationary covariance function of u is given by:

$$c(x) = Cov(u(x), u(0)) \tag{67}$$

By the definition of spectral density Equ. (55) and Equ. (65) we have:

$$S_u(k) = \frac{1}{(2\pi)^d} \frac{\mathbb{E}\left[\left|\hat{W}(k)\right|^2\right]}{(\kappa^2 + \|k\|^2)^\alpha} = \frac{1}{(\kappa^2 + \|k\|^2)^\alpha} \tag{68}$$

Then we have the variance of u:

$$c(0) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} S_u(k) dk = \frac{\Gamma(\nu)}{(4\pi)^{d/2} \kappa^{2\nu} \Gamma(\alpha)} := \sigma^2 \tag{69}$$

By Wiener-Khinchin theorem, we have:

$$\begin{aligned} c(x) = (\mathcal{F}^{-1} S_u)(x) &= \mathcal{F}^{-1}\left[\frac{1}{(\kappa^2 + \|k\|^2)^\alpha}\right] \\ &= \frac{\|x\|^\nu K_\nu(\kappa\|x\|)}{(4\pi)^{d/2} 2^{\nu-1} \kappa^\nu \Gamma(\alpha)} \\ &= \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} (\kappa\|x\|)^\nu K_\nu(\kappa\|x\|) \end{aligned} \tag{70}$$

□

Here we show some numerical cases of Matérn Field in Fig. 4. It is known that ν and ℓ controls the regularity and the correlated length.

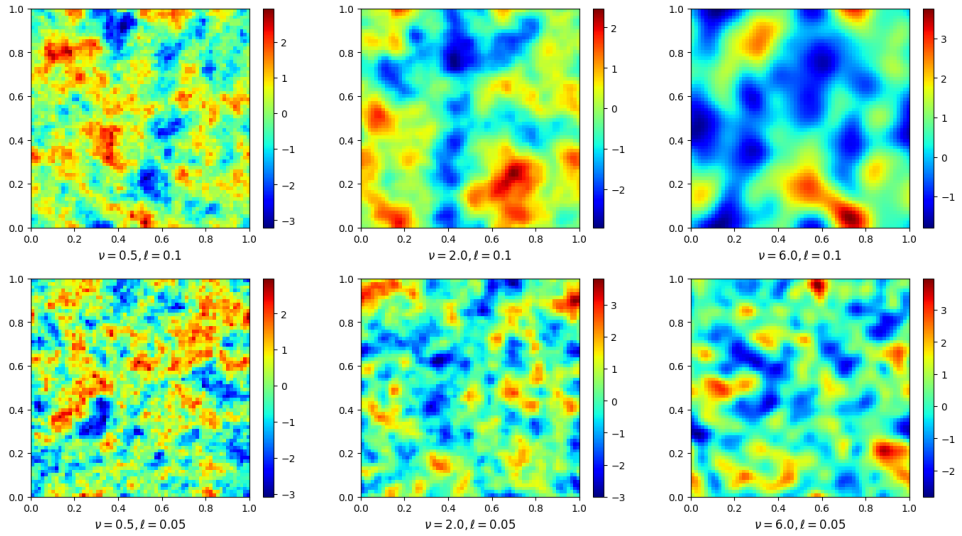


Figure 4: Matern Fields with different ν, ℓ

The smoothness parameter ν controls the regularity of the random field. Larger values of ν correspond to smoother sample paths, for which the spectral energy decays more rapidly with respect to frequency, resulting in lower energy content in the high-frequency regime. Conversely, smaller values of ν lead to rougher fields with increased high-frequency components and more pronounced local fluctuations. The correlation length parameter ℓ governs the spatial dependence structure of the field. Larger values of ℓ imply stronger correlations between distant points, giving rise to coherent structures over larger spatial regions, whereas smaller values of ℓ corresponds to rapidly decaying correlations and nearly independent local behavior. From a spectral perspective, ℓ primarily determines the frequency scale at which the spectral energy begins to decay. In Fig. 5, the spectral energy of different Matér fields is presented. It can be observed that a smaller correlation length ℓ leads to a faster decay of the covariance function, whereas stronger regularity ν results in a flatter behavior of the covariance function near zero.

Remark on anisotropic noise measures. The proposed framework can also accommodate anisotropic source noise measures. For example, one may construct Gaussian random fields with direction-dependent spectral decay or correlation lengths to match anisotropic target measures. However, such choices typically require additional problem-specific prior knowledge about dominant directions or anisotropic correlations. To keep the method broadly applicable and avoid introducing benchmark-specific design choices, we restrict our experiments to isotropic Matérn-type noise measures. Extending the noise design principle to anisotropic random fields is an interesting direction for future work.

C.3 Solutions of SPDEs

It should be noted that, in the case of stationary SPDEs, the learned SPDE solver effectively functions as an operator mapping between function spaces under randomness. For example, a stationary SPDE can be formulated as follows. Assume that $f \in L^2(\Omega, L^2(D))$ is a random field distributed according to the measure μ. The objective is to find $u : D \times \Omega \to \mathbb{R}$ in the weak sense such that, $\mathbb{P}$-almost surely:

$$\begin{cases} \mathcal{L}\left(u(x,w)\right) = f(x,w), & x \in D, \\ u(x,w) = g(x), & x \in \partial D, \end{cases} \tag{71}$$

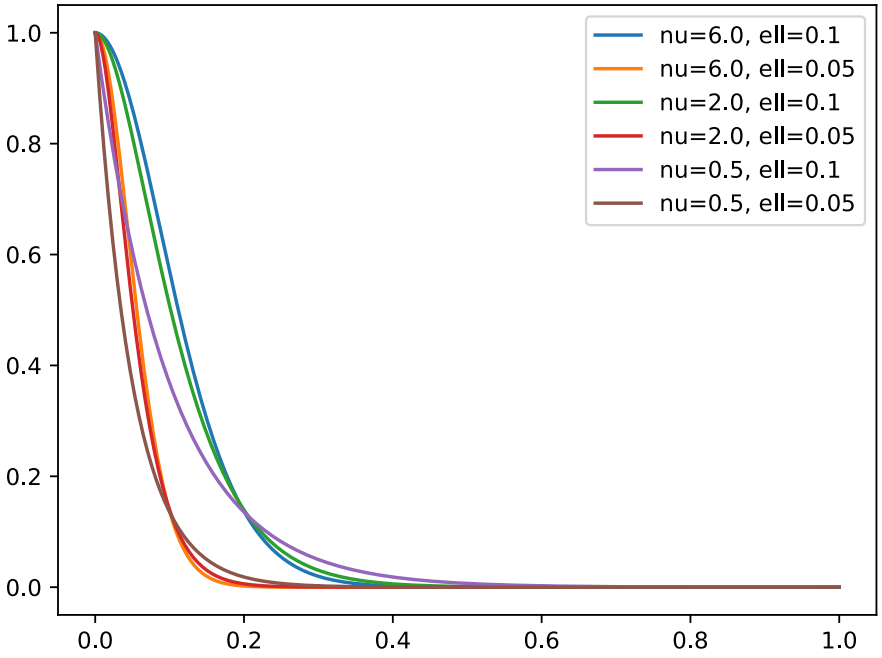


Figure 5: Covariance of Matern Fields with different ν, ℓ

where $\mathcal{L}$ is a stationary differential operator. Therefore, stationary SPDEs are conceptually aligned with neural operators in that both aim to learn mappings between function spaces.

A function $u : [0, T] \to \mathcal{H}_U$ is referred to as a *mild solution* of the semilinear evolution equation if it satisfies the following integral equation:

$$u(t) = S(t)u_0 + \int_0^t S(t-s)\big[F(u_s)\,\mathrm{d}s + G(u_s)\,\mathrm{d}W_s\big], \tag{72}$$

where $S(t) = e^{t\mathcal{L}}$ is the semigroup generated by $\mathcal{L}$. The following theorem provides the theoretical guarantee of existence and uniqueness for solutions to SPDEs.

Theorem C.19. *Assuming the global Lipschitz regularity on F and G are satisfied, the mild solution to Eq. (4) exists and is unique.*

Usually, we care about the asymptotic properties of solutions of stochastic equations as time goes to infinity [Da Prato and Zabczyk, 2014].

Theorem C.20. *Under the assumption of Theorem C.19, and the following dissipativity condition: there exists $C > 0$ such that $\forall x, y \in \mathcal{H}$, $n \in \mathbb{N}$,*

$$2\langle \mathcal{L}_n(x-y), x-y\rangle + 2\langle F(x) - F(y), x-y\rangle + \|G(x) - G(y)\|^2 \le -C\|x-y\|^2, \tag{7}$$

where $\mathcal{L}_n = n\mathcal{L}(n-\mathcal{L})^{-1}$ denotes the Yosida approximation of $\mathcal{L}$. Then there exists a unique invariant measure μ associated with Eq. (4).

D Proofs of Main Theoretical Results

D.1 Proof of Theorem 2.4

Proof. Let $(\mathcal{G}, \mathcal{B}(\mathcal{G}))$ and $(\mathcal{U}, \mathcal{B}(\mathcal{U}))$ be standard Borel spaces. For each $t \in [0, 1]$, let Γ_t be a joint probability measure on $\mathcal{G} \times \mathcal{U}$ with first marginal η_t on $\mathcal{G}$ and second marginal μ_1 on $\mathcal{U}$. By the

disintegration theorem, there exists a regular conditional distribution $\Gamma_t(\mathrm{d}g \mid f)$ such that

$$\Gamma_t(\mathrm{d}g, \mathrm{d}f) = \Gamma_t(\mathrm{d}g \mid f)\, \mu_1(\mathrm{d}f). \tag{73}$$

Assume that for μ_1-a.e. $f \in \mathcal{U}$,

$$\Gamma_t(\cdot \mid f) \ll \eta_t. \tag{74}$$

Then the Radon–Nikodym theorem gives a measurable density

$$\rho_t(g, f) := \frac{\mathrm{d}\Gamma_t(\cdot \mid f)}{\mathrm{d}\eta_t}(g), \tag{75}$$

so that

$$\Gamma_t(\mathrm{d}g \mid f) = \rho_t(g, f)\, \eta_t(\mathrm{d}g) \qquad \text{for } \mu_1\text{-a.e. } f. \tag{76}$$

Let $v_t(g \mid f)$ be a conditional velocity field satisfying the integrability condition

$$\int_0^1 \int_{\mathcal{U}} \int_{\mathcal{G}} \|v_t(g \mid f)\|^2\, \Gamma_t(\mathrm{d}g \mid f)\mu_1(\mathrm{d}f)\mathrm{d}t < \infty. \tag{77}$$

Define the marginal velocity field by

$$v_t(g) := \int_{\mathcal{U}} v_t(g \mid f)\rho_t(g, f)\, \mu_1(\mathrm{d}f), \qquad \eta_t\text{-a.e. } g. \tag{78}$$

The above integral is understood in the Bochner sense.

For any bounded measurable test function $\varphi : \mathcal{G} \to \mathbb{R}$, Fubini's theorem gives

$$\begin{aligned} \int_{\mathcal{G}} \varphi(g) v_t(g)\, \eta_t(\mathrm{d}g) &= \int_{\mathcal{U}} \int_{\mathcal{G}} \varphi(g) v_t(g \mid f)\rho_t(g, f)\, \eta_t(\mathrm{d}g)\mu_1(\mathrm{d}f) \\ &= \int_{\mathcal{U}} \int_{\mathcal{G}} \varphi(g) v_t(g \mid f)\, \Gamma_t(\mathrm{d}g \mid f)\mu_1(\mathrm{d}f). \end{aligned} \tag{79}$$

Therefore $v_t(g)$ is the conditional average of $v_t(g \mid f)$ under the disintegrated measure, and admits the representation

$$v_t(g) = \int_{\mathcal{U}} v_t(g \mid f) \frac{\mathrm{d}\Gamma_t(\cdot \mid f)}{\mathrm{d}\eta_t}(g)\, \mu_1(\mathrm{d}f), \qquad \eta_t\text{-a.e. } g. \tag{80}$$

This proves the claim. □

D.2 Proof of Theorem 2.5

Proof. Let η_t denote the marginal measure induced by the mixture

$$\eta_t(A) = \int_{\mathcal{U}} \Gamma_t(A \mid f)\, \mu_1(\mathrm{d}f), \qquad A \in \mathcal{B}(\mathcal{G}). \tag{81}$$

By assumption, there exists a measurable set $G \subset \mathcal{U}$ with $\mu_1(G) = 1$ such that for every $f, h \in G$,

$$\Gamma_t(\cdot \mid f) \sim \Gamma_t(\cdot \mid h). \tag{82}$$

We prove that $\Gamma_t(\cdot \mid f) \ll \eta_t$ for every $f \in G$. Let $A \in \mathcal{B}(\mathcal{G})$ satisfy $\eta_t(A) = 0$. Since $\mu_1(G) = 1$, the mixture identity gives

$$0 = \eta_t(A) = \int_G \Gamma_t(A \mid h)\, \mu_1(\mathrm{d}h). \tag{83}$$

Because the integrand is nonnegative, it follows that

$$\Gamma_t(A \mid h) = 0 \tag{84}$$

for μ_1-a.e. $h \in G$. Choose one such $h_0 \in G$. For any fixed $f \in G$, the equivalence $\Gamma_t(\cdot \mid f) \sim \Gamma_t(\cdot \mid h_0)$ implies

$$\Gamma_t(A \mid f) = 0. \tag{85}$$

Hence $\Gamma_t(\cdot \mid f) \ll \eta_t$ for every $f \in G$. This proves the theorem. □

D.3 Proof of Theorem 3.1

Proof. We present the proof on the periodic domain $\mathbb{T}^d$ to avoid irrelevant boundary issues. The same argument extends to bounded domains after replacing Fourier modes by the eigenfunctions of an appropriate elliptic operator.

Let u be a centered stationary Matérn Gaussian field with covariance spectrum

$$\lambda_k \asymp \left(\kappa^2 + 4\pi^2|k|^2\right)^{-(\nu+d/2)}, \qquad k \in \mathbb{Z}^d, \tag{86}$$

where $\kappa > 0$ and $\nu > 0$. Then u admits the Fourier representation

$$u = \sum_{k\in\mathbb{Z}^d} \sqrt{\lambda_k}\, \xi_k e_k, \qquad \xi_k \sim \mathcal{N}(0,1), \tag{87}$$

where $\{e_k\}_{k\in\mathbb{Z}^d}$ is the Fourier basis.

For $s \in \mathbb{R}$, the Sobolev norm on $\mathbb{T}^d$ is

$$\|u\|_{H^s}^2 = \sum_{k\in\mathbb{Z}^d} (1+|k|^2)^s |\widehat{u}_k|^2. \tag{88}$$

Taking expectation and using the covariance spectrum gives

$$\begin{aligned}\mathbb{E}\|u\|_{H^s}^2 &= \sum_{k\in\mathbb{Z}^d} (1+|k|^2)^s \lambda_k \\ &\asymp \sum_{k\in\mathbb{Z}^d} (1+|k|^2)^s \left(\kappa^2 + 4\pi^2|k|^2\right)^{-(\nu+d/2)}.\end{aligned} \tag{89}$$

For large $|k|$, the summand behaves like

$$|k|^{2s-2\nu-d}. \tag{90}$$

The series over $\mathbb{Z}^d$ converges if and only if

$$2s - 2\nu - d < -d, \tag{91}$$

which is equivalent to

$$s < \nu. \tag{92}$$

Thus,

$$\mathbb{E}\|u\|_{H^s}^2 < \infty \qquad \text{for all } s < \nu. \tag{93}$$

Since $\|u\|_{H^s}^2$ is a nonnegative random variable, this implies

$$u \in H^s \qquad \text{a.s. for all } s < \nu. \tag{94}$$

It remains to identify the Cameron–Martin space. Let $\mu = \mathcal{N}(0, C)$ be the Gaussian measure induced by u. The Cameron–Martin space is

$$\mathcal{H}_{\mathrm{CM}} = \mathrm{Ran}(C^{1/2}), \qquad \|h\|_{\mathcal{H}_{\mathrm{CM}}} = \|C^{-1/2}h\|_{L^2}. \tag{95}$$

In Fourier variables, C acts by multiplication with λ_k. Therefore

$$\|h\|_{\mathcal{H}_{\mathrm{CM}}}^2 = \sum_{k\in\mathbb{Z}^d} \lambda_k^{-1} |\widehat{h}_k|^2. \tag{96}$$

Using (86),

$$\begin{aligned}\|h\|_{\mathcal{H}_{\mathrm{CM}}}^2 &\asymp \sum_{k\in\mathbb{Z}^d} \left(\kappa^2 + 4\pi^2|k|^2\right)^{\nu+d/2} |\widehat{h}_k|^2 \\ &\asymp \|h\|_{H^{\nu+d/2}}^2.\end{aligned} \tag{97}$$

Hence

$$\mathcal{H}_{\mathrm{CM}} = H^{\nu+d/2} \tag{98}$$

as sets, with equivalent norms. This completes the proof. □

D.4 Proof of Theorem 3.2

Proof. Let g, f be independent centered Gaussian random elements in a separable Hilbert space $\mathcal{H}$ with covariance operators Q_0 and Q_1, respectively. Let

$$y := f_t = \alpha_t g + \beta_t f, \qquad \dot{f}_t = \dot{\alpha}_t g + \dot{\beta}_t f. \tag{99}$$

Since (g, f) is jointly Gaussian and $(y, \dot{f}_t)$ is obtained from (g, f) by a linear transformation, $(y, \dot{f}_t)$ is jointly Gaussian in $\mathcal{H} \times \mathcal{H}$. Therefore the conditional expectation $\mathbb{E}[\dot{f}_t \mid y]$ is linear in y:

$$\mathbb{E}[\dot{f}_t \mid y] = \mathrm{Cov}(\dot{f}_t, y)\,\mathrm{Cov}(y)^{-1} y, \tag{100}$$

where the inverse is understood on the range of $\mathrm{Cov}(y)$.

Using independence of g and f, we compute

$$\mathrm{Cov}(y) = \alpha_t^2 Q_0 + \beta_t^2 Q_1, \tag{101}$$

$$\mathrm{Cov}(\dot{f}_t, y) = \dot{\alpha}_t \alpha_t Q_0 + \dot{\beta}_t \beta_t Q_1. \tag{102}$$

Thus,

$$v_t(x) := \mathbb{E}[\dot{f}_t \mid f_t = x] = \left(\dot{\alpha}_t \alpha_t Q_0 + \dot{\beta}_t \beta_t Q_1\right) \left(\alpha_t^2 Q_0 + \beta_t^2 Q_1\right)^{-1} x. \tag{103}$$

In the simultaneously diagonalizable setting of the theorem, the relative covariance operator $Q_1 Q_0^{-1}$ acts modewise on the relevant spectral subspace. Thus (103) can be written as

$$v_t(x) = V(t)x, \tag{104}$$

with

$$V(t) = \left(\dot{\alpha}_t \alpha_t I + \dot{\beta}_t \beta_t Q_1 Q_0^{-1}\right) \left(\alpha_t^2 I + \beta_t^2 Q_1 Q_0^{-1}\right)^{-1}. \tag{105}$$

This is the claimed formula. □

D.5 Proof of Theorem 3.3

Proof. Let $f \sim \mu$, and denote by $\widehat{f_k}$ its Fourier coefficients. By definition of the spectral density,

$$\mathbb{E}|\widehat{f_k}|^2 = S_d(k). \tag{106}$$

For $s \in \mathbb{R}$, the squared Sobolev norm satisfies

$$\mathbb{E}\|f\|_{H^s}^2 = \sum_{k \in \mathbb{Z}^d} (1 + |k|^2)^s S_d(k). \tag{107}$$

Since $S_d(k) \asymp |k|^{-\beta}$ as $|k| \to \infty$, the high-frequency part of the above series is comparable to

$$\sum_{k \in \mathbb{Z}^d} |k|^{2s-\beta}. \tag{108}$$

Using the standard lattice-shell estimate

$$\#\{k \in \mathbb{Z}^d : n \le |k| < n+1\} \asymp n^{d-1},$$

this series has the same convergence behavior as

$$\sum_{n=1}^{\infty} n^{d-1+2s-\beta}. \tag{109}$$

Therefore it converges if and only if

$$d - 1 + 2s - \beta < -1, \tag{110}$$

or equivalently,

$$s < \frac{\beta - d}{2}. \tag{111}$$

Hence, for every $s < s_{\mathrm{data}}$, where

$$s_{\mathrm{data}} := \frac{\beta - d}{2},$$

we have

$$\mathbb{E}\|f\|_{H^s}^2 < \infty. \tag{112}$$

In particular,

$$f \in H^s \qquad \mu\text{-a.s. for every } s < s_{\mathrm{data}}. \tag{113}$$

In the Gaussian case with nondegenerate Fourier modes, the converse also holds: $f \notin H^s$ almost surely for $s \geq s_{\mathrm{data}}$. Thus s_{data} is the sharp Sobolev regularity threshold of the target field.

Now consider the source Matérn Gaussian noise with smoothness parameter ν_0. Its Cameron–Martin space is

$$\mathcal{H}_{\mathrm{CM}}^{(0)} = H^{\nu_0 + d/2}. \tag{114}$$

To guarantee that typical target samples belong to the source Cameron–Martin space, it is enough to require

$$\nu_0 + \frac{d}{2} < s_{\mathrm{data}}. \tag{115}$$

Indeed, under this condition, (113) gives

$$f \in H^{\nu_0 + d/2} = \mathcal{H}_{\mathrm{CM}}^{(0)} \qquad \mu\text{-a.s.}$$

Substituting $s_{\mathrm{data}} = (\beta - d)/2$ yields

$$\nu_0 + \frac{d}{2} < \frac{\beta - d}{2}. \tag{116}$$

Equivalently,

$$\nu_0 < \frac{\beta}{2} - d. \tag{117}$$

This proves the claimed condition. □

D.6 Proof of Theorem 3.4

Proof. Let μ_k denote the positive spectral covariance ratio between the target and source measures at Fourier mode k. For the linear interpolation, write

$$q_k(t) = (1-t)^2 + \mu_k t^2, \qquad \widetilde{V}_k(t) = \frac{q_k'(t)}{2q_k(t)}.$$

The extrema of $\widetilde{V}_k$ on $[0,1]$ give

$$\sup_{t\in[0,1]} |\widetilde{V}_k(t)| = \frac{1}{2}\left(\sqrt{\mu_k} + \frac{1}{\sqrt{\mu_k}}\right). \tag{118}$$

Indeed, when $0 < \mu_k < 1$, the minimum occurs at $t = (1 - \sqrt{\mu_k})/(1 + \mu_k)$ and equals the negative of the right-hand side of (118); the maximum is 1 at $t = 1$. When $\mu_k > 1$, the minimum is -1 at $t = 0$, while the maximum occurs at $t = (1 + \sqrt{\mu_k})/(1 + \mu_k)$ and equals the right-hand side. For $\mu_k = 1$, the maximum absolute value is 1 at either endpoint. Fix a stiffness threshold $L > 1$. The uniform magnitude bound

$$|\widetilde{V}_k(t)| < L, \qquad t \in [0,1],$$

is equivalent to

$$\sqrt{\mu_k} + \frac{1}{\sqrt{\mu_k}} < 2L.$$

Setting $z_k = \sqrt{\mu_k} > 0$, this becomes

$$z_k + z_k^{-1} < 2L.$$

Solving the quadratic inequality gives

$$L - \sqrt{L^2 - 1} < z_k < L + \sqrt{L^2 - 1}.$$

Therefore

$$C_L < \mu_k < C_L^{-1}, \qquad C_L := \left(L - \sqrt{L^2 - 1}\right)^2. \tag{119}$$

Here we used

$$\left(L + \sqrt{L^2 - 1}\right)^2 = C_L^{-1}.$$

Now assume that both the target and source are Matérn-type Gaussian measures with the same length-scale parameter κ. Denote their smoothness parameters by ν_1 and ν_0, respectively, and let $r_0 = \mathcal{F}_0(Q_1)/\mathcal{F}_0(Q_0) > 0$. Their spectral covariance ratio takes the form

$$\mu_k = r_0 \left(\frac{\kappa^2}{\kappa^2 + 4\pi^2\|k\|^2}\right)^{\nu_1 - \nu_0}. \tag{120}$$

By the Cameron–Martin admissibility condition, we choose

$$\nu_0 < \nu_{\mathrm{adm}} = \nu_1 - \frac{d}{2}.$$

In particular, $\nu_0 < \nu_1$. Hence $\nu_1 - \nu_0 > 0$, and μ_k is decreasing with respect to $\|k\|$.

For all resolved modes $\|k\| \le N$, define

$$M_N := \frac{\kappa^2}{\kappa^2 + 4\pi^2 N^2} \in (0, 1).$$

Since μ_k is minimized at the largest resolved frequency, we have

$$\mu_k \ge r_0 M_N^{\nu_1 - \nu_0}, \qquad \|k\| \le N. \tag{121}$$

Therefore, the lower bound $\mu_k > C_L$ in (119) is guaranteed by

$$r_0 M_N^{\nu_1 - \nu_0} > C_L. \tag{122}$$

It remains to control the upper bound $\mu_k < C_L^{-1}$. Since μ_k is largest at the zero mode, it is sufficient to impose

$$\mu_0 = r_0 < C_L^{-1}. \tag{123}$$

Combining (122) and (123) gives

$$C_L < \mu_k < C_L^{-1}, \qquad \|k\| \le N.$$

By (119), this implies

$$|\widetilde{V}_k(t)| < L, \qquad t \in [0, 1], \quad \|k\| \le N.$$

Finally, the Cameron–Martin admissibility condition from Theorem 3.3 requires

$$\nu_0 < \nu_{\mathrm{adm}} = \frac{\beta}{2} - d.$$

Choosing

$$\nu_0 < \nu_{\mathrm{adm}} - \varepsilon$$

with $\varepsilon > 0$ enforces this condition strictly. Therefore any ν_0 satisfying

$$0 < \nu_0 < \nu_{\mathrm{adm}} - \varepsilon, \qquad r_0 M_N^{\nu_1 - \nu_0} > C_L, \qquad r_0 < C_L^{-1}$$

is both Cameron–Martin admissible and satisfies the prescribed spectral stiffness bound. This proves the claim. □

D.7 Proof of Eq. 14

Proof. Let H be a separable Hilbert space. Assume $f, g \in L^2(\Omega; H)$ and, for each $t \in [0,1]$,

$$f_t = \alpha_t g + \beta_t f, \tag{124}$$

where α_t, β_t are deterministic C^1 scalar functions. Assume further that

$$\dot{\alpha}_t \neq 0 \tag{125}$$

and

$$c_t := \beta_t - \dot{\beta}_t \frac{\alpha_t}{\dot{\alpha}_t} \neq 0. \tag{126}$$

The conditional velocity is

$$v_t(f_t \mid f) := \frac{\mathrm{d}}{\mathrm{d}t} f_t = \dot{\alpha}_t g + \dot{\beta}_t f, \tag{127}$$

where the derivative is understood in $L^2(\Omega; H)$.

From (127), we solve for g:

$$g = \frac{1}{\dot{\alpha}_t} \left(v_t(f_t \mid f) - \dot{\beta}_t f \right). \tag{128}$$

Substituting this into (124) yields

$$\begin{aligned} f_t &= \alpha_t \frac{1}{\dot{\alpha}_t} \left(v_t(f_t \mid f) - \dot{\beta}_t f \right) + \beta_t f \\ &= \frac{\alpha_t}{\dot{\alpha}_t} v_t(f_t \mid f) + \left(\beta_t - \dot{\beta}_t \frac{\alpha_t}{\dot{\alpha}_t} \right) f \\ &= \frac{\alpha_t}{\dot{\alpha}_t} v_t(f_t \mid f) + c_t f. \end{aligned} \tag{129}$$

Therefore,

$$f = c_t^{-1} \left(f_t - \frac{\alpha_t}{\dot{\alpha}_t} v_t(f_t \mid f) \right). \tag{130}$$

Taking conditional expectation with respect to $\sigma(f_t)$ gives

$$\mathbb{E}[f \mid f_t] = c_t^{-1} \left(f_t - \frac{\alpha_t}{\dot{\alpha}_t} \mathbb{E}\left[v_t(f_t \mid f) \mid f_t \right] \right). \tag{131}$$

By the definition of the marginal velocity field,

$$v_t(f_t) = \mathbb{E}\left[v_t(f_t \mid f) \mid f_t \right] \qquad \text{a.s.} \tag{132}$$

Hence

$$\mathbb{E}[f \mid f_t] = c_t^{-1} \left(f_t - \frac{\alpha_t}{\dot{\alpha}_t} v_t(f_t) \right), \tag{133}$$

which proves Eq. 14. □

E Algorithm of NAGO

Why spectral analysis matters for scientific data. For data generated by PDEs and SPDEs, the frequency domain provides a natural description of both physical structure and numerical difficulty. The decay of spectral energy reflects the regularity of the underlying random fields, while the distribution of energy across modes determines how much large-scale and small-scale variation must be transported by the generative flow. This is particularly important in function-space generation: a source noise distribution with excessive high-frequency energy may look harmless after discretization, but it can induce a spectrally stiff marginal velocity field and cause resolution-dependent artifacts. Conversely, a generated sample can have a small pointwise error while still failing to reproduce the correct multiscale energy distribution. We therefore use spectral diagnostics as a central evaluation tool. They allow us to test whether the learned model preserves the target measure across spatial scales and whether the proposed regularity-aware noise design improves stability beyond visual sample quality or grid-based errors alone.

Algorithm 1 (Flow Matching Training). Algorithm 1 summarizes the flow-matching training procedure under a prescribed interpolation schedule. Given a dataset $\mathcal{D}$ of samples from the target distribution, scalar schedules α_t and β_t, and a parametric velocity field $v_\theta(\cdot, t)$, the objective is to learn a time-dependent vector field that transports samples from a source noise distribution to the target distribution. Prior to training, we estimate the spectral decay rate β_{spec} of the data and select a source noise distribution π accordingly. This step implements the proposed noise-aware principle: the source noise should be no smoother than the target distribution, while avoiding excessive high-frequency energy that would induce spectrally stiff transport dynamics.

During each iteration, a mini-batch $\{x_i\}_{i=1}^m$ is drawn from the dataset, and a matching batch of noise samples $\{\varepsilon_i\}_{i=1}^m$ is drawn independently from the chosen noise prior π. Time stamps $\{t_i\}_{i=1}^m$ are independently sampled from $\mathrm{Unif}(0,1)$, so that the model is trained across the full interpolation horizon. For each triplet $(x_i, \varepsilon_i, t_i)$, the interpolated state and the corresponding target velocity are defined by

$$x_{t_i} = \alpha_{t_i}\varepsilon_i + \beta_{t_i}x_i, \qquad u_{t_i} = \dot{\alpha}_{t_i}\varepsilon_i + \dot{\beta}_{t_i}x_i. \tag{134}$$

The network prediction $\hat{u}_{t_i} = v_\theta(x_{t_i}, t_i)$ is trained by minimizing

$$\mathcal{L}(\theta) = \frac{1}{m}\sum_{i=1}^{m} \|\hat{u}_{t_i} - u_{t_i}\|_{\mathcal{H}}^2 . \tag{135}$$

In the discrete implementation, the Hilbert-space norm is evaluated by its grid-based approximation. Repeating this procedure until convergence yields a trained velocity field for unconditional generation and guided sampling.

Algorithm 1 Flow Matching Training

Input: dataset $\mathcal{D}$, schedules α_t, β_t, model v_θ, learning rate η.
Spectral decay rate β analysis of data.
Decide noise distribution π by β.
repeat
  Sample mini-batch $\{x_i\}_{i=1}^m \sim \mathcal{D}$
  Sample noise $\{\epsilon_i\}_{i=1}^m \sim \pi$
  Sample times $\{t_i\}_{i=1}^m \sim \mathrm{Unif}(0,1)$
  for $i = 1$ **to** m **do**
    $x_{t_i} \leftarrow \alpha_{t_i}\epsilon_i + \beta_{t_i}x_i$
    $u_{t_i} \leftarrow \dot{\alpha}_{t_i}\epsilon_i + \dot{\beta}_{t_i}x_i$
    $\hat{u}_{t_i} \leftarrow v_\theta(x_{t_i}, t_i)$
  end for
  $\mathcal{L}(\theta) \leftarrow \frac{1}{m}\sum_{i=1}^m \|\hat{u}_{t_i} - u_{t_i}\|_2^2$
  $\theta \leftarrow \theta - \eta\nabla_\theta\mathcal{L}(\theta)$
  Update model parameters
until convergence

Algorithm 2 (Guided Flow Matching Sampling with DPS). Algorithm 2 describes the guided sampling procedure based on the trained velocity field $v_\theta(\cdot, t)$, where external observations are incorporated through a data-consistency term. Let A denote a measurement operator and let y be the observed measurement corrupted by noise of level σ_y. The goal is to generate a sample that remains plausible under the learned generative prior while being consistent with the observation model.

The procedure starts from an initial state $x \sim \pi$ drawn from the same noise prior used at training time. The state is evolved from $t = 0$ to $t = 1$ using a numerical ODE solver with step size Δt over N steps. At each step, the base velocity $v_\theta(x, t)$ is computed and a guidance correction is constructed following the DPS principle. An estimate $\hat{x}_1$ of the terminal (denoised) sample is formed from the current state (depending on the chosen schedule/parameterization), and a gradient signal is computed to reduce the mismatch between $A\hat{x}_1$ and y. The guidance gradient takes the form

$$g \;\propto\; \sigma\,\nabla_x\,\Phi\big(A\hat{x}_1,\, y\big),$$

where Φ denotes a data-fidelity loss (e.g., $\|A\hat{x}_1 - y\|_2^2$) and σ is the guidance strength. The update direction is obtained by combining the base velocity and the guidance correction, after which the state is advanced by one solver step. After N iterations, the resulting state is returned as the guided sample.

Algorithm 2 Guided Flow Matching Sampling with DPS Guidance

Input: trained velocity field $v_\theta(\cdot, t)$, noise prior π, measurement operator A, observation y, noise level σ_y, step size Δt, number of steps N
Sample initial state $x \sim \pi$
Set $t \leftarrow 0$
for $n = 1$ **to** N **do**
  Estimate $\hat{x}_1$ at $t = 1$.
  Compute DPS gradient: $g \leftarrow \sigma \nabla_x \Phi\big(A\hat{x}_1, y\big)$.
  Compute base velocity: $v \leftarrow v_\theta(x, t)$.
  Update state x by neural ode.
  Update time: $t \leftarrow t + \Delta t$
end for
Output: guided sample x

F Experiments Details

This section provides the experimental details needed to reproduce and interpret the reported results. All experiments were conducted on a single NVIDIA A40 GPU with 48 GB of memory. We document the full data preprocessing pipeline, model architecture, optimizer settings, learning-rate schedule, batch size, and training duration. Unless otherwise stated, all comparisons are performed under the same experimental configuration across different noise choices. In particular, we keep the backbone architecture, training budget, optimizer, and evaluation protocol fixed, and vary only the noise formulation. This controlled setup allows us to isolate the effect of noise design on training stability, sampling efficiency, and generation quality.

We also include ablation studies to examine the sensitivity of the proposed method to key hyperparameters, including the noise smoothness, correlation length, regularization strength, and sampling-step configuration. These experiments are designed to identify parameter regimes in which the method remains stable and to assess whether the observed improvements are robust to reasonable hyperparameter changes. The complete experimental results are reported in the appendix. In addition to the summary results presented in the main text, we provide per-metric comparisons, convergence curves, run-to-run variability, and qualitative visualizations. Where appropriate, we also include diagnostic plots, such as training loss curves and stability indicators across samples or time steps, to give a more complete picture of model behavior.

Overall, the results show that Matérn noise provides consistent advantages over Gaussian noise under matched experimental settings. During training, it leads to faster convergence and reduced sensitivity to hyperparameter choices. During inference, it improves numerical stability and produces more reliable samples, with fewer resolution-dependent artifacts and unstable trajectories. These findings support the central claim of this work: regularity-aware noise design improves both the optimization behavior and the inference-time robustness of functional generative models for multiscale SPDE data.

F.1 Baselines

Our baselines are constructed from standard choices commonly used in flow-matching generative models, but we report them according to their two key components: the source noise distribution and the neural backbone. This notation makes the comparison more transparent for our purpose, since the central question of this work is how the choice of noise affects training and sampling under otherwise matched settings. For example, "Gaussian + FNO", also known as DiffusionPDE [Huang et al., 2024] denotes a standard flow-matching model using Gaussian source noise and an FNO backbone, which is also known as FFM [Kerrigan et al., 2024], while "Matérn + FNO" replaces only the source noise with the proposed regularity-aware Matérn noise.

We use the Fourier Neural Operator (FNO) as the primary backbone because it is a widely used operator-learning architecture for PDE data and can efficiently capture global interactions through spectral representations. To test whether the observed improvements are tied to this specific architecture, we also evaluate U-Net as an alternative grid-based backbone. When U-Net is used, the model is instantiated as a conventional flow-matching model on Euclidean grids and is therefore treated as a standard baseline rather than as part of our operator-based formulation. Across all backbone choices,

we keep the optimizer, training budget, data preprocessing, interpolation schedule, sampling protocol, and evaluation metrics fixed unless otherwise stated. Thus, performance differences can be attributed primarily to the source noise design and, secondarily, to the chosen backbone.

Several related functional generative modeling frameworks are discussed for context. In particular, Functional Flow Matching (FFM) is closely related to our setting at a conceptual level, but its published experiments do not provide quantitative comparisons or systematic analyses of source-noise selection. We therefore include FFM as a conceptual reference rather than a direct empirical baseline.

Table 4: Summary of baseline methods.

Method	Model	Noise
Vanilla FM [Lipman et al., 2023]	U-Net	Gaussian
FFM [Kerrigan et al., 2024]	FNO	Gaussian
DiffusionPDE [Huang et al., 2024]	FNO	Gaussian
GAN [Goodfellow et al., 2020]	U-Net	–
VAE [Kingma and Welling, 2022]	U-Net	–
GANO [Rahman et al., 2022]	FNO	–
VANO [Seidman et al., 2023]	FNO	–

F.2 Introduction of Datasets

We next introduce the numerical benchmarks used to evaluate the proposed method. The experiments are designed to test whether the choice of source noise affects training stability, sampling efficiency, and resolution consistency when learning invariant measures of multiscale stochastic systems. We consider three representative stochastic-field settings with different spatial structures and spectral properties.

F.2.1 Stochastic Allen Cahn Equation (SAC)

The 1D stochastic Allen–Cahn equation is defined as:

$$\partial_t u = \partial_{xx} u - V'(u) + \sqrt{2}\eta_t, \tag{136}$$

posed on the unit interval $[0, 1]$ with Neumann boundary conditions and driven by space–time white noise η. Here, $V(u) = (1 - u^2)^2$ is a double-well potential. Under standard assumptions, the SPDE (136) admits an invariant distribution with formal density

$$\exp\left(- \int_0^1 \left(\tfrac{1}{2}(\partial_x u(x))^2 + V(u(x)) \right) dx \right). \tag{137}$$

For the Allen–Cahn equation, the data distribution is characterized by a path-integral formulation that gives rise to a bimodal invariant measure. Typical realizations of this measure exhibit nearly constant profiles around $u = \pm 1$. Owing to the pronounced nonlinearity inherent in such dynamics, this dataset is considered in our study. The spatial domain is discretized using finite-difference schemes on N equispaced grid points, resulting in an N-dimensional approximation of the invariant distribution. Reference samples x_1 from this distribution are generated using ensemble Markov chain Monte Carlo (MCMC) methods [Chen, 2025].

F.2.2 Stochastic Navier Stokes Equation (SNS)

Here we consider the dynamics of the 2D stochastic Navier-Stokes(SNS) equations with random forcing in the domain $[0, 2\pi]^2$. The vorticity formulation of SNS equations is:

$$d\omega + v\nabla\omega dt = \nu\Delta\omega - \alpha\omega + \sigma dW_t \tag{138}$$

where $v = \nabla^\perp \psi = (-\partial_y \psi,\, \partial_x \psi)$ is the velocity expressed in terms of the stream function ψ, which solves $-\Delta\psi = \omega$ and dW_t denotes white-in-time colored random forcing. The quantities ν, α, and σ are positive parameters. A pseudo-spectral scheme is used to solve Equ. (138). The generation of the stochastic Navier–Stokes (SNS) data is based on the methodology proposed by Chen [2025]. All

training and evaluation data are generated by numerically solving the two-dimensional stochastic Navier–Stokes equations under random external forcing. The viscosity is set to $\nu = 10^{-3}$, with damping and forcing amplitudes given by $\alpha = 0.1$ and $\varepsilon = 1$, respectively.

The stochastic forcing is defined as a finite-dimensional Gaussian process of the form

$$\eta(t, x, y) = \sum_{i=1}^{4} W_i(t)\, \phi_i(x, y) + \sum_{i=5}^{8} W_i(t)\, \phi_i(x, y), \tag{139}$$

where $\{W_i(t)\}_{i=1}^{8}$ are mutually independent standard Wiener processes, and the spatial basis functions are given by

$$\{\sin(6x), \cos(7x), \sin(5(x+y)), \cos(8(x+y)), \cos(6x), \sin(7x), \cos(5(x+y)), \sin(8(x+y))\}.$$

This construction induces a Gaussian random field that is stationary with respect to spatial translations and injects energy at both low and intermediate Fourier modes.

For this class of stochastic forcing, the resulting stochastic Navier–Stokes dynamics admit a unique invariant measure, as established by Hairer and Mattingly [2006]. The linear damping term $-\alpha\omega$ is introduced to suppress excessive energy accumulation at large spatial scales, thereby promoting numerical stability and enhancing ergodic mixing of the flow.

The stochastic partial differential equation is solved using a pseudo-spectral method in space combined with an Euler–Maruyama time integration scheme. All simulations are performed on a uniform 256×256 grid, which provides sufficient resolution for accurately capturing the relevant flow statistics under the considered parameter regime. The system is discretized on a 256×256 spatial grid with a time step of $\Delta t = 10^{-4}$, and vorticity snapshots are recorded every 0.5 time units. In total, 1,000 independent trajectories are simulated over the time interval $t \in [0, 100]$. To ensure convergence to stationarity, data from the initial transient period $t \in [0, 50]$ are discarded. The remaining trajectories yield approximately 10^5 snapshots, which are regarded as approximate samples from the invariant measure of Equ. (138). The main models use snapshots downsampled to 64×64; the 256×256 solver output is retained to construct the resolution-transfer evaluations.

F.3 Experiment 1: Generation at Fixed Resolution

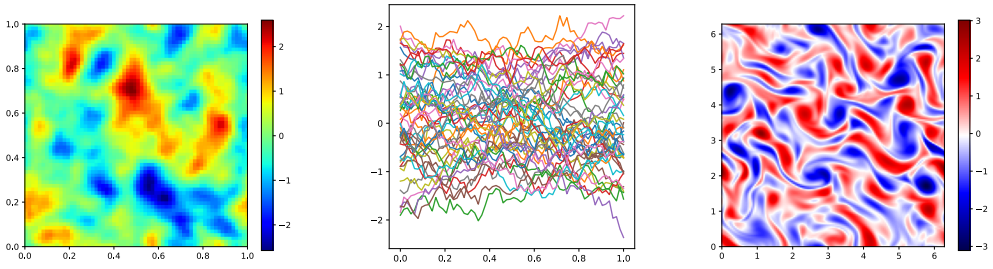
Matérn field. In the Matérn field case, the target measure is a two-dimensional Matérn Gaussian random field, $\mu_1 = \mathcal{N}\big(0,\, (400 - \Delta)^{-3}\big)$ with $(\nu_1, \ell_1) = (2, 0.1)$. The source uses $(\nu_0, \ell_0) = (0.5, 0.05)$, which satisfies the strict admissibility threshold $\nu_0 < 1$ and has the same $\kappa = 20$ as the target. This equal-κ choice is the idealized case in Thm. 3.4; the length-scale ablation in Appendix F.6 evaluates departures from it. Each random field realization is discretized on an $N \times N$ grid with $N = 64$. The resulting neural ODE is integrated using a fourth-order Runge–Kutta (RK4) scheme. To assess robustness, the effects of time-step size and noise magnitude on the generated samples are evaluated systematically, and the corresponding quantitative results are summarized in Table 2.

Architectures and training configuration. In the main experiments, the Fourier Neural Operator (FNO) employs four spectral convolution blocks. We follow the `neuralop` convention in which `n_modes` $= [m_1, m_2]$ is the total number of retained discrete Fourier coefficients per axis, including positive and negative frequencies. Thus $[64, 64]$ covers the full 64×64 Fourier grid rather than 64 modes of each sign. The hidden channel width is 64, the time embedding has dimension 64, and the projection width is 256. For the U-Net backbone, four resolution levels are used with `dim_mults` $= [1, 2, 2, 2]$, an attention head dimension of 32, and four attention heads. Training uses batch size 64, Adam, and learning rate 3×10^{-4}. All random-field data are variance-normalized. Qualitative examples are provided in Fig. 6.

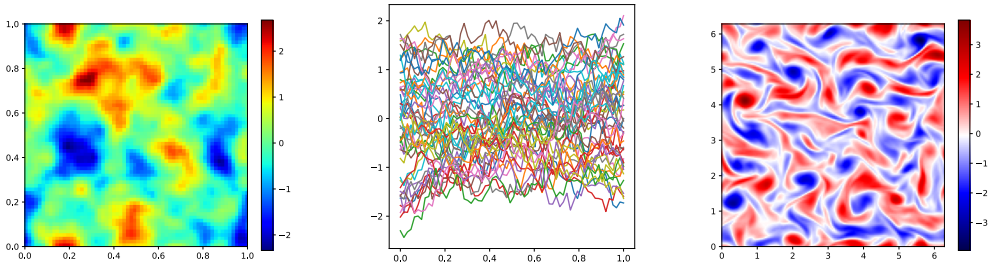
SAC. In the SAC example, the FNO backbone consists of four spectral convolution layers, with the number of retained modes set to 64 and the projection channel width set to 128. Figure 6(b, e) presents representative samples from the ground-truth dataset and the corresponding generated samples.

SNS. In the SNS example, the FNO has four spectral convolution layers, `n_modes` $= [64, 64]$ at the 64×64 training resolution, embedding and hidden widths of 128, and projection width 256. Training uses batch size 100, Adam, and learning rate 3×10^{-4}. Figure 6(c,f) shows a target and a five-step NAGO sample. Timings are reported only under the controlled protocol in Appendix F.10.

We additionally evaluate a grid-based U-Net under the same data and evaluation protocol. Matérn sources improve several U-Net entries in Table 2, showing that the source effect is not exclusive to FNO. Because only two backbones are tested and their optimal hyperparameters may differ, we do not claim architecture-independent superiority.



(a) Matérn: Ground truth (b) SAC: Ground truth (c) SNS: Ground truth



(d) Matérn: Generated by NAGO (e) SAC: Generated by NAGO (f) SNS: Generated by NAGO

Figure 6: Qualitative comparison across three cases: Matérn, SAC, and SNS. The first row shows ground-truth samples, and the second row shows samples generated by NAGO.

Remark F.1 (Non-monotonic error with respect to ODE steps)**.** *In several experiments, we observe that increasing the number of ODE integration steps does not always monotonically improve the generation error. For example, in Table 2, the Matérn FNO error increases when the number of steps is increased from a small value to a larger one, and a similar non-monotonic behavior appears in guided generation.*

This phenomenon does not contradict the standard intuition that smaller ODE time steps reduce numerical discretization error. In a learned flow model, the total sampling error is affected not only by the solver discretization error, but also by the approximation error of the learned marginal velocity field. Increasing the number of integration steps reduces the former, but it also increases the number of evaluations of the learned velocity field. Therefore, small biased errors in the learned field may accumulate along the sampling trajectory. This accumulation can be more significant in spectrally stiff regimes, where high-frequency components are more sensitive to velocity-field errors.

The effect can be further amplified in guided generation. The guidance term modifies the learned velocity field at sampling time and may introduce an additional bias that is not present during training. Although this guidance can improve conditional alignment at moderate step counts, too many integration steps may repeatedly amplify off-manifold or high-frequency components, leading to degraded metrics. Therefore, the optimal number of sampling steps reflects an empirical accuracy–stability trade-off rather than purely the numerical accuracy of the ODE solver.

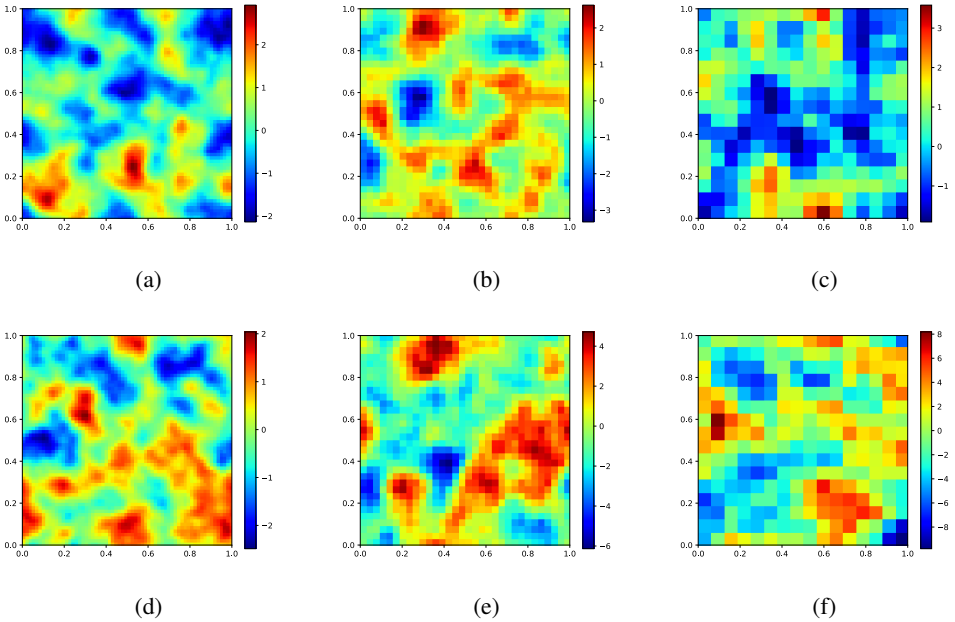
F.4 Experiment 2: Generation at Varying Resolution

Fig. 3 (C3) illustrates the spectral density curves (i.e., spectral energy distributions) estimated from samples generated by different frameworks at multiple spatial resolutions. This experiment is designed to evaluate cross-resolution consistency, namely, whether a model trained under a given discretization can preserve the characteristic frequency content of the target distribution when

generation is performed on coarser or finer grids. For the Matérn field example, NAGO employs a five-step solver during sampling, whereas all other frameworks are configured with solvers using at least ten steps, reflecting their default inference-time settings and ensuring stable generation. In all cases, 1,000 independent samples are generated per resolution to estimate the corresponding spectral energy distributions with sufficient statistical reliability. For each generated sample, a discrete Fourier transform is computed on the associated grid, and the resulting power spectrum is aggregated (e.g., via radial averaging over frequency shells) to obtain a one-dimensional spectral density curve. The curves reported in Fig. 3 (C4) are then obtained by averaging these spectral estimates across samples, which reduces Monte Carlo noise and enables a clearer comparison of frequency-domain discrepancies among methods.

For the SNS benchmark, NAGO again adopts a five-step solver, while FFM uses a ten-step solver, following the same evaluation principle while accounting for the different benchmark characteristics and computational costs. To compute the spectral energy distributions in this setting, the number of generated samples is adjusted according to resolution in order to balance estimation accuracy and runtime: 200 samples are generated at resolution 128, 500 samples at resolution 64, and 1,000 samples at resolution 32. As in the Matérn experiment, spectral densities are computed from the Fourier-domain energy and averaged across the corresponding sample sets. This protocol allows the resolution dependence of the generated frequency content to be assessed in a controlled manner, while maintaining comparable statistical confidence across resolutions.

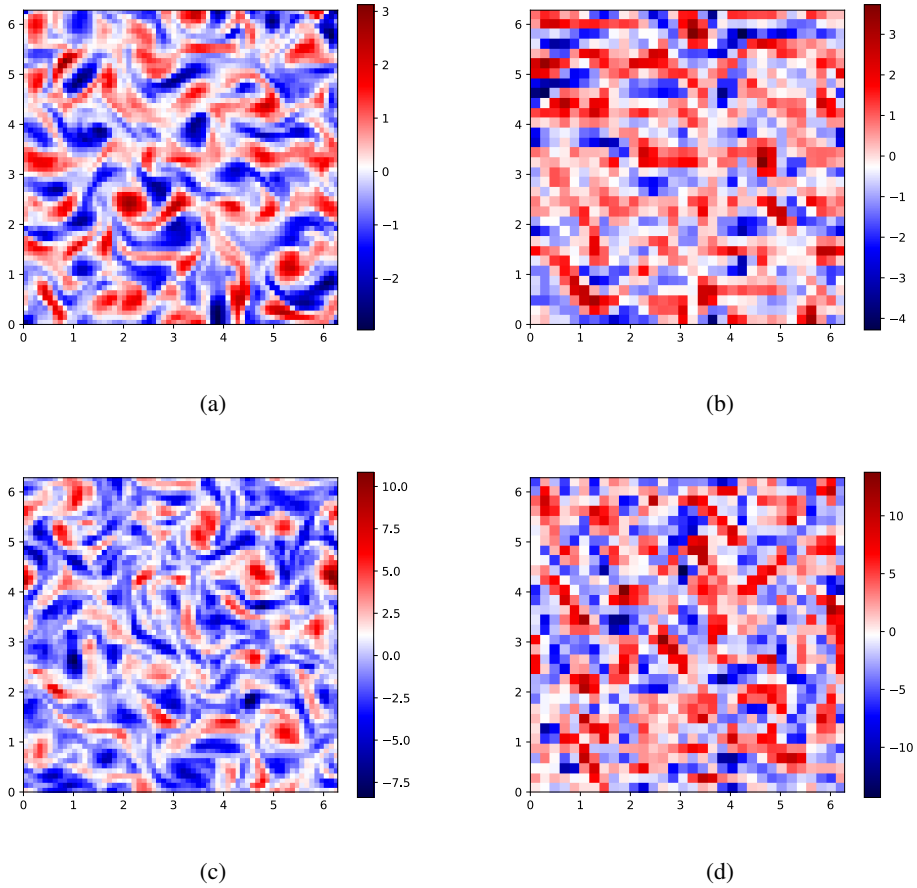
Fig. 7 and Fig. 8 present representative numerical samples of the Matérn field and the SNS benchmark, respectively, generated at different resolutions to complement the spectral-domain analysis with qualitative evidence in the physical domain. Both figures are produced using models trained in Appendix F.3. For generations, NAGO employs five sampling steps, whereas FFM uses ten steps, consistent with the solver configurations used in Fig. 3 (C3). It is observed that under the NAGO framework, both the numerical range (e.g., amplitude scale and distribution of extrema) and the spectral energy remain stable across resolutions, indicating that the generated samples preserve the target distributional characteristics under grid changes. In contrast, under the FFM framework, noticeable drift is observed as the resolution varies, manifested as changes in amplitude scaling and shifts in the frequency content, which collectively suggest a mismatch with the target distribution and reduced cross-resolution robustness.



(a) (b) (c)

(d) (e) (f)

Figure 7: Matern Field generated at different resolutions. Samples generated at resolution 64 (1st column), 32 (2nd column), 16 (3rd column) with (First row) NAGO and (Second row) FFM



(a) (b)

(c) (d)

Figure 8: Matern Field generated at different resolutions.

F.5 Experiment 3: Generation with Guidance

A detailed comparison of generation quality under guided settings is presented for generative frameworks equipped with different noise choices. In this set of experiments, guidance is provided through low-resolution observations, and the resulting conditional generation performance is evaluated both visually and quantitatively to assess how the noise design affects reconstruction fidelity and uncertainty calibration.

Figure 9 reports the guided-generation results for the Matérn Field, while Figure 10 reports the corresponding results for the SNS dataset. In both figures, panel (a) displays the full-resolution ground-truth field, and panel (b) shows the low-resolution observation that is supplied to the generative model as the guidance signal. Panels (c–e) summarize the statistics of samples produced by `NAGO`, including the pixel-wise sample mean, the pixel-wise sample standard deviation, and the reconstruction mean squared error (MSE) with respect to the ground truth. Panels (f–h) provide the same set of statistics for samples generated by FFM, enabling a direct, method-aligned comparison under identical guidance information.

All reported statistical quantities are estimated from 200 independently generated samples per method. The sample mean is used to characterize the central tendency of the conditional distribution, the sample standard deviation is used to quantify uncertainty and spatial variability under guidance, and the reconstruction MSE map is used to localize regions where systematic bias or loss of fine-scale structure is most pronounced. This evaluation protocol is intended to jointly reflect reconstruction accuracy and uncertainty consistency, thereby providing a more comprehensive assessment of guided generation quality than single-sample visualization alone.

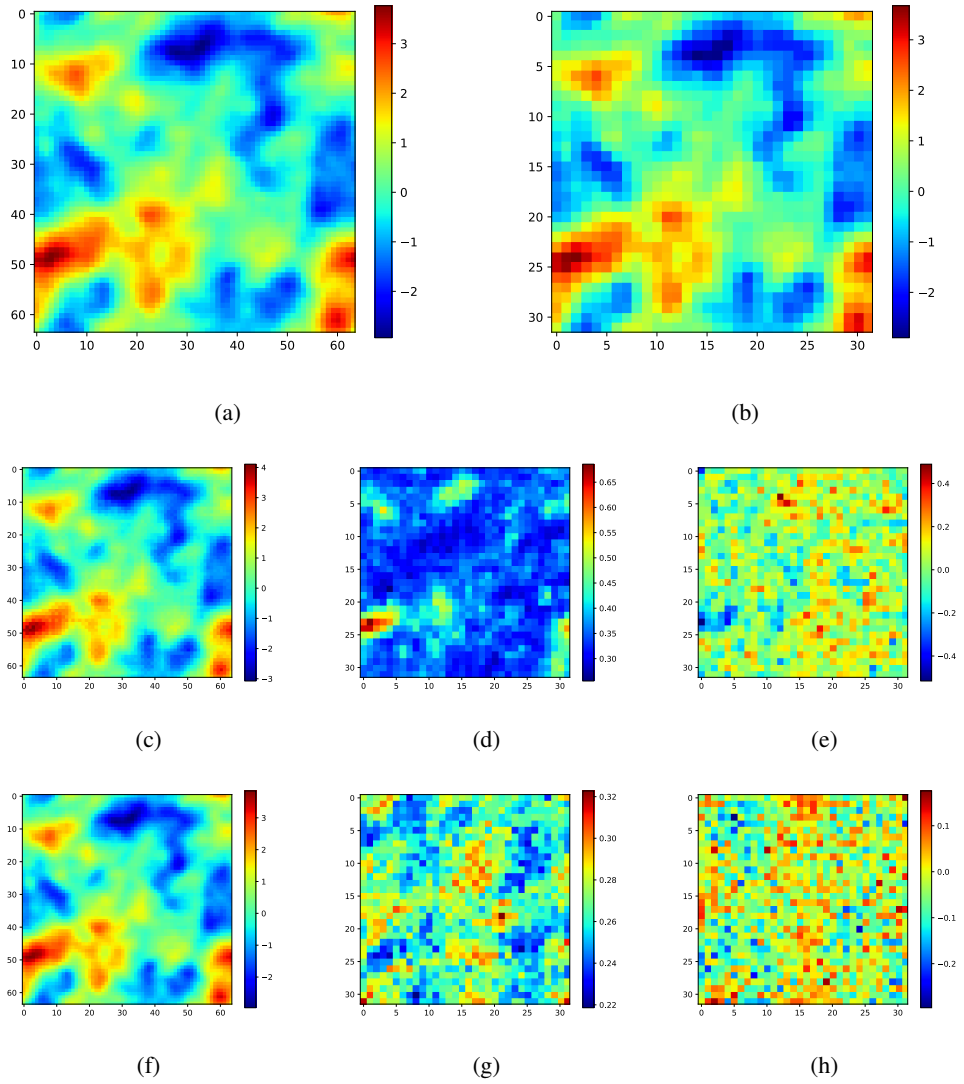


Figure 9: Matérn cases (Generation with guidance). (a) Ground Truth. (b) Low resolution with 25% observation. Mean, standard deviation, and reconstruction MSE loss of 200 samples generated with (c, d, e) NAGO (f, g, h) gaussian noise.

F.6 Ablation Experiments

The source-selection rule involves both Matérn smoothness ν_0 and correlation length ℓ_0. We vary one parameter at a time on the target $(\nu_1, \ell_1) = (2, 0.1)$ while keeping the data, FNO, training budget, and evaluation protocol fixed. The smoothness sweep includes inadmissible values to expose the boundary: the best tested choice is $\nu_0 = 0.5$, inside the strict condition $\nu_0 < 1$. Thus, the result supports choosing the smoothest effective source within a safety-margined admissible set, rather than matching target smoothness exactly.

Table 5: Source-smoothness sweep on the Matérn-field target (lower is better).

ν_0	0.5	1.0	1.5	2.0	3.0	6.0
lMSE	**1.90e-3**	2.91e-3	2.92e-3	8.94e-3	2.28e-2	6.78

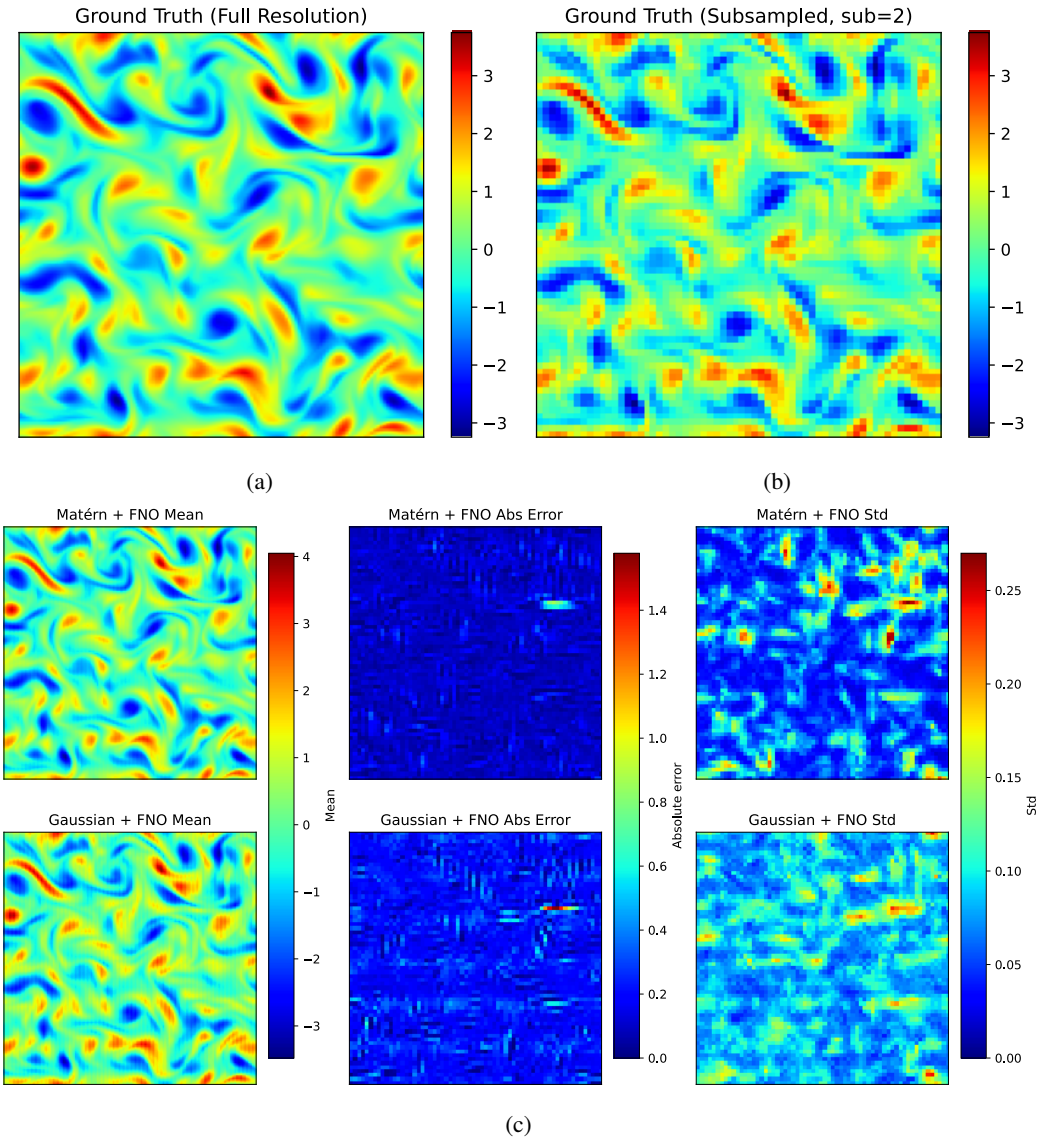


Figure 10: SNS cases (Generation with guidance). (a) Ground Truth. (b) Low resolution with 25% observation. Mean, reconstruction absolute error, and standard deviation of 200 samples generated with (First row) NAGO (Second row) gaussian noise.

With $\nu_0 = 0.5$ fixed, the length-scale sweep tests the finite-band part of Eq. (13). The equal-κ choice $\ell_0 = 0.05$ performs best; both shorter and longer correlations increase lMSE, while the discretized Gaussian source is worst. This non-monotone result shows that admissible smoothness alone does not determine a useful source.

Table 6: Source length-scale sweep at $\nu_0 = 0.5$, with a discretized Gaussian reference.

Source	$\ell_0 = 0.01$	$\ell_0 = 0.05$	$\ell_0 = 0.10$	$\ell_0 = 0.20$	Gaussian
lMSE	2.995e-3	**5.520e-4**	1.134e-3	2.241e-3	5.888e-3

F.7 Controlled RK4 Step Sweep

We compare fixed checkpoints trained on the same 64^2 Matérn target with identical FNO architecture and training protocol. Each entry uses 1,000 generated samples. The selected Matérn source reaches lMSE 1.34×10^{-3} at three steps and 6.21×10^{-4} at five steps. Gaussian FFM reaches a stable regime only around 30 steps, where its lMSE is 2.14×10^{-3}. Hence the defensible source-specific efficiency gain is a 6–10$\times$ reduction in RK4 steps and neural-function evaluations, not the larger fixed-budget ratios in individual table cells.

Table 7: Controlled RK4 sweep on the 64^2 Matérn target (lMSE, lower is better).

Steps	3	5	10	20	30	50
Gaussian	0.2673	0.09294	0.006363	0.005474	0.002138	0.002232
Matérn	**0.001339**	**0.000621**	**0.000688**	**0.000850**	**0.001247**	**0.001112**

F.8 Complementary SNS Distribution Metrics

Spectral lMSE alone does not establish distributional fidelity. We therefore evaluate 1,000 target and 1,000 generated SNS vorticity fields at 64^2, using five-step NAGO and ten-step Gaussian FFM. Kernel bandwidths, histogram bins, and normalization are fixed from held-out target samples. NAGO is better on multiscale field MMD, the pointwise marginal distances, velocity structure functions, and three physical observables, whereas Gaussian FFM is better on spatial correlations and vorticity kurtosis. These mixed results support a broad but non-universal distributional improvement.

Table 8: SNS distributional errors (lower is better).

Method	MMD^2	std. W1	JS	S_2	S_4	KE	Enstrophy	Palinstrophy	Vort. corr.
Gaussian (10)	0.003676	0.07487	0.002247	0.1895	0.3506	0.1885	0.1786	0.2262	**0.01711**
NAGO (5)	**0.001418**	**0.04575**	**0.000798**	**0.1450**	**0.2768**	**0.1330**	**0.1095**	**0.1470**	0.01977

F.9 Resolution Diagnostics

Models are trained at 64^2 and evaluated at 32^2, 64^2, and 128^2. At 128^2, NAGO assigns 93.99% of generated energy to the band shared with training and 2.39% to newly available modes; Gaussian FFM assigns 23.64% and 76.92%, respectively. This measures shared-band preservation and suppression of spurious new-band energy. It does not imply correct extrapolation of unobserved fine scales: the NAGO palinstrophy ratio to the reference remains 6.72.

F.10 Sampling Throughput Protocol

We time both samplers on one NVIDIA RTX 4090 at batch size 16 with explicit CUDA synchronization. The pseudo-spectral solver runs at 256^2, uses $\Delta t = 10^{-4}$, records every 0.5 physical-time units, and is downsampled to 64^2. During the stationary production interval it takes 2,038.139 s for 100 recorded batches, or 20.3814 s per batch. Five-step NAGO takes a median 0.228848 s per batch over ten repeats (IQR 0.002085 s), giving the reported raw ratio $89.1\times$. Checkpoint loading, serialization, dataset generation, and model training are excluded. Because solver outputs are temporally correlated and five-step NAGO does not match the finite-reference sampling floor, this is a post-training framework-throughput comparison, not an equal-fidelity or end-to-end claim.

F.11 FFT Matérn Sampling Cost

For the stationary periodic-grid setting used here, the Matérn covariance is diagonal in the Fourier basis. For a batch of B fields on an N^d grid, we draw white Fourier coefficients, multiply each coefficient by $(\kappa^2 + \|k\|^2)^{-\alpha/2}$, impose Hermitian symmetry for real fields, and apply an inverse FFT. This requires $O(BN^d \log N)$ time and $O(BN^d)$ memory. No dense $N^d \times N^d$ covariance matrix is constructed or factorized. This advantage does not directly extend to irregular domains where Fourier diagonalization is unavailable.

F.12 Stationary Anisotropic Target

We place the anisotropic experiment in the appendix because the current method selects an isotropic source. The target is a stationary anisotropic Matérn field with $\nu_1 = 2$, $\ell_x = 0.1$, $\ell_y = 0.05$, and a 30° rotation. Both 64^2 FNOs retain the full Fourier grid and are trained from scratch for 4,000 steps. At ten RK4 steps, the isotropic Matérn source improves orientation error, anisotropy-ratio error, and full 2D spectral errors over the Gaussian source (Table 9). Both sources nevertheless recover the anisotropy accurately, and paired fixed-checkpoint evaluation seeds do not show uniform superiority: at three-step Matérn versus ten-step Gaussian, 2D log-power MSE is 0.003001 ± 0.000121 versus 0.002788 ± 0.000176. The intervals quantify evaluation-sampling variability, not independent training runs.

Table 9: Stationary anisotropic Matérn target with 1,024 samples and ten RK4 steps.

Source	Angle error (°)	Ratio error	2D power L1	Log-power MSE
Gaussian	0.1310	0.01192	0.04611	0.002929
Isotropic Matérn	**0.0815**	**0.01170**	**0.03955**	**0.002563**

Figure 11 complements the aggregate spectral metrics with unpaired field realizations. The samples are drawn from the same saved 1,024-sample, ten-step evaluation used in Table 9; fixed, nonoverlapping array indices are used for the three rows. Both learned distributions reproduce the target's diagonal organization, while the quantitative differences between source choices are assessed by the distribution-level metrics above rather than by visual matching of individual draws.

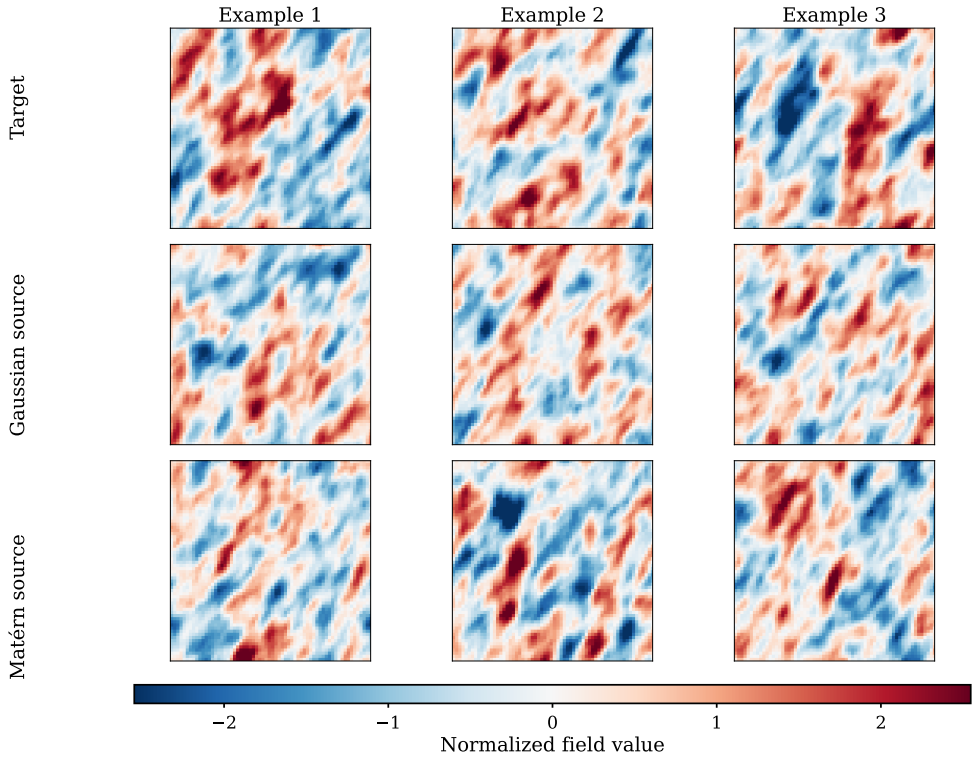
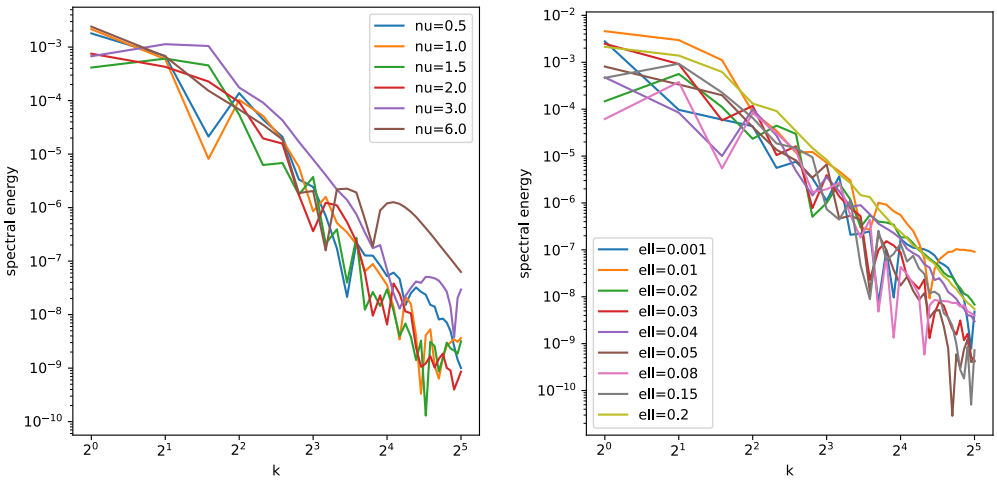


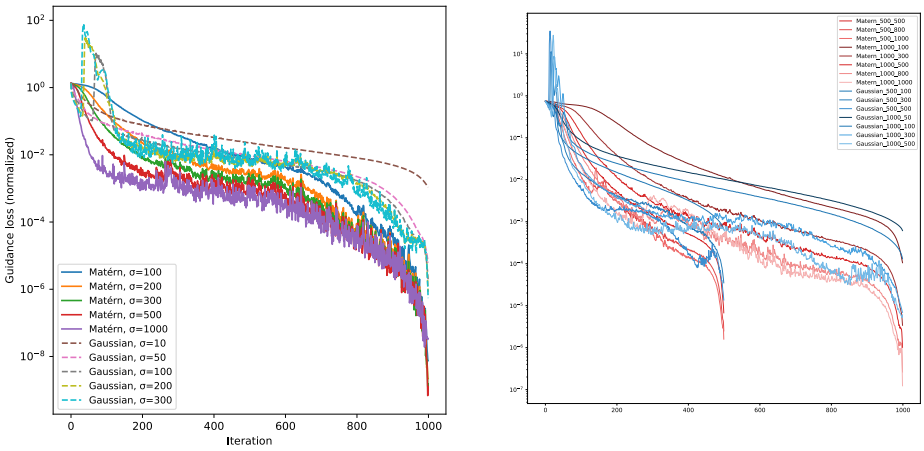
Figure 11: Qualitative samples for the stationary anisotropic Matérn target. Rows show unpaired draws from the target distribution, the Gaussian-source model, and the isotropic Matérn-source model. All panels use the same symmetric color scale.

Guided sampling introduces a second numerical intervention in addition to ODE refinement. If a per-step guidance coefficient is held fixed while the number of steps grows, the cumulative guidance



(a) Guided loss trained by different ν. (b) Guided loss trained by different ℓ

Figure 12: Comparison of guided losses under different noise settings.



(a) Guided Loss as Iteration Steps. (Matern) (b) Guided Loss as Iteration Steps. (SNS)

Figure 13: Comparison of guided losses under different noise settings.

also grows, confounding source effects with intervention strength. We therefore add a fixed-total protocol in which the per-step coefficient is scaled so that the total guidance budget remains constant.

On the 64^2 Matérn target with 25% observations, 1,000 generated samples, and total guidance budget 100, Gaussian sources yield lower observation and conditional-mean MSE at both 3 and 100 RK4 steps. The Matérn source is substantially better in spectral lMSE at three steps but not at 100 steps (Table 10). Thus, source choice changes the trade-off among observation fit, posterior spread, and spectral fidelity; no single source uniformly dominates the guided metrics. The loss trajectories in Figs. 12 and 13 remain useful diagnostics under their original parameterization, but they should not be read as causal isolation of the source.

Table 10: Fixed-total-guidance control on the Matérn target (lower is better).

Steps	Source	Observed MSE	Mean MSE	Spectral lMSE
3	Gaussian	**1.136**	**0.510**	0.2665
3	Matérn	1.817	0.907	**0.00259**
100	Gaussian	**0.977**	**0.450**	**0.02620**
100	Matérn	1.741	0.893	0.07456

G Limitations of This Study

The current theory and implementation assume stationary measures on periodic regular grids. Under these assumptions, the spectrum is diagonal in a Fourier basis and Matérn samples are generated efficiently by FFT. On irregular geometries, nonperiodic domains, or nonstationary targets, estimating a reliable spectrum and applying the same sampler are nontrivial. The finite-band selection rule also depends on the sample size and on the band used to estimate $\widehat{\beta}$; uncertainty in this estimate is not yet propagated into source selection.

The empirical evidence is also bounded. Most comparisons use one trained checkpoint per configuration, so repeated evaluation seeds measure conditional sampling variability rather than training variability. The anisotropic experiment uses an isotropic source and shows accurate recovery of a stationary directional target, but it does not establish uniform superiority or address spatially varying anisotropy. Cross-resolution tests preserve the represented frequency band but do not reconstruct an unseen spectral tail. Finally, the guidance results depend on the metric and on how guidance strength is normalized across solver steps.

We also emphasize that the scope of this work is sampling from invariant measures of SPDEs, rather than general PDE solution generation or image generation. For deterministic PDE prediction, generative models are usually less efficient than neural operator methods when the goal is to approximate the forward solution map, since iterative sampling is not naturally optimized for fast surrogate modeling. Therefore, we do not position the proposed method as a replacement for operator-learning approaches in standard PDE forecasting tasks.

Similarly, this work does not focus on image-generation benchmarks. Our resolution-consistency analysis is motivated by the function-space nature of SPDEs, where samples are discretizations of underlying random fields and should remain consistent under changes of spatial resolution. This consideration is fundamentally different from common image-generation settings, where low- and high-resolution images are usually treated as distinct finite-dimensional data distributions rather than discretizations of the same function-space object.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: We have made clear claims about contributions and scopes in the abstract and introduction.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: We have the separated limitation section in the appendix.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

Justification: We provide detailed theoretical proofs in the main paper and appendix.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: We describe the details of our model and training strategy in the paper for reproduction.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: The code is provided in supplemental material.

Guidelines:

- The answer [N/A] means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: We have introduced experiment settings and details, and even add more details in the appendix.

   Guidelines: We have introduced experiment settings and details, and even add more details in the appendix.

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: Our paper is about SPDE, so the error is all based on the mean of samples, and we also consider their spectral distribution.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).

- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: We have discussed the information about computer resources.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: We follow the NeurIPS Code of Ethics in every respect.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: Positive impacts include reducing computational costs for sampling invariant measures of SPDEs and improving the stability of scientific generative modeling. We do not anticipate specific negative societal impacts from this theoretical and computational work. The method is designed for scientific simulations of stochastic PDEs rather than human-facing decision making or deployment in sensitive domains.

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.

- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: We do not release any data or model that has high risks for misuse.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: We have cited the original paper that produced the code package and dataset.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.

- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: We provide documentation with our code.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: Our work does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: Our work does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: The core method development in this research does not involve LLMs as any important, original, or non-standard components.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.