

# Guided Unconditional and Conditional Generative Models for Super-Resolution and Inference of Quasi-Geostrophic Turbulence

Anantha Narayanan Suresh Babu<sup>1,2</sup>, Akhil Sadam<sup>2</sup>, Pierre F.J. Lermusiaux<sup>1,2</sup>

<sup>1</sup>Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, USA

<sup>2</sup>Center for Computational Science and Engineering, Massachusetts Institute of Technology, Cambridge, USA

## Key Points:

- We benchmark four diffusion models for sparse super-resolution and inference of forced 2D quasi-geostrophic turbulence in the  $\beta$ -plane
- Conditional diffusion models generate ensembles with accurate fine-scale features and show skill with corrected spectrum tails
- Guided unconditional models are much less accurate but provide inexpensive alternatives for some inversions in geophysical fluid dynamics

---

Corresponding author: Pierre F.J. Lermusiaux, [pierrel@mit.edu](mailto:pierrel@mit.edu)

## Abstract

Typically, numerical simulations of the ocean, weather, and climate are coarse, and observations are sparse and gappy. In this work, we apply four generative diffusion modeling approaches to super-resolution and inference of forced two-dimensional quasi-geostrophic turbulence on the  $\beta$ -plane from coarse, sparse, and gappy observations. Two guided approaches minimally adapt a pre-trained unconditional model: SDEdit modifies the initial condition, and Diffusion Posterior Sampling (DPS) modifies the reverse diffusion process’s score. The other two conditional approaches, a vanilla variant and classifier-free guidance, require training with paired high-resolution and observation data. We consider eight test cases spanning: two regimes, eddy and anisotropic-jet turbulence; two Reynolds numbers,  $10^3$  and  $10^4$ ; and two observation types,  $4\times$  coarse-resolution fields and coarse, sparse and gappy observations. Our comprehensive skill metrics include norms of the reconstructed vorticity fields, turbulence statistical quantities, and quantification of the super-resolved probabilistic ensembles and their errors. We also study the sensitivity to tuning parameters such as guidance strength. Results show that SDEdit generates unphysical fields, while DPS generates reasonable reconstructions at low computational cost but with smoothed fine-scale features. Both conditional approaches require re-training, but they reconstruct missing fine-scale features, are cycle-consistent with observations, and possess the correct statistics such as energy spectra. Further, their mean model errors are highly correlated with and predictable from their ensemble standard deviations. Results highlight the trade-offs between ease of implementation, fidelity (sharpness), and cycle-consistency of the diffusion models, and offer practical guidance for deployment in geophysical inverse problems.

## Plain Language Summary

Simulating the ocean, weather, or climate at high-resolution is expensive because computers solve physics equations on many tiny grid cells. As an approximation, simulations are run with a small number of large grid cells, missing important small-scale physics such as swirling eddies. Even geophysical observations with satellites, balloons, buoys, or robots can be incomplete with large gaps. We explore whether generative diffusion models, a new machine learning technique, can complete the missing details in ways that respect the laws of physics. Using turbulence simulations, we compare four diffusion models. Two *conditional* models require training from scratch to directly learn miss-

ing details, demanding more data and computation, while two *guided unconditional* models modify existing diffusion models to complete missing details. We find that conditional models create realistic details that respect turbulence physics. In challenging cases, one of the unconditional models fails, while the other provides low-cost, reasonable guesses. We show tradeoffs between resource use and realism that users can tune to their needs. Generative diffusion models can lead to sharper maps of the ocean, weather, and climate, help merge real observations with simulations, and possibly learn physics from even incomplete, indirect data.

## 1 Introduction

Higher-resolution ocean (E. P. Chassignet & Xu, 2021), weather (Lai et al., 2024), and climate simulations (Harris & Lin, 2014) can provide greater accuracy but are computationally expensive. Numerical models are limited not only by their spatio-temporal resolution but also by the processes they represent and the number of variables they simulate (Gupta & Lermusiaux, 2021; Jacobs et al., 2023). For example, simulating the full range of spatio-temporal scales in turbulent flows is computationally intractable; approximations and parameterizations, such as in Large-Eddy Simulations (LES), are thus employed to represent the unresolved scales, which can also limit utility (Leonard, 1975; Pope, 2001). Recent machine-learning-based surrogate models for the ocean and weather also contain unresolved scales due to spectral bias, which inhibits accurate prediction of fine-scale features (Lai et al., 2024). This can be amplified by dynamics such as quasi-geostrophic turbulence, leading to unstable or nonphysical long-term predictions (Chattopadhyay & Hassanzadeh, 2023). The dual-cascade of forced geophysical turbulence (Kraichnan, 1971), i.e., the inverse cascade of energy from forcing scales to large scales and the forward cascade of enstrophy to small scales, is a challenge. In the ocean, resolving sub-mesoscale processes with scales of the order of 100 m to 10 km (McWilliams, 2016; Taylor & Thompson, 2023) can contribute to energy, mass, and tracer transports (Thomas et al., 2008). In climate, simulations with scales of 10 km can improve predictions of extreme weather events and enhance risk assessment (Difffenbaugh et al., 2005; Lopez-Gomez et al., 2025). Hence, techniques that reconstruct fine-scale features from coarse or sparse data are critical to improve the utility of simulations (Buzzicotti, 2023). Various data-driven *closures* have been explored to correct coarse-scale simulations (Maulik et al., 2019; Zanna & Bolton, 2020; Gupta & Lermusiaux, 2023; Ross et al., 2023; Jakhar et al., 2024).

An alternate approach is to reconstruct fine-scale features through super-resolution (Sofos & Drikakis, 2025), which is the focus of our work.

The super-resolution of turbulent flow fields is a growing application for machine learning. Callaham et al. (2019) explored the use of sparse representations, while various approaches treat super-resolution as a point-estimation problem using deep neural networks to map the low-resolution *input* field to the high-resolution *target* field. Fukami et al. (2019); Fukami and Taira (2024) used convolutional neural networks (CNNs) and hybrid downsampled skip-connection/multi-scale networks (DSC/MS) to reconstruct laminar flows past a cylinder and isotropic decaying turbulence. Temporal data have also been incorporated for spatio-temporal super-resolution (Fukami et al., 2021; Liu et al., 2020). Other studies have introduced physics-informed neural networks (PINNs) (Gao et al., 2021), graph neural networks (GNNs) (Barwey et al., 2025), vision transformers (Xu et al., 2023), and fully differentiable solvers to compute a dynamics-based loss for training (Page, 2025). While these methods can generate plausible high-resolution fields, they provide only a single best estimate without uncertainty quantification and often struggle to preserve physics invariants such as energy spectra when input resolution is too coarse.

Super-resolution can also be interpreted as a framework for a broad class of inverse problems, including inference or data assimilation with sparse and gappy observations (Fukami et al., 2023). Since nonlinear and high-dimensional inverse problems suffer from the *curse of dimensionality* and are inherently ill-posed (Fernandez-Martinez & Fernandez-Muniz, 2020), there is a growing need for probabilistic methods that generate samples with uncertainty quantification (Lermusiaux et al., 2006). Generative Adversarial Networks (GANs) provide one such probabilistic avenue to solve inverse problems (Patel et al., 2022; Güemes et al., 2022). Kim et al. (2021) utilized cycle-consistent-GANs for super-resolution of Direct Numerical Simulation flow fields from LES fields for homogeneous isotropic and turbulent channel flows. However, training GANs is known to be unstable and susceptible to mode collapse, as it requires the simultaneous training of two competing neural networks: a generator network that generates plausible samples to fool a discriminator network (Wiatrak et al., 2019; Kossale et al., 2022). Recently, diffusion models have emerged as state-of-the-art generative tools (Y. Song et al., 2020). Diffusion models are more stable to train than GANs and have been shown to generate fields of higher quality (Dhariwal & Nichol, 2021). These models first apply a corruption process that destroys structure in the fields from a given distribution, and then train a model that

iteratively reverses this corruption process to generate fields that follow the required distribution. The reverse process could be thought of as iterative spectral regression, where the large-scale features are predicted first, with small-scale features added iteratively (B. Wang & Vastola, 2023). Diffusion models are flexible and can be either conditional (Saharia, Chan, et al., 2022), where the model is directly trained to learn the relationship between the input and the target, or guided unconditional (Daras et al., 2024), where a pre-trained model is steered towards the target using a separate guidance process.

Due to these properties, generative diffusion models have shown promise in various inverse problems in computer vision and geophysical fluid dynamics (Milanfar & Delbracio, 2025). Sardar et al. (2024) applied conditional diffusion models while Fan et al. (2025) used a guided unconditional diffusion model with a differentiable solver for super-resolution of idealized Kolmogorov flows. Conditional models were also used to reconstruct fine-scale features from coarse neural surrogate forecasts (Oommen et al., 2025; Mardani et al., 2025), infer turbulence fields from gappy observations (T. Li et al., 2023), and estimate interior ocean states from coarse surface measurements (Souza et al., 2025). Guided unconditional diffusion models were used for generative data assimilation of coarse, noisy measurements with idealized turbulent flows (Rozet & Louppe, 2023), weather fields (Manshausen et al., 2024), and surface ocean fields (Martin et al., 2025).

To the best of our knowledge, no study has focused on both super-resolution and inference of geophysically-relevant flows with  $\beta$ -plane dynamics from coarse, sparse and gappy observations. Moreover, existing works focus on a single diffusion modeling method (conditional or unconditional) in isolation without direct comparison across methods. The sensitivity of diffusion models to guidance strength also remains underexplored in geophysical inverse problems, limiting their implementation. To address these challenges, (1) We apply four diffusion modeling (2 guided unconditional and 2 conditional) approaches to super-resolution and inference of quasi-geostrophic turbulence *under the  $\beta$ -plane approximation* at different Reynolds numbers. (2) We investigate the effect of sparse and gappy coarse observations. (3) We benchmark the four score-based diffusion models using a comprehensive suite of skill metrics including norms of the generated vorticity fields and turbulence statistics. (4) We quantify and evaluate the uncertainty of our generative models using the generated ensemble of super-resolved fields and their errors. (5) We study the sensitivity of the four diffusion models to their guidance strength (tuning parameters). Our results highlight the trade-offs between ease of implementation, fidelity (sharp-

ness), and cycle-consistency when deploying diffusion models for super-resolution and inference applications in geophysical fluid dynamics (Bennett, 1992; Wunsch, 2006).

The paper is organized as follows: Sect. (2) introduces the formal problem statement for super-resolution. Details of the various guided unconditional and conditional score-based diffusion modeling methods utilized are in Sect. (3). Sect. (4) describes the quasi-geostrophic turbulence simulations used in our applications, along with details of the dynamical regimes, numerical schemes, datasets, test cases, and skill metrics. Applications of the four diffusion model methods to super-resolution and inference of quasi-geostrophic turbulence with coarse and with sparse, gappy observations are provided in Sect. (5), followed by a discussion of the comparative analyses and conclusion (Sect. 6).

## 2 Problem Statement

The inverse problem we consider is that of super-resolution (Moser et al., 2023) or the inference of a high-resolution field from low-resolution and sparse data. We aim to obtain generative diffusion models that perform such inversions in a probabilistic sense.

We start with the observation process and define a measurement model that outputs a low-fidelity field  $y \in \mathbb{R}^d$  (the low-resolution data) from a high-fidelity field  $x_0 \in \mathbb{R}^n$ , where  $d \leq n$ , through a nonlinear observation operator  $\mathcal{H} : \mathbb{R}^n \rightarrow \mathbb{R}^d$ , where  $z \sim \mathcal{N}(0, \mathbb{I}_d)$  and  $\sigma_y$  model the noise in the observation process.

$$y = \mathcal{H}(x_0) + \sigma_y z \quad (1)$$

In general, the observation operator is not restricted to just the physical space, but could also act in the Fourier domain (wavenumber space). Observations are also often partial, with only a subset of the domain being observed. We are interested in the inverse of (1), or a model  $\mathcal{M}$  that given noisy observation samples  $y$  can generate plausible reconstructed high-fidelity samples  $\hat{x}_0$  from the conditional distribution  $p(x_0|y)$ . Because this inverse mapping is not one-to-one, it is ill-posed, and perfect recovery is impossible for many applications in geophysical modeling (Bennett, 1992; Blau & Michaeli, 2018). We thus aim for probabilistic models  $\mathcal{M}$  that can sample from the posterior distribution rather than provide only point estimates such as the conditional expectation  $\mathbb{E}[x_0|y]$ . This setting has been used for generative data assimilation, in which case  $\mathcal{H}$  is a sparse observation operator in physical space (Rozet & Louppe, 2023; Manshausen et al., 2024).

For our focus on super-resolution,  $\mathcal{H}$  corresponds to filtering and down-sampling operations, and to sparse and gappy observing systems. We aim for models that reconstruct high-resolution vorticity fields using low-resolution data generated by quasi-geostrophic turbulent models under the  $\beta$ -plane approximation. We propose using generative models for this noisy inverse problem due to their ability to effectively sample an ensemble of plausible fields that follow a required distribution (Dhariwal & Nichol, 2021). We benchmark four different generative modeling strategies with guided unconditional and conditional diffusion models, with a comprehensive suite of comparative skill metrics.

### 3 Score-Based Diffusion Models

We now describe the score-based diffusion models and their use for super-resolution and inference. Diffusion models aim to generate samples from a complex distribution  $p_0$  (e.g., the distribution of all possible fields that can be simulated using the Quasi-Geostrophic equations) using a more tractable distribution  $p_T$  (e.g., a multivariate Gaussian distribution). This generative process occurs over a pseudo-time or diffusion time  $t$  within the interval  $[0, T]$  where  $T$  is the maximum diffusion time. The term pseudo-time emphasizes that this is the artificial time variable of the generative sampling process that does not correspond to a physical time. The use of these models typically involves a forward process, which consists of gradually *noising* or corrupting samples from  $p_0$  to resemble samples of  $p_T$  over pseudo-time, and a reverse process, which consists of solving a stochastic differential equation (SDE) starting from the tractable distribution  $p_T$  to generate samples from  $p_0$ . This generation process can be conditioned on some constraints, and the sampling can be very efficient on modern computers (Sohl-Dickstein et al., 2015).

Following Y. Song et al. (2020), the forward process for a diffusion model can be represented by an Itô differential equation

$$dx_t = f_t(x_t)dt + g_t dW_t \quad \forall t \in [0, T] \quad (2)$$

with initial conditions  $x_0 \sim p_0$ , i.e., samples from the complex distribution  $p_0$ . In (2),  $x_t \in \mathbb{R}^n$  is the field generated at intermediate pseudo-times  $0 \leq t \leq T$ ,  $f$  the drift coefficient,  $g$  the diffusion coefficient, and  $W$  the standard Wiener process or Brownian motion. The subscript  $t$  highlights the pseudo-time dependence. Typically, the forms of  $f$  and  $g$  are chosen a priori and hence known. This forward process has a corresponding reverse process that can be derived analytically utilizing results from stochastic calcu-

lus (Anderson, 1982),

$$dx_t = \left[ f_t(x_t) - g_t^2 \nabla_{x_t} \left( \log p_t(x_t) \right) \right] dt + g_t dW_t \quad (3)$$

where  $\log p_t(x_t)$  is the log-probability distribution at pseudo-time  $t$ . Its gradient,  $\nabla_{x_t} \log p_t(x_t)$ , is referred to as the *score*. It appears in the deterministic drift term of the reverse diffusion process and guides it, directing noisy or corrupted samples at pseudo-time  $t$  towards regions of higher probability density of the distribution  $p_0$  (Park et al., 2024). This score has no closed-form analytical expression and is typically estimated using a deep neural network. If instead we are interested in generating fields given certain constraints (e.g., high-resolution fields that correspond to a specific low-resolution field), the resulting reverse conditional diffusion process can be derived as

$$dx_t = \left[ f_t(x_t) - g_t^2 \nabla_{x_t} \left( \log p_t(x_t|y) \right) \right] dt + g_t dW_t \quad (4)$$

where  $y$  is the conditioning field and  $\log p_t(x_t|y)$  is called the *conditional score*.

In this work, we utilize the Variance-Preserving (VP-SDE) forward process (Y. Song et al., 2020), where the marginal distribution at each step during the forward diffusion process (2) is given by

$$x_t = \mu_t x_0 + \sigma_t z, \quad 0 \leq \sigma_t \leq 1, \quad \lim_{t \rightarrow T} \sigma_t \rightarrow 1 \quad (5)$$

where  $x_0 \sim p_0$  denotes samples from the complex distribution  $p_0$ , the amplitude  $\sigma_t$  defines the *noise schedule*, the normalization factor  $\mu_t = \sqrt{1 - \sigma_t^2}$  ensures the variance of noisy samples  $x_t$  is preserved over pseudo-time  $0 \leq t \leq T$ , and  $z \sim \mathcal{N}(0, \mathbb{I}_n)$ . This equation (5) is the analytical solution of (2) under the VP-SDE assumption, and provides a computationally efficient method to generate noisy or corrupted samples at intermediate pseudo-times. Using Tweedie’s formula (Robbins, 1992), the score for the unconditional reverse process is then given by

$$\nabla_{x_t} \left( \log p_t(x_t) \right) = \frac{\mu_t \mathbb{E}[X_0 | X_t = x_t] - x_t}{\sigma_t^2} \quad (6)$$

and the conditional reverse process is given by

$$\nabla_{x_t} \left( \log p_t(x_t|y) \right) = \frac{\mu_t \mathbb{E}[X_0 | X_t = x_t, Y = y] - x_t}{\sigma_t^2} \quad (7)$$

where  $\mathbb{E}$  is the expectation operator, and  $\mathbb{E}[X_0 | X_t = x_t]$  and  $\mathbb{E}[X_0 | X_t = x_t, Y = y]$  correspond to well-trained *denoiser* neural networks,  $h_\theta(., t)$ , in practice.

For the unconditional reverse process, the *denoiser* neural network  $h_{\theta,\text{uncond.}}(x_t, t)$  with trainable parameters  $\theta$  is trained to predict the noise added given a noisy field  $x_t$  as the input. The neural network is trained to minimize the following objective:

$$\min_{\theta} \mathbb{E}_{x_0, x_t} \left\| h_{\theta,\text{uncond.}}(x_t, t) - \frac{x_t - \mu_t x_0}{\sigma_t} \right\|^2 \quad (8)$$

For the conditional models, the neural network  $h_{\theta,\text{cond.}}(x_t, y, t)$  also takes in the conditioning field ( $y$ , the observations) as an additional input, and hence the training objective is modified as

$$\min_{\theta} \mathbb{E}_{x_0, x_t, y} \left\| h_{\theta,\text{cond.}}(x_t, y, t) - \frac{x_t - \mu_t x_0}{\sigma_t} \right\|^2 \quad (9)$$

The learned score,  $s_{\theta}(\cdot, t)$ , for both conditional and unconditional models can be computed using their corresponding denoiser networks  $h_{\theta}(\cdot, t)$

$$s_{\theta}(\cdot, t) = -\frac{h_{\theta}(\cdot, t)}{\sigma_t} \quad (10)$$

Finally, for sampling, the reverse diffusion process for both unconditional (3) and conditional models (4) are discretized in time and integrated using an exponential integrator scheme (Q. Zhang & Chen, 2022; J. Song et al., 2020) with Langevin Monte Carlo corrections (Y. Song et al., 2020; Rozet & Louppe, 2023) using the learned score (10).

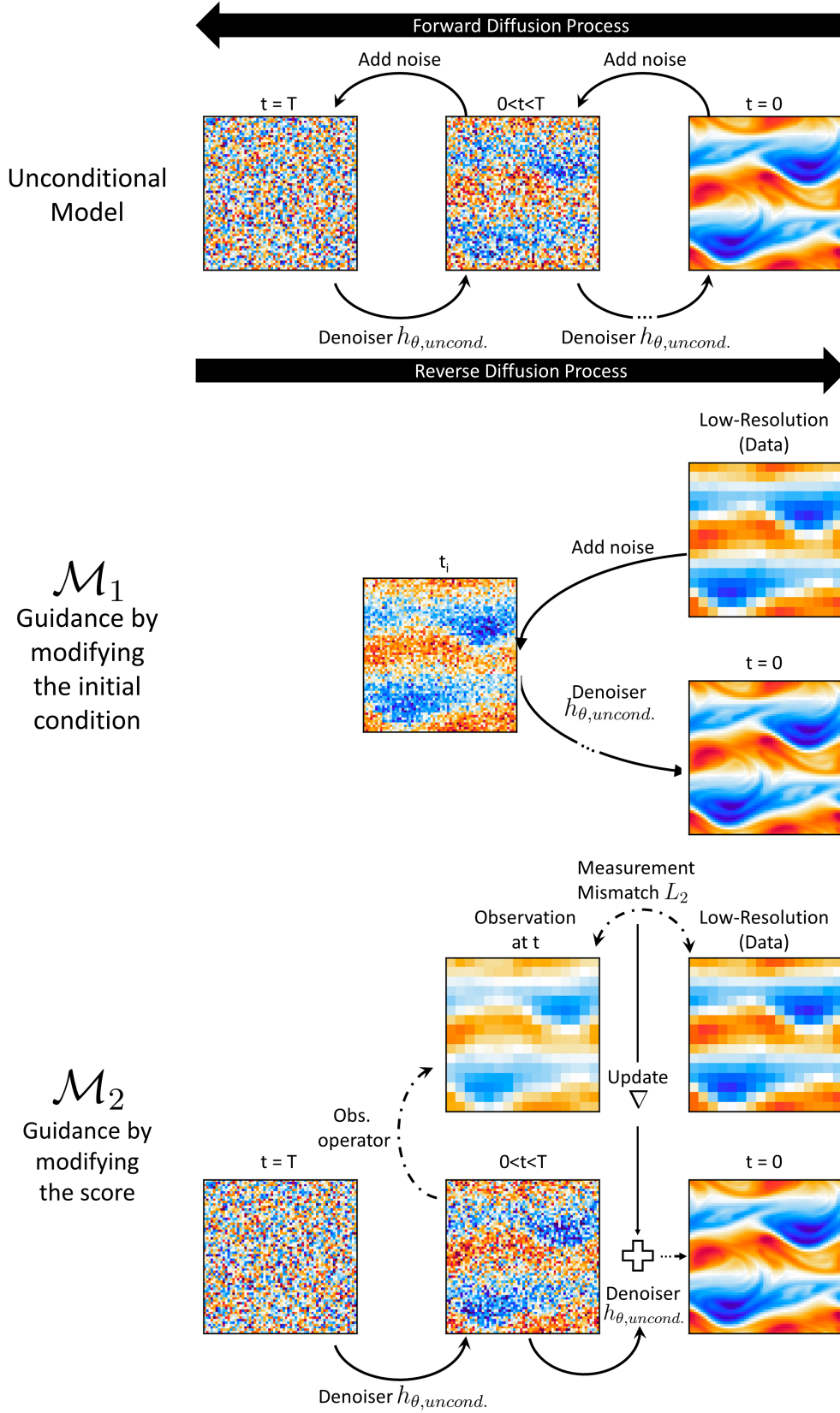
Next, we describe the four generative diffusion modeling approaches we employ and compare them inspired by the following questions: Is task-specific retraining required? Is the score modified? Is prior knowledge of the down-sampling and filtering operations required? Is the guidance controllable? Finally, we compare the number of function evaluations (neural network evaluations) per diffusion pseudo-timestep.

### 3.1 Guided Unconditional Models

One class of approaches *guides* a pre-trained unconditional diffusion model to generate conditional samples. The main advantage is that these approaches can be directly applied to a pre-trained diffusion model and require no modification to the training process, and hence save significant computational resources. Fig. (1) shows a schematic of unconditional diffusion models and two ways to guide them.

#### 3.1.1 $\mathcal{M}_1$ : *Modifying the Initial Condition (SDEdit)*

In the first approach,  $\mathcal{M}_1$ , the unconditional diffusion model is guided by simply starting the reverse diffusion process at an intermediate pseudo-time  $t_i$  and utilizing a



**Figure 1.** Schematic of the various unconditional diffusion modeling approaches. (Top row) Forward and reverse diffusion process for the unconditional model. (Middle row) Reverse process for guidance by modifying the initial condition (Sect. 3.1.1). (Bottom row) Reverse process for guidance by modifying the score (Sect. 3.1.2). -10-

new modified initial condition  $x_i$  that resembles noisy or corrupted samples generated at the intermediate pseudo-time  $t_i$ .

As we will see, this approach, also known as *SDEdit* (Meng et al., 2021), directly preserves large-scale features of interest in the forward and reverse diffusion processes. The input low-resolution data  $y \in \mathbb{R}^d$  is first interpolated to the target high-resolution to obtain  $y_{interp} \in \mathbb{R}^n$ . Next, samples from the forward process (2) with the interpolated field  $y_{interp}$  are obtained using the analytical expression (5) at the intermediate pseudo-time  $t_i$ ,  $0 < t_i < T$ ,

$$x_i = \mu_{t_i} y_{interp} + \sigma_{t_i} z \quad (11)$$

This provides initial conditions  $x_i \in \mathbb{R}^n$  for the reverse diffusion. Since this approach does not modify the score, the reverse unconditional diffusion process (3) is thus run with the unconditional score (6) (learned using (8) and (10)) from this  $x_i$  at  $t_i$  back to the initial pseudo-time  $t = 0$ . The final result is the conditional sampled high-resolution field  $\hat{x}_0$ . This approach is well-suited to super-resolution since the forward process noises out undesirable distortions that originate from the down-sampling operation while stopping it at intermediate  $t_i$  to initialize the reverse process preserves the large-scale features of the low-resolution field  $y$  that should be preserved in the generated high-resolution field. A main parameter to be fine-tuned is the intermediate time  $t_i$ , to ensure a sufficient trade-off between cycle-consistency, i.e., do the filtered generated samples  $\mathcal{H}(\hat{x}_0)$  match the low-resolution observation  $y$ , and fidelity (sharpness).

The major advantages of *SDEdit* are that it requires minimal modifications to the pre-trained diffusion model and does not require knowledge of the down-sampling operator. It does not need updates to the unconditional score since conditioning here is achieved simply by modifying the initial conditions to preserve large-scale features from the data (observations). Hence, computing additional Jacobians and gradients is not needed. In terms of computational cost, this approach requires one neural network evaluation at each step of the reverse process to compute the unconditional score. However, since the score of the unconditional model is not modified, *SDEdit* is more prone to failure compared to approaches that directly modify the score or use conditional training.

### 3.1.2 $\mathcal{M}_2$ : Modifying the Score (Diffusion Posterior Sampling)

In the second approach, the score of the unconditioned diffusion model is directly modified to generate conditional samples. The idea is to leverage Bayes' law

$$p_t(x_t|y) = \frac{p_t(y|x_t)p_t(x_t)}{p_t(y)} \quad (12)$$

and hence derive

$$\nabla_{x_t} \left( \log p_t(x_t|y) \right)_{\mathcal{M}_2} = \nabla_{x_t} \left( \log p_t(x_t) \right) + \nabla_{x_t} \left( \log p_t(y|x_t) \right) \quad (13)$$

where the first term on the right side is simply the unconditional score (6) that we have access to, and the second term is a *measurement matching* term that guides the diffusion model to generate conditional samples that are consistent with the low-resolution input field (observations). However, this measurement matching term is analytically intractable and requires further simplification for practical implementations. Daras et al. (2024) provides a comprehensive review of approaches to approximate this term. One of the most popular approaches used is Diffusion Posterior Sampling (DPS) (Chung, Kim, McCann, et al., 2023), which assumes

$$\nabla_{x_t} \left( \log p_t(y|x_t) \right) \simeq \nabla_{x_t} \left( \log p_t(y|X_0 = \mathbb{E}[X_0|X_t = x_t]) \right) \quad (14)$$

i.e., it assumes the score is the same if the conditioning is on the denoised samples instead of conditioning on the noisy or corrupted samples. Since our measurement model (1) has a Gaussian structure, the likelihood  $p_t(y|X_0 = x_0)$  can be derived as

$$p_t(y|X_0 = x_0) = \mathcal{N}(y; \mathcal{H}(x_0), \sigma_y^2 \mathcal{I}_d) \quad (15)$$

where  $\mathcal{H}$  is the nonlinear observation operator and  $\sigma_y$  models the noise in the observation process. Hence, the likelihood  $p_t(y|X_0 = \mathbb{E}[X_0|X_t = x_t])$  can be derived from (14) and (15) as

$$p_t(y|X_0 = \mathbb{E}[X_0|X_t = x_t]) \approx \mathcal{N}(y; \mathcal{H}(\mathbb{E}[X_0|X_t = x_t]), \sigma_y^2 \mathcal{I}_d) \quad (16)$$

$$\implies \nabla_{x_t} \left( \log p_t(y|x_t) \right) \simeq -\nabla_{x_t} \left( \frac{1}{2\sigma_y^2} \|y - \mathcal{H}(\mathbb{E}[X_0|X_t = x_t])\|^2 \right) \quad (17)$$

Computing (17) requires access to gradients and Jacobian-vector products if  $\mathcal{H}$  is nonlinear and if analytical closed-form expressions are inaccessible. Practically, computing (17) leads to instabilities at pseudo-times  $t$  close to 0 (Rozet & Louppe, 2023). Hence, this measurement matching term is scaled by replacing the measurement noise  $\sigma_y^2$  with a more general scaled covariance approximation  $\sigma_{y_{scaled}}^2$  that varies with the pseudo-time

through a dependence on the noise schedule  $\sigma_t$  to controllably guide the reverse diffusion process (Rozet & Louppe, 2023; Manshausen et al., 2024),

$$\sigma_{y_{scaled}}^2 = \sigma_y^2 + \frac{\sigma_t^2}{\mu_t^2} C \quad (18)$$

where  $C$  is a tunable parameter that controls the guidance strength. Hence, the final modified score is given by

$$\nabla_{x_t} \left( \log p_t(x_t|y) \right)_{\mathcal{M}_2} = \nabla_{x_t} \left( \log p_t(x_t) \right) - \frac{1}{2\sigma_{y_{scaled}}^2} \nabla_{x_t} \left( \|y - \mathcal{H}(\mathbb{E}[X_0|X_t = x_t])\|^2 \right) \quad (19)$$

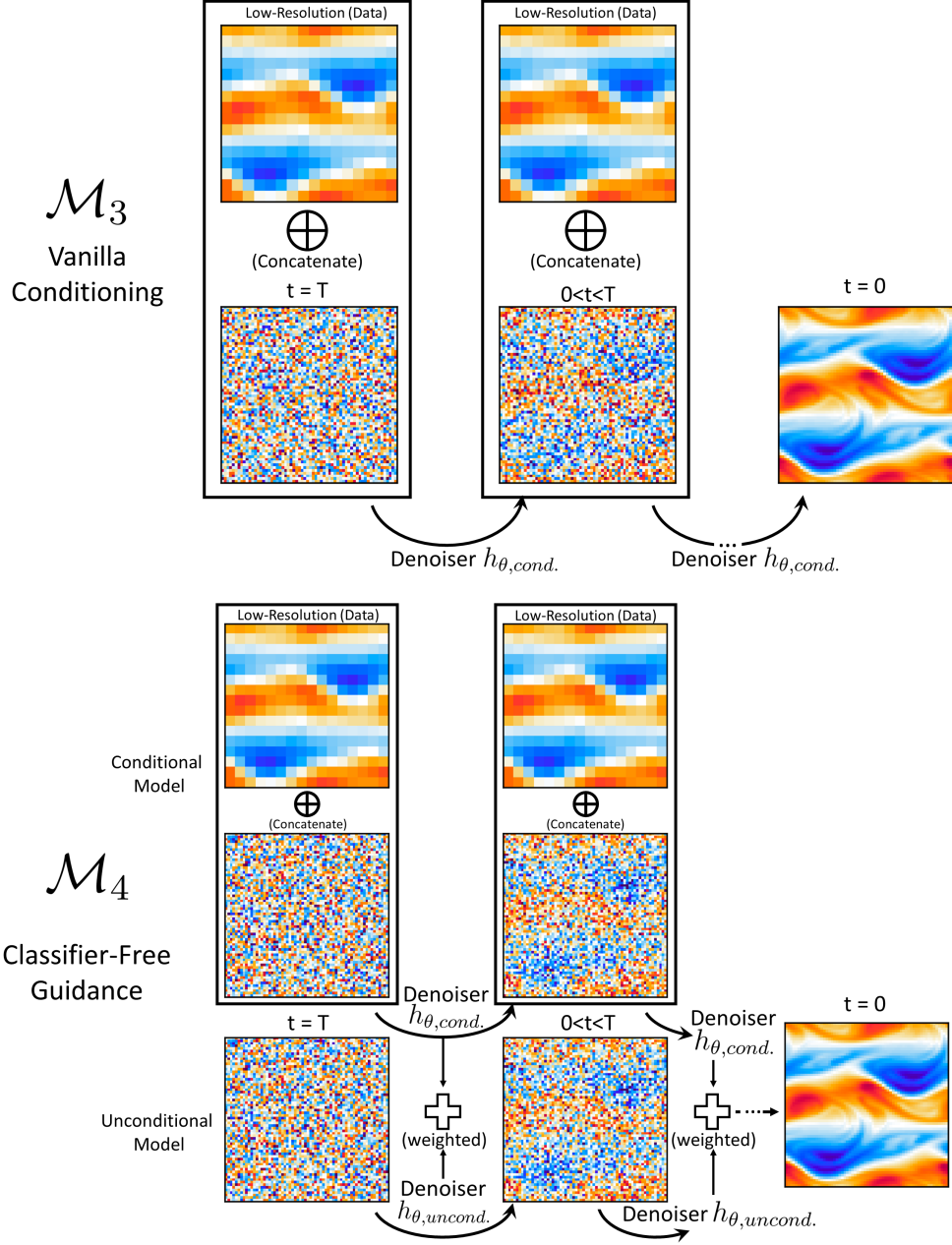
The first part of the right-hand side of (19) can be learned using (8) and (10), while the second part can be computed from (17) and (18). In summary, the second approach consists of running the reverse conditional diffusion process (4) with the score derived in (19) starting from initial conditions  $x_T \sim p_T$ , i.e, samples from the tractable distribution, and the data  $y$ , to obtain conditional sampled high-resolution fields  $\hat{x}_0$ .

The main advantage of this approach is that it provides a controllable guidance strength by directly modifying the score using the measurement matching term. However, the disadvantage is that computing this measurement matching term requires a priori knowledge of the down-sampling operator, which may not be available. This is especially true for tasks involving inference using other correlated variables, e.g., inference of ocean velocity fields from observations of sea surface height (Souza et al., 2025). In terms of cost, this approach requires one neural network evaluation at each step of the reverse process to compute the unconditional score. It also requires computing the gradients of the measurement term, which could imply training neural surrogates for the observation operator. More recently, methods such as BlindDPS have been proposed that enable posterior sampling from unknown measurement operators (Chung, Kim, Kim, & Ye, 2023).

### 3.2 Conditional Models

We now describe approaches that directly train conditional diffusion models using data pairs of the conditioning low-resolution field and target high-resolution field.

In the vanilla variant,  $\mathcal{M}_3$ , direct conditioning is typically achieved by modifying the denoiser neural network to take the low-resolution conditioning field as an additional input (9). Hence, this approach is also called *image-to-image* diffusion modeling (Saharia, Chan, et al., 2022). During sampling, the reverse conditional SDE (4) can be run using the conditional score (7) (learned using the denoiser (9) and (10)), starting from initial



**Figure 2.** Schematic of the various conditional diffusion modeling approaches (Sect. 3.2). (Top row) Reverse diffusion process for the vanilla conditional model. (Bottom row) Reverse diffusion process for classifier-free guidance.

conditions  $x_T \sim p_T$ , i.e, samples from the tractable distribution, and the data  $y$ , to obtain conditional sampled high-resolution fields  $\hat{x}_0$ . In terms of computational cost, this approach requires one neural network evaluation at each step of the reverse process to compute the conditional score.

The vanilla conditional model can further be modified with a guidance strength to further improve the quality of samples and controllably guide the diffusion model. This can be achieved using *classifier-free guidance* (CFG),  $\mathcal{M}_4$ , where the score is modified to be a linear combination of the conditional score and the unconditional score (Ho & Salimans, 2021), similarly to relaxation schemes (Lermusiaux, 2015),

$$\nabla_{x_t} \left( \log p_t(x_t|y) \right)_{\mathcal{M}_4} = \nabla_{x_t} \left( \log p_t(x_t) \right) + w \left[ \nabla_{x_t} \left( \log p_t(x_t|y) \right) - \nabla_{x_t} \left( \log p_t(x_t) \right) \right] \quad (20)$$

where the parameter  $w \geq 1$  corresponds to the guidance strength. It determines how strongly the conditional model guides the reverse process. The denoiser for classifier-free guidance can be trained from scratch by simply training a vanilla conditional model (9) with the conditioning being set randomly to a null field  $\emptyset$  (which could be a field of zeros or pure Gaussian noise) every few instances of training to resemble an unconditional diffusion model. During sampling, the reverse SDE (4) can be run using the modified conditional score (20) (learned using (8), (9), and (10)), starting from initial conditions  $x_T \sim p_T$ , i.e, samples from the tractable distribution, and the data  $y$ , to obtain conditional sampled high-resolution fields  $\hat{x}_0$ . In terms of computational cost, this approach requires two neural network evaluations at each step of the reverse process, one to compute the conditional score and another to compute the unconditional score.

The main advantage of these conditional approaches is that the model directly learns the complex relationship between the conditioning low-resolution field and the target high-resolution field. Hence, these methods are particularly powerful when the conditioning fields are mutually informative fields obtained using highly nonlinear operators. The main disadvantage is that they require task-specific re-training of the diffusion model, since the conditioning field varies based on the task (e.g., super-resolution or inference from a field obtained using a different observation operator).

Table (1) summarizes the differences among the four diffusion modeling approaches we consider and answers the questions we had asked.

**Table 1.** Comparison of implementation and sampling characteristics of the four diffusion modeling approaches considered for super-resolution and inference.

|                                             | Guided unconditional models |                 | Conditional models |                 |
|---------------------------------------------|-----------------------------|-----------------|--------------------|-----------------|
|                                             | $\mathcal{M}_1$             | $\mathcal{M}_2$ | $\mathcal{M}_3$    | $\mathcal{M}_4$ |
| Task-specific re-training                   | ✗                           | ✗               | ✓                  | ✓               |
| Modifies the score                          | ✗                           | ✓               | ✓                  | ✓               |
| Avoids need of observation operator         | ✓                           | ✗               | ✓                  | ✓               |
| Controllable guidance                       | ✓                           | ✓               | ✗                  | ✓               |
| Number of function evaluations <sup>a</sup> | $\frac{t_i}{T}$             | 1               | 1                  | 2               |

<sup>a</sup> Relative number of denoiser evaluations per pseudo-timestep during the reverse diffusion process (sampling).  
 $\mathcal{M}_1$ : Modifying the initial condition (Sect. 3.1.1),  $\mathcal{M}_2$ : Modifying the score (Sect. 3.1.2)  
 $\mathcal{M}_3$ : Vanilla conditional model (Sect. 3.2),  $\mathcal{M}_4$ : Classifier-free guidance (Sect. 3.2)

## 4 Quasi-Geostrophic Turbulence and Data

We now describe our high-resolution quasi-geostrophic turbulence simulations, including the dynamical regimes, numerical schemes, and parameters (Sects. 4.1–4.2). The measurement models we employ (Sect. 4.3) then apply Large-Eddy Simulation (LES)-style spectral filtering and sparse observation operators to obtain low-resolution data (the observations). In Sect. (4.4), we describe the super-resolution and inference test cases and the high- and low-resolution data sets that are used to train the score-based diffusion models. In Sect. (4.5), we introduce a comprehensive suite of skill-metrics used to assess the performance of the score-based diffusion models. Once trained, these models will be used to reconstruct high-resolution fields from low-resolution data (Sect. 5).

### 4.1 Quasi-Geostrophic Dynamics

As a testbed for our generative models, we use the two-dimensional (2D) incompressible one-layer Quasi-Geostrophic (QG) equations, an idealized approximation for large-scale geophysical flows on a rotating surface (Vallis, 2017). These partial differential equations (PDEs) are widely used to simulate atmospheric and oceanic flows, and have been utilized for multiple machine learning applications such as surrogate modeling (Z. Li et al., 2020) and subgrid-scale closure modeling (Frezat et al., 2022; Ross et al., 2023; Guan et al., 2024; Srinivasan et al., 2024) under various approximations.

In 2D, the non-dimensionalized QG PDE can be written in vorticity form as

$$\frac{\partial \omega}{\partial t} + J(\psi, \omega) = \frac{1}{Re} \nabla^2 \omega - \mu \omega - \beta \frac{\partial \omega}{\partial x} + F \quad (21)$$

where  $\omega$  is the vorticity, which is our 2D field of interest, and  $\psi$  is the stream function. In this governing PDE (21), the nonlinear term  $J(\psi, \omega)$  represents advection of vortic-

ity by the flow itself. The diffusion term  $\frac{1}{Re}\nabla^2\omega$  represents turbulent eddy diffusivity, where  $Re$  is the Reynolds number. The  $\beta$ -term accounts for latitudinal variation of the Coriolis force due to the Earth’s rotation, while the linear damping term  $\mu\omega$  models linear bottom drag (Stommel, 1948). The forcing  $F$  injects energy at prescribed length scales to mimic sustained geophysical driving. Since the flow is assumed to be incompressible, the velocity field  $(u, v)$  can be recovered from the stream function via

$$(u, v) = \left(-\frac{\partial\psi}{\partial y}, \frac{\partial\psi}{\partial x}\right) \quad (22)$$

$$\omega = \nabla^2\psi \quad (23)$$

To drive the system and set dominant length scales, we apply a periodic forcing,

$$F = k_f[\cos(k_f x) + \cos(k_f y)] \quad (24)$$

where  $k_f$  is the forcing wavenumber (Guan et al., 2023). We consider all terms in (21) to simulate forced-dissipative turbulence and its vorticity dynamics.

We emphasize two flow regimes: the *eddy* regime and the *jet* regime (Ross et al., 2023). The *eddy* regime corresponds to  $\beta = 0$ , where the flow is isotropic and comprised of large coherent vortices. The *jet* regime corresponds to  $\beta > 0$ , where the flow consists of alternatively banded jets. In the *jet* regime, the  $\beta$ -effect forces the flow to organize into Rossby waves, creating a barrier to the upscale turbulent cascade of energy, leading to the destruction of coherent vortices (Rhines, 1975). This interaction between Rossby waves and Quasi-Geostrophic turbulence is known as the Rhines effect, and has been observed in the Earth’s oceans (Galperin et al., 2004) and planetary circulations of Jupiter and Saturn (Williams, 1978; Theiss, 2006).

## 4.2 Numerical Schemes and Parameters

We simulate 2D vorticity fields by solving (21) on a square domain of dimensions  $[0, 2\pi] \times [0, 2\pi]$  with periodic boundary conditions. We numerically solve (21) using a pseudo-spectral method (Orszag & Israeli, 1974), with full 2/3 de-aliasing. For numerical stability, we use a semi-implicit second-order Adams-Bashforth-Crank-Nicolson (AB2-CN) scheme for time-stepping, where the diffusion, Coriolis, and bottom drag terms are treated implicitly while the advection and forcing terms are treated explicitly (Boyd, 2001).

Following Graham and Ringler (2013); Frezat et al. (2022), we use the following scales for non-dimensionalization, corresponding to ocean mesoscales: a length-scale of

$\frac{504}{\pi} \times 10^4$  m and a time-scale of  $1.2 \times 10^6$  s. We simulate fully-resolved (FR) vorticity fields using a  $N_{FR} \times N_{FR} = 512 \times 512$  numerical grid ( $\Delta_{FR} = \frac{2\pi}{N_{FR}}$ ). To initialize simulations, we follow Guan et al. (2022) and sample initial vorticity fields from a standard normal distribution  $\omega \sim \mathcal{N}(0, 1)$  with wavenumbers  $k \in [3, 10]$ , omitting small wavenumbers for numerical stability. We vary the Reynolds number between  $Re = 10^3$  and  $Re = 10^4$ . Higher values of the Reynolds number increase the *richness* of turbulence, leading to a wider range of multi-scale interactions (Pope, 2001). For the *eddy* regime, we set  $\beta = 0$ , and for the *jet* regime, we set  $\beta = 2.5$ . For time-stepping, we use a non-dimensional time-step  $dt$  a function of  $Re$  and simulate a total time of  $T = 100$ . A list of non-dimensional parameters is provided in Table (2).

**Table 2.** Parameters of the different fully-resolved (FR) forced-dissipative Quasi-Geostrophic (QG) simulations for the eddy and jet regimes. All parameters are expressed in non-dimensional units defined in Sect. 4.2 (length-scale of  $\frac{504}{\pi} \times 10^4$  m and time-scale of  $1.2 \times 10^6$  s).

| Regime | Re     | $\beta$ | dt                 | $\mu$              | $k_f$ |
|--------|--------|---------|--------------------|--------------------|-------|
| Eddy   | $10^3$ | 0       | $1 \times 10^{-3}$ | $2 \times 10^{-2}$ | 2     |
| Eddy   | $10^4$ | 0       | $5 \times 10^{-4}$ | $2 \times 10^{-2}$ | 2     |
| Jet    | $10^3$ | 2.5     | $1 \times 10^{-3}$ | $2 \times 10^{-2}$ | 2     |
| Jet    | $10^4$ | 2.5     | $5 \times 10^{-4}$ | $2 \times 10^{-2}$ | 2     |

### 4.3 Down-Sampling and Sparse Observation Operators

For down-sampling, we apply coarsening operators on the vorticity fields obtained from the fully-resolved (FR) simulations. This coarsening operator could be of various types, and could act either in the physical space or the wavenumber space (Fourier domain). Down-sampling operations in physical space, such as max-pooling or average-pooling, have been considered in previous work (Fukami et al., 2019), but these correspond to non-local box filters in the wavenumber space, leading to spurious oscillations in the kinetic energy spectrum at high wavenumbers. In this work, we use spectral filtering operators utilized in Large-Eddy Simulations (LES) (Pope, 2001). We apply our down-sampling operator ( $\mathcal{H}_{\text{full}}$ ) in wavenumber space by first applying a Gaussian filter ( $\mathcal{H}_{\text{Gaussian}}$ ), followed by a cut-off filter ( $\mathcal{H}_{\text{cut-off}}$ ), and then interpolate the resultant field to obtain the

observed field (OF) on the low-resolution grid with resolution  $\Delta_{OF} = \frac{2\pi \times \delta}{N_{FR}}$ :

$$\begin{aligned}\mathcal{H}_{\text{Gaussian}}(k_{FR}) &= \exp\left(\frac{-k_{FR}^2 (\delta \times \Delta_{FR})}{6}\right) \\ \mathcal{H}_{\text{cut-off}}(k_{FR}) &= 0, \quad \forall k_{FR} > \pi \Delta_{OF} \\ \mathcal{H}_{\text{full}} &\triangleq \text{interpolate}_{\Delta_{OF}} \circ \mathcal{H}_{\text{cut-off}} \circ \mathcal{H}_{\text{Gaussian}}\end{aligned}\tag{25}$$

where  $k_{FR}$  is the wavenumber corresponding to the fully-resolved grid and  $\delta$  is the down-sampling scale. The Gaussian filter smoothly attenuates high wavenumber content, resembling subgrid-scale energy removal, while the sharp cut-off filter enforces a strict wavenumber limit. This combination retains large-scale coherent structures while discarding small-scale features. Such filtering is crucial to pose a realistic and physically meaningful super-resolution problem for geophysical turbulence (Z. Zhou et al., 2019; Srinivasan et al., 2024).

We also study the additional effect of sparse and gappy observing systems ( $\mathcal{H}_{\text{sparse}}$ ) by defining a set of unobserved regions  $\Omega$  and using a mask operator in physical space,

$$\begin{aligned}\mathcal{H}_{\text{partial}}(x, y) &= \omega_{fill}, \quad \forall x, y \in \Omega \\ \mathcal{H}_{\text{sparse}}(x, y) &= \mathcal{H}_{\text{partial}} \circ \mathcal{H}_{\text{full}}\end{aligned}\tag{26}$$

where  $\omega_{fill}$  is a fill value corresponding to the unobserved regions. In our case, we set the fill value to 0, which is the mean of our vorticity fields over space and time. Martin et al. (2025); T. Li et al. (2023) considered a similar partial observation operator to emulate satellite observations and study flow reconstruction with *gappy* high-resolution data, respectively; however, we emphasize that here we apply these partial observations directly to the coarse-resolution output of the down-sampling operators defined in this section.

#### 4.4 Data and Test Cases

To generate data for training, we simulate 500 different trajectories of the fully-resolved (FR) QG simulations ( $N_{FR} = 512$ ) with different random initial conditions by varying the random seed. Training diffusion models is very *data-intensive*, since it requires sampling from the entire distribution of flow fields (Z. Wang et al., 2023). Hence, such large-dimensional flow fields require a lot of simulation data and computational resources for training. Therefore, we restrict our experiments to fields of dimensions  $N_{FF} \times N_{FF} = 64 \times 64$ , which are obtained by applying the down-sampling operator (25) with  $\delta = 2^3$  to these simulations. These new down-sampled fields, i.e, the *filtered fields* (FF) serve as the high-resolution ground truth for our super-resolution experiments. This re-

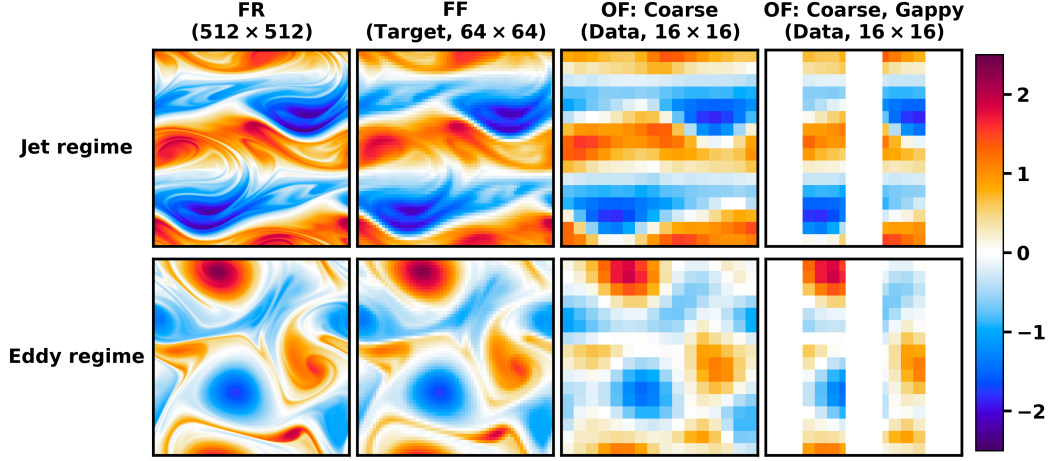
duces the number of snapshots and hence computational costs (in terms of both memory and number of operations) required for training. Alternatively, the size of the high-resolution target can also be restricted by simply selecting fields from a smaller window of the original fully-resolved fields. However, the size of this window needs to be larger than the largest eddy of the simulated vorticity field (Fukami et al., 2019).

Snapshots (frames) of the vorticity field are saved every  $\Delta T_{save} = 0.5$  non-dimensional time units, resulting in  $\frac{T}{\Delta T_{save}} = 200$  snapshots. Upon analyzing the energy spectrum, we observe self-similarity after  $t = 50$ , and only utilize snapshots obtained post this time for our training and analysis (last 100 snapshots). Hence, our generative models are trained with  $500 \times 100 = 50,000$  snapshots.

To study the performance of generative models (Table 1), we utilize test cases with two observation operators for all the simulations in Table (2): down-scaling by a scale of  $\delta = 2^2$  with full observations (25), and a similar down-scaling with partial observations (26). To study the effect of sparse, gappy observations, we restrict the observed region  $\Omega$  to cover 40% of the total domain. Utilizing the measurement model (1) with small Gaussian observation noise  $\sigma_y = 0.01$ , these test cases yield observed fields (OF) with dimensions  $N_{OF} \times N_{OF} = 16 \times 16$ , which serve as the conditioning field (input) to our diffusion models. Table (3) summarizes the 8 test cases considered. We perform super-resolution at a fixed time, utilizing observed fields obtained from the  $64 \times 64$  filtered vorticity fields of a trajectory initialized with a new initial condition, which was not utilized for training. A visual representation of test cases 2, 4, 6, and 8 is shown in Fig. (3).

**Table 3.** Main test cases utilized for super-resolution and inference with diffusion models corresponding to target dimensions ( $N_{FF} = 64 \times 64$ ) and observed field dimensions ( $N_{OF} = 16 \times 16$ ), with resultant down-sampling scale ( $\delta = 2^2$ ). The effect of other down-sampling scales ( $\delta \in 2^1, 2^3$ ) are discussed in Sect. (5.1).

| Number | Regime | Re     | Data (Observations)      |
|--------|--------|--------|--------------------------|
| 1      | Jet    | $10^3$ | Coarse-resolution field  |
| 2      | Jet    | $10^4$ | Coarse-resolution field  |
| 3      | Eddy   | $10^3$ | Coarse-resolution field  |
| 4      | Eddy   | $10^4$ | Coarse-resolution field  |
| 5      | Jet    | $10^3$ | Coarse and sparse, gappy |
| 6      | Jet    | $10^4$ | Coarse and sparse, gappy |
| 7      | Eddy   | $10^3$ | Coarse and sparse, gappy |
| 8      | Eddy   | $10^4$ | Coarse and sparse, gappy |



**Figure 3.** Test case description for super-resolution and inference in the jet (Top row) and eddy (Bottom row) regimes at  $Re = 10^4$ . Column 1: Fully-resolved (FR) vorticity field. Column 2: Filtered-field (FF), which serves as the target for super-resolution. Column 3: Observed field (OF) under coarse-resolution observations. Column 4: Observed field (OF) under coarse, sparse and gappy observations. The observed fields serve as data (input) for super-resolution.

#### 4.5 Skill Metrics

To assess the performance of the score-based diffusion models, we utilize three sets of complementary skill metrics: one is related to the generated vorticity fields, the second to statistical turbulence quantities, and the third to uncertainty quantification.

First, we assess the generated super-resolved fields using common norms such as the relative  $L_2$  norm:  $\frac{\|\omega_{FF} - \hat{\omega}_0\|_2}{\|\omega_{FF}\|_2}$  and measures of the accuracy of features of the reconstructed vorticity field (e.g., filament sharpness). We then check for cycle-consistency using the relative norm:  $\frac{\|\omega_{OF} - \mathcal{H}(\hat{\omega}_0)\|_2}{\|\omega_{OF}\|_2}$ , i.e., does the filtered high-resolution diffusion output match the low-resolution input field? Cycle-consistency is essential to ensure that the observed field can indeed be obtained from the generated super-resolved field through the measurement model (T. Zhou et al., 2016).

Second, we compare time-averaged statistical quantities relevant to 2D turbulence. Following Frezat et al. (2022), we compute the angle-averaged kinetic energy spectrum  $E(k)$  and enstrophy spectrum  $Z(k)$  in the wavenumber space using

$$E(k) = \frac{1}{2} \int_{|\mathbf{k}|=k} \hat{\mathbf{u}}_i(\mathbf{k}) \hat{\mathbf{u}}_i^*(\mathbf{k}) dS(\mathbf{k}) \quad (27)$$

$$Z(k) = \frac{1}{2} \int_{|\mathbf{k}|=k} \hat{\omega}(\mathbf{k}) \hat{\omega}^*(\mathbf{k}) dS(\mathbf{k}) \quad (28)$$

where  $k = \sqrt{k_x^2 + k_y^2}$  denotes the horizontal angular wavenumber,  $\hat{\bullet}$  the Fourier transform, and  $\bullet^*$  the complex conjugate. We also compute the relative  $L_2$  norms of the log-spectra. Because the  $\beta$ -effect induces anisotropic spectral energy transfers that are obscured by angle-averaging (Huang et al., 2001), we then compute contours of the pre-multiplied 2D kinetic energy spectrum,

$$E(k_x, k_y) = \lambda_x \lambda_y \frac{1}{2} \hat{\mathbf{u}}_i(\mathbf{k}_x, \mathbf{k}_y) \hat{\mathbf{u}}_i^*(\mathbf{k}_x, \mathbf{k}_y) \quad (29)$$

where  $\lambda_x = \frac{2\pi}{k_x}$  and  $\lambda_y = \frac{2\pi}{k_y}$  are wavelength factors to account for the logarithmic spacing in the wavenumber space (Towne et al., 2020). To study the non-Gaussian one-point statistics of turbulent flows, we compute the one-dimensional probability distribution function for the vorticity field using kernel density estimation (Wkeglarczyk, 2018).

Third, as our diffusion models can generate an ensemble of super-resolved vorticity fields, they can be utilized for uncertainty quantification (Leinonen et al., 2023; Chen et al., 2025). To quantify and verify the uncertainty, we compute the ensemble mean,  $\bar{\omega}_0 = \frac{1}{N_{ens}} \sum_{i=1}^{N_{ens}} \omega_0^i$ , ensemble standard deviation:  $\sqrt{\frac{\sum_{i=1}^{N_{ens}} (\bar{\omega}_0 - \omega_0^i)^2}{N_{ens}}}$ , and root-mean squared error (RMSE) of the generated super-resolved fields:  $\sqrt{\frac{\sum_{i=1}^{N_{ens}} \|\omega_{FF} - \omega_0^i\|_2^2}{N_{ens}}}$ , where  $N_{ens} = 16$  is the ensemble size and the superscript  $i$  indicates the  $i^{\text{th}}$  ensemble member.

## 5 Applications

To train the diffusion models, we utilize a U-Net architecture (Ronneberger et al., 2015). To provide conditioning, we concatenate the interpolated low-resolution field with the noisy field at each diffusion time-step following Saharia, Ho, et al. (2022). Details of neural architectures, training, and associated hyperparameters are in Appendix A.

Next, we present the results of the four generative diffusion models (Sect. 3) for super-resolution of turbulent vorticity fields from coarse-resolution fields (Sect. 5.1) and then from coarse, sparse, and gappy observations (Sect. 5.2). All these results are obtained using best-tuned intermediate times and guidance strengths. In Sect. (5.3), we study the sensitivity of the four diffusion models to their main tuning parameters.

## 5.1 Application 1: Super-resolution from Coarse-resolution Fields

We now discuss applications with coarse-resolution vorticity fields as the observed field (data or input), corresponding to Cases 1-4 of Table (3).

### 5.1.1 Jet Regime

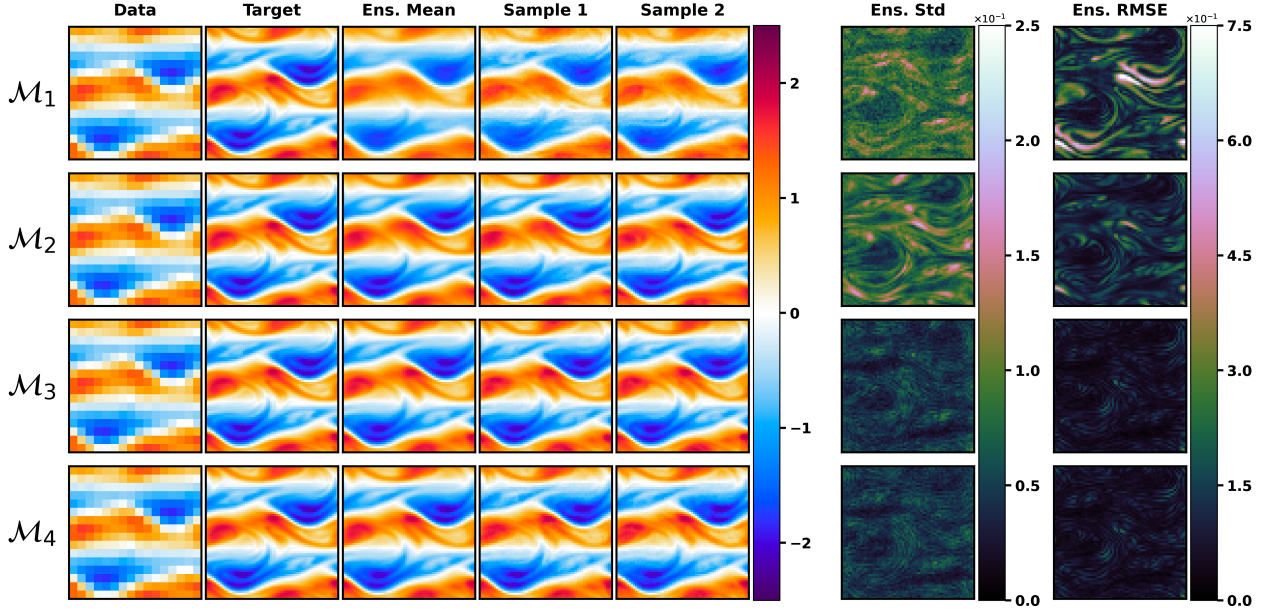
In the jet regime (Cases 1-2 of Table (3)), strong  $\beta$ -effects organize the flows into anisotropic zonal bands, leading to a more organized distribution of vorticity compared to the eddy regime. Table (4) summarizes the key quantitative skill metrics: relative error norms of the reconstructed fields, cycle-consistency error, and log-spectra, as well as the ensemble standard deviation of the reconstructed field.

**Table 4.** Comparison of the key quantitative skill metrics for super-resolution in the jet regime from coarse-resolution fields (Cases 1-2 of Table (3)).

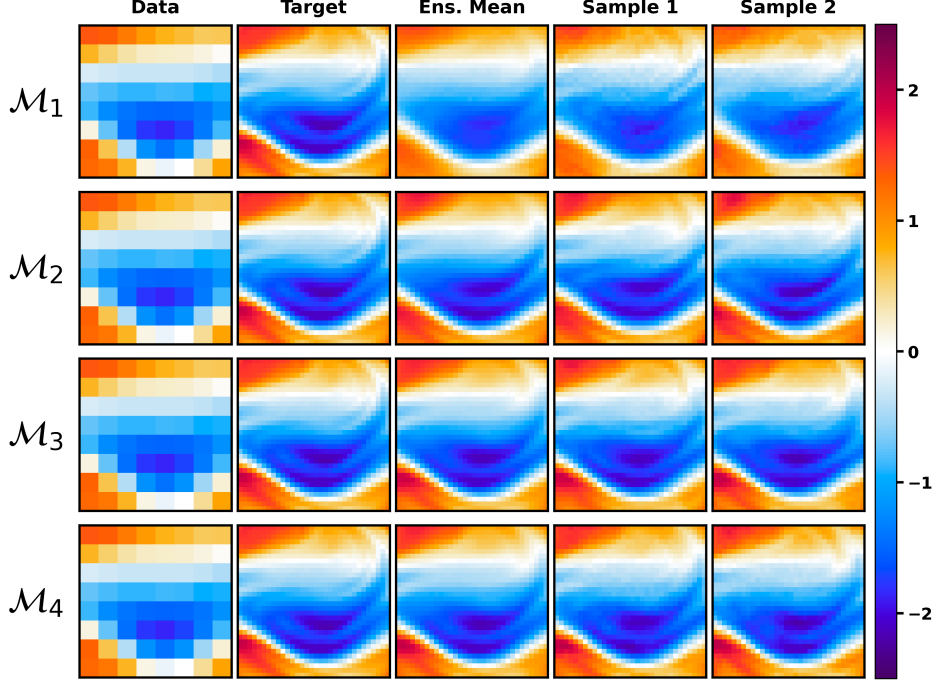
| Relative Error Metric                    | Scale              | Test Case 1 ( $Re = 10^3$ ) |                 |                                   |                 | Test Case 2 ( $Re = 10^4$ ) |                                   |                                   |                                   |
|------------------------------------------|--------------------|-----------------------------|-----------------|-----------------------------------|-----------------|-----------------------------|-----------------------------------|-----------------------------------|-----------------------------------|
|                                          |                    | $\mathcal{M}_1$             | $\mathcal{M}_2$ | $\mathcal{M}_3$                   | $\mathcal{M}_4$ | $\mathcal{M}_1$             | $\mathcal{M}_2$                   | $\mathcal{M}_3$                   | $\mathcal{M}_4$                   |
| Reconstructed vorticity ( $\downarrow$ ) | $(\times 10^{-1})$ | $1.05 \pm 0.08$             | $0.45 \pm 0.07$ | <b><math>0.19 \pm 0.01</math></b> | $0.20 \pm 0.01$ | $2.13 \pm 0.09$             | $1.29 \pm 0.07$                   | <b><math>0.69 \pm 0.05</math></b> | $0.71 \pm 0.02$                   |
| Ensemble Field Std.                      | $(\times 10^0)$    | 0.05                        | 0.04            | 0.01                              | 0.01            | 0.09                        | 0.09                              | 0.05                              | 0.05                              |
| Cycle-consistency ( $\downarrow$ )       | $(\times 10^{-1})$ | $0.77 \pm 0.05$             | $0.30 \pm 0.05$ | <b><math>0.04 \pm 0.01</math></b> | $0.05 \pm 0.01$ | $1.31 \pm 0.07$             | $0.61 \pm 0.06$                   | $0.05 \pm 0.03$                   | <b><math>0.04 \pm 0.00</math></b> |
| Log-energy spectrum ( $\downarrow$ )     | $(\times 10^{-2})$ | $11.09 \pm 1.16$            | $4.14 \pm 0.15$ | <b><math>3.41 \pm 0.40</math></b> | $3.67 \pm 0.25$ | $7.29 \pm 0.74$             | <b><math>1.81 \pm 0.18</math></b> | $2.11 \pm 0.93$                   | $2.36 \pm 0.37$                   |
| Log-ensrophy spectrum ( $\downarrow$ )   | $(\times 10^{-2})$ | $17.09 \pm 1.78$            | $6.38 \pm 0.23$ | <b><math>5.25 \pm 0.61</math></b> | $5.65 \pm 0.38$ | $11.85 \pm 1.21$            | <b><math>2.95 \pm 0.29</math></b> | $3.43 \pm 1.51$                   | $3.85 \pm 0.61$                   |

Error metrics are shown as ensemble mean  $\pm$  one standard deviation; the best-performing model for each metric is highlighted in **bold**.

*Super-resolved vorticity fields:* For both  $Re = 10^3$  and  $Re = 10^4$  (Fig. 4), the vanilla conditional model ( $\mathcal{M}_3$ ) and the classifier-free guidance approach ( $\mathcal{M}_4$ ) achieve the lowest reconstruction error, followed by the guided unconditional models: guidance by modifying the score (DPS,  $\mathcal{M}_2$ ) and guidance by modifying the initial conditions (SDEdit,  $\mathcal{M}_1$ ). All four methods capture the larger-scale forced features of the target field. From Fig. (4) and zoomed-in Fig. (5),  $\mathcal{M}_1$  leads to overly smooth fields, lacking fine-scale filamentary features of high-vorticity amplitude since it does not sample from the conditional reverse process (4), and relies on just the large-scale feature of the data (11).  $\mathcal{M}_2$  improves on this by capturing thin low-amplitude vorticity streaks between zonal bands but still lacks some fine-scale features since it only approximates the conditional distribution (19) with the assumption (14). The conditional models,  $\mathcal{M}_3$  and  $\mathcal{M}_4$ , generate super-resolved fields qualitatively very similar to the target, since they sample directly from the conditional reverse process (4), capturing vorticity streaks as well as the sharp, narrow fine-scale filaments within the zonal bands.



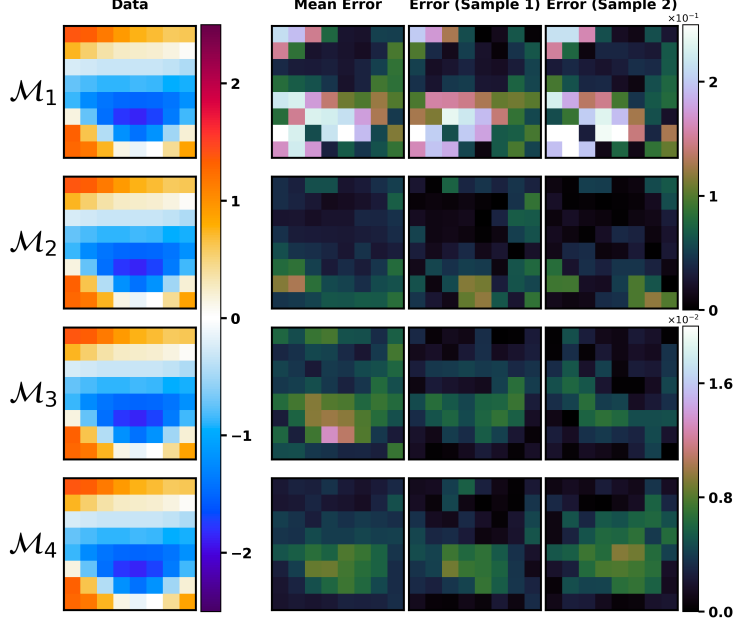
**Figure 4.** Reconstructed vorticity fields of the four approaches for super-resolution in the jet regime ( $\text{Re}=10^4$ ) from coarse-resolution fields (Test Case 2, Table 3) for one snapshot. Column 1: Coarse observed vorticity field (data, OF). Column 2: Corresponding filtered vorticity field (target, FF). Column 3: Ensemble mean of the super-resolved diffusion model outputs. Columns 4 and 5: Two representative ensemble members, highlighting the probabilistic outputs of diffusion models. Columns 6 and 7: Point-wise standard deviation (Std.) and root-mean squared error (RMSE) of the ensemble predictions, respectively. Data and target are identical for all models.



**Figure 5.** Zoomed-in view of the upper-right quadrant for the first 5 columns of Fig. (4), highlighting fine-scale features. Data and target are identical for all models.

*Uncertainty quantification:* The conditional models  $\mathcal{M}_3$  and  $\mathcal{M}_4$  exhibit the smallest standard deviation and RMSE ( $N_{ens} = 16$ ), indicating accurate predictions with tight uncertainty bounds. From Fig. (4), the standard deviation and RMSE for all models (except  $\mathcal{M}_1$ ) show similar spatial patterns, indicating that these models can serve as well-calibrated probabilistic estimators, providing both accurate mean predictions and spatially consistent uncertainty bounds. Similar spatial patterns for vanilla conditional models were observed by Souza et al. (2025).

*Cycle-consistency:*  $\mathcal{M}_1$  has region-dependent cycle-inconsistency (Fig. 6) as it samples from the unconditional (3), and not the conditional (4), reverse diffusion process.  $\mathcal{M}_2$  is half as inconsistent as  $\mathcal{M}_1$  since it approximates the conditional distribution through the measurement-matching term (17). Unfortunately,  $\mathcal{M}_2$  assumes that the conditional expectation from the noisy and denoised samples are equivalent (14), which fails near sharp image gradients generated towards the end of the reverse process (Fig. 6) and breaks cycle-consistency. Its error thus remains an order of magnitude larger than the conditional models (Table 4). Despite a lack of explicit guidance, the two conditional models  $\mathcal{M}_3$  and  $\mathcal{M}_4$  are extremely cycle-consistent, as they sample directly from the true

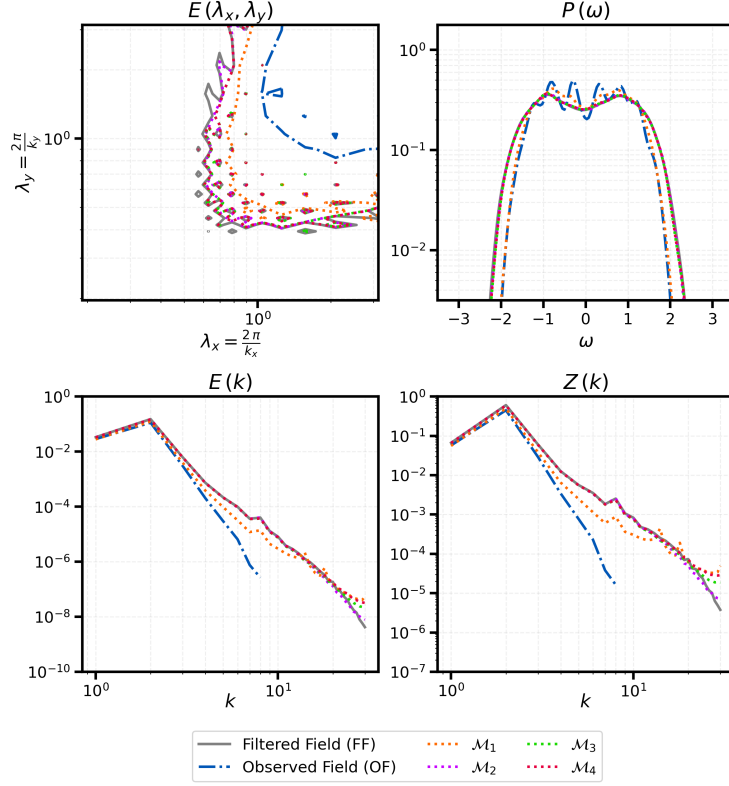


**Figure 6.** Cycle-consistency of the four diffusion modeling approaches for super-resolution in the jet regime ( $\text{Re}=10^4$ ) from coarse-resolution fields (Test Case 2, Table 3). Columns 1–4 show zoomed-in views of the upper-right quadrant. Column 1: Coarse observed vorticity field (Data). Column 2: Ensemble mean of the errors of the filtered super-resolved diffusion output. Columns 3 and 4: Errors of 2 filtered ensemble members. The upper color bar limit of the error fields for  $\mathcal{M}_1$  and  $\mathcal{M}_2$  differs from that for  $\mathcal{M}_3$  and  $\mathcal{M}_4$  by an order of magnitude.

conditional distribution, (4) and (20), respectively. While tuning the guidance strength of  $\mathcal{M}_4$  could also lead to sampling from unphysical distributions (Karras et al., 2024), additional conditioning helps avoid some artifacts seen with  $\mathcal{M}_3$ .

*Time-averaged 2D turbulence statistics:* In Fig. 7 (Top-right), the  $\beta$ -effect leads to a non-Gaussian bimodal distribution of the one-point vorticity field, which is accurately captured by  $\mathcal{M}_2$ ,  $\mathcal{M}_3$ , and  $\mathcal{M}_4$ . Model  $\mathcal{M}_1$  is skewed towards the data (since it is initialized with noisy observations), and does not capture the extremes of the distribution. Similarly, in Fig. 7 (Top-left),  $\mathcal{M}_2$ ,  $\mathcal{M}_3$ , and  $\mathcal{M}_4$  capture the characteristic *jagged* pre-multiplied 2D kinetic energy spectrum shape, but smooth the spectrum at large wavenumbers.  $\mathcal{M}_1$  loses the jaggedness and is again skewed towards the data. From Fig. 7 (Bottom),  $\mathcal{M}_2$  accurately captures the angle-averaged spectra at all wavenumbers, including smaller scales (i.e., the tails).  $\mathcal{M}_3$  and  $\mathcal{M}_4$  match the scaling of the target spectra, but with higher energy at the tails.  $\mathcal{M}_1$  is once again skewed by the data and shows steeper

scaling compared to the target spectra. For lower  $Re = 10^3$  (figures not shown for brevity), the two conditional models capture the spectra better than  $\mathcal{M}_2$  at all wavenumbers.



**Figure 7.** Time-averaged statistical quantities of the four models for super-resolution in the jet regime ( $Re=10^4$ ) from coarse-resolution fields (Test Case 2, Table 3). Top Left: Premultiplied 2D kinetic energy spectrum contours that capture 95% of the energy. Top Right: 1D point-wise probability distribution function for the vorticity field. Bottom Left and Right: Angle-averaged kinetic energy and enstrophy spectra, respectively. Solid lines correspond to the filtered field (target), dash-dot lines to the observed field (data), and dotted lines to the diffusion model outputs.

### 5.1.2 Eddy Regime

In the eddy regime (Cases 3-4 of Table (3)), the absence of  $\beta$ -effects produces nearly isotropic dynamics, leading to a more complex distribution of the vorticity fields compared to the jet regime. Table (5) summarizes the key quantitative skill metrics.

*Super-resolved vorticity fields:* As in the jet regime, for both  $Re = 10^3$  and  $Re = 10^4$  (Fig. 8),  $\mathcal{M}_3$  and  $\mathcal{M}_4$  achieve the lowest reconstruction errors, followed by  $\mathcal{M}_2$  and

**Table 5.** As Table (4) but in the eddy regime from coarse-resolution fields (Cases 3-4 of Table (3)).

| Relative Error Metric                    | Scale              | Test Case 3 ( $Re = 10^3$ ) |                                   |                                   |                                   | Test Case 4 ( $Re = 10^4$ ) |                 |                                   |                                   |
|------------------------------------------|--------------------|-----------------------------|-----------------------------------|-----------------------------------|-----------------------------------|-----------------------------|-----------------|-----------------------------------|-----------------------------------|
|                                          |                    | $\mathcal{M}_1$             | $\mathcal{M}_2$                   | $\mathcal{M}_3$                   | $\mathcal{M}_4$                   | $\mathcal{M}_1$             | $\mathcal{M}_2$ | $\mathcal{M}_3$                   | $\mathcal{M}_4$                   |
| Reconstructed vorticity ( $\downarrow$ ) | $(\times 10^{-1})$ | $3.02 \pm 0.08$             | $2.65 \pm 0.13$                   | <b><math>1.00 \pm 0.06</math></b> | $1.01 \pm 0.04$                   | $3.40 \pm 0.10$             | $3.22 \pm 0.11$ | <b><math>1.52 \pm 0.06</math></b> | $1.53 \pm 0.07$                   |
| Ensemble Field Std.                      | $(\times 10^0)$    | 0.12                        | 0.17                              | 0.06                              | 0.06                              | 0.16                        | 0.22            | 0.10                              | 0.01                              |
| Cycle-consistency ( $\downarrow$ )       | $(\times 10^{-1})$ | $1.94 \pm 0.07$             | $1.39 \pm 0.10$                   | $0.09 \pm 0.06$                   | <b><math>0.06 \pm 0.00</math></b> | $1.96 \pm 0.10$             | $1.42 \pm 0.09$ | $0.04 \pm 0.01$                   | <b><math>0.03 \pm 0.00</math></b> |
| Log-energy spectrum ( $\downarrow$ )     | $(\times 10^{-2})$ | $6.31 \pm 1.37$             | <b><math>5.83 \pm 0.20</math></b> | $6.27 \pm 1.80$                   | $5.95 \pm 0.94$                   | $5.67 \pm 0.37$             | $2.79 \pm 0.74$ | $2.63 \pm 0.32$                   | <b><math>2.51 \pm 0.30</math></b> |
| Log-ensrophy spectrum ( $\downarrow$ )   | $(\times 10^{-2})$ | $9.71 \pm 2.12$             | <b><math>8.99 \pm 0.30</math></b> | $9.67 \pm 2.77$                   | $9.18 \pm 1.46$                   | $9.20 \pm 0.60$             | $4.53 \pm 1.20$ | $4.26 \pm 0.51$                   | <b><math>4.07 \pm 0.48</math></b> |

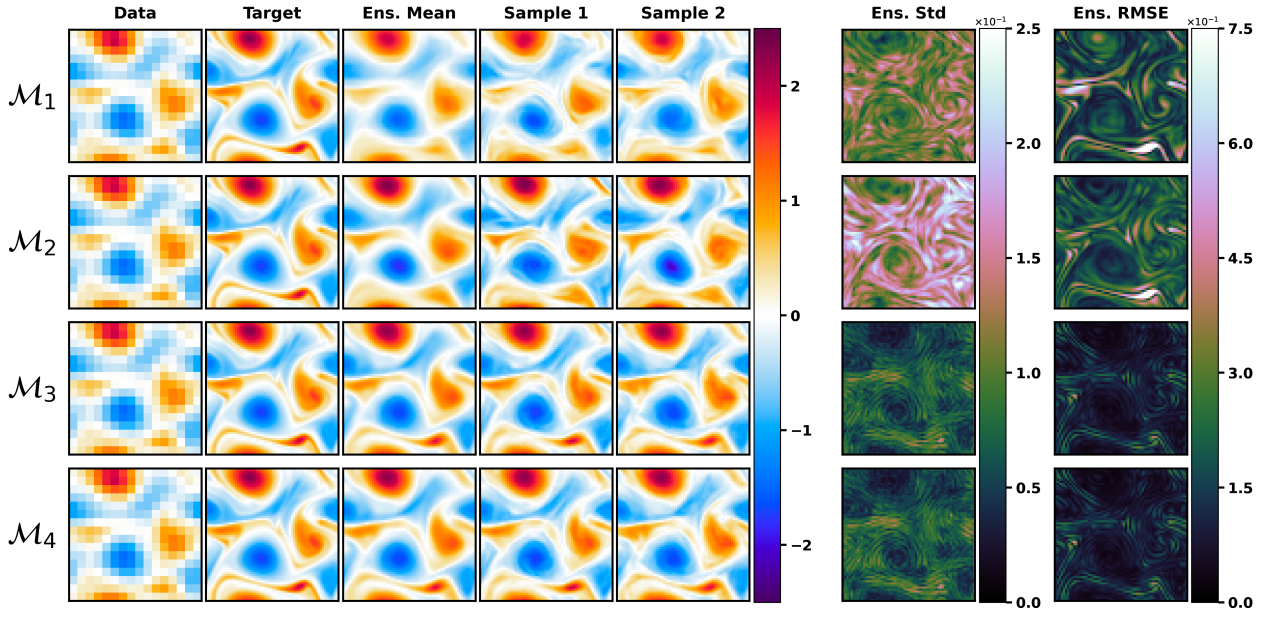
Error metrics are shown as ensemble mean  $\pm$  one standard deviation; the best-performing model for each metric is highlighted in **bold**.

$\mathcal{M}_1$ . Qualitatively, all four methods capture the larger-scale features of the target field driven by the forcing in eqs. (21)-(24). The zoomed view in the lower left quadrant shown in Fig. (9) highlights that  $\mathcal{M}_1$  smooths the thin vorticity filaments while  $\mathcal{M}_2$  slightly improves this, but with a smooth ensemble mean field.  $\mathcal{M}_3$  and  $\mathcal{M}_4$  generate super-resolved fields that are qualitatively very similar to the target, with the ensemble mean and most ensemble members accurately capturing the thin filamentary structures.

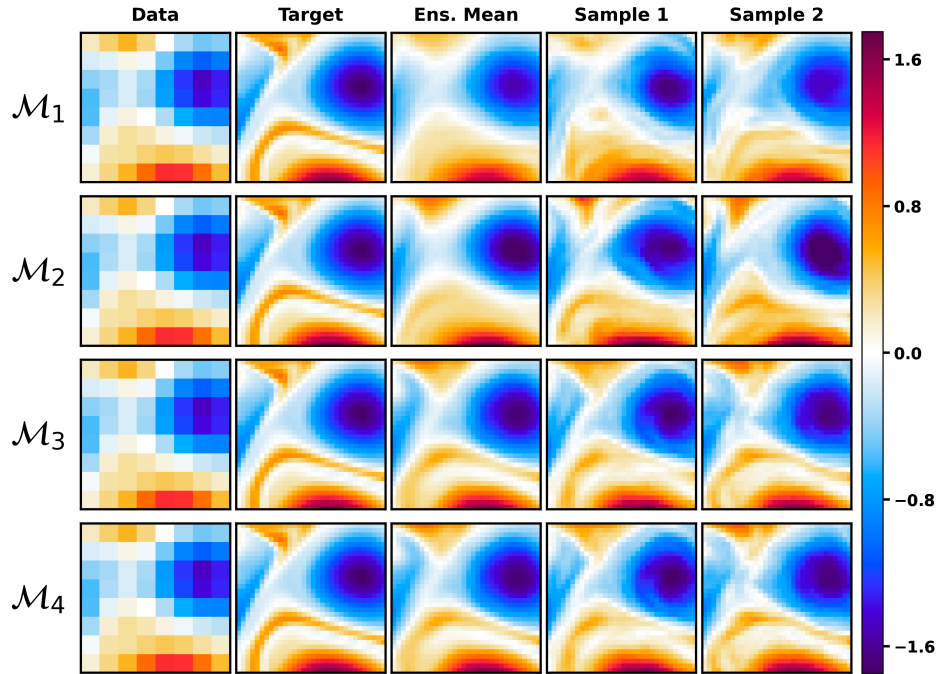
*Uncertainty quantification:* From the last two columns of Fig. (8),  $\mathcal{M}_3$  and  $\mathcal{M}_4$  exhibit the smallest standard deviation and RMSE, indicating accurate predictions with tight uncertainty bounds. As in the jet regime, the standard deviation and RMSE for all models (except  $\mathcal{M}_1$ ) show similar spatial patterns. The largest uncertainties are near the edges of larger-scale vortices. However, the standard deviations and errors for all models (Table 5) are considerably higher compared to the jet regime (Table 4), further highlighting that the complex distribution of vorticity fields in the eddy regime makes it a harder task for the diffusion models.

*Cycle-consistency:* As in the jet regime, the two conditional models,  $\mathcal{M}_3$  and  $\mathcal{M}_4$ , are extremely cycle-consistent (Fig. 10). However, the models  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are significantly less cycle-consistent, indicating that the assumptions made in these models, (11) and (14), are incorrect for these complex distributions.

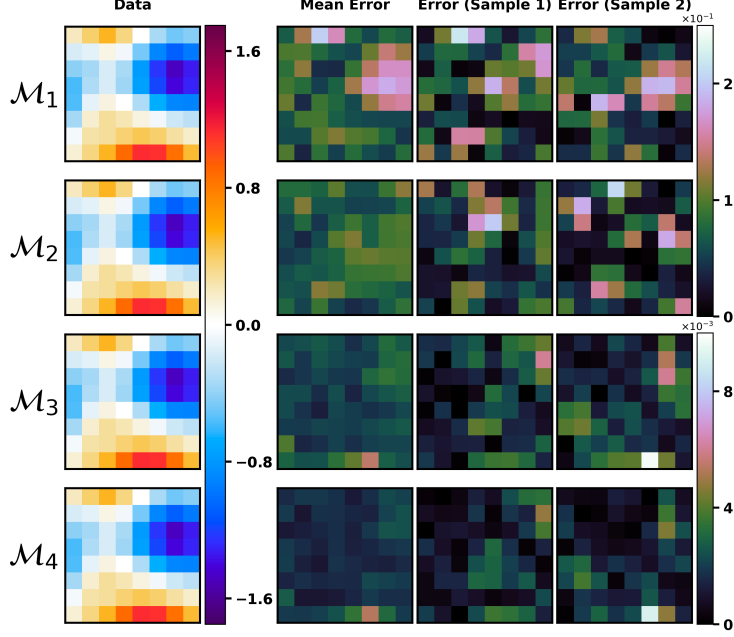
*Time-averaged 2D turbulence statistics:* Fig. 11 (Top-right) shows the 1D probability distribution function for the vorticity field, which is non-Gaussian but unimodal. This is accurately captured by  $\mathcal{M}_2$ ,  $\mathcal{M}_3$ , and  $\mathcal{M}_4$ . As in the jet case,  $\mathcal{M}_1$  is skewed towards the data and slightly underestimates larger values of vorticity. In Fig. 11 (Top-left), all four models reproduce the characteristic smooth premultiplied 2D kinetic energy spectrum shape for the eddy regime, but underestimate it across all wavenumbers.



**Figure 8.** As Fig. (4), but for super-resolution in the eddy regime ( $\text{Re}=10^4$ ) from coarse-resolution fields (Test Case 4, Table 3).



**Figure 9.** As Fig. (5), but for the lower left quadrant of Fig. (8).

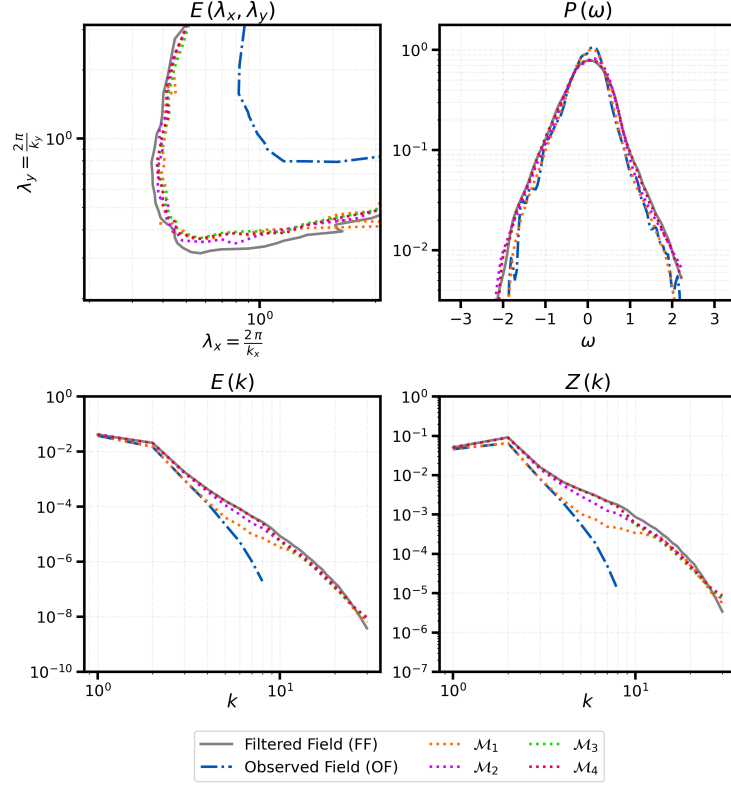


**Figure 10.** As Fig. (6), but for super-resolution in the eddy regime ( $Re=10^4$ ) from coarse-resolution fields (Test Case 4, Table 3). The upper color bar limit of the error fields for  $\mathcal{M}_1$  and  $\mathcal{M}_2$  differs from that for  $\mathcal{M}_3$  and  $\mathcal{M}_4$  by two orders of magnitude.

From Fig. 11 (Bottom),  $\mathcal{M}_2$ ,  $\mathcal{M}_3$ , and  $\mathcal{M}_4$  accurately capture the spectra at all wavenumbers, including the smaller scales (i.e., the tails).  $\mathcal{M}_1$  is again skewed by the data and shows steeper scaling compared to the target spectra at large wavenumbers.

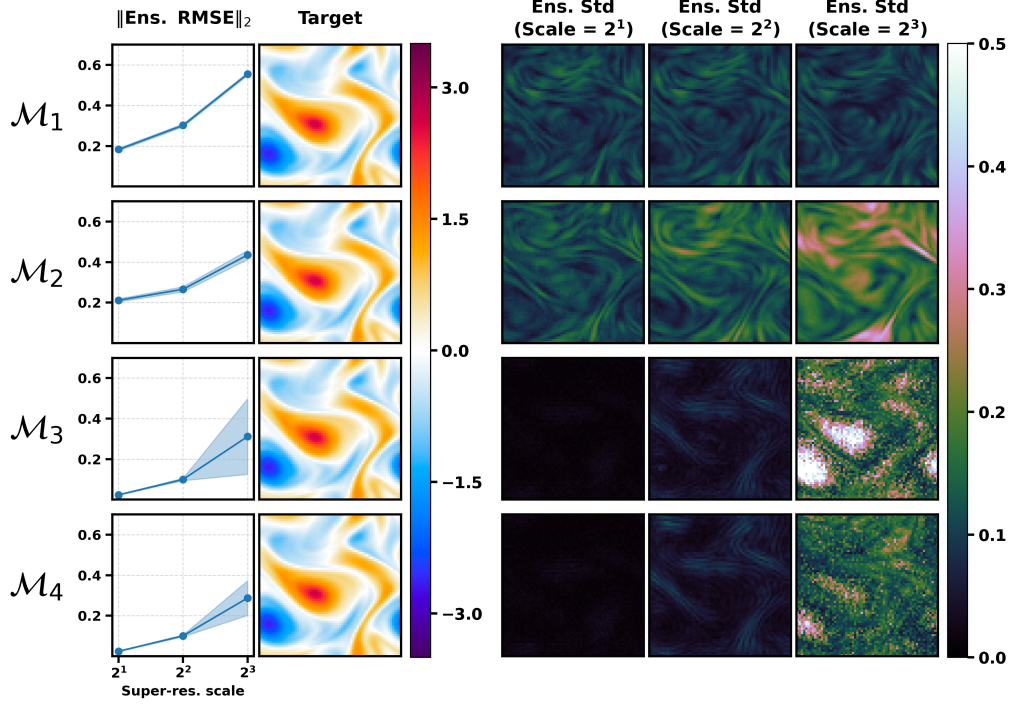
*Super-resolution at different down-sampling scales:* To assess each method’s performance under different super-resolution settings, we consider the eddy regime at  $Re = 10^3$  with down-sampling scales  $\delta = 2^1$  and  $2^3$ , in addition to  $\delta = 2^2$  used in the main test cases discussed above.  $\mathcal{M}_1$  is easy to implement and does not require any re-training since it re-uses the unconditional model with modified initial conditions that only depend on the data (11).  $\mathcal{M}_2$  also does not require re-training for each  $\delta$ , but requires access to the observation operators  $\mathcal{H}$  corresponding to each  $\delta$ , and their gradients (17).  $\mathcal{M}_3$  and  $\mathcal{M}_4$  have to be re-trained from scratch for each  $\delta$ , limiting their flexibility, but they do not require access to the gradients of each  $\mathcal{H}$ .

Fig. (12) shows that the norm of ensemble RMSE increases as  $\delta$  increases, since coarser data contains less information to aid with super-resolution, even for the large-scale vortices. For all methods except  $\mathcal{M}_1$ , the standard deviation of the norm of the ensemble



**Figure 11.** As Fig. (7), but for super-resolution in the eddy regime ( $\text{Re}=10^4$ ) from coarse-resolution fields (Test Case 4, Table 3).

RMSE increases with  $\delta$ . Similar trends are also observed in the corresponding ensemble standard deviation of the generated super-resolved fields. As shown for  $\delta = 2^2$ ,  $\mathcal{M}_1$  is inaccurate for super-resolution and generates non-physical fields, and hence its standard deviation does not vary much with  $\delta$ . The two conditional models  $\mathcal{M}_3$  and  $\mathcal{M}_4$  again show the smallest error norms, with perfect reconstruction at  $\delta = 2^1$ , followed by  $\mathcal{M}_2$ , then  $\mathcal{M}_1$ . For  $\delta = 2^3$ ,  $\mathcal{M}_3$  and  $\mathcal{M}_4$  show the largest standard deviation of error norms, predict the largest ensemble standard deviation fields, and their error patterns are directly related to the target field (FF). All these error properties are adequate and probably indicate that learning the conditional distribution between the target and the data is difficult at extremely coarse resolutions ( $N_{OF} = 8 \times 8$ ).



**Figure 12.** Super-resolution with the four diffusion modeling approaches at different down-sampling scales ( $\delta$ ) in the eddy regime ( $\text{Re}=10^3$ ) from coarse-resolution fields (Test Case 3, Table 3). Column 1: Norm of the relative ensemble root mean squared error (RMSE) with shaded regions showing  $\pm$  one standard deviation for the reconstructed field with increasing down-sampling scales. Column 2: Target vorticity field. Columns 3-5: Ensemble standard deviations at different  $\delta$ . Target is identical for all models, and all  $\delta$ 's.

## 5.2 Application 2: Super-resolution and Inference from Coarse and Sparse, Gappy Data.

We now discuss applications with coarse and sparse, gappy observations of the vorticity field, corresponding to Cases 5-8 of Table (3).

### 5.2.1 Jet Regime

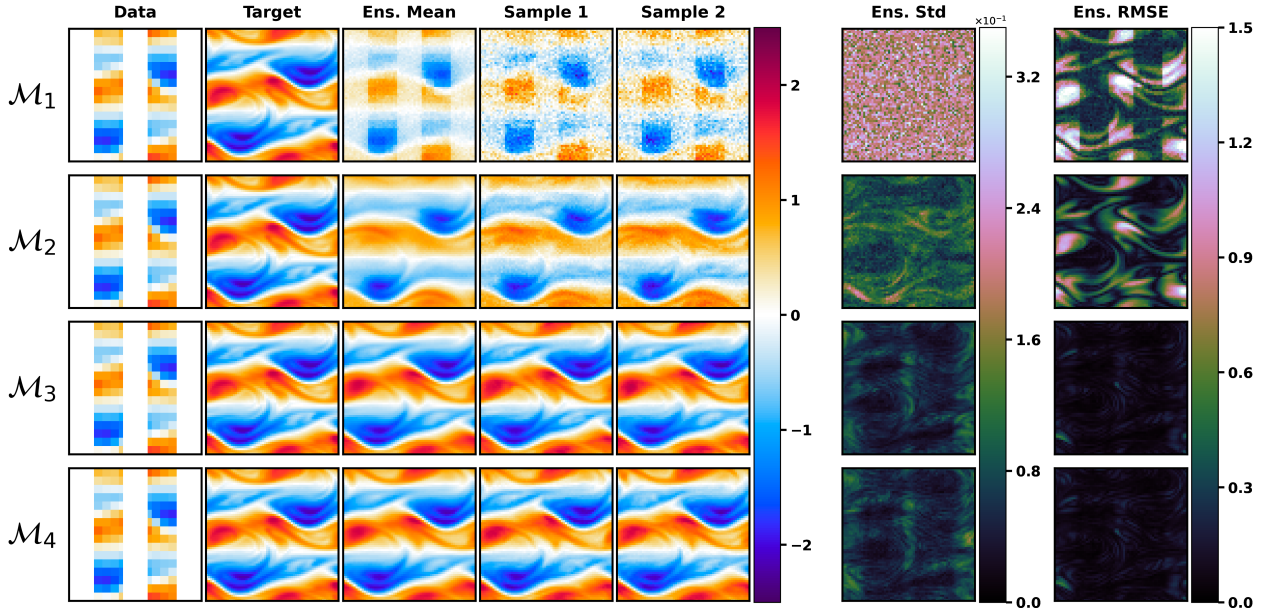
Table (6) summarizes the key quantitative skill metrics. In general, the errors of all four approaches are significantly higher (more than  $2\times$ ) than the corresponding errors of the cases with coarse-resolution fields (Table 4). This confirms that super-resolution from both coarse and sparse observations is challenging. For some methods and test cases, we will find that the coarse, sparse data are insufficient to infer the super-resolved field.

**Table 6.** Comparison of the key quantitative skill metrics for super-resolution and inference in the jet regime from coarse and sparse, gappy observations (Cases 5-6 of Table (3)).

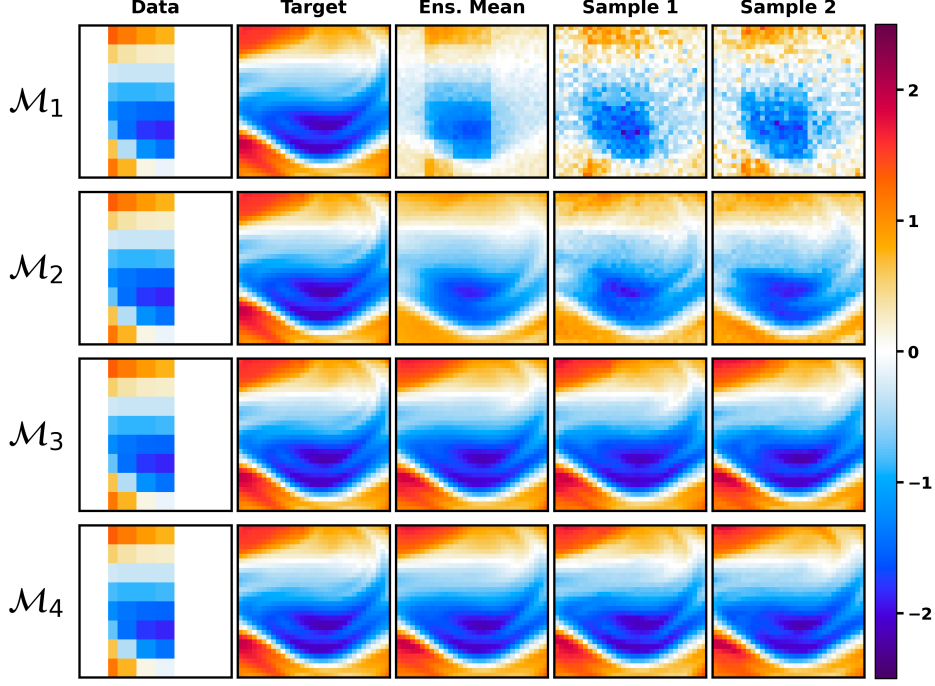
|                                          |                    | Test Case 5 ( $Re = 10^3$ ) |                  |                                    |                                   | Test Case 6 ( $Re = 10^4$ ) |                  |                                   |                                   |
|------------------------------------------|--------------------|-----------------------------|------------------|------------------------------------|-----------------------------------|-----------------------------|------------------|-----------------------------------|-----------------------------------|
| Relative Error Metric                    | Scale              | $\mathcal{M}_1$             | $\mathcal{M}_2$  | $\mathcal{M}_3$                    | $\mathcal{M}_4$                   | $\mathcal{M}_1$             | $\mathcal{M}_2$  | $\mathcal{M}_3$                   | $\mathcal{M}_4$                   |
| Reconstructed vorticity ( $\downarrow$ ) | $(\times 10^{-1})$ | $6.07 \pm 0.14$             | $3.38 \pm 0.12$  | <b><math>0.45 \pm 0.21</math></b>  | $0.48 \pm 0.13$                   | $6.68 \pm 0.10$             | $4.35 \pm 0.12$  | <b><math>0.99 \pm 0.04</math></b> | $1.02 \pm 0.05$                   |
| Ensemble Field Std.                      | $(\times 10^0)$    | 0.20                        | 0.06             | 0.04                               | 0.04                              | 0.22                        | 0.10             | 0.06                              | 0.06                              |
| Cycle-consistency ( $\downarrow$ )       | $(\times 10^{-1})$ | $3.58 \pm 0.09$             | $2.24 \pm 0.07$  | $0.13 \pm 0.13$                    | <b><math>0.10 \pm 0.06</math></b> | $3.42 \pm 0.12$             | $2.40 \pm 0.10$  | <b><math>0.10 \pm 0.01</math></b> | <b><math>0.10 \pm 0.01</math></b> |
| Log-energy spectrum ( $\downarrow$ )     | $(\times 10^{-2})$ | $24.92 \pm 0.30$            | $12.76 \pm 0.61$ | <b><math>8.63 \pm 2.35</math></b>  | $9.56 \pm 1.73$                   | $18.49 \pm 0.26$            | $7.50 \pm 0.60$  | <b><math>1.64 \pm 0.18</math></b> | $1.82 \pm 0.30$                   |
| Log-ensrophy spectrum ( $\downarrow$ )   | $(\times 10^{-2})$ | $38.41 \pm 0.47$            | $19.68 \pm 0.95$ | <b><math>13.30 \pm 3.62</math></b> | $14.73 \pm 2.67$                  | $30.09 \pm 0.43$            | $12.22 \pm 0.98$ | <b><math>2.67 \pm 0.29</math></b> | $2.95 \pm 0.50$                   |

Error metrics are shown as ensemble mean  $\pm$  one standard deviation; best-performing models are in **bold**.

*Super-resolved vorticity fields:* As in the cases with coarse-resolution fields, the two trained conditional models,  $\mathcal{M}_3$  and  $\mathcal{M}_4$ , show the lowest reconstruction error, followed by  $\mathcal{M}_2$  and  $\mathcal{M}_1$ . For  $Re = 10^4$ , Fig. (13) shows that  $\mathcal{M}_1$  fails to generate physically valid fields, since it is over-reliant on the coarse, sparse data, which is gappy and not a true physical field; qualitatively, the other models capture the right larger-scale features such as zonal jets. Fig. (14) shows that  $\mathcal{M}_3$  and  $\mathcal{M}_4$  also generate correct fine-scale thin filaments, while  $\mathcal{M}_2$  generates features with additional noisy artifacts.



**Figure 13.** As Fig. (4), but for super-resolution and inference in the jet regime ( $Re=10^4$ ) from coarse and sparse, gappy observations (Test Case 6, Table 3).

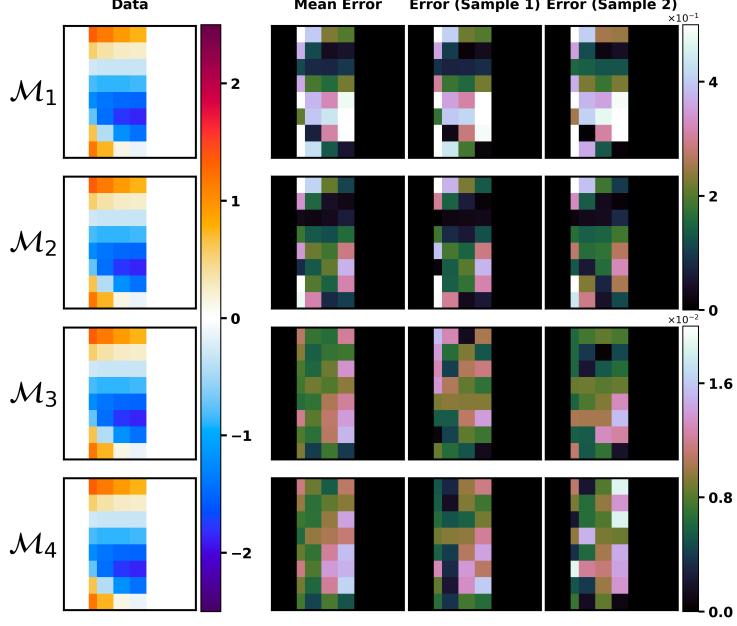


**Figure 14.** As Fig. (5), but for the lower-left quadrant of Fig. (13).

*Uncertainty quantification:* From the last two columns of Fig. (13),  $\mathcal{M}_3$  and  $\mathcal{M}_4$  achieve the smallest ensemble standard deviations and RMSE. The RMSE for all models is highest in the unobserved regions, since a lack of observations is uninformative for producing the exact super-resolved field. Hence, for all models except  $\mathcal{M}_1$ , which fails, the predicted ensemble standard deviation (uncertainty) is also correctly highest in the unobserved regions.

*Cycle-consistency:* Remarkably,  $\mathcal{M}_3$  and  $\mathcal{M}_4$  are very cycle-consistent even in this sparse observation application (Fig. 15), suggesting that using direct conditional information can enforce cycle-consistency through training. Compared to their coarse results (Table 4),  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are significantly less cycle-consistent, highlighting their shortcomings when the observation operator is gappy and less informative about the target field.

*Time-averaged 2D turbulence statistics:* From Fig. (16) (Top-right),  $\mathcal{M}_1$  completely misses the bimodal distribution of the vorticity field, instead generating a unimodal distribution.  $\mathcal{M}_2$  captures the bimodality but significantly underestimates large values of the vorticity, while  $\mathcal{M}_3$  and  $\mathcal{M}_4$  perfectly capture the distribution. Similarly, in Fig. (16) (Top-left),  $\mathcal{M}_1$  completely mispredicts the shape,  $\mathcal{M}_2$  underestimates the energy at all



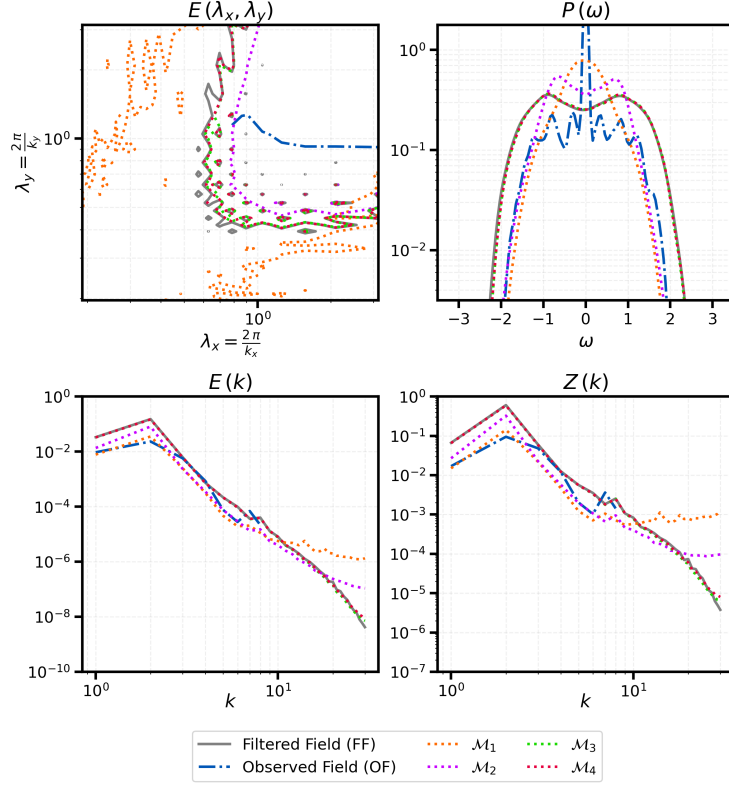
**Figure 15.** As Fig. (6), but for super-resolution and inference in the jet regime ( $Re=10^4$ ) from coarse and sparse, gappy observations (Test Case 6, Table 3). The upper color bar limit of the error fields for  $\mathcal{M}_1$  and  $\mathcal{M}_2$  differs from that for  $\mathcal{M}_3$  and  $\mathcal{M}_4$  by an order of magnitude.

wavenumbers, while  $\mathcal{M}_3$  and  $\mathcal{M}_4$  capture the shapes well at all wavenumbers. For the kinetic energy and enstrophy spectra (Fig. 16) (Bottom),  $\mathcal{M}_3$  and  $\mathcal{M}_4$  once again accurately generate super-resolved fields with the right spectra scaling, including the tails, while  $\mathcal{M}_1$  and  $\mathcal{M}_2$  shift the spectra significantly to smaller scales since they generate fields with significant fine-scale noisy artifacts.

### 5.2.2 Eddy Regime

The eddy regime from coarse and sparse, gappy observations is the most challenging of all test cases we consider: the target fields have the most complicated distribution, while the data offers the least information for super-resolution. This is reflected in the relatively higher key quantitative skill metrics of all methods, summarized in Table (7).

*Super-resolved vorticity fields:* As in the jet regime, for both  $Re = 10^3$  and  $Re = 10^4$ ,  $\mathcal{M}_3$  and  $\mathcal{M}_4$  achieve the lowest reconstruction errors, followed by  $\mathcal{M}_2$  and  $\mathcal{M}_1$ . Fig. (17) shows that  $\mathcal{M}_1$  again fails to generate physically valid fields while  $\mathcal{M}_2$  struggles to reconstruct large-scale features accurately. Fig. (18) shows that  $\mathcal{M}_3$  and  $\mathcal{M}_4$  generate thin



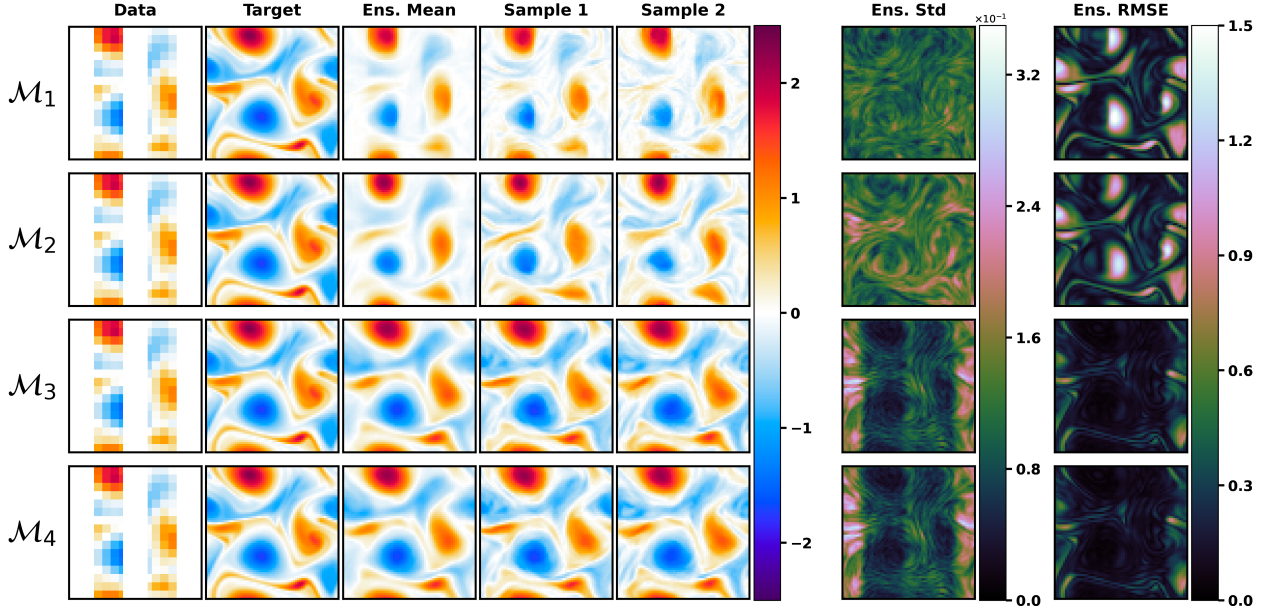
**Figure 16.** As Fig. (7), but for super-resolution and inference in the jet regime ( $Re=10^4$ ) from coarse and sparse, gappy observations (Test Case 6, Table 3).

**Table 7.** As Table (6), but in the eddy regime from coarse and sparse, gappy observations (Cases 7-8 of Table (3)).

|                                          |                    | Test Case 7 ( $Re = 10^3$ ) |                                   |                                   |                                   | Test Case 8 ( $Re = 10^4$ ) |                 |                 |                                   |
|------------------------------------------|--------------------|-----------------------------|-----------------------------------|-----------------------------------|-----------------------------------|-----------------------------|-----------------|-----------------|-----------------------------------|
| Relative Error Metric                    | Scale              | $\mathcal{M}_1$             | $\mathcal{M}_2$                   | $\mathcal{M}_3$                   | $\mathcal{M}_4$                   | $\mathcal{M}_1$             | $\mathcal{M}_2$ | $\mathcal{M}_3$ | $\mathcal{M}_4$                   |
| Reconstructed vorticity ( $\downarrow$ ) | $(\times 10^{-1})$ | $7.75 \pm 0.08$             | $4.43 \pm 0.53$                   | $2.47 \pm 0.33$                   | <b><math>2.39 \pm 0.16</math></b> | $6.83 \pm 0.11$             | $6.42 \pm 0.21$ | $3.13 \pm 0.21$ | <b><math>3.10 \pm 0.18</math></b> |
| Ensemble Field Std.                      | $(\times 10^0)$    | 0.11                        | 0.24                              | 0.14                              | 0.13                              | 0.16                        | 0.19            | 0.17            | 0.17                              |
| Cycle-consistency ( $\downarrow$ )       | $(\times 10^{-1})$ | $3.77 \pm 0.12$             | $1.80 \pm 0.17$                   | <b><math>0.16 \pm 0.01</math></b> | <b><math>0.16 \pm 0.01</math></b> | $3.08 \pm 0.11$             | $2.53 \pm 0.14$ | $0.06 \pm 0.01$ | <b><math>0.05 \pm 0.00</math></b> |
| Log-energy spectrum ( $\downarrow$ )     | $(\times 10^{-2})$ | $14.86 \pm 1.11$            | <b><math>5.82 \pm 0.27</math></b> | $5.84 \pm 0.43$                   | $6.33 \pm 0.58$                   | $6.12 \pm 0.49$             | $5.03 \pm 0.56$ | $3.13 \pm 0.50$ | <b><math>2.76 \pm 0.40</math></b> |
| Log-ensrophy spectrum ( $\downarrow$ )   | $(\times 10^{-2})$ | $22.88 \pm 1.72$            | <b><math>8.97 \pm 0.42</math></b> | $8.99 \pm 0.66$                   | $9.75 \pm 0.89$                   | $9.86 \pm 0.80$             | $8.12 \pm 0.91$ | $5.06 \pm 0.80$ | <b><math>4.46 \pm 0.65</math></b> |

Error metrics are shown as ensemble mean  $\pm$  one standard deviation; best-performing models are in **bold**.

fine-scale filaments that are accurate in the observed regions, but with large variations in the unobserved regions, leading to discrepancies with the target. In all other cases, the conditional models generated very accurate super-resolved fields. However, for the distribution of vorticity fields in this  $Re = 10^4$  eddy regime, the sparse, gappy observations are not sufficiently informative to infer all finer-scale structures in the data gaps.

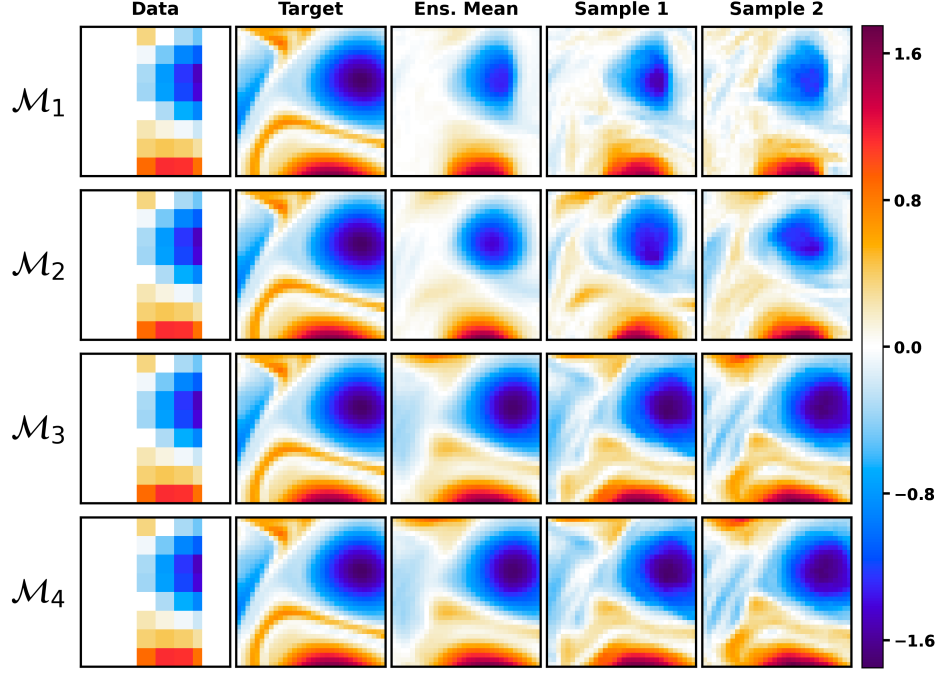


**Figure 17.** As Fig. (4), but for super-resolution and inference in the eddy regime ( $\text{Re}=10^4$ ) from coarse and sparse, gappy observations (Test Case 8, Table 3).

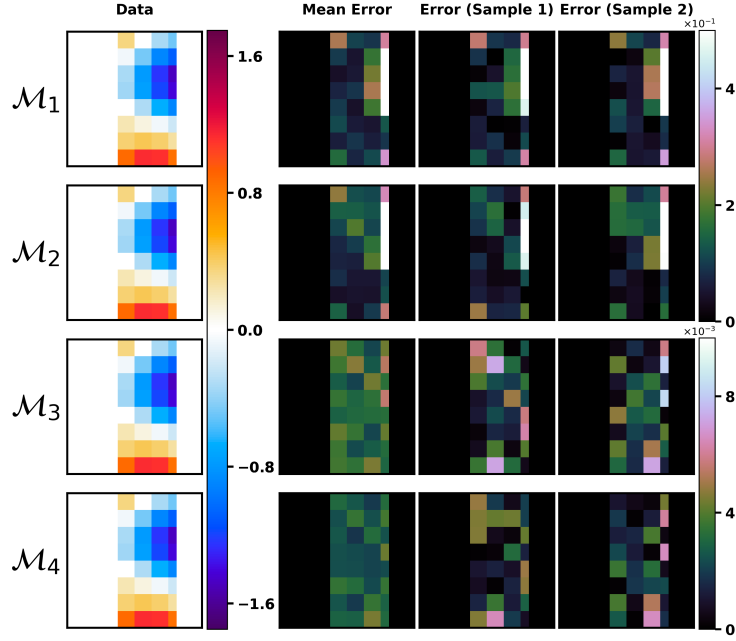
*Uncertainty quantification:* From the last two columns of Fig. (17),  $\mathcal{M}_3$  and  $\mathcal{M}_4$  achieve the smallest ensemble standard deviations and RMSE. As in the jet cases, the RMSE for all models is highest in the unobserved regions. For  $\mathcal{M}_3$  and  $\mathcal{M}_4$ , the ensemble standard deviation follows the same spatial pattern as the ensemble RMSE, suggesting that these models are well-calibrated estimators of uncertainty even in this setting.

*Cycle-consistency:* Even in this challenging setting,  $\mathcal{M}_3$  and  $\mathcal{M}_4$  are extremely cycle-consistent (Fig. 19). Once again,  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are significantly less cycle-consistent compared to the full observation applications.

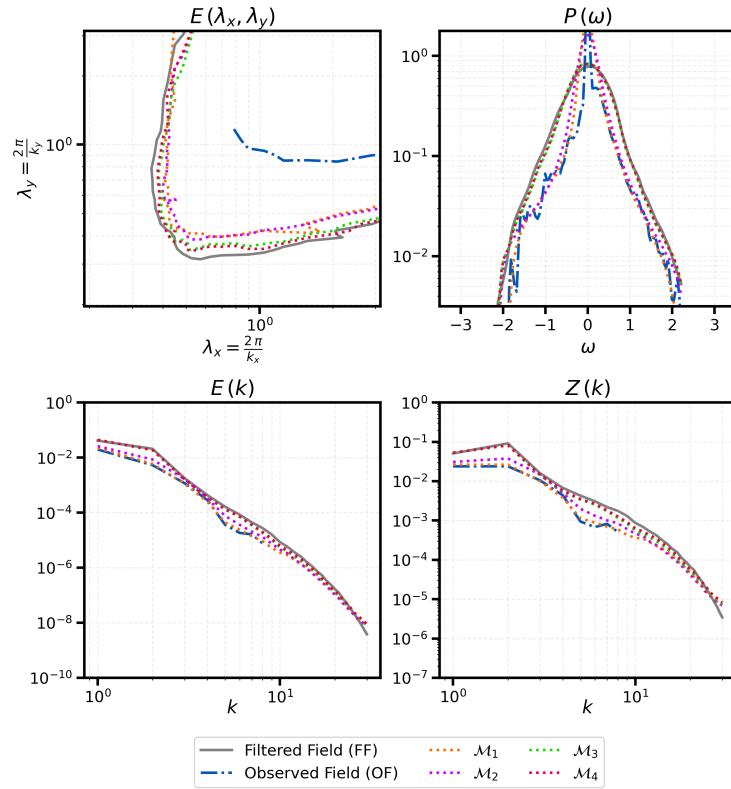
*Time-averaged 2D turbulence statistics:* From Fig. (20) (Top-right),  $\mathcal{M}_1$  and  $\mathcal{M}_2$  significantly underestimate large values of vorticity, while  $\mathcal{M}_3$  and  $\mathcal{M}_4$  perfectly capture the unimodal distribution. From Fig. (16) (Top-left), all models underestimate the kinetic energy at all wavenumbers. For the kinetic energy and enstrophy spectra (Fig. 16) (Bottom),  $\mathcal{M}_3$  and  $\mathcal{M}_4$  accurately generate super-resolved fields with the right spectra at all scales, while  $\mathcal{M}_1$  and  $\mathcal{M}_2$  generate spectra with slightly steeper scaling, again shifting the spectra to finer scales.



**Figure 18.** As Fig. (5), but for the lower-left quadrant of Fig. (13).



**Figure 19.** As Fig. (6), but for super-resolution and inference in the eddy regime ( $\text{Re}=10^4$ ) from coarse and sparse, gappy observations (Test Case 8, Table 3). The upper color bar limit of the error fields for  $\mathcal{M}_1$  and  $\mathcal{M}_2$  differs from that for  $\mathcal{M}_3$  and  $\mathcal{M}_4$  by two orders of magnitude.



**Figure 20.** As Fig. (7), but for super-resolution and inference in the eddy regime ( $\text{Re}=10^4$ ) from coarse and sparse, gappy observations (Test Case 8, Table 3).

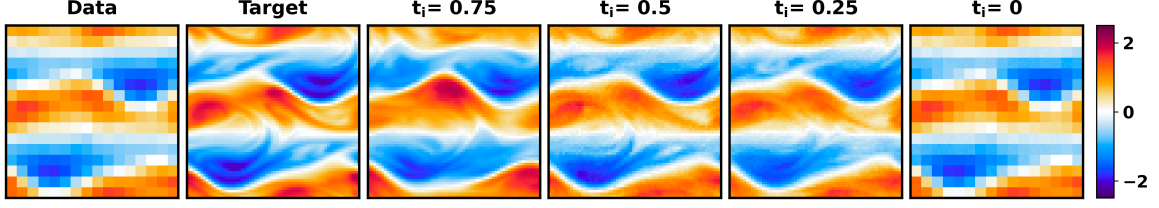
### 5.3 Sensitivity Studies: Effect of Tuning Parameters

We now study the sensitivity of the four models to their tuning parameters, and provide their optimal values. All above results used optimally tuned intermediate times (for  $\mathcal{M}_1$ ) and guidance strengths (for  $\mathcal{M}_2, \mathcal{M}_4$ ).  $\mathcal{M}_3$  does not require any tuning, it only relies on training. The choice of intermediate time and guidance strength greatly controls the quality of the generated super-resolved samples, and trades off fidelity (i.e., physical plausibility and sharpness of the generated fields) for consistency (i.e., cycle-consistency with the observations or data). Since the denoiser neural networks, (8) and (9), are trained independently of these parameters, their weights are unaffected by this fine-tuning. This allows for great flexibility, as the intermediate time and guidance strengths can be optimized and re-tuned post-training for each task (application), or even for each snapshot within an application (Chung, Kim, McCann, et al., 2023).

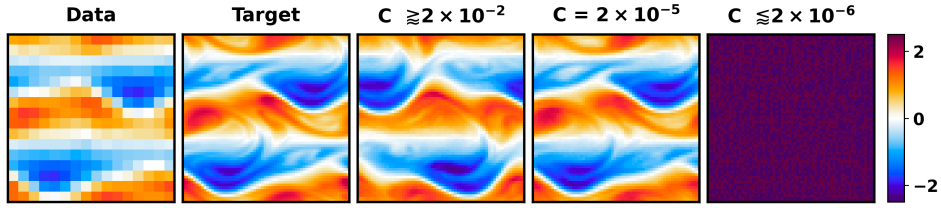
An alternate approach to improve the fidelity of the generated samples is to optimize the initial Gaussian noise used for the reverse process (Qi et al., 2024). However, this is computationally expensive, and hence, we use a fixed seed to generate Gaussian initial conditions and focus on fine-tuning only the intermediate time and guidance strengths. We optimize over all snapshots of super-resolution by sweeping over a range of values and choose the parameter value that minimizes the  $L_2$  norm of the error of the super-resolved fields. For brevity, we discuss results below for super-resolution in the jet regime ( $\text{Re}=10^4$ ) from coarse-resolution field observations. Similar trends hold for all other applications and test cases.

#### 5.3.1 $\mathcal{M}_1$ : *Modifying the Initial Condition (SDEdit)*

The main parameter to be tuned is the intermediate time  $t_i$ . Fig. (21) shows that choices of  $t_i$  close to 0 lead to unphysical super-resolved fields. Since  $t_i = 0$  corresponds to the final diffusion pseudo-timestep, the model does not get to denoise the input field (data). On the other hand, choices of  $t_i$  close to  $T = 1$  ( $t_i \gtrapprox 0.75$ ) lead to fields that lose track of large-scale features of the data, and hence generate non-cycle-consistent super-resolved fields (Meng et al., 2021). Setting  $0.1 \gtrapprox t_i \gtrapprox 0.6$  generates relatively cycle-consistent super-resolved fields, and we obtain the smallest relative error at  $t_i = 0.25$ .



**Figure 21.** Sensitivity of  $\mathcal{M}_1$  (guidance by modifying the initial condition) to the intermediate pseudo-time  $t_i$ . From left to right: Coarse observed vorticity field (Data), Filtered vorticity field (Target), Super-resolved fields generated by  $\mathcal{M}_1$  at various values of  $t_i$ .



**Figure 22.** Sensitivity of  $\mathcal{M}_2$  (guidance by modifying the score) to the guidance strength  $C$ . From left to right: Coarse observed vorticity field (Data), Filtered vorticity field (Target), Super-resolved fields generated by  $\mathcal{M}_2$  at various values of  $C$ .

### 5.3.2 $\mathcal{M}_2$ : *Modifying the Score (Diffusion Posterior Sampling)*

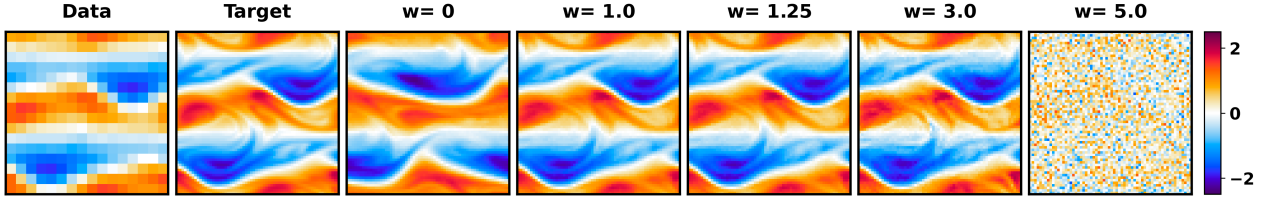
The main parameter to be tuned here is the guidance strength,  $C$ . From (19), using a value of  $C \rightarrow \infty$  is equivalent to simply utilizing an unconditional model with no guidance. Fig. (22) shows that setting  $C \gtrsim 2 \times 10^{-2}$  generates non-cycle-consistent fields. Using  $C \rightarrow 0$  corresponds to extremely strong guidance, which leads to over-saturation and unphysical noisy artifacts (Chung, Kim, McCann, et al., 2023), as seen in the figure for  $C \lesssim 2 \times 10^{-6}$ . Here, the optimal value of  $C$  was found to be  $2 \times 10^{-5}$ , which sufficiently trades off fidelity for consistency.

### 5.3.3 $\mathcal{M}_3$ : *Vanilla Conditional Model*

For this approach, the model directly learns the conditional score through training, and hence, there is no additional parameter to fine-tune.

### 5.3.4 $\mathcal{M}_4$ : Classifier-Free Guidance

The main parameter to be tuned is the guidance strength,  $w$ . Setting  $w = 0$  corresponds to sampling only from the unconditional model, while  $w = 1$  corresponds to sampling only from the conditional model (Ho & Salimans, 2021), as seen in Fig. (23). Typically, the value of  $w$  is set to be greater than 1. We obtain the smallest relative error at  $w = 1.25$ . For our applications, a value of  $w \gtrsim 2$  leads to the generation of spurious high-frequency features. This can be explained by the fact that classifier-free guidance performs a linear combination of the conditional and unconditional scores, and therefore, fields are generated from an approximated distribution that may not be physically meaningful (Karras et al., 2024). Using a very large value of  $w$  leads to unphysical noisy artifacts as shown in the last column of Fig. (23).



**Figure 23.** Sensitivity of  $\mathcal{M}_4$  (Classifier-Free Guidance) to the guidance strength  $w$ . From left to right: Coarse observed vorticity field (Data), Filtered vorticity field (Target), Super-resolved fields generated by  $\mathcal{M}_4$  at various values of  $w$ .

## 6 Conclusions and Discussion

We investigated super-resolution and inference of quasi-geostrophic turbulence under the  $\beta$ -plane approximation with four generative diffusion modeling approaches – two guided unconditional models and two conditional models. The two conditional models, vanilla ( $\mathcal{M}_3$ ) and classifier-free guidance ( $\mathcal{M}_4$ ), generate accurate ensembles of super-resolved vorticity fields across all dynamical regimes (eddy and jet) and Reynolds numbers ( $Re = 10^3, 10^4$ ). Moreover, they predict correct turbulence statistical quantities such as the 1D probability distribution function and kinetic energy spectra and enstrophy spectra (including the tails), even from only sparse and gappy coarse observations. These generated ensembles are remarkably cycle-consistent with the coarse observations, despite not being explicitly optimized for during training. These two approaches show

promise for deployment in geophysical inference applications where capturing accurate fine-scale features is critical, especially when the observation operator is nonlinear or indirect. Though these advantages come at the cost of application-specific re-training, they do not require knowledge of the observation operators for training or generation. On the other hand, if computational resources do not allow for re-training, guidance by modifying the score ( $\mathcal{M}_2$ ) trades off fidelity (sharpness) for cycle-consistency and generates smoothed fields but requires access to gradients of the observation operator at each diffusion pseudo-timestep. Although easily implemented and computationally cheap, guidance by modifying initial conditions ( $\mathcal{M}_1$ ) can not produce correct statistics, and fails in challenging applications with sparse and gappy data. Overall, our diffusion models provide computationally fast methods to correct for missing fine-scale features ( $\mathcal{M}_1$  takes  $\sim 0.1\text{ s}$ ,  $\mathcal{M}_2 \sim 2\text{ s}$ ,  $\mathcal{M}_3 \sim 0.3\text{ s}$ , and  $\mathcal{M}_4 \sim 0.7\text{ s}$  on an NVIDIA A6000 GPU, computed by averaging 30 independent evaluations).  $\mathcal{M}_2$  takes significantly longer than the other approaches, as it requires the computation of additional gradients to evaluate the measurement matching term. A major advantage of our super-resolution framework is that it does not require auto-differentiability or access to numerical solvers, and only needs either paired high- and coarse-resolution fields or access to the observation operator.

Our applications with sparse and gappy observation systems show that for conditional diffusion models, the ensemble standard deviations are good estimators of the actual errors of predictions. This capability can be leveraged for adaptive sampling, i.e., identify the most optimal types and locations of observations for inference (Lermusiaux, 2007; Lermusiaux et al., 2017). Another opportunity is the super-resolution of real ocean dynamics such as ring formation, eddy separation, and submesoscale interactions in the Gulf Stream region (E. Chassignet & Marshall, 2008; Mensa et al., 2013; Gula et al., 2019), Loop Current system (Oey et al., 2005; Bracco et al., 2019; Nickerson et al., 2022), or other regions. Related possibilities include the diagnosis of submesoscales (Z. Zhang et al., 2024; Archer et al., 2025) and the inference of subsurface fields (Lermusiaux et al., 2011; Klemas & Yan, 2014; Moore et al., 2019) using sparse and gappy, coarse or indirect observations.

However, two major limitations must be overcome for all these real ocean applications to materialize. First, training requires access to high-resolution numerical simulations and reanalyses from ocean models (Haley & Lermusiaux, 2010; Storto et al., 2019) and access to multivariate or multi-modal observations (Lermusiaux et al., 2002; Qu et

al., 2024). Next, our diffusion architecture requires extensions such as multiscale patch-wise modeling (Brenowitz et al., 2025) or modeling in a latent space (Rombach et al., 2022) to handle larger coastlines and ocean regions. Extensions to spatio-temporal super-resolution and inference can be investigated through the training of guided video diffusion (Ho et al., 2022), or hybrid-coupling with other stochastic numerical models (Lu & Lermusiaux, 2021) or neural surrogates. Our generative diffusion is also promising for stochastic data-driven closure modeling with uncertainty estimates, e.g., for stochastic subgrid forcing (Lermusiaux, 2006; Zanna et al., 2017; Perezhogan et al., 2023).

### Acknowledgments

We thank C. Daskalakis and G. Daras for insightful discussions on diffusion modeling. We thank the members of our ML-SCOPE MURI team and of our MSEAS group including V.A. Rodriguez, F.M Benfenati, and A.K. Saravanakumar for discussions on quasi-geostrophic turbulence, statistics, and figures. We are grateful to the Office of Naval Research for partial support under grant N00014-20-1-2023 (MURI ML-SCOPE) and N00014-19-1-2693 (IN-BDA) to the Massachusetts Institute of Technology. ANSB was partially supported by an MIT Mechanical Engineering MathWorks Fellowship.

### Open Research

The quasi-geostrophic simulations and data are available at Zenodo via <https://doi.org/10.5281/zenodo.15742145> under a Creative Commons Attribution 4.0 International license (Suresh Babu et al., 2025a). Codes for the diffusion model training and generation, and trained weights are preserved at <https://doi.org/10.5281/zenodo.15750243>, available via the MIT license and developed openly at <https://github.com/ananthu545/quasi-geostrophic-beta-plane-super-resolution> (Suresh Babu et al., 2025b).

## Appendix A Diffusion Model Architecture and Training

To train the unconditional diffusion model, we utilize a U-Net architecture (Ronneberger et al., 2015). We use channels of size [32, 64, 128, 256]. Following Nichol and Dhariwal

(2021); Rozet and Louppe (2023), we use the following cosine noise schedule,

$$\omega = \arccos \sqrt{10^{-3}} \quad (\text{A1})$$

$$\mu(t) = \cos^2(\omega t) \quad (\text{A2})$$

$$\sigma(t) = \sqrt{1 - \mu(t)^2} \quad (\text{A3})$$

For training and sampling, we utilize  $N_{diff} = 64$  uniform timesteps with diffusion time  $0 < t < 1$  with  $N_{corr} = 2$  Langevin Monte Carlo correction steps per diffusion pseudo-time-step, with correction size of 0.3. Hence, for sampling, the unconditional diffusion model requires  $N_{diff} \times (1 + N_{corr}) = 64 \times 3 = 192$  forward passes of the trained neural network. All the vorticity fields were normalized to have zero mean and unit variance for training. The unconditional models were trained separately for each combination of regime and Reynolds number, but were shared across the two types of observations (Table 3). Transfer learning between regimes and Reynolds numbers and its mechanism are areas of active research in subgrid modeling (Ross et al., 2023; Subel et al., 2023), but are not explored in this work.

To train the conditional model, we follow the same procedure as for the unconditional model, but incorporate conditioning by concatenating the interpolated low-resolution field to the noisy field at each diffusion time-step following Saharia, Ho, et al. (2022). We reiterate that conditional models are application-specific and were trained separately for each main test case (Table 3) and down-sampling scale ( $\delta$ ). We utilize the AdamW optimizer (Loshchilov & Hutter, 2017) with a cosine learning rate scheduler (Loshchilov & Hutter, 2016) with learning rate = 0.0002. We use a batch size of 64 and train for 250 epochs. We use NVIDIA A6000 GPUs for training, with each epoch taking  $\sim 15$  minutes of training on a single GPU. We experimented by adding attention to the bottleneck layer of the U-Net and providing the diffusion pseudo-time  $t$  as an input to the network, but did not observe noticeable performance improvement.

## References

- Anderson, B. D. (1982). Reverse-time diffusion equation models. *Stochastic Processes and their Applications*, 12(3), 313–326.
- Archer, M., Wang, J., Klein, P., Dibarboure, G., & Fu, L.-L. (2025). Wide-swath satellite altimetry unveils global submesoscale ocean dynamics. *Nature*, 640(8059), 691–696.

- Barwey, S., Pal, P., Patel, S., Balin, R., Lusch, B., Vishwanath, V., . . . Balakrishnan, R. (2025). Mesh-based super-resolution of fluid flows with multiscale graph neural networks. *Computer Methods in Applied Mechanics and Engineering*, 443, 118072.
- Bennett, A. F. (1992). *Inverse methods in physical oceanography*. Cambridge university press.
- Blau, Y., & Michaeli, T. (2018). The perception-distortion tradeoff. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 6228–6237).
- Boyd, J. P. (2001). *Chebyshev and fourier spectral methods*. Courier Corporation.
- Bracco, A., Liu, G., & Sun, D. (2019). Mesoscale-submesoscale interactions in the gulf of mexico: From oil dispersion to climate. *Chaos, Solitons & Fractals*, 119, 63–72.
- Brenowitz, N. D., Ge, T., Subramaniam, A., Gupta, A., Hall, D. M., Mardani, M., . . . Pritchard, M. S. (2025). Climate in a bottle: Towards a generative foundation model for the kilometer-scale global atmosphere. *arXiv preprint arXiv:2505.06474*.
- Buzzicotti, M. (2023). Data reconstruction for complex flows using ai: Recent progress, obstacles, and perspectives. *Europhysics Letters*, 142(2), 23001.
- Callaham, J. L., Maeda, K., & Brunton, S. L. (2019). Robust flow reconstruction from limited measurements via sparse representation. *Physical Review Fluids*, 4(10), 103907.
- Chassignet, E., & Marshall, D. (2008). Gulf stream separation in numerical ocean models. *Geophysical Monograph Series*, 177.
- Chassignet, E. P., & Xu, X. (2021). On the importance of high-resolution in large-scale ocean models. *Advances in Atmospheric Sciences*, 38, 1621–1634.
- Chattopadhyay, A., & Hassanzadeh, P. (2023). Long-term instabilities of deep learning-based digital twins of the climate system: The cause and a solution. *arXiv preprint arXiv:2304.07029*.
- Chen, N., Wiggins, S., & Andreou, M. (2025). Taming uncertainty in a complex world: The rise of uncertainty quantification—a tutorial for beginners. *Notices Of The American Mathematical Society*, 72(3).
- Chung, H., Kim, J., Kim, S., & Ye, J. C. (2023). Parallel diffusion models of opera-

- tor and image for blind inverse problems. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 6059–6069).
- Chung, H., Kim, J., McCann, M. T., Klasky, M. L., & Ye, J. C. (2023). Diffusion posterior sampling for general noisy inverse problems. In *11th international conference on learning representations, iclr 2023*.
- Daras, G., Chung, H., Lai, C.-H., Mitsufuji, Y., Ye, J. C., Milanfar, P., . . . Delbracio, M. (2024). A survey on diffusion models for inverse problems. *arXiv preprint arXiv:2410.00083*.
- Dhariwal, P., & Nichol, A. (2021). Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34, 8780–8794.
- Diffenbaugh, N. S., Pal, J. S., Trapp, R. J., & Giorgi, F. (2005). Fine-scale processes regulate the response of extreme events to global climate change. *Proceedings of the National Academy of Sciences*, 102(44), 15774–15778.
- Fan, X., Akhare, D., & Wang, J.-X. (2025). Neural differentiable modeling with diffusion-based super-resolution for two-dimensional spatiotemporal turbulence. *Computer Methods in Applied Mechanics and Engineering*, 433, 117478.
- Fernandez-Martinez, J. L., & Fernandez-Muniz, Z. (2020). The curse of dimensionality in inverse problems. *Journal of Computational and Applied Mathematics*, 369, 112571.
- Frezat, H., Le Sommer, J., Fablet, R., Balarac, G., & Lguensat, R. (2022). A posteriori learning for quasi-geostrophic turbulence parametrization. *Journal of Advances in Modeling Earth Systems*, 14(11), e2022MS003124.
- Fukami, K., Fukagata, K., & Taira, K. (2019). Super-resolution reconstruction of turbulent flows with machine learning. *Journal of Fluid Mechanics*, 870, 106–120.
- Fukami, K., Fukagata, K., & Taira, K. (2021). Machine-learning-based spatio-temporal super resolution reconstruction of turbulent flows. *Journal of Fluid Mechanics*, 909, A9.
- Fukami, K., Fukagata, K., & Taira, K. (2023). Super-resolution analysis via machine learning: a survey for fluid flows. *Theoretical and Computational Fluid Dynamics*, 37(4), 421–444.
- Fukami, K., & Taira, K. (2024). Single-snapshot machine learning for super-resolution of turbulence. *Journal of Fluid Mechanics*, 1001, A32.

- Galperin, B., Nakano, H., Huang, H.-P., & Sukoriansky, S. (2004). The ubiquitous zonal jets in the atmospheres of giant planets and earth’s oceans. *Geophysical research letters*, 31(13).
- Gao, H., Sun, L., & Wang, J.-X. (2021). Super-resolution and denoising of fluid flow using physics-informed convolutional neural networks without high-resolution labels. *Physics of Fluids*, 33(7).
- Graham, J. P., & Ringler, T. (2013). A framework for the evaluation of turbulence closures used in mesoscale ocean large-eddy simulations. *Ocean Modelling*, 65, 25–39.
- Guan, Y., Chattopadhyay, A., Subel, A., & Hassanzadeh, P. (2022). Stable a posteriori les of 2d turbulence using convolutional neural networks: Backscattering analysis and generalization to higher re via transfer learning. *Journal of Computational Physics*, 458, 111090.
- Guan, Y., Hassanzadeh, P., Schneider, T., Dunbar, O., Huang, D. Z., Wu, J., & Lopez-Gomez, I. (2024). Online learning of eddy-viscosity and backscattering closures for geophysical turbulence using ensemble kalman inversion. *arXiv preprint arXiv:2409.04985*.
- Guan, Y., Subel, A., Chattopadhyay, A., & Hassanzadeh, P. (2023). Learning physics-constrained subgrid-scale closures in the small-data regime for stable and accurate les. *Physica D: Nonlinear Phenomena*, 443, 133568.
- Güemes, A., Sanmiguel Vila, C., & Discetti, S. (2022). Super-resolution generative adversarial networks of randomly-seeded fields. *Nature Machine Intelligence*, 4(12), 1165–1173.
- Gula, J., Blacic, T. M., & Todd, R. E. (2019). Submesoscale coherent vortices in the gulf stream. *Geophysical Research Letters*, 46(5), 2704–2714.
- Gupta, A., & Lermusiaux, P. F. J. (2021, August). Neural closure models for dynamical systems. *Proceedings of The Royal Society A*, 477(2252), 1–29. doi: 10.1098/rspa.2020.1004
- Gupta, A., & Lermusiaux, P. F. J. (2023, June). Generalized neural closure models with interpretability. *Scientific Reports*, 13, 10634. doi: 10.1038/s41598-023-35319-w
- Haley, P. J., Jr., & Lermusiaux, P. F. J. (2010, December). Multiscale two-way embedding schemes for free-surface primitive equations in the “Multidisciplinary

- Simulation, Estimation and Assimilation System”. *Ocean Dynamics*, 60(6), 1497–1537. doi: 10.1007/s10236-010-0349-4
- Harris, L. M., & Lin, S.-J. (2014). Global-to-regional nested grid climate simulations in the gfdl high resolution atmospheric model. *Journal of Climate*, 27(13), 4890–4910.
- Ho, J., & Salimans, T. (2021). Classifier-free diffusion guidance. In *Neurips 2021 workshop on deep generative models and downstream applications*.
- Ho, J., Salimans, T., Gritsenko, A., Chan, W., Norouzi, M., & Fleet, D. J. (2022). Video diffusion models. *Advances in Neural Information Processing Systems*, 35, 8633–8646.
- Huang, H.-P., Galperin, B., & Sukoriansky, S. (2001). Anisotropic spectra in two-dimensional turbulence on the surface of a rotating sphere. *Physics of Fluids*, 13(1), 225–240.
- Jacobs, G., D’Addezio, J., Bartels, B., DeHaan, C., Barron, C., Carrier, M., ... Dever, M. (2023). Adapting constrained scales to observation resolution in ocean forecasts. *Ocean Modelling*, 186, 102252.
- Jakhar, K., Guan, Y., Mojgani, R., Chattopadhyay, A., & Hassanzadeh, P. (2024). Learning closed-form equations for subgrid-scale closures from high-fidelity data: Promises and challenges. *Journal of Advances in Modeling Earth Systems*, 16(7), e2023MS003874.
- Karras, T., Aittala, M., Kynkäänniemi, T., Lehtinen, J., Aila, T., & Laine, S. (2024). Guiding a diffusion model with a bad version of itself. *Advances in Neural Information Processing Systems*, 37, 52996–53021.
- Kim, H., Kim, J., Won, S., & Lee, C. (2021). Unsupervised deep learning for super-resolution reconstruction of turbulence. *Journal of Fluid Mechanics*, 910, A29.
- Klemas, V., & Yan, X.-H. (2014). Subsurface and deeper ocean remote sensing from satellites: An overview and new results. *Progress in oceanography*, 122, 1–9.
- Kossale, Y., Airaj, M., & Darouichi, A. (2022). Mode collapse in generative adversarial networks: An overview. In *2022 8th international conference on optimization and applications (icoa)* (pp. 1–6).
- Kraichnan, R. H. (1971). Inertial-range transfer in two-and three-dimensional turbulence. *Journal of Fluid Mechanics*, 47(3), 525–535.

- Lai, C.-Y., Hassanzadeh, P., Sheshadri, A., Sonnewald, M., Ferrari, R., & Balaji, V. (2024). Machine learning for climate physics and simulations. *Annual Review of Condensed Matter Physics*, 16.
- Leinonen, J., Hamann, U., Nerini, D., Germann, U., & Franch, G. (2023). Latent diffusion models for generative precipitation nowcasting with accurate uncertainty quantification. *arXiv preprint arXiv:2304.12891*.
- Leonard, A. (1975). Energy cascade in large-eddy simulations of turbulent fluid flows. In *Advances in geophysics* (Vol. 18, pp. 237–248). Elsevier.
- Lermusiaux, P. F. J. (2006). Uncertainty estimation and prediction for interdisciplinary ocean dynamics. *Journal of Computational Physics*, 217(1), 176–199. doi: 10.1016/j.jcp.2006.02.010
- Lermusiaux, P. F. J. (2007). Adaptive modeling, adaptive data assimilation and adaptive sampling. *Physica D: Nonlinear Phenomena*, 230(1), 172–196. doi: 10.1016/j.physd.2007.02.014
- Lermusiaux, P. F. J. (2015, May). *Numerical fluid mechanics*. MIT OpenCourseWare. Retrieved from <https://ocw.mit.edu/courses/mechanical-engineering/2-29-numerical-fluid-mechanics-spring-2015/lecture-notes-and-references/>
- Lermusiaux, P. F. J., Chiu, C.-S., Gawarkiewicz, G. G., Abbot, P., Robinson, A. R., Miller, R. N., ... Lekien, F. (2006). Quantifying uncertainties in ocean predictions. *Oceanography*, 19(1), 92–105. doi: 10.5670/oceanog.2006.93
- Lermusiaux, P. F. J., Haley, P. J., Leslie, W. G., Agarwal, A., Logutov, O., & Burton, L. J. (2011). Multiscale physical and biological dynamics in the Philippine Archipelago: Predictions and processes. *Oceanography*, 24(1), 70–89. (Special Issue on the Philippine Straits Dynamics Experiment) doi: 10.5670/oceanog.2011.05
- Lermusiaux, P. F. J., Robinson, A. R., Haley, P. J., & Leslie, W. G. (2002). Advanced interdisciplinary data assimilation: Filtering and smoothing via error subspace statistical estimation. In *Proceedings of the oceans 2002 mts/ieee conference* (pp. 795–802). doi: 10.1109/oceans.2002.1192071
- Lermusiaux, P. F. J., Subramani, D. N., Lin, J., Kulkarni, C. S., Gupta, A., Dutt, A., ... Jana, S. (2017, November). A future for intelligent autonomous ocean observing systems. *Journal of Marine Research*, 75(6), 765–813.

- (The Sea. Volume 17, The Science of Ocean Prediction, Part 2.) doi: 10.1357/002224017823524035
- Li, T., Lanotte, A. S., Buzzicotti, M., Bonaccorso, F., & Biferale, L. (2023). Multi-scale reconstruction of turbulent rotating flows with generative diffusion models. *Atmosphere*, 15(1), 60.
- Li, Z., Kovachki, N., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A., & Anandkumar, A. (2020). Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895*.
- Liu, B., Tang, J., Huang, H., & Lu, X.-Y. (2020). Deep learning methods for super-resolution reconstruction of turbulent flows. *Physics of fluids*, 32(2).
- Lopez-Gomez, I., Wan, Z. Y., Zepeda-Núñez, L., Schneider, T., Anderson, J., & Sha, F. (2025). Dynamical-generative downscaling of climate model ensembles. *Proceedings of the National Academy of Sciences*, 122(17), e2420288122.
- Loshchilov, I., & Hutter, F. (2016). Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*.
- Loshchilov, I., & Hutter, F. (2017). Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*.
- Lu, P., & Lermusiaux, P. F. J. (2021, December). Bayesian learning of stochastic dynamical models. *Physica D: Nonlinear Phenomena*, 427, 133003. doi: 10.1016/j.physd.2021.133003
- Manshausen, P., Cohen, Y., Harrington, P., Pathak, J., Pritchard, M., Garg, P., ... Brenowitz, N. (2024). Generative data assimilation of sparse weather station observations at kilometer scales. *arXiv preprint arXiv:2406.16947*.
- Mardani, M., Brenowitz, N., Cohen, Y., Pathak, J., Chen, C.-Y., Liu, C.-C., ... others (2025). Residual corrective diffusion modeling for km-scale atmospheric downscaling. *Communications Earth & Environment*, 6(1), 124.
- Martin, S., Manucharyan, G., & Klein, P. (2025). Generative data assimilation for surface ocean state estimation from multi-modal satellite observations. *Earth-ArXiv eprints*, X5ZT6N.
- Maulik, R., San, O., Rasheed, A., & Vedula, P. (2019). Subgrid modelling for two-dimensional turbulence using neural networks. *Journal of Fluid Mechanics*, 858, 122–144.
- McWilliams, J. C. (2016). Submesoscale currents in the ocean. *Proceedings of the*

- Royal Society A: Mathematical, Physical and Engineering Sciences*, 472(2189), 20160117.
- Meng, C., He, Y., Song, Y., Song, J., Wu, J., Zhu, J.-Y., & Ermon, S. (2021). Sdedit: Guided image synthesis and editing with stochastic differential equations. *arXiv preprint arXiv:2108.01073*.
- Mensa, J. A., Garraffo, Z., Griffa, A., Özgökmen, T. M., Haza, A., & Veneziani, M. (2013). Seasonality of the submesoscale dynamics in the gulf stream region. *Ocean Dynamics*, 63, 923–941.
- Milanfar, P., & Delbracio, M. (2025). Denoising: a powerful building block for imaging, inverse problems and machine learning. *Philosophical Transactions A*, 383(2299), 20240326.
- Moore, A. M., Martin, M., Akella, S., Arango, H., Balmaseda, M., Bertino, L., ... Weaver, A. T. (2019, March). Synthesis of ocean observations using data assimilation for operational, real-time and reanalysis systems: A more complete picture of the state of the ocean. *Frontiers in Marine Science*, 6(90), 1–6. doi: 10.3389/fmars.2019.00090
- Moser, B. B., Raue, F., Frolov, S., Palacio, S., Hees, J., & Dengel, A. (2023). Hitchhiker’s guide to super-resolution: Introduction and recent advances. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(8), 9862–9882.
- Nichol, A. Q., & Dhariwal, P. (2021). Improved denoising diffusion probabilistic models. In *International conference on machine learning* (pp. 8162–8171).
- Nickerson, A. K., Weisberg, R. H., & Liu, Y. (2022). On the evolution of the gulf of mexico loop current through its penetrative, ring shedding and retracted states. *Advances in Space Research*, 69(11), 4058–4077.
- Oey, L., Ezer, T., & Lee, H. (2005). Loop current, rings and related circulation in the gulf of mexico: A review of numerical models and future challenges. *Geophysical Monograph-American Geophysical Union*, 161, 31.
- Oommen, V., Bora, A., Zhang, Z., & Karniadakis, G. E. (2025). Integrating neural operators with diffusion models improves spectral representation in turbulence modelling. *Proceedings of the Royal Society A*, 481(2309), 20240819.
- Orszag, S. A., & Israeli, M. (1974). Numerical simulation of viscous incompressible flows. *Annual Review of Fluid Mechanics*, 6(1), 281–318.

- Page, J. (2025). Super-resolution of turbulence with dynamics in the loss. *Journal of Fluid Mechanics*, 1002, R3.
- Park, C. Y., McCann, M. T., Garcia-Cardona, C., Wohlberg, B., & Kamilov, U. S. (2024). Random walks with tweedie: A unified framework for diffusion models. *arXiv preprint arXiv:2411.18702*.
- Patel, D. V., Ray, D., & Oberai, A. A. (2022). Solution of physics-based bayesian inverse problems with deep generative priors. *Computer Methods in Applied Mechanics and Engineering*, 400, 115428.
- Perezhogin, P., Zanna, L., & Fernandez-Granda, C. (2023). Generative data-driven approaches for stochastic subgrid parameterizations in an idealized ocean model. *Journal of Advances in Modeling Earth Systems*, 15(10), e2023MS003681.
- Pope, S. B. (2001). Turbulent flows. *Measurement Science and Technology*, 12(11), 2020–2021.
- Qi, Z., Bai, L., Xiong, H., & Xie, Z. (2024). Not all noises are created equally: Diffusion noise selection and optimization. *arXiv preprint arXiv:2407.14041*.
- Qu, Y., Nathaniel, J., Li, S., & Gentine, P. (2024). Deep generative data assimilation in multimodal setting. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition* (pp. 449–459).
- Rhines, P. B. (1975). Waves and turbulence on a beta-plane. *Journal of Fluid Mechanics*, 69(3), 417–443.
- Robbins, H. E. (1992). An empirical bayes approach to statistics. In *Breakthroughs in statistics: Foundations and basic theory* (pp. 388–394). Springer.
- Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition* (pp. 10684–10695).
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention—miccai 2015: 18th international conference, munich, germany, october 5-9, 2015, proceedings, part iii 18* (pp. 234–241).
- Ross, A., Li, Z., Perezhogin, P., Fernandez-Granda, C., & Zanna, L. (2023). Benchmarking of machine learning ocean subgrid parameterizations in an idealized

- model. *Journal of Advances in Modeling Earth Systems*, 15(1).
- Rozet, F., & Louppe, G. (2023). Score-based data assimilation. *Advances in Neural Information Processing Systems*, 36, 40521–40541.
- Saharia, C., Chan, W., Chang, H., Lee, C., Ho, J., Salimans, T., . . . Norouzi, M. (2022). Palette: Image-to-image diffusion models. In *Acm siggraph 2022 conference proceedings* (pp. 1–10).
- Saharia, C., Ho, J., Chan, W., Salimans, T., Fleet, D. J., & Norouzi, M. (2022). Image super-resolution via iterative refinement. *IEEE transactions on pattern analysis and machine intelligence*, 45(4), 4713–4726.
- Sardar, M., Skillen, A., Zimoń, M., Draycott, S., & Revell, A. (2024). Spectrally decomposed denoising diffusion probabilistic models for generative turbulence super-resolution. *Physics of Fluids*, 36(11).
- Sofos, F., & Drikakis, D. (2025). A review of deep learning for super-resolution in fluid flows. *Physics of Fluids*, 37(4).
- Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., & Ganguli, S. (2015). Deep unsupervised learning using nonequilibrium thermodynamics. In *International conference on machine learning* (pp. 2256–2265).
- Song, J., Meng, C., & Ermon, S. (2020). Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*.
- Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., & Poole, B. (2020). Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*.
- Souza, A. N., Silvestri, S., Deck, K., Bischoff, T., Flierl, G., & Ferrari, R. (2025). Surface to seafloor: A generative machine learning framework for decoding the ocean interior state. *arXiv preprint arXiv:2504.15308*.
- Srinivasan, K., Chekroun, M. D., & McWilliams, J. C. (2024). Turbulence closure with small, local neural networks: Forced two-dimensional and  $\beta$ -plane flows. *Journal of Advances in Modeling Earth Systems*, 16(4), e2023MS003795.
- Stommel, H. (1948). The westward intensification of wind-driven ocean currents. *Eos, Transactions American Geophysical Union*, 29(2), 202–206.
- Storto, A., Alvera-Azcárate, A., Balmaseda, M. A., Barth, A., Chevallier, M., Counillon, F., . . . others (2019). Ocean reanalyses: recent advances and unsolved challenges. *Frontiers in Marine Science*, 6, 418.

- Subel, A., Guan, Y., Chattopadhyay, A., & Hassanzadeh, P. (2023). Explaining the physics of transfer learning in data-driven turbulence modeling. *PNAS nexus*, 2(3), pgad015.
- Suresh Babu, A. N., Sadam, A., & Lermusiaux, P. F. J. (2025a, June). *Guided unconditional and conditional generative models for super-resolution and inference of quasi-geostrophic turbulence [Dataset]*. Zenodo. doi: 10.5281/zenodo.15742146
- Suresh Babu, A. N., Sadam, A., & Lermusiaux, P. F. J. (2025b, June). *Guided unconditional and conditional generative models for super-resolution and inference of quasi-geostrophic turbulence [Software]*. Zenodo. doi: 10.5281/zenodo.15750244
- Taylor, J. R., & Thompson, A. F. (2023). Submesoscale dynamics in the upper ocean. *Annual Review of Fluid Mechanics*, 55(1), 103–127.
- Theiss, J. (2006). A generalized rhines effect and storms on jupiter. *Geophysical research letters*, 33(8).
- Thomas, L. N., Tandon, A., & Mahadevan, A. (2008). Submesoscale processes and dynamics. *Ocean modeling in an Eddying Regime*, 177, 17–38.
- Towne, A., Lozano-Durán, A., & Yang, X. (2020). Resolvent-based estimation of space-time flow statistics. *Journal of Fluid Mechanics*, 883, A17.
- Vallis, G. K. (2017). *Atmospheric and oceanic fluid dynamics*. Cambridge University Press.
- Wang, B., & Vastola, J. J. (2023). Diffusion models generate images like painters: an analytical theory of outline first, details later. *arXiv preprint arXiv:2303.02490*.
- Wang, Z., Jiang, Y., Zheng, H., Wang, P., He, P., Wang, Z., ... others (2023). Patch diffusion: Faster and more data-efficient training of diffusion models. *Advances in neural information processing systems*, 36, 72137–72154.
- Wiatrak, M., Albrecht, S. V., & Nystrom, A. (2019). Stabilizing generative adversarial networks: A survey. *arXiv preprint arXiv:1910.00927*.
- Williams, G. P. (1978). Planetary circulations: 1. barotropic representation of jovian and terrestrial turbulence. *Journal of Atmospheric Sciences*, 35(8), 1399–1426.
- Wkeglarczyk, S. (2018). Kernel density estimation and its application. In *Itm web of*

- conferences* (Vol. 23, p. 00037).
- Wunsch, C. (2006). *Discrete inverse and state estimation problems: with geophysical fluid applications*. Cambridge University Press.
- Xu, Q., Zhuang, Z., Pan, Y., & Wen, B. (2023). Super-resolution reconstruction of turbulent flows with a transformer-based deep learning framework. *Physics of Fluids*, 35(5).
- Zanna, L., & Bolton, T. (2020). Data-driven equation discovery of ocean mesoscale closures. *Geophysical Research Letters*, 47(17), e2020GL088376.
- Zanna, L., Mana, P. P., Anstey, J., David, T., & Bolton, T. (2017). Scale-aware deterministic and stochastic parametrizations of eddy-mean flow interaction. *Ocean Modelling*, 111, 66–80.
- Zhang, Q., & Chen, Y. (2022). Fast sampling of diffusion models with exponential integrator. *arXiv preprint arXiv:2204.13902*.
- Zhang, Z., Miao, M., Qiu, B., Tian, J., Jing, Z., Chen, G., ... Zhao, W. (2024). Submesoscale eddies detected by swot and moored observations in the north-western pacific. *Geophysical Research Letters*, 51(15), e2024GL110000.
- Zhou, T., Krahenbuhl, P., Aubry, M., Huang, Q., & Efros, A. A. (2016). Learning dense correspondence via 3d-guided cycle consistency. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 117–126).
- Zhou, Z., He, G., Wang, S., & Jin, G. (2019). Subgrid-scale model for large-eddy simulation of isotropic turbulent flows using an artificial neural network. *Computers & Fluids*, 195, 104319.