
NUTS: Eddy-Robust Reconstruction of Surface Ocean Nutrients via Two-Scale Modeling

Hao Zheng¹* Shiyu Liang¹*† Yuting Zheng¹ Chaofan Sun¹ Lei Bai² Enhui Liao¹

¹Shanghai Jiao Tong University, China ²Shanghai Artificial Intelligence Laboratory, China

{hubert.zheng, lsy18602808513, zhengyt058, scf024, ehliao}@sjtu.edu.cn
baisanshi@gmail.com

Abstract

Reconstructing ocean surface nutrients from sparse observations is critical for understanding long-term biogeochemical cycles. Most prior work focuses on reconstructing atmospheric fields and treats the reconstruction problem as image inpainting, assuming smooth, single-scale dynamics. In contrast, nutrient transport follows advection–diffusion dynamics under nonstationary, multiscale ocean flow. This mismatch leads to instability, as small errors in unresolved eddies can propagate through time and distort nutrient predictions. To address this, we introduce NUTS, a two-scale reconstruction model that decouples large-scale transport and mesoscale variability. The homogenized solver captures stable, coarse-scale advection under filtered flow. A refinement module then restores mesoscale detail conditioned on the residual eddy field. NUTS is stable, interpretable, and robust to mesoscale perturbations, with theoretical guarantees from homogenization theory. NUTS outperforms all data-driven baselines in global reconstruction and achieves site-wise accuracy comparable to numerical models. On real observations, NUTS reduces NRMSE by 79.9% for phosphate and 19.3% for nitrate over the best baseline. Ablation studies validate the effectiveness of each module.

1 Introduction

Reconstructing historical nutrient concentrations in the surface ocean is essential for understanding long-term biogeochemical cycles, ecosystem variability, and anthropogenic influence Stüeken et al. [2024]. However, nutrient observations are extremely sparse, especially before the bio-Argo era when data came from irregular ship-based campaigns. Even today, nutrient data remain far less available than satellite-measured variables like sea surface temperature (SST) or chlorophyll Mishonov et al. [2024], Locarnini et al. [2018].

Recent deep learning advances have driven progress in forecasting and reconstructing high-dimensional atmospheric fields. Transformer-based models Pathak et al. [2022], Bi et al. [2023], Lam et al. [2023] achieve state-of-the-art short-term forecasts by capturing temporal dependencies in data-rich regimes with complete initial conditions. In contrast, climate field reconstruction operates in sparse settings and is often framed as a spatial in-painting task Bochow et al. [2025], Plésiat et al. [2024], Kadow et al. [2020]. Early models Ronneberger et al. [2015], Dosovitskiy et al. [2020], Gao et al. [2022] focus on spatial correlations, while recent hybrids Li et al. [2020], Wang et al. [2025], Beauchamp et al. [2023] add physical constraints for greater consistency. However, these methods are mainly designed for smooth, single-scale wind fields with well-resolved large-scale structure.

*Equal contribution.

†Corresponding author.

Reconstructing ocean nutrients demands a fundamentally different approach. Unlike atmospheric fields, nutrient transport follows advection–diffusion dynamics driven by a nonstationary, multiscale velocity field. Surface currents consist of a slowly evolving large-scale mean flow overlaid with rapidly fluctuating mesoscale eddies. These eddies—coherent vortices spanning 10–100 km—govern most lateral nutrient transport Vallis [2017], McWilliams [2016], Chelton et al. [2011], yet are poorly resolved in numerical circulation models due to limited resolution and inherent uncertainty. As a result, reconstruction models that rely directly on such flow fields are fragile: even small perturbations in the eddy component can degrade predictions.

Robust nutrient reconstruction presents a core modeling dilemma. Filtering the input velocity field improves stability by suppressing high-frequency eddy perturbations. However, it also removes fine-scale structures essential for capturing local nutrient gradients. Retaining all scales introduces instability; over-filtering sacrifices resolution. A principled solution must separate scales—preserving large-scale transport while reintroducing mesoscale variability in a controlled manner.

We propose NUTS, a novel and robust two-scale model that, for the first time, resolves the reconstruction challenge through a structured decomposition. At its core is a *homogenized advection–diffusion solver* that models nutrient transport under the filtered large-scale flow. By replacing unresolved mesoscale variability with an effective diffusion term, this formulation captures the net impact of fine-scale dynamics without tracking unstable eddy fluctuations. The *coarse module* leverages this framework to propagate nutrient fields with stability and physical consistency. To recover fine-scale structure, the *refinement module* models localized redistribution conditioned on the residual mesoscale flow and the coarse prediction. This coarse-to-refined architecture preserves large-scale transport patterns while restoring spatial detail in dynamically active regions. NUTS is robust to mesoscale perturbations, respects scale separation, and generalizes effectively under sparse observational coverage. We establish accuracy and stability guarantees under standard assumptions from homogenization theory, and empirically demonstrate that NUTS consistently outperforms prior baselines on both simulated and real-world datasets. Our contributions are as follows:

- We formulate nutrient reconstruction as a spatiotemporal advection–diffusion problem and reveal the vulnerability of naive methods to mesoscale perturbations.
- We propose NUTS, a two-scale model that combines a homogenized PDE solver with adaptive diffusion and a refinement module conditioned on normalized eddy flow. We provide theoretical justification of its effectiveness under standard homogenization assumptions.
- We empirically demonstrate that NUTS outperforms all data-driven baselines in global nutrient reconstruction on both simulated and real-world datasets, achieving site-wise accuracy comparable to physics-based numerical models. On the WOD dataset of real observations, NUTS reduces NRMSE by 79.9% for phosphate and 19.3% for nitrate relative to the best baseline.
- Ablation studies highlight the contribution of each component and offer empirical guidance for designing robust reconstruction architectures.

2 Related Work

This section outlines key related work and we provide a comprehensive review with extended background and references in Appendix A.

Nutrient Data. Ocean nutrient data are typically derived from observational datasets and simulation-based products. Raw observational archive, such as WOD Mishonov et al. [2024], provide high-quality and in-situ measurements but suffer from sparse and uneven distribution. In contrast, simulation-based products like the CMEMS Global Ocean Biogeochemical Hindcast (GOBH) Perruche [2018], ECCO-Darwin Carroll et al. [2020], MOM6-COBALT2 Griffies et al. [2012] can offer global coverage data product with coupled physical and biogeochemical dynamics, but require extensive calibration. Furthermore, most of these simulation-based products do not incorporate biogeochemical data assimilation and often employ simplified parameterization of biogeochemical processes, resulting in regional biases and uncertainties in nutrient fields.

Reconstruction Approaches. Traditional methods such as optimal interpolation Conkright et al. [2002], 3D/4D-Var Courtier et al. [1994], ensemble Kalman filters Nerger and Gregg [2008], and variational inverse models Brasseur and Haus [1991] rely on data assimilation and inverse modeling to integrate sparse observations with physical dynamics, but are often limited by computational cost

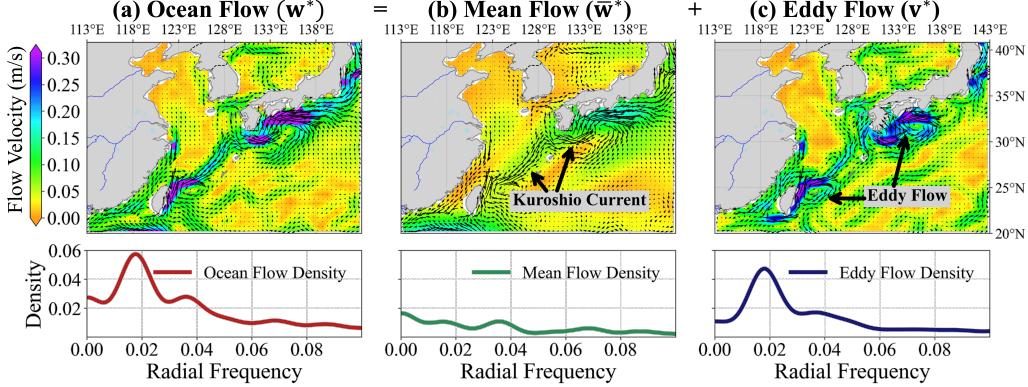


Figure 1: **Decomposition of surface ocean flow into mean and eddy components.** (a) Full velocity field containing both large-scale and mesoscale structures. (b) Mean flow obtained via low-pass filtering, capturing large-scale structures such as the Kuroshio Current. (c) Eddy flow computed as the residual, representing high-frequency mesoscale variability. Bottom panels display the radial frequency spectra corresponding to each flow component, with energy concentrated at low frequencies for the mean flow and at higher frequencies for the eddy flow, illustrating effective scale separation.

and data sparsity. Recent advances in deep learning offer alternative solutions for spatiotemporal reconstruction. CNN-based models (e.g., U-Net Ronneberger et al. [2015]) and transformers (e.g., ViT Dosovitskiy et al. [2020], Earthformer Gao et al. [2022]) capture spatial structures but lack physical grounding. Physics-informed approaches—such as neural operators Li et al. [2020], Wang et al. [2025], implicit neural representations Luo et al. [2024], and 4DVarNet Beauchamp et al. [2023]—embed governing equations or physical constraints into the learning process to improve physical consistency but are limited in capacity. Foundation models (e.g., Prithvi Schmude et al. [2024], AtmoRep Lessig et al. [2023]) show promise in meteorology but remain untested in marine biogeochemistry. General-purpose inpainting methods using GANs Zhao et al. [2021] and diffusion models Lugmayr et al. [2022] perform well in vision tasks but lack physical constraints and robustness to sparse data.

3 Methodology

Notations. Let \mathbb{S}^2 denote the unit surface in \mathbb{R}^3 , parameterized by latitude-longitude coordinates $\mathbf{x} = (\theta, \phi) \in \Omega = [-\frac{\pi}{2}, \frac{\pi}{2}] \times [-\pi, \pi]$. For a time-dependent function $\varphi(\theta, \phi, t)$, define $\dot{\varphi} = \frac{\partial \varphi}{\partial t}$. The divergence and spherical Laplacian operators are denoted $\nabla \cdot$ and ∇^2 , respectively.

Problem Setup. The nutrient concentration φ follow the advection-diffusion equation: $\mathcal{L}_{\mathbf{w}, \eta}[\varphi] = \dot{\varphi} + \nabla \cdot (\mathbf{w}\varphi) - \eta \nabla^2 \varphi = s$, where $\eta = \eta(\theta, \phi)$ denotes the time-invariant diffusion coefficient and $s = s(\theta, \phi, t)$ represents the external source and sink terms. These terms account for biological uptake and remineralization through photosynthesis, respiration, and demineralization, as well as physical downwelling and upwelling. Given sparse nutrient measurements on $\mathcal{Z} \times \mathcal{T} \subset \Omega \times [0, T]$ and perturbed ocean flow estimates \mathbf{w} , our goal is reconstructing nutrient concentrations by solving the constrained PDE: $\mathcal{L}_{\mathbf{w}, \eta}[\varphi] = s$, subject to $\varphi|_{\mathcal{Z} \times \mathcal{T}} = f|_{\mathcal{Z} \times \mathcal{T}}$, where f represents observed nutrient concentrations and $f|_{\mathcal{Z} \times \mathcal{T}}$ denotes its restriction to the subset $\mathcal{Z} \times \mathcal{T}$.

3.1 A Naive Spatial-Temporal Reconstruction Model

In this subsection, we introduce a naive spatiotemporal reconstruction model and discuss its advantage over image-inpainting-based methods. We then demonstrate its sensitivity to mesoscale perturbations in the eddy component of the ocean velocity field.

Naive Model. Given an interval $[t_0, t_1]$, the naive model first uses a data-driven initializer \mathcal{F}_0 to estimate the initial nutrient field $\hat{\varphi}(\mathbf{x}, t_0)$ through the velocity field \mathbf{w} , auxiliary variables Φ , and sparse observations f . The estimate is then propagated by solving the advection-diffusion equation $\mathcal{L}_{\mathbf{w}, \eta}[\hat{\varphi}] = s$, where both the diffusion coefficient η and source term s are learned to match the true field. Prior work Schiesser [2012] has demonstrated that this propagation can be implemented via the method of lines (MOL), which discretizes the PDE into a system of first-order ODEs at spatial locations $\{\mathbf{x}_k\}_k$: $\hat{\varphi}(\mathbf{x}_k, t) = \hat{\varphi}(\mathbf{x}_k, t_0) + \int_{t_0}^t [-\nabla \cdot (\mathbf{w}\varphi) + \eta \nabla^2 \varphi + s] (\mathbf{x}_k, \tau) d\tau$, where the forward solution can be solved approximately using numerical solvers such as Runge–Kutta LeVeque [2007]. During the model training, all components, i.e., \mathcal{F}_0 , s and η are jointly optimized to minimize

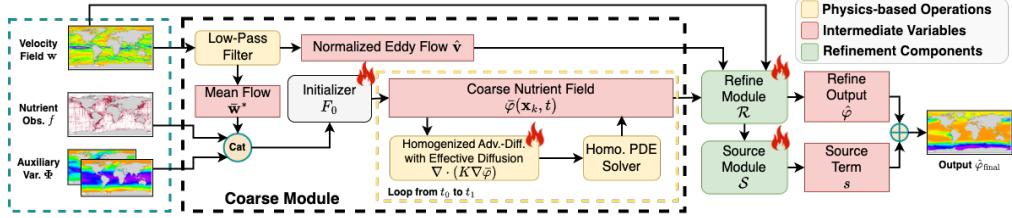


Figure 2: **Overview of NUTS.** NUTS is a two-scale model that combines a data-driven initializer, a homogenized PDE solver with learned effective diffusion, and a refinement module to reconstruct ocean nutrients under sparse observations. The velocity field is decomposed into mean and eddy components for scale separation. A learnable source module captures unresolved inputs. Trainable modules are marked with fire icons. Cat denotes channel concatenation and \oplus denotes element-wise addition.

the mean squared error between the prediction $\hat{\varphi}$ and the ground-truth φ , i.e., $\min_{\mathcal{F}_0, s, \eta} \mathcal{L}_{\text{MSE}} \triangleq \|\varphi - \hat{\varphi}\|_2^2$.

Advantages over Existing Image In-painting Approach. (1) *Physical consistency and mass conservation.* The naive model evolves nutrient fields through an advection–diffusion PDE, ensuring temporally consistent reconstructions that follow physical transport processes and conserve mass. In contrast, image in-painting methods rely purely on spatial interpolation, lacking temporal dynamics and physical grounding. (2) *Effective use of sparse observations.* By jointly learning the initializer, source term, and diffusion coefficient, the naive model directly integrates observational data to constrain transport dynamics, leading to more data-consistent estimates in sparsely sampled regions.

Sensitivity to Mesoscale Perturbations. The naive reconstruction approach evolves nutrient estimates using velocity fields from numerical circulation models, which accurately capture large-scale mean currents but often misrepresent mesoscale eddies due to limited resolution and structural uncertainties. Mesoscale eddies are small-scale (10–100 km), high-energy structures that play a dominant role in nutrient transport. As illustrated in Figure 1, the true velocity field w^* can be decomposed into a smooth mean component \bar{w}^* and a rapidly fluctuating eddy component v^* . When the MOL uses a perturbed velocity field $w_\delta = w^* + \delta$, structural errors δ in the eddy component introduce an additional transport term into the advection–diffusion dynamics:

$$\hat{\varphi}(x_k, t_1) = \hat{\varphi}(x_k, t_0) + \int_{t_0}^{t_1} \underbrace{[-\nabla \cdot (w^* \varphi) + \eta \nabla^2 \varphi + s]}_{\text{true advection-diffusion}}(x_k, \tau) - \underbrace{[\nabla \cdot (\delta \varphi)](x_k, \tau)}_{\text{error from flow perturbation}} d\tau. \quad (1)$$

The perturbation term scales with $\|\delta\|$, which can be large, as mesoscale eddies typically carry more energy than the mean flow (Figure 1). By Equation (1), such perturbations induce significant transport errors that accumulate and propagate over time. This underscores the need for reconstruction models that are robust to mesoscale flow inaccuracies.

3.2 NUTS: Eddy-Robust Nutrient Reconstruction via Two-Scale Modeling

We introduce **NUTS**, a principled two-scale model that reconstructs surface ocean nutrients from sparse observations and noisy velocity inputs (see Figure 2). Unlike prior approaches, NUTS separates nutrient transport into stable mean dynamics and unstable mesoscale variability. It applies a homogenized PDE solver for large-scale propagation and a refinement module for controlled recovery of fine-scale structure. This decomposition improves robustness and generalization in multiscale ocean flows. All architectural details are provided in Appendix B.

Coarse Module Part I: Robust Initializer. The coarse stage begins by estimating the nutrient field $\bar{\varphi}(x_k, t_0)$ at the start of the reconstruction interval. To suppress mesoscale noise, we apply a Fourier-based low-pass spatial filter to the input velocity field and extract the mean flow \bar{w}^* . This filtered flow, along with sparse nutrient observations and auxiliary variables, is encoded by a spatiotemporal transformer that captures long-range dependencies across space and time. The initializer is designed to be robust to flow perturbations and produces a stable starting point for physical propagation.

Coarse Module Part II: Homogenized PDE Solver. To evolve the field forward, NUTS applies a homogenized advection–diffusion equation:

$$\bar{\varphi}(x_k, t_1) = \bar{\varphi}(x_k, t_0) + \int_{t_0}^{t_1} \left[\underbrace{-\nabla \cdot (\bar{w}^* \hat{\varphi})}_{\text{mean flow advection}} + \underbrace{\nabla \cdot (K \nabla \hat{\varphi})}_{\text{effective diffusion}} \right] (x_k, \tau) d\tau. \quad (2)$$

This formulation replaces unresolved mesoscale effects with an effective diffusion tensor $K(\mathbf{x})$, which is predicted by a hypernetwork conditioned on $\bar{\varphi}$. We discretize the system using the method of lines and numerically integrate it over time. This structured PDE solver ensures stable and physically grounded transport under filtered dynamics.

Refinement Module. The refinement stage corrects residual errors and restores mesoscale variability. It takes as input the coarse prediction $\bar{\varphi}$, mean flow $\bar{\mathbf{w}}^*$, normalized eddy velocity $\hat{\mathbf{v}} = (\mathbf{w} - \bar{\mathbf{w}}^*)/\|\mathbf{w} - \bar{\mathbf{w}}^*\|_\infty$, sparse observations, and static covariates. These inputs are tokenized and passed through a vision transformer \mathcal{R} that produces the refined estimate $\hat{\varphi}(\mathbf{x}, t)$, i.e., $\hat{\varphi}(\mathbf{x}, t) = \mathcal{R}[\bar{\varphi}, \bar{\mathbf{w}}^*, \hat{\mathbf{v}}, \Phi, f](\mathbf{x}, t)$. Refinement is performed independently at each timestep and learns localized spatial redistribution driven by eddy structures.

Source Term and Conservation Loss. To account for unresolved sources and sinks, we introduce a learnable correction term $s = \mathcal{S}(\hat{\varphi})$, where \mathcal{S} is parameterized by a ResNet. The final prediction is $\hat{\varphi}_{\text{final}} = \hat{\varphi} + s$. To enforce physical realism, we define total nutrient mass as $M[\varphi](t) = \sum_k \varphi(\mathbf{x}_k, t)$, and penalize mass drift through the conservation loss:

$$\mathcal{L}_{\text{cons.}} = \int_{t_0}^{t_1} |M[\bar{\varphi}](\tau) - M[\bar{\varphi}](t_0)|^2 + |M[\hat{\varphi}](\tau) - M[\bar{\varphi}](t_0)|^2 d\tau.$$

The final training objective is: $\mathcal{L}_{\text{total}} = \|\hat{\varphi}_{\text{final}} - \varphi\|_2^2 + \lambda \mathcal{L}_{\text{cons.}}$, which governs the optimization of all learnable components in NUTS.

Core Insight: Why Two-Scale Modeling Works. The key challenge in reconstructing ocean nutrient fields lies in the dual nature of the underlying dynamics: large-scale currents govern basin-wide transport, while mesoscale eddies induce localized variability and dominate error sensitivity. NUTS addresses this by explicitly separating these two regimes. The coarse module filters out unstable mesoscale fluctuations and models stable transport via a homogenized PDE with learnable diffusion. This prevents error accumulation from uncertain eddy inputs. The refinement module then selectively reintroduces mesoscale information—not as direct forcing, but as spatial corrections conditioned on the residual flow. This two-stage architecture mirrors the physical structure of ocean transport and enables both robustness and resolution in a way that single-scale models cannot.

Advantages over the Naive Approach. NUTS preserves the physical grounding of the naive model, including advection–diffusion transport and the effective use of sparse observations. But it adds two critical improvements: (1) *Scale-aware architecture*. By decoupling mean and eddy-driven dynamics, NUTS reconstructs both broad circulation and localized nutrient features with greater fidelity. (2) *Built-in robustness*. Homogenization shields the system from mesoscale perturbation errors, while spatial refinement restores resolution without destabilizing temporal evolution.

Context and Relation to Prior Work. While prior hybrid models such as FNO Li et al. [2020], 4DVarNet Beauchamp et al. [2023], and GraphCast Lam et al. [2023] embed physical priors into data-driven forecasting pipelines, they typically rely on direct PDE application or learn-to-solve strategies that do not explicitly separate stable and unstable components. In contrast, NUTS reformulates the transport equation itself: it applies homogenization to eliminate mesoscale instability at the PDE level and delegates high-frequency recovery to a separate spatial refinement module. This scale-aware decomposition is essential for robustness in noisy flow regimes.

3.3 Theoretical Analysis: Effectiveness of NUTS under Eddy Perturbations

We adopt a standard multiscale formulation for ocean velocity Pavliotis and Stuart [2008], modeling $\mathbf{w}^*(\mathbf{x}, t) = \bar{\mathbf{w}}^*(\mathbf{x}, t) + \frac{1}{\varepsilon} \mathbf{v}^*(\mathbf{x}, t; \mathbf{y}, \tau)$, where $\varepsilon \ll 1$ characterizes the scale separation between slow large-scale transport and fast mesoscale variability, and $\mathbf{y} = \mathbf{x}/\varepsilon$, $\tau = t/\varepsilon^2$ are fast space-time variables that resolve high-frequency eddy dynamics. The mean flow $\bar{\mathbf{w}}^*$ governs large-scale advection, while \mathbf{v}^* captures mesoscale eddies with rapid, oscillatory fluctuations. This parabolic scaling is standard in homogenization theory for advection–diffusion systems Pardoux and Veretennikov [2005], ensuring that mesoscale variability mixes locally without inducing net large-scale transport. We further assume that both \mathbf{v}^* and the perturbation δ satisfy the same structural form: periodic and mean-zero in the fast variables (\mathbf{y}, τ) . This assumption is classical in homogenization theory Jikov et al. [2012] and reflects the physical behavior of mesoscale eddies—highly energetic but oscillatory and net-zero under space-time averaging.

Theorem 1 (Informal; Accuracy and Robustness under Eddy Perturbations). *Suppose that both the true velocity field and the perturbation satisfy the periodic, mean-zero eddy flow assumption. Then, under mild regularity conditions, the NUTS prediction $\hat{\varphi}$ differs from the true solution φ^* by at most $\mathcal{O}(\varepsilon)$, independent of the perturbation strength $\|\delta\|_\infty$.*

Remark: The formal statement and proof of this result are provided in Appendix C.

Interpretation. This theorem establishes two key properties of NUTS. First, the error is $\mathcal{O}(\varepsilon)$ and independent of the perturbation strength $\|\delta\|_\infty$, ensuring robustness: fast, high-amplitude eddy perturbations have negligible impact on the coarse-scale reconstruction. Second, the result guarantees accuracy when the true eddy field varies on small spatial and temporal scales ($\varepsilon \ll 1$). This is nontrivial, as the eddy field enters the dynamics with magnitude $1/\varepsilon$; despite being mean-zero, its local influence is large. The bound confirms that the homogenized model captures the correct large-scale behavior, justifying the use of coarse dynamics in this regime.

4 Experiment

In this section, we answer the following research questions:

RQ1. How does NUTS perform in reconstructing global surface ocean nutrient concentrations compared to existing baselines, using both simulated and real-world observations?

RQ2. Does the proposed two-scale modeling framework enhance robustness to mesoscale perturbations? How do filtering strategies and diffusion implementations influence this robustness?

RQ3. How do individual design choices—such as model architecture, auxiliary inputs, conservation loss, and reconstruction interval—affect reconstruction accuracy?

4.1 Experimental Setup

We present the experimental setup, including datasets, baselines and evaluation metrics. Implementation and training details are in Appendix D. Code and data are available at URL.

Data. We conduct experiments using two datasets for global surface nutrient reconstruction. **Simulation Dataset.** To support high-quality long-term reconstruction, we release two data products generated by the numerical physical-biogeochemical model MOM6-COBALT2 Liu et al. [2022], referred to as MOM6 (Daily) and MOM6 (Monthly). The simulations were conducted on 1000 CPU cores of AMD EPYC 9654 96-Core Processors over an 11-day period, spanning 1959 to 2022 at a global nominal resolution of 0.5° (576×720). The model output is subsequently regressed to a uniform 0.5° grid (360×720) using bilinear interpolation. Each data product includes surface nitrate and phosphate concentrations, along with auxiliary variables such as temperature, salinity, and horizontal velocities (u, v). Compared to GOBH (Monthly) Perruche [2018], our MOM6 (Monthly) data product show improved agreement with in-situ observations from WOD, achieving approximately 60% lower NRMSE on a $0.5^\circ \times 0.5^\circ$ grid (Table 1). Additional details are provided in Appendix D.1. **Real Observations.** We use in-situ nutrient measurements from the World Ocean Database (WOD) Mishonov et al. [2024], which contains nitrate and phosphate records from 1959 to 2022. These observations are extremely sparse, covering only **0.16%** of the full spatio-temporal grid. All measurements are regressed to match the spatial and temporal resolution of the MOM6 data product.

Tasks. We evaluate the model on two resolution-specific nutrient reconstruction tasks. **Daily Average Reconstruction.** Sparse observations are simulated by randomly sampling nutrient values from the MOM6 (Daily) dataset at sparsity levels of 0.1%, 1%, and 10%. The

0.1% level reflects the sparsity of real-world observations, while 10% aligns with settings used in prior work Luo et al. [2024]. The model reconstructs full daily nitrate and phosphate fields using these samples together with MOM6 daily flow and auxiliary variables. **Monthly Average Reconstruction.** Real-world nutrient measurements from WOD and monthly flow and auxiliary variables from MOM6 are used to reconstruct complete monthly averages of nutrient fields. Dataset partitions are summarized in Table 2, with sampling details in Appendix D.4.

Table 1: NRMSEs (\downarrow) of MOM6 (Monthly) and GOBH (Monthly) data compared to real observations from WOD.

Data Source	Nitrate	Phosphate
MOM6	0.463	0.301
GOBH	1.444	1.335

Table 2: Overview of dataset divisions by year.

Task	Train	Validation	Test
Daily Avg.	2019, 2020	2021	2022
Monthly Avg.	1959–1998	1999–2010	2011–2022

Table 3: NRMSE (↓) of different models for reconstructing (1) global daily average nutrient concentrations from the MOM6 simulation under sampling ratios of 0.1%, 1%, and 10%, and (2) global monthly average concentrations from WOD observations. *Params* denotes the number of model parameters. The numbers after \pm are standard errors under 3 trials.

Methods	Params	MOM6 (Daily)						WOD (Monthly)	
		Phosphate			Nitrate			Phosphate	Nitrate
		0.1%	1%	10%	0.1%	1%	10%	–	–
Kriging(Exp.)	–	0.535 \pm 0.022	0.262 \pm 0.015	0.184 \pm 0.023	0.642 \pm 0.020	0.368 \pm 0.025	0.256 \pm 0.019	1.275 \pm 0.130	1.495 \pm 0.091
Kriging(Sph.)	–	0.537 \pm 0.019	0.276 \pm 0.022	0.192 \pm 0.020	0.649 \pm 0.017	0.399 \pm 0.018	0.272 \pm 0.021	1.270 \pm 0.086	1.517 \pm 0.057
4D-VarNet	0.3M	0.151 \pm 0.008	0.154 \pm 0.012	0.156 \pm 0.010	0.168 \pm 0.006	0.170 \pm 0.007	0.161 \pm 0.008	0.187 \pm 0.008	0.203 \pm 0.009
Marble	0.6M	0.397 \pm 0.051	0.227 \pm 0.044	0.232 \pm 0.069	0.441 \pm 0.078	0.222 \pm 0.044	0.297 \pm 0.047	0.363 \pm 0.058	0.326 \pm 0.056
FNO	4.8M	0.251 \pm 0.015	0.227 \pm 0.016	0.229 \pm 0.014	0.261 \pm 0.012	0.256 \pm 0.013	0.257 \pm 0.014	0.244 \pm 0.015	0.276 \pm 0.017
U-Net	31.0M	0.151 \pm 0.008	0.148 \pm 0.013	0.149 \pm 0.011	0.169 \pm 0.007	0.166 \pm 0.012	0.167 \pm 0.013	0.174 \pm 0.012	0.187 \pm 0.008
ViT	77.7M	0.257 \pm 0.032	0.242 \pm 0.044	0.359 \pm 0.048	0.311 \pm 0.046	0.256 \pm 0.044	0.256 \pm 0.052	0.263 \pm 0.034	0.260 \pm 0.002
AtmoRep	0.7B	0.196 \pm 0.010	0.194 \pm 0.011	0.192 \pm 0.010	0.190 \pm 0.009	0.219 \pm 0.011	0.218 \pm 0.013	0.206 \pm 0.013	0.260 \pm 0.013
Prithvi	2.3B	0.216 \pm 0.055	0.197 \pm 0.043	0.208 \pm 0.054	0.279 \pm 0.049	0.274 \pm 0.057	0.275 \pm 0.036	0.222 \pm 0.042	0.338 \pm 0.046
NUTS	125.6M	0.014 \pm 0.002	0.015 \pm 0.001	0.022 \pm 0.002	0.143 \pm 0.003	0.136 \pm 0.003	0.142 \pm 0.004	0.035 \pm 0.002	0.151 \pm 0.003
Promotion	–	90.7%	89.9%	85.2%	14.9%	18.1%	11.8%	79.9%	19.3%

Baselines. We compare our model against a wide range of baselines grouped into six categories: (1) Kriging interpolation with exponential and spherical variogram models; (2) CNN-based model U-Net Ronneberger et al. [2015]; (3) transformer-based model ViT Dosovitskiy et al. [2020]; (4) neural operator Fourier Neural Operator (FNO) Li et al. [2020]; (5) implicit representation method Marble model Wang et al. [2025]; (6) foundation models pretrained on climate data, including Prithvi WxC Schmude et al. [2024] and AtmoRep Lessig et al. [2023]; (7) physics-guided hybrid assimilation model such as 4DVarNet Beauchamp et al. [2023]. All baselines except Marble and Kriging reconstruct each frame independently using static inputs—observations, auxiliary variables, and velocity fields at a single time step. Marble leverages temporal observations but excludes auxiliary variables and flow inputs. Kriging uses only static observations. In contrast, our model takes temporal sequences of all inputs and generates spatiotemporal nutrient reconstructions. See Appendix B.2 and D.5 for details.

Metrics. We use Normalized Root Mean Squared Error (NRMSE) to evaluate model performance, which ensures scale independence Shcherbakov et al. [2013]. We first calculate the latitude-weighted RMSE between the reconstructed values and the corresponding ground-truth, while NRMSE is obtained by normalizing RMSE using the mean of the ground-truth.

4.2 Main Results (RQ1)

We compare the reconstruction performance of our model on simulation and observation data as summarized in Table 3 and Figure 4.

Obs 1: NUTS achieves the lowest NRMSE across all daily and monthly reconstruction tasks. We evaluate performance under varying observation sparsity across simulated and real-world datasets. As shown in Table 3, NUTS consistently outperforms all baselines. On the daily task with 0.1% sparsity, it reduces NRMSE by 90.7% for phosphate and 14.9% for nitrate compared to U-Net. On the monthly WOD dataset, it achieves 79.9% and 19.3% improvement, respectively. The gain is more substantial for phosphate, which exhibits smoother temporal variation and is easier to model dynamically. Figure 3 supports this, showing the spatial distribution of the phosphate-to-nitrate ratio of coefficients of variation (CVs), where each CV is defined as the temporal standard deviation divided by the mean concentration. Lower values indicate weaker phosphate fluctuations, which NUTS captures more reliably.

Among baselines, U-Net and 4D-VarNet perform best. U-Net extracts multiscale features via skip-connected encoders Ronneberger et al. [2015], while 4D-VarNet enforces physical consistency through advection-aware design Beauchamp et al. [2023]. NUTS combines both principles—multiscale modeling and physics-based dynamics—yielding consistent improvements across sparsity levels. These gains are especially pronounced under low observation density, where auxiliary physical variables become essential for accurate reconstruction. Baselines that lack such inputs—such as Kriging (Exp.) and (Sph.)—exhibit large accuracy drops. In contrast, NUTS remains robust by

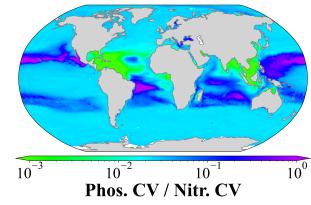


Figure 3: Spatial distribution of the phosphate-to-nitrate ratio of coefficients of variation.

Table 4: **Model Analysis (NRMSE ↓).** (a) Comparison of different low-pass filter types; (b) Evaluation of cutoff ratios for frequency filtering; (c) Comparison of advection–diffusion implementations, including advection-only, fixed diffusion matrix, and learned diffusion network. Unless otherwise specified, the target nutrient is nitrate, the low-pass filter is Fourier-based with a cutoff ratio of 0.1, and the diffusion module is implemented using a 6-layer ResNet.

(a) Filter Type.			(b) Filter Cutoff Ratio.			(c) Implementation of Advection Diffusion.		
filter	Daily	Monthly	param.	Daily	Monthly	case	Daily	Monthly
Fourier	0.136	0.151	0.1	0.136	0.151	advection-only	0.138	0.176
Wavelet	0.144	0.197	0.2	0.145	0.194	adv. + diffusion matrix	0.142	0.165
Gaussian	0.143	0.169	0.5	0.145	0.189	adv. + diffusion network	0.136	0.151
Moving Avg.	0.154	0.167	1.0	0.171	0.189			

leveraging oceanographic drivers like sea surface temperature, as further confirmed in our ablation study in Section 5.

Obs 2: In reconstructing real observation site records, our model outperforms data-driven baselines and matches the performance of traditional numerical methods. We evaluate the site-wise reconstruction accuracy by training on 75% of WOD sites and testing on the remaining 25%. As shown in Figure 4, NUTS achieves site-wise NRMSEs of 1.32 for phosphate and 2.18 for nitrate, outperforming all data-driven baselines. Its performance is comparable to traditional numerical models, including MOM6 and GOBH, demonstrating strong generalization under real-world sparsity.

4.3 Component Analysis: Contribution of the Coarse and Refinement Modules (RQ2)

We evaluate the contribution of two key coarse-stage components—low-pass filtering and effective diffusion—as well as the refinement module. NUTS is compared against U-Net and three ablated variants, each omitting a specific component while keeping all other settings fixed. The structural details of these variants are summarized in Table 5, and all variants are parameter-matched with NUTS for a fair comparison. All ablation results reported in this section use nitrate as the reconstruction target. Results for the daily task are reported under a 1% sparsity ratio. Full hyperparameter configurations are provided in Appendix D.5.

Obs 3: Our model achieves both robustness and accuracy; filtering alone improves stability but sacrifices mesoscale information. We assess robustness by perturbing the eddy component \mathbf{v}^* using Fourier-based scaling, generating $\delta = \gamma \mathbf{v}^*$, and injecting it into the velocity field. As shown in Figure 5, NUTS maintains low NRMSE across all perturbation levels, demonstrating strong resilience to mesoscale variability. *Naive-B*, which directly propagates the full velocity field without filtering, suffers large errors—especially at $\gamma = \pm 1$ —highlighting its sensitivity to unresolved eddy perturbations. *Naive-F* improves robustness by suppressing high-frequency noise but exhibits degraded accuracy due to the removal of informative mesoscale signals. In contrast, NUTS combines the strengths of both: the coarse stage stabilizes dynamics through filtering, while the refinement stage recovers fine-scale nutrient structure conditioned on residual eddy flow.

Obs 4: Effective diffusion enhances filtered transport, but refinement is essential for recovering mesoscale structure. As shown in Figure 5, *Naive-(F+D)*—which combines flow filtering with the effective diffusion module—achieves lower RMSE than *Naive-F* and remains robust under mesoscale perturbations. This validates the use of homogenized advection–diffusion dynamics to stabilize transport and retain partial mesoscale effects. However, despite comparable architecture and parameter count, *Naive-(F+D)* still underperforms our full model, highlighting the importance of the refinement module.

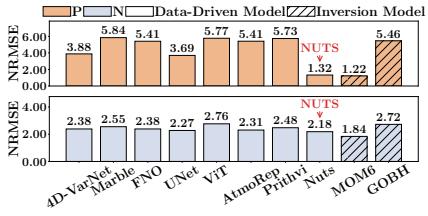


Figure 4: Site-wise NRMSE (↓) of different methods evaluated on WOD real observation.

Table 5: Overview of Model Ablation Variants. “B”, “F” and “F+D” represent the base model, the base model with filtering, and the base model with both filtering and diffusion, respectively. ✓ denotes inclusion; ✗ denotes exclusion.

Variants	Params Count	Low-pass Filter	Effective Diffusion	Refine Module
Naive-B	131.9M	✗	✗	✗
Naive-F	131.9M	✓	✗	✗
Naive-(F+D)	131.7M	✓	✓	✗
NUTS	125.6M	✓	✓	✓

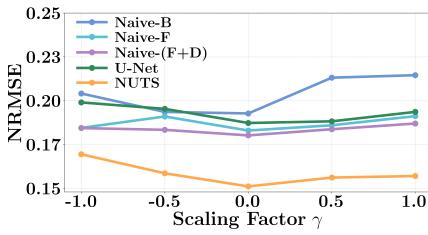


Figure 5: NRMSE of different models under varying mesoscale perturbation levels.

Table 6: **Ablation Study (NRMSE \downarrow).** (a) Comparison of coarse-stage initializers, including static and dynamic architectures; (b) Analysis of model depth in the coarse module; (c) Analysis of model depth in the refinement module; (d) Evaluation of source and conservation loss terms; (e) Quantification of the impact of auxiliary input variables; (f) Assessment of sensitivity to temporal interval length. All experiments use the default setting: coarse/refine depth of 12/6, all loss terms and inputs included, and interval length set to 4.

(a) Coarse Model Structure.			(b) Depth of Coarse Module.			(c) Depth of Refine Module.		
case	Daily	Monthly	depth	Daily	Monthly	depth	Daily	Monthly
2D CNN	0.206	0.161	6	0.185	0.160	2	0.142	0.164
ViT	0.159	0.157	8	0.165	0.153	4	0.146	0.210
3D CNN	0.162	0.153	12	0.136	0.151	6	0.136	0.151
NUTS	0.136	0.151	16	0.148	0.177	8	0.180	0.170

(d) Source and Conservation Loss.			(e) Auxiliary Variables.			(f) Interval Length.		
case	Daily	Monthly	removed var.	Daily	Monthly	length	Daily	Monthly
w/ src, w/ cons.	0.136	0.151	temp.	1.059	1.014	1	0.159	0.157
w/ src, w/o cons.	0.153	0.151	salt	0.170	0.168	2	0.166	0.151
w/o src, w/ cons.	0.156	0.155	u	0.164	0.155	4	0.136	0.151
w/o src, w/o cons.	0.155	0.154	v	0.142	0.160	8	0.220	0.158

ing the importance of the refinement module in reconstructing fine-scale nutrient variability lost during filtering.

Obs 5: Filter design in the coarse module is critical; Fourier filtering with strong high-frequency suppression yields the best performance. We ablate the design of the low-pass filter used in the coarse module of NUTS. Among several options, the Fourier filter achieves the lowest NRMSE on both daily (0.136) and monthly (0.151) tasks, outperforming wavelet, Gaussian, and moving average filters (Table 4a). This result is consistent with prior work in ocean modeling and geophysical fluid dynamics Abernathey and Marshall [2013], Callies and Ferrari [2013], where spectral (Fourier-based) filtering is widely adopted to separate large-scale flow from unresolved mesoscale variability. We further vary the cutoff ratio of the Fourier filter, which determines the extent of high-frequency suppression. Lower ratios—removing more unresolved eddy components—consistently improve reconstruction accuracy, while higher ratios degrade performance (Table 4b). These results highlight that principled filtering in the coarse module is essential for stabilizing nutrient transport, while fine-scale variability is later recovered by the refinement stage.

Obs 6: Incorporating a learnable diffusion module improves accuracy; state-dependent designs further enhance performance. We ablate the diffusion design in the advection–diffusion solver of NUTS. We compare three variants: (1) advection-only, (2) with a trainable, time-invariant diffusion matrix $K = UU^\top$, and (3) the state-dependent formulation used in **NUTS**, where $K = GG^\top$ and $G = G(\bar{\varphi})$ is produced by a hyper-network conditioned on the coarse prediction $\bar{\varphi}$. As shown in Table 4c, both diffusion-enhanced variants outperform the advection-only baseline on daily and monthly tasks, confirming the benefit of modeling unresolved subgrid dispersion. The state-dependent design used in NUTS further improves accuracy over the time-invariant variant (0.151 vs. 0.165 on the monthly task), consistent with the theoretical expectation that effective diffusion depends on the tracer state McWilliams [2006], McDougall and McIntosh [2001]. The improvement is more substantial in the monthly setting, where longer temporal scales allow diffusion to play a more dominant role in shaping nutrient transport.

5 Discussions

Ablation Study (RQ3). We evaluate the impact of architectural selection of both modules, source module, loss design, auxiliary variables and temporal interval length on model performance. Additional ablation results on spatial and temporal resolution, as well as the conservation loss weight coefficient, are provided in Appendix E.1. All ablation results in this section use nitrate as the target variable. **• Coarse and Refine Module Structure.** We evaluate architecture and depth for both the coarse and refinement modules. As shown in Table 6a, static 2D CNNs underperform due to the lack of temporal modeling, while dynamic architectures—3D CNN and spatiotemporal ViT—achieve lower errors. NUTS, which uses a spatiotemporal transformer, yields the best NRMSE of 0.151. Depth analysis (Tables 6b, 6c) shows that performance peaks with 12 layers in the coarse module and 6 layers in the refinement module. Shallower models underfit, while deeper ones degrade due to over-smoothing or training instability. These results highlight the importance of both dynamic structure and moderate depth. **•Source and Conservation Loss.** Incorporating the source module

and conservation loss enhances reconstruction accuracy (Table 6d). Additionally, the conservation loss contributes to preserving total nutrient mass, as detailed in Appendix E.2. •**Auxiliary Variables.** Sea surface temperature is the most influential auxiliary input, with its removal causing the largest increase in reconstruction error (Table 6e). This highlights its essential role in guiding nutrient reconstruction and is consistent with prior findings in related work such as 4DVarNet Beauchamp et al. [2023]. •**Reconstruction Interval Length.** Model performance is sensitive to the choice of reconstruction interval length, with both short and long intervals resulting in higher error relative to intermediate settings (Table 6f). In the daily task, a 4-step interval yields the lowest NRMSE (0.136), balancing informative temporal context and noise from redundancy or uncorrelated variability.

Conclusion and Broader Impact. We present NUTS, a two-stage, physics-informed framework for reconstructing global surface ocean nutrients from sparse observations. By combining coarse advection–diffusion dynamics with data-driven refinement, NUTS achieves state-of-the-art performance on both simulated and real-world datasets. While our experiments focus on nitrate and phosphate, the framework is grounded in general transport physics and naturally extends to other passive tracers. Preliminary results (Appendix F) show promising generalization, supporting broader applications in environmental reconstruction, climate monitoring, and Earth system science.

Future Work. Future directions include extending NUTS in three areas: spatial coverage, biogeochemical complexity, and air–sea exchange. A 3D extension will capture vertical transport and subsurface gradients. Adding processes like remineralization and nutrient uptake will improve modeling of regeneration and biological consumption. Air–sea gas exchange will enable reconstruction of gas tracers for carbon and oxygen cycle monitoring.

Acknowledgements

This research is supported by the National Natural Science Foundation of China (No. 62306179), the National Key Research and Development Program of China (2023YFC2808802), Southern Marine Science and Engineering Guangdong Laboratory (Zhuhan) (nos. SML2023SP219), the Ocean Negative Carbon Emissions (ONCE) Program.

References

Eva E Stüeken, Alice Pellerin, Christophe Thomazo, Benjamin W Johnson, Samuel Duncanson, and Shane D Schoepfer. Marine biogeochemical nitrogen cycling through earth’s history. *Nature Reviews Earth & Environment*, 5(10):732–747, 2024.

Alexey V Mishonov, Tim P Boyer, Olga K Baranova, Courtney N Bouchard, Scott L Cross, Hernan E Garcia, Ricardo A Locarnini, Christopher R Paver, James R Reagan, Zhankun Wang, et al. World ocean database 2023. 2024.

MM Locarnini, AV Mishonov, OK Baranova, TP Boyer, MM Zweng, HE Garcia, D Seidov, KW Weathers, CR Paver, I Smolyar, et al. World ocean atlas 2018, volume 1: Temperature. 2018.

Jaideep Pathak, Aditya Subramanian, Peter Harrington, et al. Fourcastnet: A global data-driven high-resolution weather model using adaptive fourier neural operators. *Advances in Neural Information Processing Systems*, 2022.

Kaifeng Bi, Lingxi Xie, Hengheng Zhang, Xin Chen, Xiaotao Gu, and Qi Tian. Accurate medium-range global weather forecasting with 3d neural networks. *Nature*, 619(7970):533–538, 2023.

Remi Lam, Alvaro Sanchez-Gonzalez, Matthew Willson, Peter Wirsberger, Meire Fortunato, Ferran Alet, Suman Ravuri, Timo Ewalds, Zach Eaton-Rosen, Weihua Hu, et al. Learning skillful medium-range global weather forecasting. *Science*, 382(6677):1416–1421, 2023.

Nils Bochow, Anna Poltronieri, Martin Rypdal, and Niklas Boers. Reconstructing historical climate fields with deep learning. *Science Advances*, 11(14):eadp0558, 2025. doi: 10.1126/sciadv.adp0558. URL <https://www.science.org/doi/abs/10.1126/sciadv.adp0558>.

Étienne Plésiat, Robert JH Dunn, Markus G Donat, and Christopher Kadow. Artificial intelligence reveals past climate extremes by reconstructing historical records. *Nature Communications*, 15(1):9191, 2024.

Christopher Kadow, David Matthew Hall, and Uwe Ulbrich. Artificial intelligence reconstructs missing climate information. *Nature Geoscience*, 13(6):408–413, Jun 2020. ISSN 1752-0908. doi: 10.1038/s41561-020-0582-5. URL <https://doi.org/10.1038/s41561-020-0582-5>.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015. URL <https://arxiv.org/abs/1505.04597>.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv:2010.11929*, 2020.

Zhihan Gao, Xingjian Shi, Hao Wang, Yi Zhu, Yuyang Bernie Wang, Mu Li, and Dit-Yan Yeung. Earthformer: Exploring space-time transformers for earth system forecasting. *Advances in Neural Information Processing Systems*, 35:25390–25403, 2022.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. *arXiv:2010.08895*, 2020.

Honghui Wang, Shiji Song, and Gao Huang. Gridmix: Exploring spatial modulation for neural fields in pde modeling. In *The Thirteenth International Conference on Learning Representations*, 2025.

Maxime Beauchamp, Quentin Febvre, Hugo Georghenm, and Ronan Fablet. 4dvarnet-ssh: end-to-end learning of variational interpolation schemes for nadir and wide-swath satellite altimetry. *Geoscientific Model Development*, 16(8):2119–2147, 2023.

Geoffrey K Vallis. *Atmospheric and Oceanic Fluid Dynamics*. Cambridge University Press, 2017.

James C McWilliams. Submesoscale currents in the ocean. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 472(2189):20160117, 2016.

Dudley B Chelton, Michael G Schlax, Roger M Samelson, and R A de Szoeke. The influence of nonlinear mesoscale eddies on near-surface oceanic chlorophyll. *Science*, 334(6054):328–332, 2011.

Coralie Perruche. Product user manual for the global ocean biogeochemistry hindcast global_reanalysis_bio_001_029. version 1. 2018.

D Carroll, D Menemenlis, JF Adkins, KW Bowman, H Brix, S Dutkiewicz, I Fenty, MM Gierach, C Hill, O Jahn, et al. The ecco-darwin data-assimilative global ocean biogeochemistry model: Estimates of seasonal to multidecadal surface ocean pco₂ and air-sea co₂ flux. *Journal of Advances in Modeling Earth Systems*, 12(10):e2019MS001888, 2020.

Stephen M Griffies et al. Elements of the modular ocean model (mom). *GFDL Ocean Group Tech. Rep*, 7(620):47, 2012.

Margarita E Conkright, Ricardo A Locarnini, Hernan E Garcia, Todd D O’Brien, Timothy P Boyer, C Stephens, and John I Antonov. World ocean atlas 2001: Objective analyses, data statistics, and figures: Cd-rom documentation. 2002.

PHILIPPE Courtier, J-N Thépaut, and Anthony Hollingsworth. A strategy for operational implementation of 4d-var, using an incremental approach. *Quarterly Journal of the Royal Meteorological Society*, 120(519):1367–1387, 1994.

Lars Nerger and Watson W Gregg. Improving assimilation of seawifs data by the application of bias correction with a local seik filter. *Journal of marine systems*, 73(1-2):87–102, 2008.

Pierre P Brasseur and Jacques A Haus. Application of a 3-d variational inverse model to the analysis of ecohydrodynamic data in the northern bering and southern chukchi seas. *Journal of Marine Systems*, 1(4):383–401, 1991.

Xihai Luo, Wei Xu, Yihui Ren, Shinjae Yoo, and Balu Nadiga. Continuous field reconstruction from sparse observations with implicit neural networks. *arXiv:2401.11611*, 2024.

Johannes Schmude, Sujit Roy, Will Trojak, Johannes Jakubik, Daniel Salles Civitarese, Shraddha Singh, Julian Kuehnert, Kumar Ankur, Aman Gupta, Christopher E Phillips, et al. Prithvi wxc: Foundation model for weather and climate. *arXiv:2409.13598*, 2024.

Christian Lessig, Ilaria Luise, Bing Gong, Michael Langguth, Scarlet Stadtler, and Martin Schultz. Atmorep: A stochastic model of atmosphere dynamics using large scale representation learning. *arXiv:2308.13280*, 2023.

Shengyu Zhao, Jonathan Cui, Yilun Sheng, Yue Dong, Xiao Liang, Eric I Chang, and Yan Xu. Large scale image completion via co-modulated generative adversarial networks. *arXiv preprint arXiv:2103.10428*, 2021.

Andreas Lugmayr, Martin Danelljan, Andres Romero, Fisher Yu, Radu Timofte, and Luc Van Gool. Repaint: Inpainting using denoising diffusion probabilistic models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11461–11471, 2022.

William E Schiesser. *The numerical method of lines: integration of partial differential equations*. Elsevier, 2012.

Randall J. LeVeque. *Finite Difference Methods for Ordinary and Partial Differential Equations*. Society for Industrial and Applied Mathematics, 2007. doi: 10.1137/1.9780898717839. URL <https://pubs.siam.org/doi/abs/10.1137/1.9780898717839>.

Grigoris Pavliotis and Andrew Stuart. *Multiscale methods: averaging and homogenization*. Springer Science & Business Media, 2008.

E. Pardoux and A. Yu. Veretennikov. On the poisson equation and diffusion approximation 3. *The Annals of Probability*, 33(3), May 2005. ISSN 0091-1798. doi: 10.1214/009117905000000062. URL <http://dx.doi.org/10.1214/009117905000000062>.

Vasili Vasilievitch Jikov, Sergei M Kozlov, and Olga Arsenievna Oleinik. *Homogenization of differential operators and integral functionals*. Springer Science & Business Media, 2012.

Xiao Liu, Charles Stock, John Dunne, Minjin Lee, Elena Shevliakova, Sergey Malyshev, Paul C.D. Milly, and Matthias Büchner. Isimip3a ocean physical and biogeochemical input data [gfdl-mom6-cobalt2 dataset], 2022. URL <https://doi.org/10.48364/ISIMIP.920945>.

Maxim Vladimirovich Shcherbakov, Adriaan Brebels, Nataliya Lvovna Shcherbakova, Anton Pavlovich Tyukov, Timur Alexandrovich Janovsky, Valeriy Anatol'evich Kamaev, et al. A survey of forecast error measures. *World applied sciences journal*, 24(24):171–176, 2013.

Ryan Abernathey and John Marshall. Global surface eddy diffusivities derived from satellite altimetry. *Journal of Geophysical Research: Oceans*, 118(2):901–916, 2013. doi: 10.1002/jgrc.20066.

Jörn Callies and Raffaele Ferrari. Interpreting energy and tracer spectra of upper-ocean turbulence in the submesoscale range (1–200 km). *Journal of Physical Oceanography*, 43(11):2456 – 2474, 2013. doi: 10.1175/JPO-D-13-063.1. URL <https://journals.ametsoc.org/view/journals/phoc/43/11/jpo-d-13-063.1.xml>.

James C. McWilliams. *Fundamentals of Geophysical Fluid Dynamics*. Cambridge University Press, 2006.

Trevor J. McDougall and Peter C. McIntosh. The temporal-residual-mean velocity. part ii: Isopycnal interpretation and the tracer equation. *Journal of Physical Oceanography*, 31(5):1222–1246, 2001. doi: 10.1175/1520-0485(2001)031<1222:TTRMVP>2.0.CO;2.

Alistair Adcroft, Jean-Michel Campin, E Doddridge, S Dutkiewicz, C Evangelinos, D Ferreira, M Follows, G Forget, B Fox-Kemper, P Heimbach, et al. Mitgcm documentation. *Release checkpoint67a-12-gbf23121*, 19, 2018.

Alexander F Shchepetkin and James C McWilliams. The regional oceanic modeling system (roms): a split-explicit, free-surface, topography-following-coordinate oceanic model. *Ocean modelling*, 9(4):347–404, 2005.

Malek Belgacem, Katrin Schroeder, Alexander Barth, Charles Troupin, Bruno Pavoni, and Jacopo Chiggiato. Climatological distribution of dissolved inorganic nutrients in the western mediterranean sea (1981–2017). *Earth System Science Data Discussions*, 2021:1–49, 2021.

Ariane Verdy and Matthew R Mazloff. A data assimilating model for estimating southern ocean biogeochemistry. *Journal of Geophysical Research: Oceans*, 122(9):6968–6988, 2017.

Cecile S Rousseaux and Watson W Gregg. Climate variability and phytoplankton composition in the pacific ocean. *Journal of Geophysical Research: Oceans*, 117(C10), 2012.

Wanqin Zhong, Xin Ma, Tianqi Shi, Ge Han, Haowei Zhang, and Wei Gong. Reconstruction of global ocean surface pco₂ and air-sea co₂ flux: Based on multigrained cascade forest model. *Journal of Geophysical Research: Oceans*, 130(2):e2024JC021483, 2025.

Maziar Raissi, Paris Perdikaris, and George Em Karniadakis. Physics informed deep learning (part i): Data-driven solutions of nonlinear partial differential equations. *arXiv:1711.10561*, 2017.

Eugenio Cutolo, Ananda Pascual, Simon Ruiz, Nikolaos D Zarokanellos, and Ronan Fablet. Cloinet: ocean state reconstructions through remote-sensing, in-situ sparse observations and deep learning. *Frontiers in Marine Science*, 11:1151868, 2024.

Bin Lu, Ze Zhao, Luyu Han, Xiaoying Gan, Yuntao Zhou, Lei Zhou, Luoyi Fu, Xinbing Wang, Chenghu Zhou, and Jing Zhang. Oxygenerator: Reconstructing global ocean deoxygenation over a century with deep learning. *arXiv:2405.07233*, 2024.

Kamyar Nazeri, Eric Ng, Tony Joseph, Faisal Z Qureshi, and Mehran Ebrahimi. Edgeconnect: Generative image inpainting with adversarial edge learning. *arXiv:1901.00212*, 2019.

Wenbo Li, Zhe Lin, Kun Zhou, Lu Qi, Yi Wang, and Jiaya Jia. Mat: Mask-aware transformer for large hole image inpainting. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10758–10768, 2022.

Tobias Pfaff, Meire Fortunato, Alvaro Sanchez-Gonzalez, and Peter Battaglia. Learning mesh-based simulation with graph networks. In *International conference on learning representations*, 2020.

Dan Hendrycks and Kevin Gimpel. Gaussian error linear units (gelus). *arXiv:1606.08415*, 2016.

Alain Bensoussan, Jacques-Louis Lions, and George Papanicolaou. *Asymptotic analysis for periodic structures*. American Mathematical Society, 2011.

Andrew J Majda and Peter R Kramer. Simplified models for turbulent diffusion: Theory, numerical modelling, and physical phenomena. *Physics Reports*, 314(4-5):237–574, 1999.

Gary Froyland, Kathrin Padberg, Matthew H England, and Anne Marie Treguier. Detection of coherent oceanic structures via transfer operators. *Physical review letters*, 98(22):224503, 2007.

Peter B Rhines. Geostrophic turbulence. *Annual Review of Fluid Mechanics*, 11(1):401–441, 1979.

Kevin Sieck and Daniela Jacob. Influence of the boundary forcing on the internal variability of a regional climate model. *American Journal of Climate Change*, 5(3):373–382, 2016.

Grigoris A. Pavliotis. *Homogenization Theory for Advection–Diffusion Equations with Mean Flow*. PhD thesis, Rensselaer Polytechnic Institute, 2002.

Thierry Goudon and Frédéric Poupaud. Homogenization of transport equations: Weak mean field approximation. *SIAM Journal on Mathematical Analysis*, 36(3):856–881, 2005.

Alistair Adcroft, Whit Anderson, V. Balaji, Chris Blanton, Mitchell Bushuk, Carolina O. Dufour, John P. Dunne, Stephen M. Griffies, Robert Hallberg, Matthew J. Harrison, Isaac M. Held, Malte F. Jansen, Jasmin G. John, John P. Krasting, Amy R. Langenhorst, Sonya Legg, Zhi Liang, Colleen McHugh, Aparna Radhakrishnan, Brandon G. Reichl, Tony Rosati, Bonita L. Samuels, Andrew Shao, Ronald Stouffer, Michael Winton, Andrew T. Wittenberg, Baoqiang Xiang, Niki Zadeh, and Rong Zhang. The gfdl global ocean and sea ice model om4.0: Model description and simulation features. *Journal of Advances in Modeling Earth Systems*, 11(10):3167–3211, 2019. doi: 10.1029/2019ms001726. URL <https://doi.org/10.1029/2019ms001726>.

CA Stock, JP Dunne, and JG John. Drivers of trophic amplification of ocean productivity trends in a changing climate. *Biogeosciences*, 11(24):7125–7135, 2014.

Hiroyuki Tsujino, Shogo Urakawa, Hideyuki Nakano, R Justin Small, Who M Kim, Stephen G Yeager, Gokhan Danabasoglu, Tatsuo Suzuki, Jonathan L Bamber, Mats Bentsen, et al. Jra-55 based surface dataset for driving ocean–sea-ice models (jra55-do). *Ocean Modelling*, 130:79–139, 2018.

Fortunat Joos and Renato Spahni. Rates of change in natural and anthropogenic radiative forcing over the past 20,000 years. *Proceedings of the National Academy of Sciences*, 105(5):1425–1430, 2008.

H. E. Garcia, R. A. Locarnini, T. P. Boyer, J. I. Antonov, O. K. Baranova, M. M. Zweng, ..., and J. R. Reagan. *World Ocean Atlas 2013, Volume 3: Dissolved Oxygen, Apparent Oxygen Utilization, and Oxygen Saturation*. Number 75 in NOAA Atlas NESDIS. NOAA, 2013a.

H. E. Garcia, T. P. Boyer, O. K. Baranova, C. Coleman, C. R. Paver, R. A. Locarnini, ..., and J. R. Reagan. *World Ocean Atlas 2013, Volume 4: Dissolved Inorganic Nutrients (phosphate, nitrate, silicate)*. Number 76 in NOAA Atlas NESDIS. NOAA, 2013b.

R. A. Locarnini, A. V. Mishonov, J. I. Antonov, T. P. Boyer, H. E. Garcia, O. K. Baranova, ..., and J. R. Reagan. *World Ocean Atlas 2013, Volume 1: Temperature*. Number 73 in NOAA Atlas NESDIS. NOAA, 2013.

M. M. Zweng, J. R. Reagan, J. I. Antonov, R. A. Locarnini, A. V. Mishonov, T. P. Boyer, ..., and D. Seidov. *World Ocean Atlas 2013, Volume 2: Salinity*. Number 74 in NOAA Atlas NESDIS. NOAA, 2013.

Are Olsen, Robert M Key, Steven Van Heuven, Siv K Lauvset, Anton Velo, Xiaohua Lin, Carsten Schirnick, Alex Kozyr, Toste Tanhua, Mario Hoppema, et al. The global ocean data analysis project version 2 (glodapv2)—an internally consistent data product for the world ocean. *Earth System Science Data*, 8(2):297–323, 2016.

Samar Khatiwala, Francois Primeau, and T Hall. Reconstruction of the history of anthropogenic co2 concentrations in the ocean. *Nature*, 462(7271):346–349, 2009.

I Loshchilov. Decoupled weight decay regularization. *arXiv:1711.05101*, 2017.

Vincent Sitzmann, Julien Martel, Alexander Bergman, David Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. *Advances in neural information processing systems*, 33:7462–7473, 2020.

Gao Huang, Yu Sun, Zhuang Liu, Daniel Sedra, and Kilian Q Weinberger. Deep networks with stochastic depth. In *European conference on computer vision*, pages 646–661. Springer, 2016.

NeurIPS Paper Checklist

1. Claims

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

Answer: [\[Yes\]](#)

Justification: The claims in Abstract and Introduction 1 reflect the contributions of this paper.

Guidelines:

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

2. Limitations

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

Answer: [\[Yes\]](#)

Justification: We discuss the limitations of this paper in Appendix G.

Guidelines:

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

3. Theory assumptions and proofs

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

Answer: [\[Yes\]](#)

Justification: We provide complete proofs for the proposed theorem in Appendix C.

Guidelines:

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

4. Experimental result reproducibility

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

Answer: [Yes]

Justification: We provide implementation details of the model in Appendix B and hyper-parameter configurations in Appendix D.3. We also open-source the code and data required to conduct the experiments in this anonymized URL.

Guidelines:

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

5. Open access to data and code

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

Answer: [\[Yes\]](#)

Justification: We provide code and data in this anonymized URL.

Guidelines:

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

6. Experimental setting/details

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

Answer: [\[Yes\]](#)

Justification: We provide dataset divisions in Table 2, and hyper-parameter configuration details in Appendix D.3.

Guidelines:

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

7. Experiment statistical significance

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

Answer: [\[Yes\]](#)

Justification: We report standard deviations of experimental results in Table 3.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer “Yes” if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).

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

8. Experiments compute resources

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

Answer: [\[Yes\]](#)

Justification: We provide the compute resources used to generate the simulation data in Section 4, and the compute resources used to conduct experiments in Appendix D.3.

Guidelines:

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

9. Code of ethics

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

Answer: [\[Yes\]](#)

Justification: The research conducted in this paper conform with the NeurIPS Code of Ethics in every respect.

Guidelines:

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

10. Broader impacts

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

Answer: [\[Yes\]](#)

Justification: We discuss broader impacts of this paper in Section 5. We also provide some preliminary results in Appendix F.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.

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

11. Safeguards

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

Answer: [NA]

Justification: This paper poses no such risks.

Guidelines:

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

12. Licenses for existing assets

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

Answer: [No]

Justification: All code implementations are cited with license details in Appendix D.5. For the WOD dataset, although we were unable to locate the specific license, we have cited the official source and provided the official link.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.

- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, paperswithcode.com/datasets has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. New assets

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

Answer: [\[Yes\]](#)

Justification: We release two datasets under the CC-BY 4.0 licenses and code implementation under the MIT license. Datasets and code can be found in this anonymized URL.

Guidelines:

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

14. Crowdsourcing and research with human subjects

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

Answer: [\[NA\]](#)

Justification: This paper does not involve any crowdsourcing nor research with human subjects.

Guidelines:

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

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

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

Answer: [\[NA\]](#)

Justification: This paper does not involve any crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.

- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. Declaration of LLM usage

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

Answer: [NA]

Justification: The core methodology of this paper does not involve LLMs as any important, original, nor non-standard components.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (<https://neurips.cc/Conferences/2025/LLM>) for what should or should not be described.