

## Appendix

### A Diffusion model framework

Score-based generative modeling via stochastic differential equations (SDEs) provides a unified framework that generalizes denoising score matching with Langevin dynamics (SMLD) and denoising diffusion probabilistic models (DDPM). The forward diffusion process gradually perturbs data  $p_0(\mathbf{x})$  into a tractable prior  $p_T(\mathbf{x})$  via the Itô SDE:

$$d\mathbf{x} = f(\mathbf{x}, t) dt + g(t) d\mathbf{w}, \quad t \in [0, T], \quad (1)$$

where  $f$  is the drift term,  $g$  is the diffusion coefficient, and  $\mathbf{w}$  is a standard Wiener process.

By Anderson's reverse-time diffusion theorem, the reverse process is also an SDE:

$$d\mathbf{x} = [f(\mathbf{x}, t) - g(t)^2 \nabla_{\mathbf{x}} \log p_t(\mathbf{x})] dt + g(t) d\bar{\mathbf{w}}, \quad (2)$$

where  $\bar{\mathbf{w}}$  denotes a Wiener process in reversed time. Given the time-dependent score function  $\nabla_{\mathbf{x}} \log p_t(\mathbf{x})$ , learned via continuous-time denoising score matching

We can simulate the reverse SDE to generate samples. A key special case is the probability flow ODE, a deterministic process sharing the same marginals  $p_t(\mathbf{x})$  as the forward SDE:

$$d\mathbf{x} = \left[ f(\mathbf{x}, t) - \frac{1}{2} g(t)^2 \nabla_{\mathbf{x}} \log p_t(\mathbf{x}) \right] dt, \quad (3)$$

which becomes a neural ODE when  $\nabla_{\mathbf{x}} \log p_t$  is replaced by the trained score network  $s_{\theta}(\mathbf{x}, t)$ .

**Discrete Probability Flow ODE Sampling.** In practice, can be discretized for generation by dividing the time horizon  $[0, T]$  into  $N$  steps  $t_0 > t_1 > \dots > t_N$  and updating

$$\mathbf{x}_{i-1} = \mathbf{x}_i + \left[ f(\mathbf{x}_i, t_i) - \frac{1}{2} g(t_i)^2 s_{\theta}(\mathbf{x}_i, t_i) \right] \Delta t_i, \quad (4)$$

where  $\Delta t_i = t_{i-1} - t_i < 0$  is the reverse-time step. This predictor-only scheme requires no stochastic term, enabling fast and deterministic sample generation. In particular, for the Variance Exploding (VE) SDE formulation, the drift term is identically zero:  $f(\mathbf{x}, t) \equiv 0$ , and the diffusion coefficient depends solely on the prescribed noise schedule,  $g(t) = \sigma(t)$ , where  $\sigma(t)$  is a monotonically increasing function that controls the standard deviation of the injected Gaussian perturbations. This setting recovers the continuous-time limit of SMLD, where  $\sigma(t)$  typically follows a geometric progression between  $\sigma_{\min}$  and  $\sigma_{\max}$ , ensuring that  $\mathbf{x}_N$  at  $t = T$  is well-approximated by an isotropic Gaussian prior.

### B Temporal Constraints

This appendix presents the proofs and derivations of the mathematical results employed in the temporal physical constraints. Section B.1 describes how the time-domain solution of the viscous fluid equation can be formulated as a convolution with a Gaussian kernel. Sections B.2 through B.4 analyze the upper bounds of sample space variations under Lipschitz continuity for three distinct types of initialization noise.

### B.1 Viscous-Informed Gaussian kernel

**Proposition 1.** *The fundamental solution (Green's function) of the two-dimensional diffusion equation*

$$\frac{\partial u}{\partial t} = \nu \nabla^2 u, \quad (x, y) \in \mathbb{R}^2, \quad t > 0, \quad (5)$$

with initial condition

$$u(x, y, 0) = \delta(x, y), \quad (6)$$

is given by

$$G(x, y, t) = \frac{1}{4\pi\nu t} e^{-\frac{x^2+y^2}{4\nu t}}. \quad (7)$$

**Proof:** The governing equation for viscous fluid flow is given by:

$$\frac{\partial u}{\partial t} = \nu \nabla^2 u \quad (8)$$

In a two-dimensional Cartesian coordinate system, the Laplace operator  $\nabla^2$  expands to:

$$\frac{\partial u}{\partial t} = \nu \left( \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right) \quad (9)$$

Here,  $u = u(x, y, t)$  is a function defined on  $\mathbb{R}^2 \times [0, \infty)$ , and  $\nu$  denotes the kinematic viscosity. We consider an initial value problem with the initial condition specified as a two-dimensional point source located at the origin. This is represented by the 2D Dirac delta function:

$$\begin{cases} \frac{\partial u}{\partial t} = \nu \nabla^2 u & \text{for } (x, y) \in \mathbb{R}^2, \quad t > 0 \\ u(x, y, 0) = \delta(x, y) & \text{for } (x, y) \in \mathbb{R}^2 \end{cases} \quad (10)$$

where  $\delta(x, y) = \delta(x)\delta(y)$ , and it satisfies the property:

$$\iint_{\mathbb{R}^2} f(x, y) \delta(x, y) dx dy = f(0, 0) \quad (11)$$

The solution  $u(x, y, t)$  to this problem is the fundamental solution (or Green's function) of the 2D diffusion equation, which we denote as  $G(x, y, t)$ .

We apply a two-dimensional Fourier transform to the spatial variables  $(x, y)$ . Let  $\mathbf{x} = (x, y)$  denote the spatial coordinate vector, and  $\mathbf{k} = (k_x, k_y)$  the corresponding frequency (wavenumber) vector.

The 2D Fourier transform is defined as:

$$\hat{u}(\mathbf{k}, t) = \mathcal{F}\{u(\mathbf{x}, t)\} = \iint_{\mathbb{R}^2} u(\mathbf{x}, t) e^{-i\mathbf{k} \cdot \mathbf{x}} dx \quad (12)$$

where  $\mathbf{k} \cdot \mathbf{x} = k_x x + k_y y$ , and  $d\mathbf{x} = dx dy$ .

We begin by applying the Fourier transform to the partial derivative  $\frac{\partial u}{\partial x}$ . By definition, we have:

$$\mathcal{F}\left\{\frac{\partial u}{\partial x}\right\} = \iint_{\mathbb{R}^2} \frac{\partial u(x, y)}{\partial x} e^{-i(k_x x + k_y y)} dx dy. \quad (13)$$

This can be rewritten as:

$$= \int_{-\infty}^{\infty} e^{-ik_y y} \left( \int_{-\infty}^{\infty} \frac{\partial u(x, y)}{\partial x} e^{-ik_x x} dx \right) dy. \quad (14)$$

We focus on evaluating the inner integral:

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{\partial u}{\partial x} e^{-ik_x x} dx &= [u(x, y) e^{-ik_x x}]_{x=-\infty}^{x=\infty} \\ &\quad - \int_{-\infty}^{\infty} u(x, y) (-ik_x e^{-ik_x x}) dx \end{aligned} \quad (15)$$

For the Fourier transform to be well-defined, we require the boundary term  $[u(x, y) e^{-ik_x x}]_{x=-\infty}^{x=\infty} = 0$ . Thus, the integral becomes:

$$\begin{aligned} &= -(-ik_x) \int_{-\infty}^{\infty} u(x, y) e^{-ik_x x} dx \\ &= ik_x \int_{-\infty}^{\infty} u(x, y) e^{-ik_x x} dx \end{aligned} \quad (16)$$

Substituting this back into the full double integral:

$$\begin{aligned} \mathcal{F}\left\{\frac{\partial u}{\partial x}\right\} &= \int_{-\infty}^{\infty} e^{-ik_y y} \left( ik_x \int_{-\infty}^{\infty} u(x, y) e^{-ik_x x} dx \right) dy \\ &= ik_x \iint_{\mathbb{R}^2} u(x, y) e^{-i(k_x x + k_y y)} dx dy \end{aligned} \quad (17)$$

The expression in parentheses is the definition of the 2D Fourier transform of  $u(x, y)$ , denoted as  $\hat{u}(k_x, k_y)$ . Hence, we arrive at the key identity:

$$\mathcal{F}\left\{\frac{\partial u}{\partial x}\right\} = ik_x \hat{u}(k_x, k_y) \quad (18)$$

We now derive the Fourier transform of the second-order derivative:

$$\mathcal{F}\left\{\frac{\partial^2 u}{\partial x^2}\right\} = \mathcal{F}\left\{\frac{\partial}{\partial x} \left(\frac{\partial u}{\partial x}\right)\right\} \quad (19)$$

Treating  $\frac{\partial u}{\partial x}$  as a new function and applying the same rule gives:

$$= ik_x \cdot \mathcal{F}\left\{\frac{\partial u}{\partial x}\right\} \quad (20)$$

Substituting the result of the first-order derivative transform:

$$= ik_x \cdot (ik_x \hat{u}(k_x, k_y)) = (ik_x)^2 \hat{u}(k_x, k_y) \quad (21)$$

Substituting the previous results into the original PDE yields:

$$\frac{\partial \hat{u}(\mathbf{k}, t)}{\partial t} = \nu (-|\mathbf{k}|^2 \hat{u}(\mathbf{k}, t)) \frac{d\hat{u}}{dt} = -\nu |\mathbf{k}|^2 \hat{u} \quad (22)$$

This transforms the partial differential equation (PDE) into an ordinary differential equation (ODE) in time  $t$ . The general solution to this ODE is:

$$\hat{u}(\mathbf{k}, t) = C(\mathbf{k}) e^{-\nu |\mathbf{k}|^2 t} \quad (23)$$

The constant  $C(\mathbf{k})$  is determined by the initial condition. Taking the Fourier transform of the initial condition  $u(\mathbf{x}, 0) = \delta(\mathbf{x})$  gives:

$$\hat{u}(\mathbf{k}, 0) = \mathcal{F}\{\delta(\mathbf{x})\} = \iint_{\mathbb{R}^2} \delta(\mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x}} d\mathbf{x} = 1 \quad (24)$$

Therefore,  $C(\mathbf{k}) = 1$ , and the solution in the Fourier domain becomes:

$$\hat{u}(\mathbf{k}, t) = e^{-\nu |\mathbf{k}|^2 t} = e^{-\nu (k_x^2 + k_y^2) t} \quad (25)$$

This expression represents a 2D Gaussian function in the frequency domain  $(k_x, k_y)$ . We now apply the inverse Fourier transform to recover the solution in the spatial domain. Denoting this as the fundamental solution  $G(\mathbf{x}, t)$ , we have:

$$G(\mathbf{x}, t) = \frac{1}{(2\pi)^2} \iint_{\mathbb{R}^2} e^{-\nu |\mathbf{k}|^2 t} e^{i\mathbf{k}\cdot\mathbf{x}} d\mathbf{k} \quad (26)$$

Since

$$\begin{aligned} e^{-\nu |\mathbf{k}|^2 t} e^{i\mathbf{k}\cdot\mathbf{x}} &= e^{-\nu (k_x^2 + k_y^2) t} e^{i(k_x x + k_y y)} \\ &= \left(e^{-\nu t k_x^2} e^{ik_x x}\right) \left(e^{-\nu t k_y^2} e^{ik_y y}\right), \end{aligned} \quad (27)$$

the double integral can be separated into the product of two independent one-dimensional integrals:

$$G(x, y, t) = \quad (28)$$

$$\frac{1}{(2\pi)^2} \left( \int_{-\infty}^{\infty} e^{-\nu t k_x^2} e^{ik_x x} dk_x \right) \left( \int_{-\infty}^{\infty} e^{-\nu t k_y^2} e^{ik_y y} dk_y \right) \quad (29)$$

It is known that:

$$\int_{-\infty}^{\infty} e^{-(\nu t) k^2} e^{ikx} dk = \sqrt{\frac{\pi}{\nu t}} e^{-\frac{x^2}{4\nu t}} \quad (30)$$

Substituting this into the expression above gives:

$$\begin{aligned} G(x, y, t) &= \frac{1}{(2\pi)^2} \left( \sqrt{\frac{\pi}{\nu t}} e^{-\frac{x^2}{4\nu t}} \right) \left( \sqrt{\frac{\pi}{\nu t}} e^{-\frac{y^2}{4\nu t}} \right) \\ &= \frac{1}{4\pi^2} \cdot \frac{\pi}{\nu t} e^{-\frac{x^2+y^2}{4\nu t}} \end{aligned} \quad (31)$$

Simplifying, we obtain the 2D Gaussian kernel:

$$G(x, y, t) = \frac{1}{4\pi\nu t} e^{-\frac{x^2+y^2}{4\nu t}} \quad (32)$$

For an arbitrary initial condition  $u(x, y, 0) = f(x, y)$ , the linearity of the PDE implies that the solution  $u(x, y, t)$  is given by the 2D convolution of the initial condition  $f$  with the fundamental solution  $G$ .

Any function  $f(\mathbf{x})$  can be represented via:

$$f(\mathbf{x}) = \iint_{\mathbb{R}^2} f(\mathbf{y}) \delta(\mathbf{x} - \mathbf{y}) d\mathbf{y} \quad (33)$$

This confirms that solving the viscous process is equivalent to performing a convolution with a Gaussian kernel.

## B.2 Independent Noise Analyze

Assume the dimensionality of the noise vector is  $d$ . For a sequence of independent noise samples  $z_k \sim \mathcal{N}(0, I)$ , each sample is drawn independently from a standard multivariate Gaussian distribution. The difference between two independent Gaussian noise vectors  $z_{k+1} - z_k$  also follows a Gaussian distribution. Specifically, since both  $z_k$  and  $z_{k+1}$  have zero mean and identity covariance, we have:

$$\mathbb{E}[z_{k+1} - z_k] = \mathbb{E}[z_{k+1}] - \mathbb{E}[z_k] = 0, \quad (34)$$

$$\text{Var}(z_{k+1} - z_k) = \text{Var}(z_{k+1}) + \text{Var}(z_k) = I + I = 2I. \quad (35)$$

Thus,  $z_{k+1} - z_k \sim \mathcal{N}(0, 2I)$ . The expected squared norm of this difference is given by:

$$\mathbb{E}[|z_{k+1} - z_k|^2] = \mathbb{E}\left[\sum_{i=1}^d (z_{k+1,i} - z_{k,i})^2\right]. \quad (36)$$

By the linearity of expectation, this simplifies to:

$$\sum_{i=1}^d \mathbb{E}[(z_{k+1,i} - z_{k,i})^2]. \quad (37)$$

Since each component  $z_{k+1,i} - z_{k,i}$  is a univariate Gaussian with variance 2, we have:

$$\sum_{i=1}^d \mathbb{E}[(z_{k+1,i} - z_{k,i})^2] = \sum_{i=1}^d 2 = 2d. \quad (38)$$

### B.3 Ornstein-Uhlenbeck (OU) Process Noise Analyze

The discrete-time formulation of the Ornstein-Uhlenbeck (OU) process is given by:

$$z_{k+1} = z_k + \theta(\mu - z_k)\Delta t + \sigma\sqrt{\Delta t}\mathcal{W}_k, \quad (39)$$

where  $\mathcal{W}_k \sim \mathcal{N}(0, I)$  denotes a standard Gaussian noise vector. Without loss of generality, we assume the mean  $\mu = 0$ , leading to:

$$z_{k+1} = (1 - \theta\Delta t)z_k + \sigma\sqrt{\Delta t}\mathcal{W}_k. \quad (40)$$

Thus, the difference between successive steps becomes:

$$z_{k+1} - z_k = -\theta\Delta t \cdot z_k + \sigma\sqrt{\Delta t} \cdot \mathcal{W}_k. \quad (41)$$

We are interested in the expected squared norm of this difference:

$$\mathbb{E}[|z_{k+1} - z_k|^2] = \mathbb{E}[|-\theta\Delta t \cdot z_k + \sigma\sqrt{\Delta t} \cdot \mathcal{W}_k|^2]. \quad (42)$$

Expanding the squared norm, we obtain:

$$\begin{aligned} & \mathbb{E}\left[(-\theta\Delta t z_k + \sigma\sqrt{\Delta t}\mathcal{W}_k)^\top (-\theta\Delta t z_k + \sigma\sqrt{\Delta t}\mathcal{W}_k)\right] \\ &= \mathbb{E}\left[(\theta\Delta t)^2 |z_k|^2 - 2\theta\Delta t \sigma\sqrt{\Delta t} (z_k^\top \mathcal{W}_k) + (\sigma\sqrt{\Delta t})^2 |\mathcal{W}_k|^2\right]. \end{aligned} \quad (43)$$

We now analyze each term individually.

#### Cross term

$$\mathbb{E}[-2\theta\Delta t \sigma\sqrt{\Delta t} (z_k^\top \mathcal{W}_k)] \quad (44)$$

Since  $\mathcal{W}_k$  is independent of  $z_k$  and  $\mathbb{E}[\mathcal{W}_k] = 0$ , the expectation of the cross term vanishes:  $\mathbb{E}[z_k^\top \mathcal{W}_k] = 0$ .

#### Stochastic term:

$$\mathbb{E}\left[(\sigma\sqrt{\Delta t})^2 |\mathcal{W}_k|^2\right] \quad (45)$$

Since  $\mathcal{W}_k \sim \mathcal{N}(0, I)$ , it follows that  $\mathbb{E}[|\mathcal{W}_k|^2] = d$ . Thus, this term evaluates to  $\sigma^2 d \Delta t$ .

#### Drift term:

$$\mathbb{E}[(\theta\Delta t)^2 |z_k|^2]. \quad (46)$$

We now analyze  $\mathbb{E}[|z_k|^2]$ . From the discretized Ornstein-Uhlenbeck (OU) process, we have:

$$\mathbb{E}[|z_k|^2] = \mathbb{E}[|(1 - \theta\Delta t)z_{k-1} + \sigma\sqrt{\Delta t}\mathcal{W}_k|^2]. \quad (47)$$

This leads to the following recurrence relation:

$$\mathbb{E}[|z_k|^2] = (1 - \theta\Delta t)^2 \mathbb{E}[|z_{k-1}|^2] + \sigma^2 d \Delta t. \quad (48)$$

Solving this recurrence yields:

$$\mathbb{E}[|z_k|^2] = \frac{\sigma^2 d}{2\theta - \theta^2 \Delta t} + \left(d - \frac{\sigma^2 d}{2\theta - \theta^2 \Delta t}\right) (1 - \theta\Delta t)^{2k}. \quad (49)$$

For sufficiently small step size  $\Delta t$ , the term  $\Delta t^2$  is of higher order relative to  $\Delta t$ , and can thus be neglected. Consequently, the variation is predominantly driven by the stochastic term. That is,

$$\mathbb{E}[|z_{k+1} - z_k|^2]_{\text{OU}} \approx \sigma^2 d \Delta t. \quad (50)$$

### B.4 Viscous-Informed Noise Analyze

Given an initial noise sample  $z_0 \sim \mathcal{N}(0, I_d)$ , we generate noise at a sequence of discrete time points  $t_0, t_1, t_2, \dots, t_N$ . The evolution of the signal in the Fourier domain is described as:

$$\hat{z}_k(\omega) = \hat{z}_0(\omega) \cdot e^{-D||\omega||^2 t_k} \quad (51)$$

We now compute the difference between two consecutive frames,  $z_k$  and  $z_{k+1}$ . Since

$$\hat{z}_k(\omega) = \hat{z}_0(\omega) \cdot e^{-D||\omega||^2 t_k} \quad (52)$$

$$\begin{aligned} \hat{z}_{k+1}(\omega) &= \hat{z}_0(\omega) \cdot e^{-D||\omega||^2 t_{k+1}} \\ &= \hat{z}_0(\omega) \cdot e^{-D||\omega||^2 (t_k + \Delta t)} \end{aligned} \quad (53)$$

the difference in the Fourier domain between adjacent samples is given by:

$$\begin{aligned} \Delta \hat{z}_k(\omega) &= \hat{z}_0(\omega) \left( e^{-D||\omega||^2 (t_k + \Delta t)} - e^{-D||\omega||^2 t_k} \right) \\ &= \hat{z}_0(\omega) \cdot e^{-D||\omega||^2 t_k} \left( e^{-D||\omega||^2 \Delta t} - 1 \right) \end{aligned} \quad (54)$$

In the 2D case, this becomes:

$$\Delta \hat{z}_k(m, n) = \hat{z}_0(m, n) \cdot e^{-D\lambda_{m,n} t_k} \left( e^{-D\lambda_{m,n} \Delta t} - 1 \right) \quad (55)$$

For a 2D grid of size  $d = L \times K$ , Parseval's theorem in two dimensions yields:

$$|\Delta z_k|^2 = \sum_{i=0}^{L-1} \sum_{j=0}^{K-1} |\Delta z_k(i, j)|^2 = \frac{1}{d} \sum_{m=0}^{L-1} \sum_{n=0}^{K-1} |\Delta \hat{z}_k(m, n)|^2 \quad (56)$$

Taking expectation on both sides leads to:

$$\mathbb{E}[|\Delta z_k|^2] = \frac{1}{d} \sum_{m=0}^{L-1} \sum_{n=0}^{K-1} \mathbb{E}[|\Delta \hat{z}_k(m, n)|^2] \quad (57)$$

Substituting the expression from Eq. 55, we obtain:

$$\begin{aligned} \mathbb{E}[|\Delta z_k|^2] &= \frac{1}{d} \sum_{m=0}^{L-1} \sum_{n=0}^{K-1} \mathbb{E}[|\hat{z}_0(m, n)|^2] \cdot e^{-2D\lambda_{m,n}t_k} (e^{-D\lambda_{m,n}\Delta t} - 1)^2 \\ &= \sum_{m=0}^{L-1} \sum_{n=0}^{K-1} e^{-2D\lambda_{m,n}t_k} (e^{-D\lambda_{m,n}\Delta t} - 1)^2 \end{aligned} \quad (58)$$

where

$$\lambda_{m,n} = \|\omega\|^2 = \omega_m^2 + \omega_n^2 = \left(\frac{2\pi m}{L}\right)^2 + \left(\frac{2\pi n}{K}\right)^2. \quad (59)$$

Applying a first-order Taylor expansion for small  $\Delta t$ , we approximate:

$$\mathbb{E}[|\Delta z_k|^2] \approx \sum_{m=0}^{L-1} \sum_{n=0}^{K-1} e^{-2D\lambda_{m,n}t_k} (-D\lambda_{m,n}\Delta t)^2 + O(\Delta t^3) \quad (60)$$

Define the summation term  $S_k$  in the 2D case as:

$$S_k = \sum_{m=0}^{L-1} \sum_{n=0}^{K-1} \lambda_{m,n}^2 e^{-2D\lambda_{m,n}t_k} \quad (61)$$

Therefore, the expected squared difference can be expressed as:

$$\mathbb{E}[|\Delta z_k|^2] = D^2 S_k \Delta t^2 + O(\Delta t^3) \quad (62)$$

We now analyze the term  $S_k$ . Substituting the expression for  $\lambda_{m,n}$ , we obtain:

$$\begin{aligned} S_k &= \sum_{m=0}^{L-1} \sum_{n=0}^{K-1} \left[ \left( \frac{2\pi m}{L} \right)^2 + \left( \frac{2\pi n}{K} \right)^2 \right]^2 \cdot \\ &\quad e^{-2D[(\frac{2\pi m}{L})^2 + (\frac{2\pi n}{K})^2]t_k} \end{aligned} \quad (63)$$

Introducing continuous variables  $x_m = \frac{m}{L}$  and  $y_n = \frac{n}{K}$ , we can rewrite the sum as:

$$\begin{aligned} S_k &= d \cdot \sum_{m=0}^{L-1} \sum_{n=0}^{K-1} [(2\pi x_m)^2 + (2\pi y_n)^2]^2 \cdot \\ &\quad e^{-2D[(2\pi x_m)^2 + (2\pi y_n)^2]t_k} \Delta x \Delta y \end{aligned} \quad (64)$$

This discrete summation converges to a double integral over the unit square  $[0, 1] \times [0, 1]$ :

$$\begin{aligned} S_k &\approx d \cdot \int_0^1 \int_0^1 [(2\pi x)^2 + (2\pi y)^2]^2 \\ &\quad \cdot e^{-2D[(2\pi x)^2 + (2\pi y)^2]t_k} dx dy \end{aligned} \quad (65)$$

The value of this integral does not depend on the total dimension  $d$ ; it is a constant determined solely by  $D$  and  $t_k$ . We denote this constant as  $C_{2D}(D, t_k)$ .

Therefore, we obtain an approximate relationship between  $S_k$  and the total dimensionality  $d$ :

$$S_k \approx d \cdot C_{2D}(D, t_k) \quad (66)$$

where

$$\begin{aligned} C_{2D}(D, t_k) &= \\ &\int_0^1 \int_0^1 (4\pi^2(x^2 + y^2))^2 \cdot e^{-8D\pi^2(x^2 + y^2)t_k} dx dy \end{aligned} \quad (67)$$

To simplify the analysis, we consider the one-dimensional case. In 1D, the summation term  $S_k$  can be expressed as:

$$S_k \approx d \cdot \int_0^1 (2\pi x)^4 e^{-2D(2\pi x)^2 t_k} dx \quad (68)$$

## C Metrics

**CRPS** The Continuous Ranked Probability Score (CRPS) is employed to evaluate the accuracy of probabilistic forecasts by directly comparing the cumulative distribution functions (CDFs) of predictions and observations. Unlike pointwise metrics, CRPS jointly penalizes discrepancies in location, spread, and overall shape between the forecast distribution and the reference distribution, thereby providing a single, interpretable scalar score. A lower CRPS indicates better probabilistic calibration and sharpness.

Formally, given  $N$  forecast initializations, a spatial index set  $I$  with  $|I|$  locations, and  $n_{ens}$  ensemble members, the CRPS is estimated as

$$\begin{aligned} \text{CRPS} &= \frac{1}{N} \sum_{n=1}^N \frac{1}{|I|} \sum_{i \in I} a_i \\ &\quad \left( \frac{1}{n_{ens}} \sum_{k=1}^{n_{ens}} |x_{i,n}^k - y_{i,n}| \cdot \frac{1}{2n_{ens}^2} \sum_{k=1}^{n_{ens}} \sum_{k'=1}^{n_{ens}} |x_{i,n}^k - x_{i,n}^{k'}| \right). \end{aligned} \quad (69)$$

Here,  $x_{i,n}^k$  denotes the prediction from the  $k$ -th ensemble member at spatial location  $i$  and initialization  $n$ ,  $y_{i,n}$  is the corresponding ground truth, and  $a_i$  is a spatial weighting factor (e.g., proportional to grid-cell area). The first term measures the mean absolute error of ensemble members relative to the observation, while the second term penalizes excessive spread within the ensemble.

**SSR** Spread quantifies the variability within the ensemble and is defined as the root mean square of the ensemble variance over both spatial locations and forecast initializations. Formally, it is given by

$$\text{Spread} = \frac{1}{N} \sum_{n=1}^N \sqrt{\frac{1}{|I|} \sum_{i \in I} \frac{1}{n_{ens}-1} \sum_{k=1}^{n_{ens}} a_i (x_{i,n}^k - \bar{x}_{i,n})^2}. \quad (70)$$

Here,  $N$  denotes the number of forecast initializations,  $|I|$  is the total number of spatial locations,  $n_{ens}$  is the ensemble size, and  $a_i$  is a spatial weighting factor (e.g., grid-cell area).  $x_{i,n}^k$  represents the prediction from the  $k$ -th ensemble member at location  $i$  and initialization  $n$ , and  $\bar{x}_{i,n}$  is the corresponding ensemble mean.

To assess the reliability of the ensemble forecast, we employ the spread/skill ratio (SSR), which compares the ensemble spread to the root mean square error (RMSE) of the ensemble mean forecast relative to observations, with a finite-ensemble correction:

$$\text{SSR} = \sqrt{\frac{n_{\text{ens}} + 1}{n_{\text{ens}}} \frac{\text{Spread}}{\text{RMSE}}}. \quad (71)$$

An SSR value close to 1 indicates a well-calibrated ensemble, where the predicted uncertainty (spread) is commensurate with the actual forecast error (skill). Values less than 1 imply under-dispersion, whereas values greater than 1 indicate over-dispersion.

## D Performance Evaluation for Prediction

### D.1 Performance Evaluation for Velocity Prediction

To provide a clearer illustration of our model's performance over the entire year, we present line plots of prediction errors at different forecasting time points throughout the year Fig. 5. To further demonstrate the error coverage of our ensemble prediction model, we plot the  $2\sigma$  coverage error curves for the entire year Fig. 6.

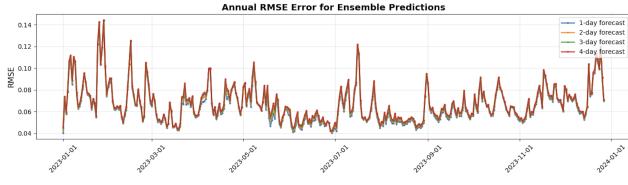


Figure 5: Flow velocity prediction errors at different forecasting time steps over the entire year.

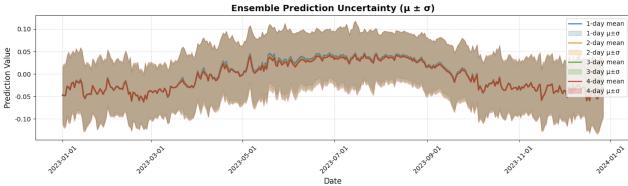


Figure 6:  $\mu \pm \sigma$  uncertainty bands for flow velocity predictions over the entire year

### D.2 Performance Evaluation for Eddy Prediction

The Fig. 7 presents a comparison of the predicted and actual numbers of vortices across different categories over the entire year, along with scatter plots Fig. 8 illustrating the spatial distribution of vortices from a subset of sampled instances throughout the year.

## E Ablation Study

To quantify the effect of incorporating spatial physical constraints, we compare models trained with and without such constraints, while adopting identical independent sampling strategies Tab. 4. To provide a fine-grained analysis of the

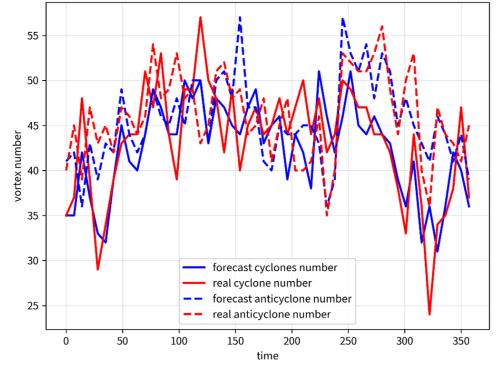


Figure 7: Vortex category prediction results over the entire year.

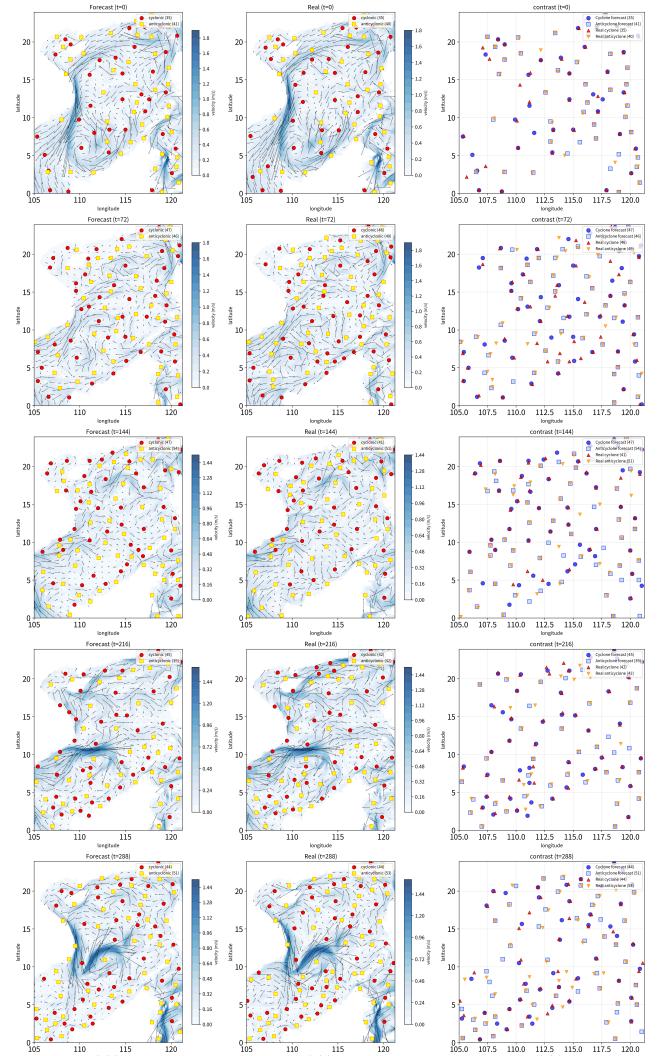


Figure 8: Scatter plots of sampled vortex category and location predictions over the entire year.

Var	Metric	Method	Lead Time			
			1	2	3	4
<b>u</b>	RMSE	spatial-constraints	<b>6.89</b>	<b>8.03</b>	<b>9.18</b>	<b>10.09</b>
		Non	7.00	8.17	9.25	10.18
	SSR	spatial-constraints	<b>1.23</b>	<b>1.05</b>	0.92	0.83
		Non	1.27	1.09	<b>0.96</b>	<b>0.86</b>
<b>v</b>	RMSE	spatial-constraints	<b>6.09</b>	<b>7.08</b>	<b>8.17</b>	<b>9.14</b>
		Non	6.40	7.29	8.32	9.27
	SSR	spatial-constraints	1.18	1.01	0.88	0.78
		Non	<b>1.16</b>	1.01	0.88	0.78

Table 4: Final prediction results with and without spatial physical constraints, non denotes the method without incorporating spatial physical constraints.

training process, we examine the convergence trajectory of each individual loss component. Fig. 9 illustrates the relative change rate of the **continuity**, **geostrophic**, and **energy** terms, normalized by their values at the first epoch. This normalization is crucial as it decouples the analysis from the disparate initial scales of the loss terms, enabling a fair assessment of their relative convergence speeds. To exam-

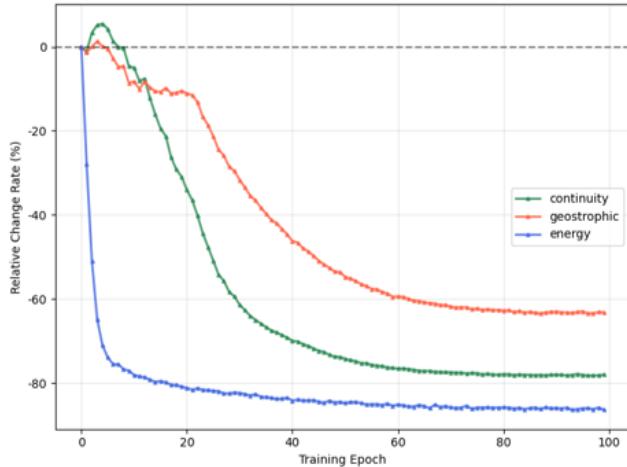


Figure 9: Cumulative variation rates of three loss functions relative to their initial values.

ine the impact of temporal physical constraints, we evaluate four initialization schemes: (1) identical initialization noise, (2) independent initialization noise, (3) Ornstein–Uhlenbeck process noise, and (4) viscous-informed noise. Performance on the target problem is assessed using RMSE and SSR metrics, with the results summarized in Tab. 5.

For the method employing an exponential kernel decay in the Fourier domain, we experimented with different choices of  $D$  and the temporal variation function, and found that the best performance is achieved when  $D < 1$  and the temporal variation function is logarithmic.

Var	Metric	Method	Lead Time			
			1	2	3	4
<b>u</b>	RMSE	same noise	6.94	8.05	9.17	10.05
		independent noise	6.89	8.03	9.18	10.09
		OU noise	6.93	8.03	9.17	10.07
		Vicious noise	<b>6.87</b>	<b>8.03</b>	<b>9.16</b>	<b>10.04</b>
	SSR	same noise	<b>1.22</b>	<b>1.05</b>	0.92	0.83
		independent noise	1.23	1.05	0.92	0.83
		OU noise	1.23	1.06	0.92	0.83
		Vicious noise	1.25	1.12	<b>1.00</b>	<b>0.92</b>
<b>v</b>	RMSE	same noise	6.10	7.08	8.17	9.13
		independent noise	6.09	7.08	8.17	9.14
		OU noise	6.11	7.09	8.17	9.13
		Vicious noise	<b>6.07</b>	<b>7.06</b>	<b>8.13</b>	<b>9.08</b>
	SSR	same noise	1.18	1.01	0.88	0.78
		independent noise	1.18	<b>1.01</b>	0.88	0.78
		OU noise	1.18	1.01	0.87	0.78
		Vicious noise	1.18	1.05	<b>0.93</b>	<b>0.84</b>

Table 5: Prediction results of different initialization noise methods with spatial physical constraints.