LEARNING FIELD RECONSTRUCTION FROM INCOMPLETE DATA BY GLOBALLY CORRECTING LOCAL ESTIMATES

Anonymous authors
Paper under double-blind review

ABSTRACT

Reconstructing physical fields from training samples that are always incomplete requires learning spatial structure from fragmented observations. Existing context–query work establishes how held-out observations provide valid training targets, but this does not make the complete-field distribution identifiable when every training field is incomplete. With finite data, weak evidence of sharp transitions and localized variations can further favor averaged predictions that attenuate local detail. A structural prior is therefore needed to favor plausible completions; local spatial relationships offer one grounded in the observations. We propose a locally constructed, globally revisable estimator that explicitly learns local field estimates and subsequently corrects them using full-domain observations. A shared coordinate-conditioned predictor learns from incomplete patches, allowing relatively well-observed neighborhoods to provide direct supervision of local structure. Its overlapping predictions are reconciled into an observation-conditioned consensus field. A full-domain estimator retains the original observations and learns a residual correction around this frozen field estimate, allowing locally constructed structure to be revised by broader evidence. The local estimate serves as both an explicit input, accompanied by its discrepancies with the observations, and a prediction starting point that the global model can revise. On three real-world ocean datasets with authentic observation gaps, our estimator achieves the lowest MSE and highest PSNR on withheld source-supported values, reducing MSE by 28.9%–34.5% against the strongest external baseline.

1 INTRODUCTION

Learning from never-complete fields. Environmental monitoring relies on spatially resolved physical fields, yet satellite observations contain irregular gaps caused by sampling and retrieval constraints (Barth et al., 2024; Ma et al., 2026). Reconstructing the unobserved regions requires inferring their values from measurements elsewhere in the field and spatial relationships learned across samples (Beckers & Rixen, 2003; Alvera-Azcárate et al., 2005). In many practical datasets, however, every recorded field is incomplete: each training example contains only the measurements on its observed support, with no corresponding complete field available as a target (Zhou et al., 2026). The learning task is therefore to acquire a reconstruction rule from a collection of incomplete fields and apply it to the missing regions of a new, partially observed field.

Limited supervision can obscure local detail. Missingness limits both the evidence available for each prediction and the supervision available to learn that prediction. A model cannot be trained directly against the unknown values in a sample's missing regions; it must instead learn from relationships among the values that are observed together. Context–query methods such as IDCD and OAMP create such learning examples by withholding some available measurements as targets and using the remaining observations as context (Zhou et al., 2026; Ma et al., 2026). This provides a training signal, but its coverage of spatial relationships remains limited by the original observations. The full-domain direct estimator considered here relies on these examples to learn local structure and the use of distant observations jointly. With extensive missingness and finite data, sharp transitions and localized variations may be sparsely represented in the context–target pairs. Where

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

the available context leaves several configurations plausible, squared-error prediction favors their conditional average, potentially attenuating local detail (Lehtinen et al., 2018; Barth et al., 2024). These difficulties motivate giving local structure a dedicated learning role within the reconstruction process.

Local learning provides structure for global revision. Local views offer a useful starting point: even a heavily incomplete field can contain patches with relatively dense observations, providing more direct supervision of spatial variation within those neighborhoods. A shared patch predictor can pool these local learning examples across fields, while coordinate conditioning accommodates variation in their spatial statistics (Wang et al., 2023; Hu et al., 2024). Our hypothesis is that organizing the available supervision around smaller prediction problems helps finite models learn local relationships explicitly. Yet local estimates can remain ambiguous, disagree across overlapping views, or miss information outside their support. This suggests constructing a field estimate from local predictions and then learning its revision from full-domain observations. The intended division of labor is to establish locally supported detail first, reconcile overlapping predictions, and use broader evidence to correct their errors and inconsistencies.

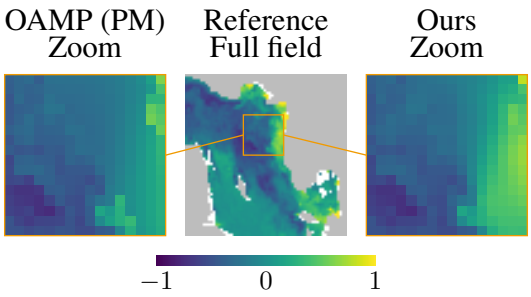

Figure 1: Fine-scale detail in full-mask Baltic Sea NANO reconstruction at 64×64. OAMP (PM) is the parameter-matched OAMP control, denoted Full Direct. The yellow box marks the held-out region enlarged in the side panels. Gray denotes outside-domain locations; white in Reference denotes source-missing values.

A consensus prior anchors the full-domain estimator. The proposed estimator first trains a coordinate-conditioned patch predictor and freezes it to construct an observation-conditioned local consensus prior—an intermediate field estimate for the current sample. Overlapping predictions are reconciled by least squares, building on position-aware patch modeling (Wang et al., 2023; Hu et al., 2024) and view reconciliation (Bar-Tal et al., 2023). A full-domain model retains the original observation context and learns a residual correction around this estimate under the same context–query supervision. Compared with OAMP (PM), our reconstruction preserves more distinct fine-scale variations in the enlarged Baltic Sea NANO region (Figure 1).

Component controls test the role of the prior. The reference field and its discrepancies reorganize the retained observations into an explicit representation for full-domain prediction. Our hypothesis concerns how finite models learn with this local-to-global organization; estimator comparisons and component controls test the contributions of prior conditioning and residual correction.

Results support revisable local estimates. Experiments on three real-world ocean products yield the lowest MSE and highest PSNR among the compared methods at 64×64, with metrics computed on withheld source-supported values. On the ocean products, prior-conditioned direct prediction reduces MSE by 8.3%–9.0% relative to parameter-matched no-prior direct prediction, and residual correction yields a further 9.8%–16.2% reduction. These comparisons show that a local estimate can be valuable as a revisable intermediate representation even when it is a weaker standalone predictor.

Contributions. We make three contributions:

- We introduce a locally constructed, globally revisable estimator that explicitly learns local field structure from incomplete patches and uses full-domain observations to revise the resulting consensus estimate.
- We construct an explicit reference field by overlap consensus and couple it with prior–observation discrepancy features and a residual output anchored at the frozen reference, using only context–query supervision.

- We demonstrate reconstruction gains on ocean fields with authentic observation gaps and show, through estimator comparisons and targeted controls, the value of local estimates as revisable intermediate representations.

2 PRELIMINARIES

Appendix A reviews related work on incomplete-data learning, observation-aligned reconstruction, and local–global estimation (Sections A.1–A.3). We now define the observation and supervision interfaces used throughout the paper.

2.1 INCOMPLETE OBSERVATIONS

Following the notation of IDCD (Zhou et al., 2026), let $\boldsymbol{x}_0$ denote a complete physical field and $\boldsymbol{M}$ a binary mask indicating its observed locations. The available data are $\boldsymbol{x}_{\mathrm{obs}} = \boldsymbol{M} \odot \boldsymbol{x}_0$, where $\odot$ denotes elementwise multiplication. Each field has d_x channels on a spatial grid Ω; masks apply to all channels, and a validity mask $\boldsymbol{V}$ excludes locations outside the physical domain, such as land. Training uses only incomplete pairs $(\boldsymbol{x}_{\mathrm{obs}}, \boldsymbol{M})$ with $\boldsymbol{M} \leq \boldsymbol{V}$ and $\boldsymbol{M} \neq \boldsymbol{V}$. Each sample's available support stays fixed across training visits. Our goal is point reconstruction of the valid missing values from these incomplete training pairs.

2.2 OBSERVATION-ALIGNED SUPERVISION

Context–query supervision. Originally missing values cannot serve as supervision, so we temporarily withhold some available measurements as targets. Following IDCD and OAMP (Zhou et al., 2026; Ma et al., 2026), we divide the available observations into a context mask $\boldsymbol{M}_{\mathrm{ctx}} \leq \boldsymbol{M}$ and a nonempty query mask $\boldsymbol{M}_{\mathrm{qry}} = \boldsymbol{M} \odot (1 - \boldsymbol{M}_{\mathrm{ctx}})$. Context values are supplied to the model, while query values are withheld as training targets. The internal partition varies across visits; the original support $\boldsymbol{M}$ remains fixed.

Training context–query construction. The training partition respects the acquisition structure: individual sites for pixel observations and whole blocks for block observations (Zhou et al., 2026). For Black Sea and Baltic, dataset-specific frozen mask-only Bayesian Flow Networks (BFNs; Graves et al., 2025) generate binary proposals using OAMP's observation-aligned guidance (Ma et al., 2026); Global Ocean SSS instead uses an empirical authentic-track mask bank. Before feasibility handling, each proposal $\widetilde{\boldsymbol{M}}$ is intersected with the available support:

$$\boldsymbol{M}_{\mathrm{ctx}} = \boldsymbol{M} \odot \widetilde{\boldsymbol{M}}, \qquad \boldsymbol{M}_{\mathrm{qry}} = \boldsymbol{M} \odot (1 - \widetilde{\boldsymbol{M}}). \tag{1}$$

The constructor selects observations without supplying field values or pseudo-labels. Appendices C.1–C.3 specify the partition requirements, dataset-specific constructors, and feasibility handling.

Denoising query prediction. At diffusion time t, we perturb the observations as

$$\boldsymbol{x}_{\mathrm{obs},t} = \boldsymbol{M} \odot (\alpha_t \boldsymbol{x}_{\mathrm{obs}} + \sigma_t \boldsymbol{\epsilon}), \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(0, I). \tag{2}$$

A clean-field predictor $\boldsymbol{x}_{\boldsymbol{\theta}}$ receives the context values $\boldsymbol{M}_{\mathrm{ctx}} \odot \boldsymbol{x}_{\mathrm{obs},t}$ and their mask, and is trained by

$$\mathcal{L}_{\mathrm{qry}}(\boldsymbol{\theta}) = \mathbb{E}\left[\left\| \boldsymbol{M}_{\mathrm{qry}} \odot \left(\boldsymbol{x}_{\boldsymbol{\theta}}(t, \boldsymbol{M}_{\mathrm{ctx}} \odot \boldsymbol{x}_{\mathrm{obs},t}, \boldsymbol{M}_{\mathrm{ctx}}) - \boldsymbol{x}_{\mathrm{obs}} \right) \right\|_2^2 \right]. \tag{3}$$

Here $\boldsymbol{\theta}$ denotes the parameters of the chosen predictor; fixed geometry is suppressed in this generic signature. The query mask selects the loss terms without entering the model; predictions are constrained on states with positive query exposure. Both stages use this supervision, restricted to patches for local training and applied over the full domain for global training. Appendices B.1 and D.2 specify the diffusion parameterization and query-conditioned target; Section 3.3 describes reconstruction from the supplied observations.

3 METHOD

Never-complete observations need not identify an unrestricted complete-field distribution (Theorem D.1; Appendix D.1). We focus on point reconstruction with finite data and model capacity,

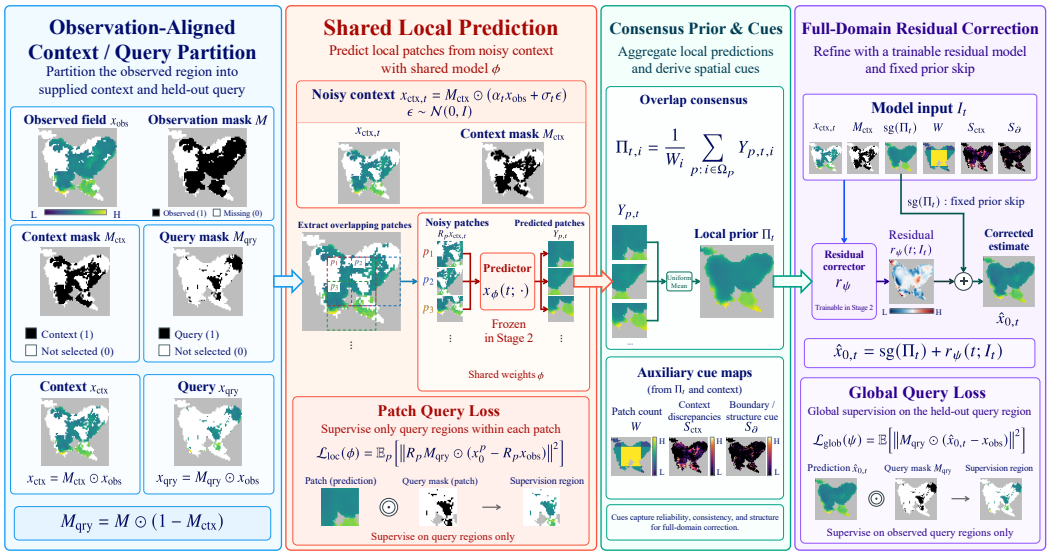

Figure 2: Local prior construction and full-domain correction for one context. Overlapping patch predictions form a consensus prior that conditions the corrector and is added to its learned residual. The patch predictor is frozen during global training; query values enter only the training losses.

learning recurring local relationships through shared patch-level tasks and revising their estimates using full-domain observations.

Both stages use the observation-aligned supervision from Section 2. Given noisy context $\boldsymbol{x}_{\mathrm{ctx},t} = \boldsymbol{M}_{\mathrm{ctx}} \odot \boldsymbol{x}_{\mathrm{obs},t}$, we express reconstruction as a local field prior $\boldsymbol{\Pi}_t$ plus a learned full-domain revision:

$$\widehat{\boldsymbol{x}}_{0,t} = \mathrm{sg}(\boldsymbol{\Pi}_t) + r_{\boldsymbol{\psi}}(t; I_t), \tag{4}$$

Here I_t contains the context, prior, and discrepancy features, while sg denotes stop-gradient through the frozen local predictor. The prior both conditions the corrector and anchors its output (Figure 2).

3.1 Learning and Constructing a Local Prior

The first stage learns to infer withheld values from the measurements within a patch, then assembles these local estimates into a field. Cropping reorganizes the available supervision into smaller prediction tasks without adding independent observations.

Coordinate-conditioned local learning. We share one predictor across patches so that local relationships learned from one neighborhood can help reconstruct another. Absolute coordinates tell the predictor where each patch lies, allowing its predictions to reflect spatially varying field statistics. Let R_p crop patch p with support Ω_p from a fixed covering $\mathcal{P}$, and let $\boldsymbol{\Gamma}$ contain normalized absolute coordinates. A shared predictor receives the patch's noisy context, mask, and coordinate crop:

$$\widehat{\boldsymbol{x}}_0^p = \boldsymbol{x}_{\boldsymbol{\phi}}(t, R_p \boldsymbol{x}_{\mathrm{ctx},t}, R_p \boldsymbol{M}_{\mathrm{ctx}}, R_p \boldsymbol{\Gamma}) \,. \tag{5}$$

This follows coordinate-aware patch modeling (Liu et al., 2018; Wang et al., 2023; Hu et al., 2024). Within each patch, we ask the predictor to recover the withheld measurements from its supplied context, using the shared query objective restricted to that patch:

$$\mathcal{L}_{\mathrm{loc}}(\boldsymbol{\phi}) = \mathbb{E}\left[\left\| R_p \boldsymbol{M}_{\mathrm{qry}} \odot \left(\widehat{\boldsymbol{x}}_0^p - R_p \boldsymbol{x}_{\mathrm{obs}} \right) \right\|_2^2 \right] . \tag{6}$$

For each training item, a patch is sampled uniformly from those with nonempty query support, and only its withheld observed values contribute to the loss. Each patch prediction uses its own context and coordinates, with no exchange of current-sample information between patch evaluations. Together, parameter sharing and coordinate conditioning let the model pool local examples without treating all regions as statistically interchangeable. This gives local variation a dedicated learning task; the full-domain stage can subsequently revise its estimates using evidence outside the patch.

From patch predictions to a reference field. After local training, we freeze the trained predictor $\boldsymbol{x}_{\phi}$ and evaluate it on every patch in the covering, including patches with no query values. The parameters ϕ remain fixed throughout global training and reconstruction. The resulting predictions $\boldsymbol{Y}_{p,t}$ may disagree at overlapping locations because they use different neighborhoods. We reconcile them by averaging all predictions that cover each valid location:

$$\boldsymbol{\Pi}_{t,i} = \frac{1}{\boldsymbol{W}_i} \sum_{p:i\in\Omega_p} \boldsymbol{Y}_{p,t,i}, \qquad \boldsymbol{W}_i = \sum_{p\in\mathcal{P}} \mathbf{1}[i \in \Omega_p], \qquad \boldsymbol{V}_i = 1. \tag{7}$$

Every valid location is covered, and $\boldsymbol{\Pi}_t$ is set to zero outside the valid domain. This uniform consensus is the unique solution of an equal-weight least-squares reconciliation problem (Appendix B.2), following the use of overlap reconciliation in regional prediction (Bar-Tal et al., 2023). The resulting field has the same spatial domain, channels, and units as the target, so subsequent correction can operate directly in field space. We call $\boldsymbol{\Pi}_t$ the *local field prior*: an observation-conditioned estimate that provides a reference for reconstruction, distinct from a probability distribution or the mask prior used to select context and query.

Consensus produces a full-sized field, but each value still uses only the neighborhoods that contributed to it. The next stage lets observations beyond those neighborhoods revise the estimate; Appendix D.3 formalizes this local-support boundary.

3.2 Global Residual Correction

The local prior is a starting estimate with two reasons to need revision: distant observations may favor a different completion, and the local predictor may make errors even where nearby evidence is available. The global stage learns both kinds of revision through Equation (4). Its design connects three elements: the evidence available for correction, an output anchored at the prior, and a supervised target for the required change.

Give the corrector the estimate and the evidence. We retain the full-domain context so that the corrector can use observations outside any contributing patch, and supply the prior so that it can inspect the current estimate. We also expose where this estimate disagrees with the available measurements. The context discrepancy $\boldsymbol{S}_{\mathrm{ctx}}$ measures absolute differences between the prior and noisy values at context locations; the boundary discrepancy $\boldsymbol{S}_{\partial}$ makes the analogous comparison at adjacent non-context locations using the mean of neighboring noisy context values. The overlap count $\boldsymbol{W}$ records how many patch predictions contribute to each location. Together these inputs form

$$I_t = (\boldsymbol{x}_{\mathrm{ctx},t}, \boldsymbol{M}_{\mathrm{ctx}}, \mathrm{sg}(\boldsymbol{\Pi}_t), \boldsymbol{W}, \boldsymbol{S}_{\mathrm{ctx}}, \boldsymbol{S}_{\partial}) \,. \tag{8}$$

These maps make disagreement and construction geometry explicit, helping a finite model learn where and how to revise the field. The discrepancy maps are observable cues rather than clean error targets, and $\boldsymbol{W}$ is not a calibrated uncertainty estimate. All features use the same realized noisy context, without access to query values; Appendix B.3 gives their definitions.

Make the output a revision of the prior. Conditioning on the prior tells the corrector what the local stage predicts; adding its output to that prior makes the learning task explicitly one of revision. We freeze the local predictor and initialize the correction head at zero, so the initial full-field prediction equals $\boldsymbol{\Pi}_t$. Training then learns departures from this context-specific reference, with the correction free to change every predicted value. At each query location, the required change is simply the withheld measurement minus the prior estimate. Applying the shared query loss to Equation (4) therefore gives

$$\mathcal{L}_{\mathrm{glob}}(\boldsymbol{\psi}) = \mathbb{E}\Big[\big\|\boldsymbol{M}_{\mathrm{qry}} \odot \big(r_{\boldsymbol{\psi}}(t; I_t) - \big[\boldsymbol{x}_{\mathrm{obs}} - \mathrm{sg}(\boldsymbol{\Pi}_t)\big]\big)\big\|_2^2\Big] \,. \tag{9}$$

Thus the observed queries directly teach the model how to correct its starting estimate, without requiring complete-field targets. Appendix D.4 formalizes the two roles of this residual under squared loss: incorporating evidence beyond local support and correcting remaining local-estimation error.

Separating the design choices. Section 4.3 compares prior conditioning, prior-anchored residual prediction, and alternative references; Section 4.4 examines local-prior construction and correction cues. These are choices about finite-model learning: the prior and features are deterministic functions of the retained context, so unrestricted direct and residual predictors have the same optimal query risk (Appendix D.4).

3.3 Training and Reconstruction

Two-stage training. We first establish the local reconstruction rule by optimizing $\mathcal{L}_{\mathrm{loc}}$, then freeze it and learn its revision by optimizing $\mathcal{L}_{\mathrm{glob}}$ for the corrector. Both stages use the same stored incomplete observations and observation-aligned partition rule. For each forward pass, the frozen predictor rebuilds the prior from that pass's context. Query values enter only the losses. Algorithm 1 summarizes the procedure.

Reconstruction from supplied observations. Following IDCD's single-step procedure (Zhou et al., 2026), inference uses a fixed near-zero diffusion time $0 < \delta \ll 1$. Given $\boldsymbol{x}_{\mathrm{in}} = \boldsymbol{M}_{\mathrm{in}} \odot \boldsymbol{x}_0$, we draw one perturbation $\boldsymbol{x}_{\mathrm{in},\delta} = \alpha_\delta \boldsymbol{x}_{\mathrm{in}} + \sigma_\delta \boldsymbol{\epsilon}$. Full-mask inference sets $K = 1$ and $\boldsymbol{M}_{\mathrm{ctx}}^{(1)} = \boldsymbol{M}_{\mathrm{in}}$, using all supplied observations rather than a complete reference field. Context-ensemble inference instead uses K subcontexts sampled within $\boldsymbol{M}_{\mathrm{in}}$ by the training partition rule, with the same perturbation shared across members. For each selected context, we construct its prior and discrepancy features and apply the full-domain correction.

Algorithm 1: Local prior learning, global correction, and reconstruction.

Stage 1: learn the local prior
1: Sample context–query partitions and perturb contexts.
2: Train the coordinate-conditioned patch predictor with $\mathcal{L}_{\mathrm{loc}}$; freeze it.
Stage 2: learn global correction
3: For each sampled context, fuse local predictions into $\boldsymbol{\Pi}_t$ and compute I_t.
4: Train the zero-initialized residual corrector with $\mathcal{L}_{\mathrm{glob}}$.
Reconstruction
5: Draw one near-zero-noise perturbation; select the full supplied mask ($K = 1$) or K subcontexts according to the inference mode.
6: **for** each selected context k **do**
7: Construct its prior $\boldsymbol{\Pi}_\delta^{(k)}$ and features; predict $\boldsymbol{\Pi}_\delta^{(k)} + r_{\boldsymbol{\psi}}(\delta; I_\delta^{(k)})$.
8: **end for**
9: Average the corrected fields and restore supplied observations.

Reconstruction combines the resulting predictions and restores the supplied observations:

$$\widetilde{\boldsymbol{x}}_\delta = \frac{1}{K} \sum_{k=1}^{K} \widehat{\boldsymbol{x}}_{0,\delta}^{(k)}, \qquad \widehat{\boldsymbol{x}}_{\mathrm{out}} = \boldsymbol{x}_{\mathrm{in}} + (\boldsymbol{V} - \boldsymbol{M}_{\mathrm{in}}) \odot \widetilde{\boldsymbol{x}}_\delta. \tag{10}$$

Validation selects full-mask inference for ocean data and context ensembling for PDE data (Appendix B.4). Appendix B.5 gives the complete training and reconstruction procedure in Algorithm 2.

4 Experiments

We evaluate reconstruction under authentic ocean gaps, compare matched estimators and alternative references, and isolate prior construction and corrector inputs. We then intervene outside local support to test whether global correction responds to evidence unavailable to the local predictor.

4.1 Baselines

On the ocean products, we compare with DINEOF (Beckers & Rixen, 2003; Alvera-Azcárate et al., 2005), AmbientDiff (Daras et al., 2023), DINDiff (Barth et al., 2024), MissDiff (Ouyang et al., 2023), MSM (Park et al., 2026), and OAMP (Ma et al., 2026). Controlled PDE comparisons use AmbientDiff, DINDiff, MissDiff, MSM, and IDCD (Zhou et al., 2026). Methods share the available training observations, evaluation items, supplied supports, and scoring coordinates, while external baselines retain their method-specific objectives. Full Direct denotes the parameter-matched OAMP and IDCD controls on ocean and PDE data, respectively (Appendix E.4). Appendix F.1 compares MSM's reconstruction mechanism with ours and discusses its behavior under Pixel 30%.

4.2 Datasets and Evaluation

The ocean products preserve realistic, structured gaps and reveal reference values only where the source product observed them.

Table 1: Ocean reconstruction at 64×64 on withheld source-supported values. Learned-model results are means $\pm$ standard deviations across independent training seeds; DINEOF results are point estimates. Bold marks the best result in each column.

Method	Black Sea CHL		Baltic Sea NANO		Global Ocean SSS	
	MSE↓ ($\times 10^{-3}$)	PSNR↑ (dB)	MSE↓ ($\times 10^{-3}$)	PSNR↑ (dB)	MSE↓ ($\times 10^{-3}$)	PSNR↑ (dB)
DINEOF	21.457	16.89	37.398	14.37	44.451	13.68
AmbientDiff	12.204 $\pm$0.156	19.17 $\pm$0.03	28.304 $\pm$0.206	15.72 $\pm$0.03	47.458 $\pm$1.718	14.07 $\pm$0.12
DINDiff	43.830 $\pm$1.515	14.77 $\pm$0.09	53.153 $\pm$1.775	13.22 $\pm$0.04	61.082 $\pm$0.960	12.73 $\pm$0.07
MissDiff	25.636 $\pm$0.300	15.92 $\pm$0.08	47.818 $\pm$1.704	13.29 $\pm$0.16	41.935 $\pm$0.173	14.01 $\pm$0.03
MSM	15.077 $\pm$0.184	18.27 $\pm$0.04	30.206 $\pm$1.170	15.37 $\pm$0.19	40.901 $\pm$0.277	14.13 $\pm$0.03
OAMP	10.263 $\pm$0.162	19.84 $\pm$0.10	27.220 $\pm$0.458	15.60 $\pm$0.24	44.082 $\pm$0.472	14.04 $\pm$0.05
Ours	**7.164** $\pm$**0.060**	**21.79** $\pm$**0.04**	**19.345** $\pm$**0.741**	**17.37** $\pm$**0.07**	**26.798** $\pm$**0.533**	**15.93** $\pm$**0.05**

Ocean products. We use daily Black Sea CHL, Baltic Sea NANO, and Global Ocean SSS observations, retaining their cloud-related and track- or swath-structured gaps. Fields and masks are jointly resampled to 64×64 and 128×128, with every retained field remaining incomplete on its valid ocean domain. Samples are split chronologically into 70%/15%/15% training, validation, and test partitions before resolution-specific filtering.

Controlled PDE fields. We additionally evaluate Shallow Water height and Navier–Stokes vorticity under controlled pixel- and block-observation protocols. Complete PDE references permit evaluation on valid unsupplied coordinates, while training uses the stored incomplete observations. Appendices E.1 and E.2 give the data preparation and observation protocols.

Evaluation. Ocean validation and testing follow OAMP's real-mask overlay protocol (Ma et al., 2026). A mask from another retained sample of the same product, split, and resolution is intersected with the current source support $\boldsymbol{M}_{\mathrm{raw}}$. After feasibility handling, the input and scoring supports are cached and shared across methods (Appendix E.2). Given $\boldsymbol{M}_{\mathrm{in}}$, we score only withheld source-supported values:

$$\boldsymbol{M}_{\mathrm{eval}} = \boldsymbol{M}_{\mathrm{raw}} \odot (1 - \boldsymbol{M}_{\mathrm{in}}). \tag{11}$$

Source-missing ocean locations are not scored. MSE pools squared errors over scored spatial–channel values, whereas PSNR averages sample-level scores; both use the normalized, clipped evaluation space (Appendix E.3). Preprocessing statistics are fitted only to available training observations (Appendix E.1).

4.3 Reconstruction Accuracy and Estimator Organization

Main results. Our first question is whether local-to-global reconstruction improves accuracy across three authentic ocean products with different observation geometries. At 64×64, our estimator achieves the lowest MSE and highest PSNR among the compared methods on all three ocean products (Table 1). MSE decreases by 28.9%–34.5% relative to the lowest-MSE external baseline for each product. Complete comparisons under the controlled PDE observation settings of prior work (Zhou et al., 2026) are reported in Tables 8 and 9 in Appendix F.1.

Matched estimator organization. Table 2 compares Local Prior with three full-domain estimators trained on the same incomplete pairs and observation-aligned query supervision. Full Direct predicts from the context and coordinates and matches the total parameter count of Prior Direct and Prior Residual; its training time is comparable to our combined two-stage training time (Appendix E.4). Prior Direct and Prior Residual share the frozen local predictor, full-domain backbone, and augmented input I_t. Their zero-output heads give initial full-field predictions of zero and $\mathbf{\Pi}_t$, respectively, so their comparison evaluates residual parameterization together with its induced initialization (Appendix B.3).

Table 2: Estimator organization and prior replacement for full-mask ocean reconstruction at 64×64. Results are means $\pm$ standard deviations across independent training seeds. Bold marks the best result in each column.

Estimator	Black Sea CHL		Baltic Sea NANO		Global Ocean SSS	
	MSE↓ ($\times 10^{-3}$)	PSNR↑ (dB)	MSE↓ ($\times 10^{-3}$)	PSNR↑ (dB)	MSE↓ ($\times 10^{-3}$)	PSNR↑ (dB)
Local Prior	15.227 $\pm$0.103	18.15 $\pm$0.03	30.576 $\pm$0.479	15.17 $\pm$0.04	32.053 $\pm$0.276	15.13 $\pm$0.04
Full Direct	9.350 $\pm$0.306	20.32 $\pm$0.12	23.559 $\pm$0.834	16.41 $\pm$0.16	32.895 $\pm$0.967	14.99 $\pm$0.15
Low-rank2Global	11.067 $\pm$0.243	19.79 $\pm$0.11	22.841 $\pm$0.171	16.43 $\pm$0.04	30.282 $\pm$0.649	15.46 $\pm$0.08
Global2Global	9.175 $\pm$0.499	20.55 $\pm$0.16	25.097 $\pm$1.442	16.23 $\pm$0.21	41.695 $\pm$0.670	14.30 $\pm$0.04
Prior Direct	8.546 $\pm$0.062	20.82 $\pm$0.03	21.445 $\pm$0.417	16.89 $\pm$0.09	30.151 $\pm$0.197	15.40 $\pm$0.03
Prior Residual (Ours)	**7.164** $\pm$0.060	**21.79** $\pm$0.04	**19.345** $\pm$0.741	**17.37** $\pm$0.07	**26.798** $\pm$0.533	**15.93** $\pm$0.05

The Local Prior alone has higher MSE than Full Direct on Black Sea CHL and Baltic Sea NANO, yet Prior Direct reduces MSE relative to Full Direct by 8.3%–9.0% across the three products. Prior Residual yields a further 9.8%–16.2% reduction over Prior Direct. These results distinguish the value of a local estimate as an intermediate representation from its standalone prediction accuracy.

Prior replacement. We next ask whether the result follows from any two-stage correction pipeline or from the particular locally learned reference. We replace the learned local consensus prior with a full-domain OAMP prediction (Global2Global) or a low-rank DINEOF reconstruction (Low-rank2Global), preserving the prior–corrector connection design and retraining a separate global residual corrector for each replacement. Relative to the standalone OAMP and DINEOF results in Table 1, their respective cascades improve both MSE and PSNR (Table 2). Prior Residual nevertheless achieves lower MSE and higher PSNR than both cascades on all three products, reducing MSE by 21.9%–35.7% relative to Global2Global and 11.5%–35.3% relative to Low-rank2Global. Thus a corrector can improve alternative references, but the choice of reference still matters within the tested two-stage designs.

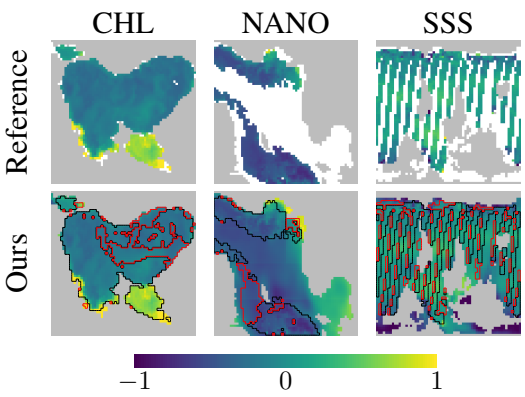

Figure 3: Full-mask ocean reconstructions at 64×64, with supplied observations restored in Ours. Black and red outlines mark supplied observations and withheld source-supported scoring locations, respectively. Gray denotes outside-domain locations; white in Reference denotes source-missing values.

Figure 3 shows full-mask reconstructions of the three ocean products. Figure 5 in Appendix H.1 decomposes these predictions into local priors and learned residuals, while Figures 6–8 compare Ours with Full Direct.

Fine-scale structure under block missingness. In the highlighted missing block of Navier–Stokes Block 8/9, Ours retains a more distinct elongated vorticity structure than IDCD (PM), with local contrast closer to the reference (Figure 4). Appendices H.3–H.4 provide additional PDE reconstruction examples.

Inference cost. At the evaluated common reconstruction-latency ceilings, our estimator retains lower MSE than parameter-matched Full Direct on Black Sea CHL and Shallow Water Pixel 30% (Table 18). Appendix F.5 reports the inference-mode comparison and the corresponding observation-retention and spatial-coverage statistics (Table 19).

Resolution. With each learned method trained and its checkpoint selected separately at 128×128, our estimator achieves lower MSE and higher PSNR than all six ocean baselines under the same source-supported scoring protocol (Appendix F.2, Table 11). Appendix H.2 shows reconstructions at this resolution (Figures 9–11).

4.4 Prior Construction and Connection Ablations

We examine how the local reference is constructed and connected to the corrector. Removing local coordinates retrains both stages, whereas input-map controls fix the prior constructor and retrain only the corrector (Appendix E.4). Both controls evaluate the final full-mask reconstruction.

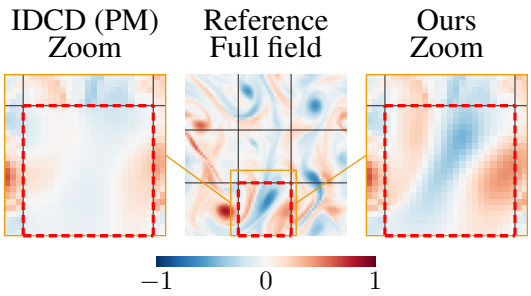

Figure 4: Navier–Stokes Block 8/9 reconstruction at 64×64. IDCD (PM) is the parameter-matched IDCD control, denoted Full Direct. Red dashes mark the held-out block, black lines the acquisition partition, and orange boxes the zoomed region.

Constructing the local prior. Removing local coordinates raises final MSE on all three products (Table 12, Panel A). For the local prior constructed from the full supplied mask before correction, uniform fusion gives the lowest MSE among tested weighting rules applied to identical overlapping predictions, while half-stride overlap gives lower MSE than non-overlap (Table 13). Final reconstruction favors 32×32 patches on CHL and NANO but 16×16 patches on SSS; patch scale is selected per product on validation data (Table 14). Appendix F.3 gives the fusion definitions and explains the coupled changes in the covering comparison.

Connecting the prior to the corrector. Removing any tested input map raises final MSE on all three products. Removing both discrepancy maps gives the largest increase among these controls: 36.8%, 22.9%, and 17.4% on CHL, NANO, and SSS, respectively (Appendix F.3, Table 12, Panel B). These results support explicit prior–evidence features for a finite corrector, not access to new measurements. Because each corrector is retrained, the ablations do not isolate an independent causal contribution for each map.

4.5 Does Global Correction Use Evidence Beyond Local Support?

Does global correction respond to observations that cannot affect the local prior at a target? For a scored block B, $U(B)$ is the union of contributing patch supports, so deleting observations outside it cannot change the prior on B (Appendix D.3). We pair full-mask ocean reconstruction with deletion of the farthest third of supplied observations outside $U(B)$, fixing the checkpoint and perturbation and recomputing the prior, features, and final reconstruction.

The prior on B remains unchanged for every eligible pair, while the corrected prediction has nonzero mean squared displacement on all three products (Table 16). This establishes responsiveness beyond local support, but not whether the removed evidence improves accuracy. We test that separately using matched-count deletions from three remote distance bands and a Local Near control, comparing paired block MSE of the final reconstructions. Remote Mid and Remote Far deletion increase mean error on CHL and NANO, with 95% confidence intervals above zero. On SSS, Remote Near and Local Near deletion increase mean error, Remote Far deletion decreases it, and the Remote Mid interval crosses zero (Table 17). Thus the accuracy effect of evidence beyond local support depends on the product and distance band. Appendix F.4 specifies the diagnostic supports and paired deletion protocols.

5 Conclusion

We studied an explicit structural organization for learning from incomplete physical fields: construct a local consensus estimate, then learn its full-domain residual correction. Estimator comparisons and retrained controls support this organization, alongside reconstruction gains on withheld source-supported ocean values. The central finding is that a local estimate can be valuable as a revisable intermediate representation even when it is not the strongest standalone predictor, allowing local structure to be learned explicitly while retaining full-domain observations for correction.

Remaining boundary artifacts under block missingness are discussed in Appendix G.

AI USE STATEMENT

Generative AI tools were used for language editing, manuscript organization, LaTeX assistance, and preliminary feedback on mathematical exposition and experimental design. They did not execute experiments, access unreleased data, or generate reported measurements. The authors independently verified all definitions, proofs, citations, protocols, and numerical claims and take full responsibility for the final manuscript.

REPRODUCIBILITY STATEMENT

To support reproducibility, we will make the experimental code publicly available upon publication. The release will include the model implementations, training and evaluation procedures, experimental configurations, and documentation. The datasets used in our experiments are available from their original providers and remain subject to the providers' access and licensing terms. The ocean data are provided by the E.U. Copernicus Marine Environment Monitoring Service (Letraon et al., 2017). Appendix E.1 records the corresponding product identifiers and DOIs and describes the preprocessing used in our experiments. The use of generative AI tools is documented in the AI Use Statement.

REFERENCES

A. Alvera-Azcárate, A. Barth, M. Rixen, and J.M. Beckers. Reconstruction of incomplete oceanographic data sets using empirical orthogonal functions: application to the adriatic sea surface temperature. *Ocean Modelling*, 9(4):325–346, 2005. ISSN 1463-5003. doi: https://doi.org/10.1016/j.ocemod.2004.08.001. URL https://www.sciencedirect.com/science/article/pii/S1463500304000472.

Omer Bar-Tal, Lior Yariv, Yaron Lipman, and Tali Dekel. MultiDiffusion: Fusing diffusion paths for controlled image generation. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (eds.), *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 1737–1752. PMLR, 23–29 Jul 2023. URL https://proceedings.mlr.press/v202/bar-tal23a.html.

A. Barth, J. Brajard, A. Alvera-Azcárate, B. Mohamed, C. Troupin, and J.-M. Beckers. Ensemble reconstruction of missing satellite data using a denoising diffusion model: application to chlorophyll a concentration in the black sea. *Ocean Science*, 20(6):1567–1584, 2024. doi: 10.5194/os-20-1567-2024. URL https://os.copernicus.org/articles/20/1567/2024/.

J. M. Beckers and M. Rixen. Eof calculations and data filling from incomplete oceanographic datasets. *Journal of Atmospheric and Oceanic Technology*, 20(12):1839 – 1856, 2003. doi: 10.1175/1520-0426(2003)020<1839:ECADFF>2.0.CO;2. URL https://journals.ametsoc.org/view/journals/atot/20/12/1520-0426_2003_020_1839_ecadff_2_0_co_2.xml.

Shuhao Cao. Navier-stokes dataset of isotropic turbulence in a periodic box, 2024. URL https://huggingface.co/datasets/scaomath/navier-stokes-dataset. Funded by National Science Foundation: NSF award DMS-2309778.

Shuhao Cao, Francesco Brarda, Ruipeng Li, and Yuanzhe Xi. Spectral-refiner: Accurate fine-tuning of spatiotemporal fourier neural operator for turbulent flows. In Y. Yue, A. Garg, N. Peng, F. Sha, and R. Yu (eds.), *International Conference on Learning Representations*, volume 2025, pp. 60348–60384, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/file/972eea97af7fe9a7588f5af136edef0c-Paper-Conference.pdf.

Giannis Daras, Kulin Shah, Yuval Dagan, Aravind Gollakota, Alex Dimakis, and Adam Klivans. Ambient diffusion: Learning clean distributions from corrupted data. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), *Advances in Neural Information Processing Systems*, volume 36, pp. 288–313. Curran Associates, Inc., 2023. doi: 10.52202/075280-0015. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/012af729c5d14d279581fc8a5db975a1-Paper-Conference.pdf.

Josh Givens, Song Liu, and Henry Reeve. Score matching with missing data. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu (eds.), *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 19523–19561. PMLR, 13–19 Jul 2025. URL https://proceedings.mlr.press/v267/givens25a.html.

Alex Graves, Rupesh Kumar Srivastava, Timothy Atkinson, and Faustino Gomez. Bayesian flow networks, 2025. URL https://arxiv.org/abs/2308.07037.

Jason Hu, Bowen Song, Xiaojian Xu, Liyue Shen, and Jeffrey Fessler. Learning image priors through patch-based diffusion models for solving inverse problems. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (eds.), *Advances in Neural Information Processing Systems*, volume 37, pp. 1625–1660. Curran Associates, Inc., 2024. doi: 10.52202/079017-0052. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/03261886741f1f21f52f2a2d570616a2-Paper-Conference.pdf.

Satoshi Iizuka, Edgar Simo-Serra, and Hiroshi Ishikawa. Globally and locally consistent image completion. *ACM Trans. Graph.*, 36(4), July 2017. ISSN 0730-0301. doi: 10.1145/3072959.3073659. URL https://doi.org/10.1145/3072959.3073659.

Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. Noise2Noise: Learning image restoration without clean data. In Jennifer Dy and Andreas Krause (eds.), *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 2965–2974. PMLR, 10–15 Jul 2018. URL https://proceedings.mlr.press/v80/lehtinen18a.html.

Py Letraon, A. Ali, E. Alvarez Fanjul, L. Aouf, L. Axell, R. Aznar, M. Ballarotta, A. Behrens, M. Benkiran, A. Bentamy, L. Bertino, P. Bowyer, V. Brando, L. A. Breivik, B. Buongiorno Nardelli, S. Cailleau, S. A. Ciliberti, S. Colella, N. Mc Connell, G. Coppini, G. Cossarini, T. Dabrowski, M. de Alfonso Alonsomuñoyerro, E. O'dea, C. Desportes, F. Dinessen, M. Drevillon, Y. Drillet, M. Drudi, R. Dussurget, Y. Faugère, V. Forneris, C. Fratianni, O. Le Galloudec, I. Garcia Hermosa, M. García Sotillo, P. Garnesson, G. Garric, I. Golbeck, J. Gourrion, M. L. Grégoire, S. Guinehut, E. Gutknecht, C. Harris, F. Hernandez, V. Huess, J.A. Johannessen, S. Kay, R. Killick, R. King, J. De Kloe, G. Korres, P. Lagemaa, R. Lecci, J.F. Legeais, J. M. Lellouche, B. Levier, P. Lorente, A. Mangin, M. Martin, A. Melet, J. Murawski, E. Özsoy, A. Palazov, S. Pardo, L. Parent, A. Pascual, J. Paul, E. Peneva, C. Perruche, D. Peterson, L. Petit de La Villeon, N. Pinardi, S. Pouliquen, M. I. Pujol, R. Rainaud, Pierre Rampal, G. Reffray, C. Regnier, A. Reppucci, A. Ryan, S. Salon, A. Samuelsen, R. Santoleri, Andrew Saulter, J. She, C. Solidoro, E. Stanev, J. Staneva, A. Stoffelen, A. Storto, P. Sykes, T. Szekely, G. Taburet, B. Taylor, J. Tintore, C. Toledano, M. Tonani, L. Tuomi, G. Volpe, H. Wedhe, T. Williams, L. Vandendbulcke, D. Van Zanten, K. Von Schuckmann, J. Xie, A. Zacharioudaki, and H. Zuo. THE COPERNICUS MARINE ENVIRONMENTAL MONITORING SERVICE: MAIN SCIENTIFIC ACHIEVEMENTS AND FUTURE PROSPECTS. Research Report 56, MERCATOR OCEAN, September 2017. URL https://univ-grenoble-alpes.hal.science/hal-03405376.

Zhenyu Li, Shariq Farooq Bhat, and Peter Wonka. Patchfusion: An end-to-end tile-based framework for high-resolution monocular metric depth estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10016–10025, June 2024.

Zhenyu Li, Shariq Farooq Bhat, and Peter Wonka. Patchrefiner: Leveraging synthetic data for real-domain high-resolution monocular metric depth estimation. In Aleš Leonardis, Elisa Ricci, Stefan Roth, Olga Russakovsky, Torsten Sattler, and Gül Varol (eds.), *Computer Vision – ECCV 2024*, pp. 250–267, Cham, 2025. Springer Nature Switzerland. ISBN 978-3-031-72855-6.

Zhenyu Li, Wenqing Cui, Shariq Bhat, and Peter Wonka. Patchrefiner v2: Fast and lightweight real-domain high-resolution metric depth estimation. In C. Vondrick, B. Hariharan, C. Raffel, L. Pinto, D. Yang, and A. Faust (eds.), *International Conference on Learning Representations*, volume 2026, pp. 117738–117753, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/file/bf2c36a946443a6f5b9e696a7b224ffb-Paper-Conference.pdf.

Rosanne Liu, Joel Lehman, Piero Molino, Felipe Petroski Such, Eric Frank, Alex Sergeev, and Jason Yosinski. An intriguing failing of convolutional neural networks and the coordconv solution. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL https://proceedings.neurips.cc/paper_files/paper/2018/file/60106888f8977b71e1f15db7bc9a88d1-Paper.pdf.

Chiyuan Ma, Zihan Zhou, and Tianshu Yu. Observation-aligned mask priors for learning physical dynamics from authentic occlusions, 2026. URL https://arxiv.org/abs/2605.16818.

Kamyar Nazeri, Eric Ng, Tony Joseph, Faisal Qureshi, and Mehran Ebrahimi. Edgeconnect: Structure guided image inpainting using edge prediction. In *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pp. 3265–3274, 2019. doi: 10.1109/ICCVW.2019.00408.

Yidong Ouyang, Liyan Xie, Chongxuan Li, and Guang Cheng. Missdiff: Training diffusion models on tabular data with missing values. In *ICML 2023 Workshop on Structured Probabilistic Inference & Generative Modeling*, 2023. URL https://openreview.net/forum?id=S435pkeAdT.

Chicago Y. Park, Shirin Shoushtari, Hongyu An, and Ulugbek Kamilov. Measurement score-based diffusion model. In C. Vondrick, B. Hariharan, C. Raffel, L. Pinto, D. Yang, and A. Faust (eds.), *International Conference on Learning Representations*, volume 2026, pp. 96428–96461, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/file/9bcc5e97c6343b52dc4ff7b034cc640c-Paper-Conference.pdf.

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

Makoto Takamoto, Timothy Praditia, Raphael Leiteritz, Daniel MacKinlay, Francesco Alesiani, Dirk Pflüger, and Mathias Niepert. Pdebench: An extensive benchmark for scientific machine learning. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (eds.), *Advances in Neural Information Processing Systems*, volume 35, pp. 1596–1611. Curran Associates, Inc., 2022. doi: 10.52202/068431-0117. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/0a9747136d411fb83f0cf81820d44afb-Paper-Datasets_and_Benchmarks.pdf.

Zhendong Wang, Yifan Jiang, Huangjie Zheng, Peihao Wang, Pengcheng He, Zhangyang "Atlas" Wang, Weizhu Chen, and Mingyuan Zhou. Patch diffusion: Faster and more data-efficient training of diffusion models. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), *Advances in Neural Information Processing Systems*, volume 36, pp. 72137–72154. Curran Associates, Inc., 2023. doi: 10.52202/075280-3158. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/e4667dd0a5a54b74019b72b677ed8ec1-Paper-Conference.pdf.

Qianxun Xu and Zuchuan Li. Partial physics informed diffusion model for ocean chlorophyll concentration reconstruction. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume 38, Main Conference, pp. 156490–156507. Curran Associates, Inc., 2025. doi: 10.52202/085713-5230. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/e57cd5735565b737b2c2afe56daca99b-Paper-Conference.pdf.

Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S. Huang. Generative image inpainting with contextual attention. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

Hengrui Zhang, Liancheng Fang, Qitian Wu, and Philip Yu. Diffputer: Empowering diffusion models for missing data imputation. In Y. Yue, A. Garg, N. Peng, F. Sha, and R. Yu (eds.), *International Conference on Learning Representations*, volume 2025, pp. 63164–63185, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/file/9e2ad37b8899366767df9319efc6f4de-Paper-Conference.pdf.

Zihan Zhou, Chenguang Wang, Hongyi Ye, Yongtao Guan, and Tianshu Yu. Incomplete data, complete dynamics: A diffusion approach. In C. Vondrick, B. Hariharan, C. Raffel, L. Pinto, D. Yang, and A. Faust (eds.), *International Conference on Learning Representations*, volume 2026, pp. 10918–10952, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/file/123a18dfd821c8b440f42a00a27648d6-Paper-Conference.pdf.

M. Zupančič Muc, V. Zavrtanik, A. Barth, A. Alvera-Azcarate, M. Ličer, and M. Kristan. Criter 1.0: a coarse reconstruction with iterative refinement network for sparse spatio-temporal satellite data. *Geoscientific Model Development*, 18(17):5549–5573, 2025. doi: 10.5194/gmd-18-5549-2025. URL https://gmd.copernicus.org/articles/18/5549/2025/.

A Related Work

A.1 Learning from Incomplete Training Data

Learning from incomplete training samples is distinct from filling test-time gaps with a model learned from complete data (Zhou et al., 2026). AmbientDiff introduces additional measurement corruption and trains predictions against the available measurements (Daras et al., 2023). MissDiff adapts denoising score matching by masking its regression loss to observed entries (Ouyang et al., 2023), while MSM learns partial measurement scores from noisy, subsampled observations and combines them through stochastic sampling (Park et al., 2026).

Alternative formulations include DiffPuter's expectation–maximization procedure, which alternates diffusion-model learning and imputation (Zhang et al., 2025), and the importance-weighted and variational score-matching objectives of Givens et al. (2025). These approaches develop ways to learn and reconstruct without complete training samples. Our study concerns the organization of a point estimator around a locally constructed field and its full-domain revision. Unlike MSM's sampling-time composition of partial measurement scores (Park et al., 2026), we reconcile patchwise clean-value predictions into a frozen reference and train a separate residual corrector under held-out-query supervision. Appendix F.1 ("MSM under pixel observations") provides a mechanism-level comparison and discusses MSM under the Pixel 30% protocol.

A.2 Physical-Field Reconstruction and Observation-Aligned Supervision

Geophysical reconstruction has long used structural assumptions to infer missing values. DINEOF reconstructs incomplete ocean fields through low-rank empirical orthogonal functions (Beckers & Rixen, 2003; Alvera-Azcárate et al., 2005). DINDiff trains a denoising diffusion model directly on cloud-obscured satellite chlorophyll, using additionally withheld observations to supervise noise prediction (Barth et al., 2024). PPIDM incorporates a trusted partial physics operator into diffusion training for ocean chlorophyll reconstruction (Xu & Li, 2025).

Within context–query learning, the observation pattern determines which relationships can receive supervision. IDCD develops a PDE-oriented framework that aligns context–query sampling with known pixel- or block-structured observation patterns and analyzes the role of query exposure (Zhou et al., 2026). OAMP is designed for ocean reconstruction under authentic occlusions: it learns a prior over binary observation masks with a Bayesian Flow Network (Graves et al., 2025) and uses guided sampling and mask intersection to construct context–query pairs (Ma et al., 2026). We build on their observation-aligned supervision and adopt IDCD's single-step prediction, with inference modes specified in Section 3.3. The inherited mask construction selects available observations for

Table 3: Original settings, reconstruction mechanisms, and experimental roles of the evaluated baselines.

Method	Original setting	Principal mechanism	Use in this paper
Evaluated baselines			
DINEOF (Beckers & Rixen, 2003; Alvera-Azcárate et al., 2005)	Incomplete ocean fields	Low-rank EOF reconstruction	Ocean baseline; low-rank-prior control
AmbientDiff (Daras et al., 2023)	Corrupted images	Further corruption; prediction supervised on available measurements	Ocean and PDE baseline
DINDiff (Barth et al., 2024)	Satellite ocean chlorophyll	Noise prediction on additionally withheld observations	Ocean and PDE baseline
MissDiff (Ouyang et al., 2023)	Incomplete tabular data	Masked denoising score matching	Ocean and PDE baseline
MSM (Park et al., 2026)	Natural and medical imaging	Partial measurement scores; stochastic sampling	Ocean and PDE baseline
IDCD (Zhou et al., 2026)	PDE-oriented physical dynamics	Observation-structured context–query prediction	PDE baseline
OAMP (Ma et al., 2026)	Authentically occluded ocean fields	Learned mask prior; aligned context–query prediction	Ocean baseline; full-domain-prior control

Table 4: Comparison with related local–global and staged reconstruction methods in their original training settings.

Method	Original training setting	Construction and revision	Distinction from our estimator
GLCIC (Iizuka et al., 2017)	Artificially masked complete images	A completion network trained with global and local discriminators	Local–global consistency is imposed through training losses, rather than a local-value consensus followed by a learned full-domain corrector.
Contextual Attention (Yu et al., 2018)	Artificially masked complete images	Coarse completion followed by contextual-attention refinement	Uses a coarse-to-fine image-completion pipeline rather than a reference learned from incomplete spatial patches.
EdgeConnect (Nazeri et al., 2019)	Images and derived edge targets	Predicted edges condition subsequent image completion	The intermediate representation is an edge map, not a field-valued estimate used as a residual anchor.
Patch Diffusion (Wang et al., 2023)	Clean images with coordinate-conditioned, multiscale training	Learns scores from patch and full-image training views	Patch learning serves diffusion-model training rather than held-out-query prediction followed by a separate full-domain corrector.
PaDIS (Hu et al., 2024)	Clean image patches for prior learning	Patch scores define a prior used by an inverse-problem solver	Global measurement constraints enter the solver; there is no separately trained full-domain residual corrector around an observed-query-trained local consensus.
MultiDiffusion (Bar-Tal et al., 2023)	A pretrained text-to-image diffusion model	Reconciles regional diffusion updates during sampling	Fusion operates on sampling updates, rather than constructing a frozen clean-value reference for a learned corrector.
PatchFusion (Li et al., 2024)	Image–depth training data	Fuses global coarse and tiled depth predictions	Global information guides high-resolution tile fusion, rather than revising a consensus constructed from incomplete local field observations.
PatchRefiner (Li et al., 2025)	Image–depth data with synthetic-to-real supervision	Frozen global coarse depth predictor followed by patchwise residual refinement	Shares a frozen reference and residual refinement; the reference origin, depth supervision, and global-coarse-to-patch direction differ.
PatchRefiner V2 (Li et al., 2026)	Image–depth data with synthetic-to-real supervision	Lightweight patch refinement and cross-scale fusion; coarse network and refiner jointly fine-tuned	Jointly fine-tunes global-coarse-to-patch refinement; ours freezes the local predictor before learning full-domain correction.
CRITER (Zupančič Muc et al., 2025)	Sparse spatiotemporal satellite fields	Global low-frequency reconstruction followed by iterative refinement	Begins with global spatiotemporal reconstruction, whereas ours begins with spatially local estimates and then applies full-domain revision.
Ours	**Never-complete field observations with context–query supervision**	**Coordinate-conditioned local predictions → overlap consensus → learned full-domain residual correction**	**The same observation-conditioned field is used as an explicit input and the residual output anchor.**

context and supervision; our local field prior is a different object, supplying sample-conditioned values for a full-domain estimator to revise.

Table 3 summarizes the original settings, principal mechanisms, and experimental roles of the evaluated baselines.

A.3 Patch Priors and Local–Global Estimation

Local–global organization and staged prediction have a long history in image completion. Globally and Locally Consistent Image Completion uses global and local discriminators to train a completion network (Iizuka et al., 2017). Contextual Attention employs a coarse-to-fine completion pipeline (Yu et al., 2018), while EdgeConnect predicts edges before conditioning image completion on them (Nazeri et al., 2019).

Patch Diffusion learns position-conditioned patch scores (Wang et al., 2023), and PaDIS constructs image priors from position-aware patch scores for inverse-problem solvers (Hu et al., 2024). MultiDiffusion reconciles regional diffusion updates through least-squares fusion during sampling (Bar-Tal et al., 2023). Our construction also uses positional information and overlap reconciliation, but its intermediate object is a clean-value field conditioned on the current observations. The patch predictor is trained on query targets drawn only from observed entries of incomplete fields, and predictions from the frozen local model are reconciled before learned full-domain correction.

Other multiscale estimators use a global coarse prediction to support local detail. PatchFusion combines a globally consistent coarse depth estimate with tiled predictions (Li et al., 2024). PatchRefiner freezes a global coarse depth predictor and learns patchwise residual refinements (Li et al., 2025).

PatchRefiner V2 introduces lightweight encoding and cross-scale feature fusion; after separate pre-training, the coarse network and refiner are jointly fine-tuned end to end (Li et al., 2026). For satellite fields, CRITER first reconstructs low-frequency structure using global spatiotemporal correlations and then iteratively refines the estimate (Zupančič Muc et al., 2025). Our estimator proceeds from local construction to full-domain revision: a shared local predictor produces a consensus reference, and a separate corrector retains the original context to revise that reference. The reference field serves both as an input, accompanied by observation-discrepancy features, and as the anchor of the residual output. When context ensembling is used, construction and correction are performed separately for each context before averaging the corrected fields. Table 4 compares the original supervision, intermediate representations, and reconstruction organization of related local–global and staged methods.

B Method Implementation and Reconstruction

This appendix specifies the diffusion parameterization, local-prior construction, estimator interfaces, reconstruction settings, and complete training and reconstruction procedure.

B.1 Diffusion Parameterization and Noise Schedule

We use the variance-preserving corruption

$$\boldsymbol{x}_t = \alpha_t \boldsymbol{x}_0 + \sigma_t \boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(0, I), \qquad \alpha_t^2 + \sigma_t^2 = 1, \tag{12}$$

and parameterize both local and full-domain networks as clean-data predictors. Given context $\boldsymbol{M}_{\mathrm{ctx}}$,

$$\boldsymbol{x}_{\mathrm{ctx},t} = \boldsymbol{M}_{\mathrm{ctx}} \odot (\alpha_t \boldsymbol{x}_{\mathrm{obs}} + \sigma_t \boldsymbol{\epsilon}). \tag{13}$$

The diffusion conventions follow Song et al. (2021). We use clean-data prediction so that the local prior, full-domain output, and residual correction are represented in physical-field space. Training samples diffusion times across the prescribed interval, while inference uses the near-zero time δ with the reconstruction modes in Section 3.3.

The noise schedule is

$$\alpha_t = \exp\left[-\frac{1}{4}(\beta_{\max} - \beta_{\min})t^2 - \frac{1}{2}\beta_{\min} t\right], \qquad \sigma_t^2 = 1 - \alpha_t^2. \tag{14}$$

B.2 Local Prediction and Overlap Consensus

We use the main-text noisy context $\boldsymbol{x}_{\mathrm{ctx},t} = \boldsymbol{M}_{\mathrm{ctx}} \odot \boldsymbol{x}_{\mathrm{obs},t}$ and abbreviate its observable state by $O_t = (\boldsymbol{x}_{\mathrm{ctx},t}, \boldsymbol{M}_{\mathrm{ctx}})$. The crop R_p, support Ω_p, and covering $\mathcal{P}$ are defined in Section 3.1; let $\boldsymbol{\Gamma}_{\mathcal{P}} = (R_p \boldsymbol{\Gamma})_{p \in \mathcal{P}}$ collect the coordinate crops. The partition law is denoted by $(\boldsymbol{M}_{\mathrm{ctx}}, \boldsymbol{M}_{\mathrm{qry}}) \sim \mathcal{P}_{\mathrm{align}}(\cdot \mid \boldsymbol{M})$. The compact local input is

$$J_{p,t} = (R_p \boldsymbol{x}_{\mathrm{ctx},t}, R_p \boldsymbol{M}_{\mathrm{ctx}}, R_p \boldsymbol{\Gamma}). \tag{15}$$

For view p, set $\boldsymbol{M}_{\mathrm{ctx},p} = R_p \boldsymbol{M}_{\mathrm{ctx}}$ and $\boldsymbol{M}_{\mathrm{qry},p} = R_p \boldsymbol{M}_{\mathrm{qry}}$. The local denoiser receives the corrupted context values, context mask, and normalized absolute coordinates:

$$\boldsymbol{Z}_{p,t}^{(1)} = \mathrm{cat}[R_p \boldsymbol{x}_{\mathrm{ctx},t}, R_p \boldsymbol{M}_{\mathrm{ctx}}, R_p \boldsymbol{\Gamma}]. \tag{16}$$

The coordinate atlas $\boldsymbol{\Gamma}_{\mathcal{P}}$ is fixed by the covering geometry. A training item uses one perturbation at its sampled time, while an inference item uses one perturbation to form $\boldsymbol{x}_{\mathrm{in},\delta}$ before applying either the full supplied mask or the ensemble subcontext masks. Within a context evaluation, all local views, discrepancy features, and the full-domain estimator use the same realized corrupted values.

With this notation, $\boldsymbol{x}_{\boldsymbol{\phi}}(t; J_{p,t})$ is the prediction in Equation (5), trained with Equation (6). After local training, $\boldsymbol{\phi}$ is held fixed during prior construction and global correction. Local training samples views with $\|R_p \boldsymbol{M}_{\mathrm{qry}}\|_1 > 0$; prior construction evaluates the full covering.

The frozen predictor and overlap consensus define

$$\boldsymbol{\Pi}_t = \mathcal{P}_{\boldsymbol{\phi}}(t; O_t; \boldsymbol{\Gamma}_{\mathcal{P}}). \tag{17}$$

Here $\mathcal{P}_{\boldsymbol{\phi}}$ is a deterministic construction operator, not a probability law. For fixed patch predictions $\boldsymbol{Y}_{p,t} = \boldsymbol{x}_{\boldsymbol{\phi}}(t; J_{p,t})$, overlap consensus solves

$$\boldsymbol{\Pi}_t = \arg \min_{\boldsymbol{Z}=\boldsymbol{V}\odot\boldsymbol{Z}} \sum_{p\in\mathcal{P}} \|R_p\boldsymbol{Z} - \boldsymbol{Y}_{p,t}\|_2^2 . \tag{18}$$

At each valid coordinate i, differentiating with respect to $\boldsymbol{Z}_i$ gives $\boldsymbol{W}_i\boldsymbol{Z}_i = \sum_{p:i\in\Omega_p} \boldsymbol{Y}_{p,t,i}$. Every valid coordinate is covered, so $\boldsymbol{W}_i > 0$ and the unique solution is the pointwise average in Equation (7); the validity constraint fixes the remaining coordinates to zero.

Default patch geometry. At 64×64, Black Sea CHL, Baltic Sea NANO, and the PDE models use 32×32 patches with stride 16, giving nine overlapping views. The local-prior models for Global Ocean SSS use 16×16 patches with stride 8, giving 49 views, throughout the main, estimator-organization, component, fusion, and intervention experiments. The ocean patch scales are selected on validation data; Appendix F.3 (Table 14) reports the comparison between 32×32/stride-16 and 16×16/stride-8 coverings. At 128×128, the ocean models use 64×64 patches with stride 32, giving nine views. Every covering includes all valid output coordinates, and the frozen local predictor makes one clean-field prediction per view for each evaluated context. Each patch is evaluated independently from its local values, mask, and coordinate crop, using fixed preprocessing statistics and no operation that exchanges current-sample information across patch calls.

B.3 Prior–Evidence Features and Estimator Interfaces

The feature operator returns $\boldsymbol{D}_t = (\boldsymbol{W}, \boldsymbol{S}_{\mathrm{ctx}}, \boldsymbol{S}_{\partial})$, where $\boldsymbol{W}$ is the overlap multiplicity. The overlap count $\boldsymbol{W}$ is defined in Equation (7). At supplied coordinates, define

$$\boldsymbol{S}_{\mathrm{ctx},i} = \boldsymbol{M}_{\mathrm{ctx},i} \frac{1}{d_x} \sum_{c=1}^{d_x} |\boldsymbol{\Pi}_{t,c,i} - \boldsymbol{x}_{\mathrm{ctx},t,c,i}| . \tag{19}$$

Let $N_i(\boldsymbol{M}_{\mathrm{ctx}}) = \{j \in \Omega : \|j - i\|_1 = 1,\ \boldsymbol{M}_{\mathrm{ctx},j} = 1\}$ and define the neighboring corrupted-context mean

$$\bar{\boldsymbol{x}}_{\mathrm{ctx},t,c,i} = \begin{cases} \dfrac{1}{|N_i(\boldsymbol{M}_{\mathrm{ctx}})|} \displaystyle\sum_{j\in N_i(\boldsymbol{M}_{\mathrm{ctx}})} \boldsymbol{x}_{\mathrm{ctx},t,c,j}, & |N_i(\boldsymbol{M}_{\mathrm{ctx}})| > 0, \\ 0, & |N_i(\boldsymbol{M}_{\mathrm{ctx}})| = 0. \end{cases}$$

At a non-context coordinate adjacent to supplied observations, define

$$\boldsymbol{S}_{\partial,i} = (1 - \boldsymbol{M}_{\mathrm{ctx},i})\mathbf{1}[|N_i(\boldsymbol{M}_{\mathrm{ctx}})| > 0] \frac{1}{d_x} \sum_{c=1}^{d_x} |\boldsymbol{\Pi}_{t,c,i} - \bar{\boldsymbol{x}}_{\mathrm{ctx},t,c,i}| . \tag{20}$$

The time dependence of the discrepancy maps is suppressed in their notation, and all comparisons use the realized corrupted context of the corresponding model evaluation.

The full-domain input and its deterministic features can be abbreviated as

$$\begin{aligned} (\boldsymbol{\Pi}_t, \boldsymbol{D}_t) &= h(t, O_t), \\ \boldsymbol{D}_t &= d_{\mathcal{P}}(O_t, \boldsymbol{\Pi}_t), \\ I_t &= (O_t, \mathrm{sg}(\boldsymbol{\Pi}_t), \boldsymbol{D}_t). \end{aligned} \tag{21}$$

Here $\boldsymbol{D}_t = (\boldsymbol{W}, \boldsymbol{S}_{\mathrm{ctx}}, \boldsymbol{S}_{\partial})$, $d_{\mathcal{P}}$ is the feature constructor defined above, and sg denotes stop-gradient. The nested tuple I_t denotes exactly the full input in Equation (8). The residual objective in Equation (9) is equivalently the query loss on $\mathrm{sg}(\boldsymbol{\Pi}_t) + r_{\boldsymbol{\psi}}(t; I_t)$; the algorithms denote its minibatch version by $\mathcal{L}_{\mathrm{glob}}$. The implementation operator $\mathrm{Cues}_{\mathcal{P}}(\boldsymbol{\Pi}_t, O_t)$ returns the deterministic feature tuple defined above. The no-prior input tensor is

$$\boldsymbol{Z}_t^{\mathrm{base}} = \mathrm{cat}[\boldsymbol{x}_{\mathrm{ctx},t}, \boldsymbol{M}_{\mathrm{ctx}}, \boldsymbol{\Gamma}] . \tag{22}$$

The two prior-conditioned estimators receive the channel-ordered tensor

$$\boldsymbol{Z}_t^{\mathrm{PC}} = \mathrm{cat}[\boldsymbol{x}_{\mathrm{ctx},t}, \boldsymbol{M}_{\mathrm{ctx}}, \mathrm{sg}(\boldsymbol{\Pi}_t), \boldsymbol{W}, \boldsymbol{S}_{\mathrm{ctx}}, \boldsymbol{S}_{\partial}] . \tag{23}$$

The no-prior full-domain direct estimator receives $\boldsymbol{Z}_t^{\mathrm{base}}$. The prior-conditioned full-domain direct estimator and residual corrector receive $\boldsymbol{Z}_t^{\mathrm{PC}}$. These channel orders are used during training and for every inference context. With a single-channel mask, single-channel W, and channel-averaged discrepancy maps, the input channel counts are

$$\dim_{\mathrm{channel}} \boldsymbol{Z}_t^{\mathrm{base}} = d_x + 3,$$
$$\dim_{\mathrm{channel}} \boldsymbol{Z}_t^{\mathrm{PC}} = 2d_x + 4.$$

For a univariate field ($d_x = 1$), these are 4 and 6 channels, respectively; $\boldsymbol{\Gamma}$ contributes two normalized absolute-coordinate channels.

Their output forms are

$$\widehat{\boldsymbol{x}}_t^{\mathrm{PD}} = g_{\boldsymbol{\theta}}(t; I_t),$$
$$g_{\boldsymbol{\theta}_0}(t; I_t) = 0,$$
$$\widehat{\boldsymbol{x}}_t^{\mathrm{PR}} = \mathrm{sg}(\boldsymbol{\Pi}_t) + r_{\boldsymbol{\psi}}(t; I_t),$$
$$r_{\boldsymbol{\psi}_0}(t; I_t) = 0.$$

Zero-output initialization therefore gives initial full-field predictions of zero for Prior Direct and $\boldsymbol{\Pi}_t$ for Prior Residual. Under the shared query loss, the direct branch predicts field values, whereas the residual branch predicts their discrepancy from the frozen reference.

B.4 Full-Mask and Context-Ensemble Reconstruction

For our estimator and internal controls, the validation-selected default settings are full-mask reconstruction on ocean data at both resolutions and $K = 8$ context ensembling on PDE data. Appendix F.5 varies the inference mode and ensemble size. Full-mask inference sets $K = 1$ and $\boldsymbol{M}_{\mathrm{ctx}}^{(1)} = \boldsymbol{M}_{\mathrm{in}}$; a one-member context ensemble instead uses one subsampled context. The same near-zero diffusion time δ is used across datasets, resolutions, validation, and testing.

Algorithm 2 defines the states $O_\delta^{(k)}$, local priors $\boldsymbol{\Pi}_\delta^{(k)}$, and augmented inputs $I_\delta^{(k)}$ for the selected contexts. The local-prior reference and the two direct estimators are evaluated as

$$\widetilde{\boldsymbol{\Pi}}_\delta = \frac{1}{K}\sum_{k=1}^{K} \boldsymbol{\Pi}_\delta^{(k)},$$
$$\widetilde{\boldsymbol{x}}_\delta^{\mathrm{NP}} = \frac{1}{K}\sum_{k=1}^{K} f_{\boldsymbol{\omega}}\left(\delta; O_\delta^{(k)}, \boldsymbol{\Gamma}\right),$$
$$\widetilde{\boldsymbol{x}}_\delta^{\mathrm{PD}} = \frac{1}{K}\sum_{k=1}^{K} g_{\boldsymbol{\theta}}\left(\delta; I_\delta^{(k)}\right).$$

One near-zero perturbation is shared by all contexts for each evaluation item. Each context independently completes local-prior construction and full-domain prediction before the resulting fields are averaged. Each prior-conditioned prediction uses the prior and discrepancy features constructed from its own context. The residual estimator combines corrected predictions as in Equation (10), after which the original supplied observations are restored.

B.5 Training and Reconstruction Algorithm

Algorithm 2 specifies local training, frozen-prior construction, full-domain residual training, and full-mask or context-ensemble reconstruction.

Training accumulates query-masked squared errors over minibatches, while evaluation uses the collection-level metric definitions in Appendix E.3.

Algorithm 2 Overall Two-Stage Training and Reconstruction

Require: Incomplete dataset $\mathcal{D}$, partition rule $\mathcal{P}_{\mathrm{align}}$, covering $\mathcal{P}$, coordinates $\mathbf{\Gamma}$, valid mask $\boldsymbol{V}$; inference input $(\boldsymbol{x}_{\mathrm{in}}, \boldsymbol{M}_{\mathrm{in}})$, inference mode, K, δ
Ensure: Trained parameters $(\boldsymbol{\phi}, \boldsymbol{\psi})$ and reconstruction $\widehat{\boldsymbol{x}}_{\mathrm{out}}$

Stage 1: learn local structure
while local training has not converged **do**
  Sample $(\boldsymbol{x}_{\mathrm{obs}}, \boldsymbol{M}) \sim \mathcal{D}$ and obtain $(\boldsymbol{M}_{\mathrm{ctx}}, \boldsymbol{M}_{\mathrm{qry}}) \sim \mathcal{P}_{\mathrm{align}}(\cdot \mid \boldsymbol{M})$ using the dataset-specific feasibility handling in Appendix C
  Draw $t \sim \mathrm{Unif}[t_\varepsilon, 1]$ and $\boldsymbol{\epsilon} \sim \mathcal{N}(0, I)$
  $\boldsymbol{x}_{\mathrm{ctx},t} \leftarrow \boldsymbol{M}_{\mathrm{ctx}} \odot (\alpha_t \boldsymbol{x}_{\mathrm{obs}} + \sigma_t \boldsymbol{\epsilon})$
  Sample p uniformly from $\{p \in \mathcal{P} : \|R_p \boldsymbol{M}_{\mathrm{qry}}\|_1 > 0\}$
  Predict $\widehat{\boldsymbol{x}}_0^p$ using Eq. (5); update $\boldsymbol{\phi}$ with the patch query loss (6)
end while
Freeze $\boldsymbol{\phi}$; initialize the residual head so that $r_{\boldsymbol{\psi}_0} = 0$
Stage 2: learn full-domain revision
while global training has not converged **do**
  Sample an incomplete pair, a context–query partition, t, and $\boldsymbol{\epsilon}$ as in Stage 1
  $\boldsymbol{x}_{\mathrm{ctx},t} \leftarrow \boldsymbol{M}_{\mathrm{ctx}} \odot (\alpha_t \boldsymbol{x}_{\mathrm{obs}} + \sigma_t \boldsymbol{\epsilon})$
  $(\mathbf{\Pi}_t, I_t) \leftarrow \textsc{BuildReference}(t, \boldsymbol{x}_{\mathrm{ctx},t}, \boldsymbol{M}_{\mathrm{ctx}})$
  $\widehat{\boldsymbol{x}}_{0,t} \leftarrow \mathrm{sg}(\mathbf{\Pi}_t) + r_{\boldsymbol{\psi}}(t; I_t)$; update only $\boldsymbol{\psi}$ with the query loss (9)
end while
Fix $\boldsymbol{\psi}$ for reconstruction
Reconstruction: full-mask or context ensemble
Draw one $\boldsymbol{\epsilon} \sim \mathcal{N}(0, I)$
Set $\boldsymbol{x}_{\mathrm{in},\delta} \leftarrow \alpha_\delta \boldsymbol{x}_{\mathrm{in}} + \sigma_\delta \boldsymbol{\epsilon}$
if inference mode is full-mask **then**
  Set $K \leftarrow 1$ and $\boldsymbol{M}_{\mathrm{ctx}}^{(1)} \leftarrow \boldsymbol{M}_{\mathrm{in}}$
else
  Use the first K masks from the cached context sequence sampled from the context marginal of $\mathcal{P}_{\mathrm{align}}(\cdot \mid \boldsymbol{M}_{\mathrm{in}})$
end if
for $k = 1, \ldots, K$ **do**
  $\boldsymbol{x}_{\mathrm{ctx},\delta}^{(k)} \leftarrow \boldsymbol{M}_{\mathrm{ctx}}^{(k)} \odot \boldsymbol{x}_{\mathrm{in},\delta}$
  $(\mathbf{\Pi}_\delta^{(k)}, I_\delta^{(k)}) \leftarrow \textsc{BuildReference}(\delta, \boldsymbol{x}_{\mathrm{ctx},\delta}^{(k)}, \boldsymbol{M}_{\mathrm{ctx}}^{(k)})$
  $\widehat{\boldsymbol{x}}_{0,\delta}^{(k)} \leftarrow \mathbf{\Pi}_\delta^{(k)} + r_{\boldsymbol{\psi}}(\delta; I_\delta^{(k)})$
end for
$\widehat{\boldsymbol{x}}_{\mathrm{out}} \leftarrow \boldsymbol{x}_{\mathrm{in}} + (\boldsymbol{V} - \boldsymbol{M}_{\mathrm{in}}) \odot K^{-1} \sum_{k=1}^{K} \widehat{\boldsymbol{x}}_{0,\delta}^{(k)}$
Shared context-only reference construction
function $\textsc{BuildReference}(t, \boldsymbol{x}_{\mathrm{ctx},t}, \boldsymbol{M}_{\mathrm{ctx}})$
  Evaluate the frozen local predictor on every $p \in \mathcal{P}$ using Eq. (5)
  Fuse the patch predictions into $\mathbf{\Pi}_t$ by Eq. (7); set $\mathbf{\Pi}_t = 0$ outside $\boldsymbol{V}$
  Compute $\boldsymbol{S}_{\mathrm{ctx}}, \boldsymbol{S}_\partial$ from this same noisy context; form I_t by Eq. (8)
  return $(\mathbf{\Pi}_t, I_t)$
end function

C Observation-Aligned Context–Query Construction

This appendix specifies training context–query construction and the subcontext rule used for context-ensemble reconstruction. Ocean validation/test input and scoring supports are constructed separately through the empirical real-mask overlays in Appendix E.2.

C.1 Partition Requirements

The estimator forms share acquisition-aligned context–query construction: sitewise partitions for pixel observations, whole-block partitions for block observations, and mask-only proposals for authentic ocean observations (Zhou et al., 2026; Ma et al., 2026). Ocean proposals need not lie within $\boldsymbol{M}$; the intersection in Equation (1) restricts context and query to available measurements. The dataset-specific constructors are specified below. Marginal query frequencies describe the resulting supervision, while Appendix D.2 defines the conditional target on states with positive query exposure.

C.2 Controlled Pixel and Block Partitions

At each training visit, candidate context–query partitions use independent retention proposals within the stored outer support:

$$\begin{aligned} B_i^{(n,k)} &\sim \text{Bernoulli}(0.7), \\ \boldsymbol{M}_{\text{ctx}}^{(n,k)} &= \boldsymbol{M}^{(n)} \odot B^{(n,k)}, \qquad \boldsymbol{M}_{\text{qry}}^{(n,k)} = \boldsymbol{M}^{(n)} \odot (1 - B^{(n,k)}). \end{aligned} \tag{24}$$

Before rejecting empty queries, the expected context and query fractions conditional on $\boldsymbol{M}$ are $0.7\rho(\boldsymbol{M})$ and $0.3\rho(\boldsymbol{M})$, respectively, giving nominal fractions of approximately 21% and 9%. Empty queries are handled by resampling the inner partition. Context-ensemble inference reuses this inner context-retention rule.

For Block 3/9, two of the three available blocks form the context and the remaining block forms the query. For Block 8/9, seven of the eight available blocks form the context and the remaining block forms the query.

C.3 Authentic-Mask Constructors

Guided BFN partitions for Black Sea and Baltic. Training context–query partitions for Black Sea and Baltic use dataset-specific mask-only Bayesian Flow Networks (Graves et al., 2025) with the observation-aligned guided sampler of OAMP (Ma et al., 2026). Each BFN is trained only on masks from the corresponding training split, selected by validation-mask prediction accuracy, and frozen before physical-field training. The current observation mask is intersected with the guided sample to produce context and query supports.

We expand the inherited OAMP sampling procedure here to distinguish its mask generation state from the physical-field diffusion state. Let $\boldsymbol{z}_\tau$ denote the continuous mask-sampling logits at time τ, and let $p_{\text{BFN},i}(\tau, \boldsymbol{z}_\tau)$ denote the frozen BFN's predicted probability of the observed class at spatial location i (with the softmax mapping of the logits included in this notation). For each proposal, draw independent $b_i \sim \text{Bernoulli}(\rho)$ and set $y_i = M_i b_i$. Following Ma et al. (2026, Section 3.4), the guidance objective is

$$\mathcal{L}_{\text{guide}}(\boldsymbol{z}_\tau, \boldsymbol{y}) = -\frac{1}{|\Omega|} \sum_{i \in \Omega} \left[y_i \log p_{\text{BFN},i}(\tau, \boldsymbol{z}_\tau) + (1 - y_i) \log(1 - p_{\text{BFN},i}(\tau, \boldsymbol{z}_\tau)) \right]. \tag{25}$$

The normalization is over the entire spatial grid, not only the retained anchor sites. Given the base reverse-sampling update $\boldsymbol{z}_{\tau_{k+1}}^{\text{base}}$, apply

$$\boldsymbol{z}_{\tau_{k+1}} = \boldsymbol{z}_{\tau_{k+1}}^{\text{base}} - w_g \nabla_{\boldsymbol{z}_{\tau_k}} \mathcal{L}_{\text{guide}}(\boldsymbol{z}_{\tau_k}, \boldsymbol{y}). \tag{26}$$

Here w_g controls guidance strength; gradients update the sampling state while the BFN remains fixed. The anchor remains fixed within a sampling trajectory. Final categorical decoding produces a binary proposal, which is then intersected with $\boldsymbol{M}$ as in Equation (1). Neither anchor retention nor soft guidance fixes the final intersection size.

The retry loop requires both context and query to be nonempty; the subsequent minimum-size check also applies after an early exit. Anchor retention does not set the final context-retention fraction, which is determined by categorical decoding and intersection with the available support. The run-level random-number state advances across proposals without reseeding. The mask sampler's reverse steps are distinct from the near-zero-time, single-step field prediction used for each evaluated context. Table 5 records the sampling configuration.

Empirical authentic-track partitions for Global Ocean SSS. Global Ocean SSS draws a proposal from the empirical authentic-track mask bank and intersects it with the current available support as in Equation (1), producing the context and the remaining query support.

D Theoretical Analysis and Proofs

We first establish a distributional limitation of never-complete observations, then characterize the query-supervised target, the support of the local consensus prior, and the role of full-domain residual revision.

Algorithm 3 Guided BFN context–query partition for Black Sea CHL and Baltic Sea NANO

Require: current mask $\boldsymbol{M}$, frozen BFN, $\rho = 0.7$, $w_g = 80$, $S = 20$; uniform time grid $\tau_0 = 10^{-3} < \cdots < \tau_S = 1$; run-level random-number state initialized with seed 42

Set $m_{\mathrm{ctx}} = m_{\mathrm{qry}} = \max(16, \lceil 0.05\|\boldsymbol{M}\|_1 \rceil)$
for $r = 0, \ldots, 8$ **do** ▷ Initial proposal and at most eight additional retries
  Draw $b_i \sim \mathrm{Bernoulli}(\rho)$ and set $y_i \leftarrow M_i b_i$
  Initialize mask logits $\boldsymbol{z}_{\tau_0} \sim \mathcal{N}(0, I)$
  for $k = 0, \ldots, S-1$ **do**
    Set $w_g^{(k)} \leftarrow 0.4 w_g$ if $k < 0.1S$, and $w_g^{(k)} \leftarrow w_g$ otherwise
    Compute the base BFN ODE update $\boldsymbol{z}_{\tau_{k+1}}^{\mathrm{base}}$
    Evaluate $\mathcal{L}_{\mathrm{guide}}(\boldsymbol{z}_{\tau_k}, \boldsymbol{y})$ by Equation (25)
    Update $\boldsymbol{z}_{\tau_{k+1}}$ by Equation (26) using $w_g^{(k)}$
  end for
  Evaluate the frozen BFN once more at τ_S and decode its categorical argmax to obtain $\widetilde{\boldsymbol{M}}$
  $\boldsymbol{M}_{\mathrm{ctx}} \leftarrow \boldsymbol{M} \odot \widetilde{\boldsymbol{M}}$
  $\boldsymbol{M}_{\mathrm{qry}} \leftarrow \boldsymbol{M} \odot (1 - \widetilde{\boldsymbol{M}})$
  if $\|\boldsymbol{M}_{\mathrm{ctx}}\|_1 > 0 \ \wedge \ \|\boldsymbol{M}_{\mathrm{qry}}\|_1 > 0$ **then**
    break
  end if
end for
if $\|\boldsymbol{M}_{\mathrm{ctx}}\|_1 < m_{\mathrm{ctx}}$ or $\|\boldsymbol{M}_{\mathrm{qry}}\|_1 < m_{\mathrm{qry}}$ **then**
  Reassign pixels between context and query within $\boldsymbol{M}$ to meet both minimum sizes
end if
if $\|\boldsymbol{M}_{\mathrm{ctx}}\|_1 < m_{\mathrm{ctx}}$ or $\|\boldsymbol{M}_{\mathrm{qry}}\|_1 < m_{\mathrm{qry}}$ **then**
  Raise an error
end if
return $(\boldsymbol{M}_{\mathrm{ctx}}, \boldsymbol{M}_{\mathrm{qry}})$

Table 5: Guided BFN sampling configuration for Black Sea CHL and Baltic Sea NANO.

Setting	Value
Anchor retention ρ	**0.7**
Guidance scale w_g	**80**; **32** for the first two updates
Sampling steps S	**20**
Reverse solver / schedule	Guided BFN ODE; uniform time grid $t : 10^{-3} \to 1$
Sampling seed	**42**; run-level random-number state
Empty-support handling	Initial proposal plus at most 8 retries; post-loop minimum-size repair

D.1 Distributional Non-Identifiability

This subsection states and proves the distributional limitation discussed in Section 3. The result holds even under a fixed, known mask law independent of the field. This independence assumption applies only to the theorem; ocean experiments retain authentic acquisition masks. Let $\boldsymbol{x}_0 \in \mathbb{R}^d$ denote a field vectorized over its valid domain.

Theorem D.1 (Non-identifiability under never-complete masking)**.** *Suppose that $\boldsymbol{M}$ has a fixed, known distribution independent of $\boldsymbol{x}_0$, and that $\Pr(\boldsymbol{M} = \mathbf{1}_d) = 0$. Over the unrestricted class of probability laws on $\mathbb{R}^d$, the complete-field distribution is not identifiable from $(\boldsymbol{M} \odot \boldsymbol{x}_0, \boldsymbol{M})$. Consequently, no algorithm can consistently recover that distribution for every law in this class.*

The proof constructs two distinct complete-field laws whose marginals agree on every proper subset of coordinates, making their incomplete observations indistinguishable.

Observation operator and identifiability. Write $\boldsymbol{Y} = \boldsymbol{M} \odot \boldsymbol{x}_0$ and let q be the fixed, known mask law, independent of $\boldsymbol{x}_0$. The observation operator is

$$\mathcal{T}_q(P_{\boldsymbol{x}_0}) = \mathrm{Law}_{P_{\boldsymbol{x}_0}, q}(\boldsymbol{M} \odot \boldsymbol{x}_0, \boldsymbol{M}). \tag{27}$$

Complete-law identifiability over a class $\mathcal{C}$ means

$$\mathcal{T}_q(P_1) = \mathcal{T}_q(P_2) \quad \Longrightarrow \quad P_1 = P_2 \qquad \text{for all } P_1, P_2 \in \mathcal{C}. \tag{28}$$

Parity construction and proper-subset marginals. Let $\boldsymbol{x}_{0,i} \in \{-1, +1\}$ and define the two parity supports

$$\mathcal{X}_+ = \left\{ \boldsymbol{x} : \prod_{i=1}^{d} \boldsymbol{x}_i = +1 \right\}, \qquad \mathcal{X}_- = \left\{ \boldsymbol{x} : \prod_{i=1}^{d} \boldsymbol{x}_i = -1 \right\}. \tag{29}$$

Let P_+ and P_- be the uniform distributions on $\mathcal{X}_+$ and $\mathcal{X}_-$, respectively. The supports are disjoint, so $P_+ \neq P_-$, and each support contains 2^{d-1} points.

Fix any proper subset $S \subsetneq [d]$ and any assignment $\boldsymbol{x}_S$. At least one coordinate remains unassigned. Under either required parity, choose all but one of the remaining signs freely; the final sign is then uniquely determined by the parity constraint. Hence $\boldsymbol{x}_S$ has exactly $2^{d-|S|-1}$ valid completions under each parity law. Dividing by the support size gives

$$P_+(\boldsymbol{x}_{0,S} = \boldsymbol{x}_S) = P_-(\boldsymbol{x}_{0,S} = \boldsymbol{x}_S) = \frac{2^{d-|S|-1}}{2^{d-1}} = 2^{-|S|}. \tag{30}$$

Therefore every proper-subset marginal is uniform and identical under P_+ and P_-, even though the complete laws have opposite d-way parity.

Equality of the observed laws. For a mask value m with $q(m) > 0$, let $S_m = \{i : m_i = 1\}$. The assumption $\Pr_q(\boldsymbol{M} = \mathbf{1}_d) = 0$ implies $S_m \subsetneq [d]$. Conditional on $\boldsymbol{M} = m$, the nonzero coordinates of $\boldsymbol{Y} = \boldsymbol{M} \odot \boldsymbol{x}_0$ have the $\boldsymbol{x}_{0,S_m}$ marginal, while all remaining coordinates are zero. Equation (30) therefore yields

$$\mathrm{Law}_{P_+}(\boldsymbol{Y} \mid \boldsymbol{M} = m) = \mathrm{Law}_{P_-}(\boldsymbol{Y} \mid \boldsymbol{M} = m). \tag{31}$$

Both worlds use the same mask law q, and $\boldsymbol{M}$ is independent of $\boldsymbol{x}_0$ in each. Thus, for every measurable set A and every m,

$$\begin{aligned} \Pr_{P_+,q}(\boldsymbol{Y} \in A, \boldsymbol{M} = m) &= q(m) \Pr_{P_+}(\boldsymbol{Y} \in A \mid \boldsymbol{M} = m) \\ &= q(m) \Pr_{P_-}(\boldsymbol{Y} \in A \mid \boldsymbol{M} = m) = \Pr_{P_-,q}(\boldsymbol{Y} \in A, \boldsymbol{M} = m). \end{aligned} \tag{32}$$

Consequently, $\mathcal{T}_q(P_+) = \mathcal{T}_q(P_-)$ while $P_+ \neq P_-$, proving the non-identifiability claim.

Consequence for universal recovery algorithms. Suppose an algorithm estimates the complete-field law by $\widehat{P}_n$ from n i.i.d. observed pairs. Its inputs have identical distributions under P_+ and P_- for every n. Coupling any internal randomization identically gives the same output distribution in both cases. Thus $\widehat{P}_n$ cannot converge to P_+ under one law and the distinct P_- under the other, ruling out universal consistency.

Two-dimensional example and missing-coordinate prediction. The ambiguity already appears when $d = 2$. Let $\boldsymbol{M} = (1, 0)$ or $(0, 1)$, each with probability $1/2$ independently of $\boldsymbol{x}_0$, and define

$$P_1(\boldsymbol{x}_0 = (1, 1)) = P_1(\boldsymbol{x}_0 = (-1, -1)) = \tfrac{1}{2}, \tag{33}$$

$$P_2(\boldsymbol{x}_0 = (1, -1)) = P_2(\boldsymbol{x}_0 = (-1, 1)) = \tfrac{1}{2}. \tag{34}$$

The laws have opposite dependence: $\boldsymbol{x}_{0,1}\boldsymbol{x}_{0,2} = +1$ almost surely under P_1 and -1 under P_2. Yet each coordinate is uniform on $\{-1, +1\}$ in both laws. Since each mask reveals only one coordinate and has the same probability in both cases, $\mathcal{T}_q(P_1) = \mathcal{T}_q(P_2)$.

When only $\boldsymbol{x}_{0,1}$ is observed, the two observationally indistinguishable laws imply different missing-coordinate conditional means: $\mathbb{E}_{P_1}[\boldsymbol{x}_{0,2} \mid \boldsymbol{x}_{0,1}] = \boldsymbol{x}_{0,1}$ and $\mathbb{E}_{P_2}[\boldsymbol{x}_{0,2} \mid \boldsymbol{x}_{0,1}] = -\boldsymbol{x}_{0,1}$. This ambiguity concerns prediction without joint exposure of $\boldsymbol{x}_{0,1}$ and $\boldsymbol{x}_{0,2}$, whereas the supervised conditional mean in Appendix D.2 is defined on states with positive query exposure.

D.2 Query-Supervised Conditional Targets

The no-prior direct, prior-conditioned direct, and prior-conditioned residual estimators in Section 4.3 share the observation-aligned query support and pooled masked squared-error objective. Their output forms determine whether the trainable network predicts the total field or its discrepancy from the frozen prior. The partition constructor is fixed across these comparisons, and the unrestricted target on positive-exposure states is the observed-query conditional mean defined below.

For a predictor receiving (t, O_t), with $O_t = (\boldsymbol{x}_{\mathrm{ctx},t}, \boldsymbol{M}_{\mathrm{ctx}})$, the unrestricted squared-loss target is

$$\mu_{O,i}(t, O_t) = \mathbb{E}_{\mathrm{train}}[\boldsymbol{x}_{0,i} \mid t, O_t, \boldsymbol{M}_{\mathrm{qry},i} = 1], \tag{35}$$

on states with $\Pr(\boldsymbol{M}_{\mathrm{qry},i} = 1 \mid t, O_t) > 0$, assuming finite relevant second moments. The query-selection event remains in this expression: the empirical authentic-mask setting is not assumed to satisfy additional conditions that would allow dropping it. Here $\boldsymbol{x}_{0,i}$ denotes the clean field value (or channel vector) at coordinate i. When query exposure is zero, the loss does not constrain the prediction at that state and coordinate.

D.3 Local Support of the Consensus Prior

Fix the trained predictors, patch covering, coordinate atlas, validity mask, and preprocessing operators. Although consensus assembles a full-sized field, the context used by each output is determined by the supports of its contributing patches. For a block B, the union of contributing patch supports is

$$U(B) = \bigcup_{p:\Omega_p \cap B \neq \varnothing} \Omega_p. \tag{36}$$

For a valid coordinate i, write $U_i = U(\{i\})$. Each contributing prediction depends only on the context inside its patch, so their uniform average $\boldsymbol{\Pi}_{t,i}$ depends only on the context in U_i. Consequently, two states at the same diffusion time that agree on $U(B)$ produce identical reference fields on B. Overlap reconciles the local predictions but leaves observations beyond their combined support unused for this prediction. The full-domain corrector can use those observations to revise the estimate and can also correct errors in the local predictions themselves. This gives the support boundary for the paired remote-observation interventions in Section 4.5 and Appendix F.4; in particular, deleting supplied observations outside $U(B)$ leaves $\boldsymbol{\Pi}_{\delta,B}$ unchanged.

D.4 Residual Revision and Representation

Let $\mathcal{L}_i = \sigma(t, R_{U_i}\boldsymbol{x}_{\mathrm{ctx},t}, R_{U_i}\boldsymbol{M}_{\mathrm{ctx}})$ and $\mathcal{G} = \sigma(t, \boldsymbol{x}_{\mathrm{ctx},t}, \boldsymbol{M}_{\mathrm{ctx}})$ denote local and full-domain information. For a coordinate with positive query exposure, assume finite conditional second moments and define

$$\mu_i^L = \mathbb{E}[\boldsymbol{x}_{0,i} \mid \mathcal{L}_i, \boldsymbol{M}_{\mathrm{qry},i} = 1], \qquad \mu_i^G = \mathbb{E}[\boldsymbol{x}_{0,i} \mid \mathcal{G}, \boldsymbol{M}_{\mathrm{qry},i} = 1]. \tag{37}$$

For a fixed square-integrable reference $\boldsymbol{\Pi}_{t,i}$, the conditional mean minimizes squared error under the measure conditioned on $\boldsymbol{M}_{\mathrm{qry},i} = 1$. Adding and subtracting μ_i^L from the optimal unrestricted full-domain correction gives

$$r_i^\star = \mu_i^G - \boldsymbol{\Pi}_{t,i} = \underbrace{\mu_i^G - \mu_i^L}_{\text{adjustment beyond local support}} + \underbrace{\mu_i^L - \boldsymbol{\Pi}_{t,i}}_{\text{remaining local-estimation error}}. \tag{38}$$

The first term accounts for evidence unavailable to the local construction; the second accounts for error that can remain even within its support. Thus the prior need not be an optimal local predictor to serve as a useful reference for correction. The query-selection condition is retained because authentic acquisition and partition rules need not expose all prediction states in the same way. This identity explains that revision can address both missing nonlocal evidence and local-estimation error; it does not guarantee that a trained finite corrector attains this target or outperforms direct prediction.

Scope of the representation. With the trained constructor fixed, the prior and discrepancy maps are deterministic functions of (t, O_t), and O_t is retained in I_t. Thus $\sigma(t, I_t) = \sigma(t, O_t) = \mathcal{G}$: unrestricted direct and residual predictors have the same optimal total-field prediction and query risk. The empirical comparisons in Sections 4.3 and 4.4 test how finite models learn with this representation and initialization, while the remote interventions diagnose responses to observations beyond local support.

E EXPERIMENTAL SETUP AND EVALUATION

This appendix specifies the datasets, preprocessing, observation and scoring protocols, evaluation metrics, and estimator controls used in Section 4.

E.1 DATASETS AND PREPROCESSING

Ocean Products. We use daily Level-3 products from the E.U. Copernicus Marine Environment Monitoring Service (CMEMS) (Letraon et al., 2017) and derive observation supports from their quality-controlled source masks. Black Sea CHL and Baltic Sea NANO are multi-sensor ocean-colour observations at approximately 1-km native spacing; Global Ocean SSS uses SMOS salinity observations on the global cylindrical EASE-Grid 2.0, with approximately 25-km spacing at 30° latitude. Longitude increments are approximately 0.259°, while latitude increments vary with latitude. The reconstructed variables are chlorophyll-a, nanophytoplankton expressed as chlorophyll (both in $\mathrm{mg\,m^{-3}}$), and practical sea-surface salinity, respectively. The regional products contain cloud-related gaps, while SSS exhibits orbital-track and inter-swath gaps. Table 6 identifies the sources and requested geographic extents; Black Sea CHL uses the full regional product domain, and Global Ocean SSS uses the ascending-orbit dataset. The records span approximately 26 years for Black Sea CHL, 28 years for Baltic Sea NANO, and 16 years for Global Ocean SSS.

This study has been conducted using E.U. Copernicus Marine Service Information; `10.48670/moi-00303`, `10.48670/moi-00296`, and `10.48670/mds-00368`.

Table 6: Ocean data products and geographic subsets.

	Black Sea CHL	Baltic Sea NANO	Global Ocean SSS
Product ID	`OCEANCOLOUR_BLK_BGC_L3_MY_009_153`	`OCEANCOLOUR_BAL_BGC_L3_MY_009_133`	`MULTIOBS_GLO_PHY_SSS_L3_MYNRT_015_014`
DOI	`10.48670/moi-00303`	`10.48670/moi-00296`	`10.48670/mds-00368`
Variable	`CHL`	`NANO`	`Sea_Surface_Salinity`
Spatial subset	Full Black Sea domain; approximately 26.5°–42°E, 40°–48°N	15°–25°E, 54°–59°N	180°W–179.8°E, 83.6°S–83.6°N
Native grid	Regional grid; approximately 1-km spacing	Regional grid; approximately 1-km spacing	Global cylindrical EASE-Grid 2.0

After source-level quality control and before empty-field or sparse-field filtering, the mean observed-ocean fractions over all dates, including empty fields, are 38.30%, 26.50%, and 30.60% for Black Sea CHL, Baltic Sea NANO, and Global Ocean SSS, respectively. For N fields, we compute the per-field average

$$\bar{r} = \frac{1}{N} \sum_{n=1}^{N} \frac{|\boldsymbol{M}_n \cap \boldsymbol{V}_n|}{|\boldsymbol{V}_n|},$$

where $\boldsymbol{M}_n$ is the quality-controlled observation support and $\boldsymbol{V}_n$ is the empirical ocean support inferred from locations with valid data at least once in the temporal record. This support is defined by data availability rather than an independent coastline map. Each date receives equal weight, and the fractions are computed before context–query partitioning.

Ocean samples are assigned chronologically to 70/15/15 training, validation, and test splits before resolution-specific filtering. We remove empty samples and retain fields with at least 128 observed pixels at 64×64 or 512 at 128×128. Fields and masks are resampled jointly using value-weighted area resampling, and every retained ocean field remains incomplete on its valid domain. SSS retains source values with `Sea_Surface_Salinity_QC` equal to zero. Black Sea CHL and Baltic Sea NANO use $\log(\boldsymbol{x} + 10^{-6})$, while SSS uses the original salinity. At each resolution, normalization statistics are fitted to the available training observations and held fixed for validation, testing, and patch evaluation. Values withheld by the evaluation protocol or unavailable under the observation masks are outside this fitting support. Normalized values are clipped to three standard deviations and divided by three, and invalid locations are set to zero.

Controlled PDE systems. Shallow Water uses water-height fields from the PDEBench two-dimensional radial dam-break benchmark (Takamoto et al., 2022). Navier–Stokes uses the released 64×64 vorticity fields of forced two-dimensional Kolmogorov turbulence at $\mathrm{Re} = 1000$ (Cao, 2024; Cao et al., 2025), specifically the file `Kolmogorov2d_fp32_64x64_N1152_Re1000_T100.pt`.

For Shallow Water, we retain one representative from each of the 81 distinct rasterized initial-state classes in the 1,000-trajectory source and assign 57/12/12 trajectories to training/validation/test splits. Navier–Stokes uses all 1,152 released trajectories with an 806/173/173 split. Trajectory selection and splitting use seed 42. Each trajectory is assigned to a single split before all stored frames are treated as individual spatial reconstruction samples, with no temporal windows or further temporal subsampling. Shallow Water contributes 101 frames per trajectory and uses height channel 0, bilinearly resized from 128×128 to 64×64 with antialiasing; Navier–Stokes contributes 100 frames per trajectory at its released 64×64 resolution. PDE normalization and clipping statistics are fitted to observed values within the stored training acquisition masks and held fixed for subsequent patch and full-domain evaluations.

E.2 Observation and Scoring Protocols

Ocean authentic-mask protocol. Training supplies every method with the source-supported observations of each sample. Our context–query training follows Appendix C.3: Black Sea CHL and Baltic Sea NANO use frozen guided BFN constructors, whereas Global Ocean SSS uses an empirical authentic-track mask bank. Ocean validation and testing instead follow OAMP's real-mask overlay protocol (Ma et al., 2026), drawing authentic observation masks directly from their respective splits rather than using randomly placed blocks.

For a fixed ocean product at resolution $L \times L$, let $\mathcal{I}_{s,L}$ index the samples retained after the resolution-specific filtering in Appendix E.1, where $s \in \{\mathrm{val}, \mathrm{test}\}$. For the current sample n, each proposal $\widetilde{\boldsymbol{M}}^{(r)}$ is drawn from the observation masks indexed by $\mathcal{I}_{s,L} \setminus \{n\}$. Thus validation and test proposals come from their respective splits, and the current sample's own mask is explicitly excluded. Suppressing the current sample index, the candidate supports are

$$\boldsymbol{M}_{\mathrm{in}}^{(r)} = \boldsymbol{M}_{\mathrm{raw}} \odot \widetilde{\boldsymbol{M}}^{(r)}, \qquad \boldsymbol{M}_{\mathrm{eval}}^{(r)} = \boldsymbol{M}_{\mathrm{raw}} \odot \big(1 - \widetilde{\boldsymbol{M}}^{(r)}\big). \tag{39}$$

The donor sample contributes only its observation mask; all input and scoring values come from the current sample.

A proposal is rejected if either its input or scoring support is empty. Starting from the initial proposal $r = 0$, we allow at most eight additional retries, for at most nine proposals in total, and stop the retry loop as soon as both supports are nonempty.

After the retry loop, including early termination, we apply the same minimum-size check and pixel-reassignment procedure as in Algorithm 3, with

$$m_{\mathrm{in}} = m_{\mathrm{eval}} = \max(16,\ \lceil 0.05 \|\boldsymbol{M}_{\mathrm{raw}}\|_1 \rceil)\,.$$

If either support is smaller than its minimum, observed pixels are reassigned between the input and scoring supports within $\boldsymbol{M}_{\mathrm{raw}}$. The finalized pair is checked against both minima before caching and remains a disjoint partition of the current sample's source-supported observations.

For each evaluation sample, the finalized pair $(\boldsymbol{M}_{\mathrm{in}}, \boldsymbol{M}_{\mathrm{eval}})$ is cached and reused across all compared methods and internal controls. Each method receives $\boldsymbol{x}_{\mathrm{in}} = \boldsymbol{M}_{\mathrm{in}} \odot \boldsymbol{x}_{\mathrm{obs}}$ and $\boldsymbol{M}_{\mathrm{in}}$, where $\boldsymbol{x}_{\mathrm{obs}} = \boldsymbol{M}_{\mathrm{raw}} \odot \boldsymbol{x}_0$. Scoring uses $\boldsymbol{M}_{\mathrm{eval}} = \boldsymbol{M}_{\mathrm{raw}} \odot (1 - \boldsymbol{M}_{\mathrm{in}})$. This protocol is used for all reported ocean evaluations at both resolutions. In the default ocean reconstruction protocol, our estimator and internal controls use the entire cached $\boldsymbol{M}_{\mathrm{in}}$, without a further internal context split.

Controlled Pixel protocol. Each stored pixel-acquisition mask is constructed from Bernoulli proposals with mask-only coverage control, excludes complete observation, and retains approximately 30% of valid coordinates.

$$\boldsymbol{M}^{(n)} \sim q_{0.3}^{\mathrm{pix}}, \qquad q_{0.3}^{\mathrm{pix}}(\boldsymbol{M} = V) = 0, \qquad \rho(\boldsymbol{M}^{(n)}) = \frac{\|\boldsymbol{M}^{(n)}\|_1}{\|V\|_1} \approx 0.3. \tag{40}$$

Training partitions are sampled within each fixed acquisition support. Validation and test supports are independently constructed from the same acquisition law and then fixed for evaluation. Internal partitions follow Appendix C.2.

Controlled Block protocols. Block 3/9 and Block 8/9 retain three and eight blocks of a regular 3×3 partition, respectively. Grid lines in the block-protocol visualizations mark this acquisition partition, whereas the local predictor uses the overlapping covering specified in Appendix B.2. Their internal context–query partitions preserve whole-block units (Appendix C.2). For both Pixel and Block protocols, evaluation scores all valid complete-reference coordinates outside $\boldsymbol{M}_{\mathrm{in}}$, while our estimator and internal controls use the default $K = 8$ context ensemble specified in Appendix B.4. Table 7 summarizes the observation and scoring protocols. Appendix C.3 specifies the training context–query constructors; the ocean validation/test overlays are defined above.

Ocean evaluation scores source-supported values withheld from the estimator, whereas controlled PDE evaluation scores valid unsupplied coordinates using complete references. Reconstructions at authentically source-missing ocean locations lack independent quantitative references.

E.3 Evaluation Metrics and Seed Aggregation

Training objectives and reconstruction comparisons use masked squared errors, with evaluation MSE pooled over the scored spatial-channel values. For a finite collection $\mathcal{D}$ with estimator states $\mathcal{S}^{(b)}$, target fields $\boldsymbol{x}_0^{(b)}$, and nonempty scoring masks $A^{(b)}$, define

$$\mathrm{MSE}_A(f;\mathcal{D}) = \frac{\sum_{b\in\mathcal{D}} \left\| A^{(b)} \odot \left(f(\mathcal{S}^{(b)}) - \boldsymbol{x}_0^{(b)} \right) \right\|_2^2}{d_x \sum_{b\in\mathcal{D}} \left\| A^{(b)} \right\|_1}. \tag{41}$$

Evaluation uses $A = \boldsymbol{M}_{\mathrm{eval}}$, and pooling over the complete collection gives equal weight to every scored spatial-channel value. For our estimator and internal controls, the scored prediction is the final reconstruction under the dataset-specific inference protocol in Section 3.3 and Appendix B.4.

Table 7: Available training observations, training context–query partitions, validation/test inputs, and scoring supports by dataset and protocol.

Setting	Available training support	Training context rule	Training query	Validation/test input	Scoring support
Black Sea CHL	All source-supported values	Guided-mask intersection within the current support	Remaining available values	Cached support from a same-split authentic-mask overlay	Withheld source-supported values, $\boldsymbol{M}_{\mathrm{raw}} \odot (1 - \boldsymbol{M}_{\mathrm{in}})$
Baltic Sea NANO	All source-supported values	Guided-mask intersection within the current support	Remaining available values	Cached support from a same-split authentic-mask overlay	Withheld source-supported values, $\boldsymbol{M}_{\mathrm{raw}} \odot (1 - \boldsymbol{M}_{\mathrm{in}})$
Global Ocean SSS	All source-supported values	Empirical-track intersection within the current support	Remaining available values	Cached support from a same-split authentic-mask overlay	Withheld source-supported values, $\boldsymbol{M}_{\mathrm{raw}} \odot (1 - \boldsymbol{M}_{\mathrm{in}})$
Pixel 30%	Fixed $q_{0.3}^{\mathrm{pix}}$ mask; approximately 30%	Retention probability 0.7; approximately 21% of the field	Approximately 9% of the field	Fixed $q_{0.3}^{\mathrm{pix}}$ mask; approximately 30%	$V \setminus \boldsymbol{M}_{\mathrm{in}}$
Block $3/9$	Three fixed blocks	Two available blocks	One available block	Three blocks	Remaining six blocks
Block $8/9$	Eight fixed blocks	Seven available blocks	One available block	Eight blocks	Remaining one block

PSNR is computed for each sample and then averaged over the evaluation collection. For sample n, define

$$e_n = \frac{\left\| A^{(n)} \odot \left(\widehat{\boldsymbol{x}}^{(n)} - \boldsymbol{x}_0^{(n)} \right) \right\|_2^2}{d_x \left\| A^{(n)} \right\|_1}.$$

The reported PSNR is

$$\mathrm{PSNR} = \frac{1}{N} \sum_{n=1}^{N} 10 \log_{10} \left(\frac{R_n^2}{e_n} \right), \qquad R_n = \begin{cases} R_n^{\mathrm{ocean}}, & \text{ocean data,} \\ 2, & \text{PDE data.} \end{cases} \tag{42}$$

Here R_n^{ocean} is the sample-specific dynamic range of the ocean reference in the normalized, clipped evaluation space. Lower MSE and higher PSNR indicate better reconstruction within a setting. The two metrics use the same scoring support but differ in aggregation, so PSNR is not a direct transformation of pooled MSE.

Our estimator and internal controls share supplied observations and perturbations within each comparison; context-ensemble comparisons additionally reuse paired subcontext masks.

The learned-model entries in Tables 1, 2, and 8 report means and standard deviations of the evaluation metrics across models independently trained with different random seeds. The intervention diagnostics in Appendix F.4 compute paired measurements on fixed scored blocks of the final full-mask reconstructions and summarize them over evaluation items and training-seed checkpoints.

E.4 Estimator Controls and Prior Replacement

Table 2 compares the Local Prior reference with Full Direct, Prior Direct, and Prior Residual, and includes two prior-replacement controls. On the ocean products, Full Direct is a parameter-matched version of OAMP. The Navier–Stokes Full Direct control in Figure 4 and the Shallow Water Full Direct control in Appendix F.5 are defined analogously using IDCD. In each case, the total parameter count matches the frozen local predictor plus the trainable full-domain network in our two-stage estimator. Full Direct receives $(\boldsymbol{x}_{\mathrm{ctx},t}, \boldsymbol{M}_{\mathrm{ctx}}, \boldsymbol{\Gamma})$ and uses the shared observation-aligned query supervision. The OAMP and IDCD entries in the external-baseline tables are distinct from their parameter-matched Full Direct controls.

For the ocean comparisons and the Shallow Water Pixel 30% comparison in Appendix F.5, Full Direct's training time is comparable to the combined training time of our two stages. The second-stage time includes frozen-prior construction. Shared mask-model pretraining, where applicable, is excluded from both sides of this comparison.

Prior Direct and Prior Residual share the frozen local predictor, full-domain backbone, input-construction procedure, and checkpoint-selection procedure. Appendix B.3 specifies their output forms and initialization; Section 4.3 interprets the comparison.

Prior-replacement controls. Global2Global uses a full-domain OAMP prediction as the reference, and Low-rank2Global uses a DINEOF reconstruction; their standalone results are reported in Table 1. We retain the same input channels and prior–corrector connection design, recompute $\boldsymbol{S}_{\mathrm{ctx}}$ and $\boldsymbol{S}_{\partial}$ from each replacement prior, and independently retrain the corresponding full-domain residual corrector. Global2Global tests a learned two-stage cascade, while Low-rank2Global replaces dedicated local-prior learning with a low-rank reconstruction.

All rows in Table 2 use full-mask reconstruction under Appendix B.4 and the shared-input and metric conventions in Appendix E.3.

Coordinate and input-feature controls. Table 12 in Appendix F.3 reports the final-reconstruction results. Its two panels answer different questions. In Panel A, we remove the absolute-coordinate crop from the local predictor, learn a new local model under the same patch-level query objective, freeze it, and learn a new corrector. This tests whether position information in the shared local learning task helps the complete estimator. In Panel B, the patch predictor and consensus construction remain fixed. We remove one or both deterministic input maps and learn a new residual corrector for each variant, preserving the retained context and prior itself. All variants use full-mask reconstruction under the pairing conventions in Appendix E.3.

F SUPPLEMENTARY EXPERIMENTS

This appendix reports controlled PDE reconstruction, higher-resolution ocean results, prior-construction and corrector-input ablations, remote-evidence interventions, and inference-cost and observation-retention analyses.

F.1 CONTROLLED PDE RECONSTRUCTION

These experiments complement the authentic ocean evaluation with controlled pixel- and block-observation settings and complete references. Data preparation, observation protocols, and evaluation metrics are specified in Appendix E. We compare with AmbientDiff, DINDiff, MissDiff, MSM, and IDCD (Zhou et al., 2026), using shared training observations, evaluation items, supplied supports, and scoring coordinates.

Primary pixel and block settings. Our estimator has the lowest MSE and highest PSNR among the compared methods in Shallow Water Pixel 30%, Shallow Water Block 3/9, and Navier–Stokes Pixel 30% (Table 8).

Table 8: PDE reconstruction at 64×64, evaluated on valid unsupplied coordinates. Results are means $\pm$ standard deviations across independent training seeds. Bold marks the best result in each column.

Method	Shallow Water Pixel 30%		Shallow Water Block 3/9		Navier–Stokes Pixel 30%	
	MSE↓ ($\times 10^{-3}$)	PSNR↑ (dB)	MSE↓ ($\times 10^{-3}$)	PSNR↑ (dB)	MSE↓ ($\times 10^{-3}$)	PSNR↑ (dB)
AmbientDiff	0.903 $\pm$0.018	37.168 $\pm$0.092	13.777 $\pm$0.025	26.516 $\pm$0.035	6.338 $\pm$0.448	28.146 $\pm$0.311
DINDiff	0.866 $\pm$0.033	37.233 $\pm$0.005	17.656 $\pm$0.106	26.170 $\pm$0.017	7.995 $\pm$0.017	27.074 $\pm$0.009
MissDiff	1.177 $\pm$0.089	35.779 $\pm$0.388	16.839 $\pm$0.164	25.798 $\pm$0.106	8.740 $\pm$0.467	26.728 $\pm$0.260
MSM	4.025 $\pm$0.094	30.352 $\pm$0.056	18.660 $\pm$0.023	25.017 $\pm$0.003	17.756 $\pm$0.017	23.613 $\pm$0.005
IDCD	0.880 $\pm$0.111	37.223 $\pm$0.275	8.001 $\pm$0.055	29.279 $\pm$0.241	7.545 $\pm$0.085	27.359 $\pm$0.056
Ours	**0.035** $\pm$**0.007**	**51.628** $\pm$**0.957**	**4.647** $\pm$**0.456**	**34.554** $\pm$**0.343**	**4.623** $\pm$**0.046**	**29.603** $\pm$**0.041**

The Shallow Water Block 3/9 result tests a contiguous missing region: with only three of nine spatial blocks supplied, MSE decreases from 8.001 for the strongest external baseline to 4.647 for ours (both in units of 10^{-3}). Figures 12 and 13 show Navier–Stokes and Shallow Water Pixel 30% reconstructions, respectively, and Figure 14 shows Shallow Water Block 3/9.

Additional block settings. For Block 8/9, eight blocks of a regular 3×3 partition are supplied and the remaining block is scored.

Our estimator achieves the lowest MSE and highest PSNR in both settings (Table 9). Relative to IDCD, MSE decreases by 26.7% on Shallow Water and 11.0% on Navier–Stokes. Qualitative comparisons with DINDiff on Shallow Water and AmbientDiff on Navier–Stokes appear in Figures 15 and 16, respectively.

Table 9: PDE reconstruction under Block 8/9 observations at 64×64. Entries are point estimates. Bold marks the best result in each column.

Method	Shallow Water Block 8/9		Navier–Stokes Block 8/9	
	MSE↓ ($\times 10^{-3}$)	PSNR↑ (dB)	MSE↓ ($\times 10^{-3}$)	PSNR↑ (dB)
AmbientDiff	8.608	38.859	33.725	21.260
DINDiff	5.530	42.014	34.068	21.360
MissDiff	16.510	35.989	53.201	19.207
MSM	16.099	31.509	45.947	20.013
IDCD	4.751	41.694	28.444	22.054
Ours	**3.481**	**45.044**	**25.326**	**22.591**

MSM under pixel observations. The two Pixel 30% PDE settings in Table 8 show lower reconstruction accuracy for MSM. Table 10 contrasts its partial-measurement denoising and sampling pathway with our query-supervised reconstruction pathway.

Table 10: Mechanistic comparison of MSM and our estimator.

Aspect	MSM (Park et al., 2026)	Ours
Training observations	Partial measurements, including noisy or subsampled measurement settings	Incomplete physical-field observations
Supervised coordinates	Denoising targets lie on the measurement support supplied in noisy form	Query values are withheld from the network input and supervise local and full-domain predictions
Learned object	Partial-measurement denoisers and associated scores	Local clean-value predictions and a full-domain residual corrector
Aggregation object	Lifted partial scores or corresponding denoising estimates	Overlapping spatial-patch predictions reconciled into a field-valued consensus
Role of full-domain information	Propagated through sampling and observation-consistency updates	Supplied to a separately trained corrector together with the local reference and discrepancy features
Reconstruction procedure	Iterative stochastic sampling for generation or conditional reconstruction	A local-to-global reconstruction path per selected context at near-zero diffusion time; final fields are averaged when context ensembling is used

Pixel and block observations provide different spatial information. In the training-mask summaries, approximately 94% of missing sites under Pixel 30% have an observed eight-neighbor, compared with approximately 5% under Block 3/9. This geometry provides widespread local evidence for conditional reconstruction. At the same time, pixel masks fragment jointly visible neighborhoods, whereas observed blocks retain contiguous local structure that may facilitate learning joint denoising relationships. Pixel observations can therefore favor reconstruction without providing equally coherent partial-denoising examples. This also differs from MSM's original face-image experiments, which used 32×32 block masks with approximately 40% missing entries (Park et al., 2026).

During inference, our MSM implementation applies a sampled mask S to the current noisy iterate and aggregates predictions on that support. The denoiser receives S, while the actual observation mask M_{in} enters through observation-consistency updates. With both masks retaining approximately 30% of coordinates and independent of one another, $\Pr(S_i = 1, M_{\text{in},i} = 1) \approx 0.09$. This quantity describes support overlap; the network input remains a noisy iterate rather than clean measurements. An observation-consistency update directly changes observed coordinates, and its influence on missing coordinates is propagated by later denoising calls. Pixel subsampling further fragments the jointly selected observation support.

The published MSM formulation interprets its aggregated score as a composite-likelihood/product-of-experts construction (Park et al., 2026). Its stochastic-approximation guarantee compares finite-mask sampling with the corresponding full MSM composition, rather than establishing equality with the true joint-field score. Increasing the number of partial-score evaluations can improve coverage and stochastic approximation, but does not by itself guarantee exact conditional dependencies. Together, fragmented co-observation and this denoising-and-composition pathway may limit how effectively the distributed observations constrain point reconstruction in the pixel protocol.

F.2 Higher-Resolution Ocean Reconstruction

Each learned method is trained and its checkpoint selected separately at 128×128. Our estimator uses the default patch geometry in Appendix B.2, and evaluation follows the source-supported scoring protocol in Appendix E.2.

Table 11: Ocean reconstruction at 128×128 on withheld source-supported values. Entries are point estimates. Bold marks the best result in each column.

Method	Black Sea CHL		Baltic Sea NANO		Global Ocean SSS	
	MSE↓ ($\times 10^{-3}$)	PSNR↑ (dB)	MSE↓ ($\times 10^{-3}$)	PSNR↑ (dB)	MSE↓ ($\times 10^{-3}$)	PSNR↑ (dB)
DINEOF	21.806	17.17	38.067	14.57	51.323	13.02
AmbientDiff	12.594	19.37	30.786	15.45	55.249	13.52
DINDiff	43.359	15.22	42.796	14.27	72.747	12.06
MissDiff	31.779	15.27	51.621	12.94	50.593	13.19
MSM	30.211	16.06	34.102	15.07	45.435	13.71
OAMP	10.282	20.20	26.825	15.81	37.698	14.43
Ours	**6.935**	**22.00**	**21.102**	**16.87**	**30.230**	**15.42**

Table 11 shows the lowest MSE and highest PSNR for our estimator on all three ocean products. Relative to OAMP, the lowest-MSE baseline on each product, MSE decreases by 32.5%, 21.3%, and 19.8% on Black Sea CHL, Baltic Sea NANO, and Global Ocean SSS, respectively.

F.3 Prior Construction and Connection Ablations

This subsection examines local-coordinate conditioning, corrector inputs, fusion weights, covering geometry, and patch scale. The comparisons reuse the same full supplied mask and Gaussian perturbation for each evaluation item.

Coordinate and input-feature controls. Table 12 reports the final-reconstruction effects of local-coordinate conditioning and explicit corrector inputs; Appendix E.4 specifies the corresponding retraining procedures. Removing local coordinates raises final MSE by 1.7%, 2.9%, and 14.2% on Black Sea CHL, Baltic Sea NANO, and Global Ocean SSS, respectively (Panel A). Section 4.4 interprets the input-feature results in Panel B.

Consensus-rule comparison. The consensus operator reconciles patch estimates that may disagree at overlapping locations. We ask whether unequal weights improve that reconciliation, and separately whether using overlapping views is preferable to tiling the field without overlap. These comparisons evaluate the local prior constructed from the full supplied mask before full-domain correction, so they isolate construction quality rather than the ability of a retrained corrector to compensate for a changed prior. Panel A fixes the overlapping views and their predictions under the full supplied mask and varies only the fusion weights. Panel B compares non-overlapping and half-stride coverings under uniform fusion; unlike Panel A, it changes which patch predictions are available at each location

Table 12: Local-coordinate and corrector-input ablations for final full-mask reconstruction on 64×64 ocean fields. Bold marks the best result in each column within each panel.

Variant	Black Sea CHL		Baltic Sea NANO		Global Ocean SSS	
	MSE↓ ($\times 10^{-3}$)	PSNR↑ (dB)	MSE↓ ($\times 10^{-3}$)	PSNR↑ (dB)	MSE↓ ($\times 10^{-3}$)	PSNR↑ (dB)
A: Local coordinate conditioning						
With local coordinates	**7.164**	**21.79**	**19.345**	**17.37**	**26.798**	**15.93**
No local coordinates	7.286	21.66	19.910	17.17	30.614	15.33
B: Explicit prior–evidence representation						
Full state	**7.164**	**21.79**	**19.345**	**17.37**	**26.798**	**15.93**
No $\boldsymbol{W}$	8.007	21.14	20.996	16.99	30.077	15.47
No $\boldsymbol{S}_{\mathrm{ctx}}$	8.160	21.05	21.355	16.89	29.432	15.48
No $\boldsymbol{S}_{\partial}$	8.227	21.10	22.046	16.79	30.202	15.38
No $\boldsymbol{S}_{\mathrm{ctx}}$ or $\boldsymbol{S}_{\partial}$	9.800	20.14	23.772	16.26	31.467	15.20

and which observations can influence them. Fusion choices and weighting settings are selected on validation data before testing. For rule f, the scored prior is the consensus field $\mathbf{\Pi}_\delta^f$ constructed with $\boldsymbol{M}_{\mathrm{ctx}} = \boldsymbol{M}_{\mathrm{in}}$.

For a covering $\mathcal{P}^{(s)}$, let $\mathcal{P}_i^{(s)} = \{p \in \mathcal{P}^{(s)} : i \in \Omega_p\}$. For a field of side length L and patch width h, non-overlap origins are $\{0, h, \dots, L-h\}^2$ and half-stride overlap origins are $\{0, h/2, \dots, L-h\}^2$. For $L = 64, h = 32$ and $L = 128, h = 64$, these give four non-overlapping and nine overlapping views. For the default Global Ocean SSS setting $L = 64, h = 16$, the non-overlap origins are $\{0, 16, 32, 48\}^2$ (16 views), and the overlap origins are $\{0, 8, 16, 24, 32, 40, 48\}^2$ (49 views). The crop operators are in-bounds direct slices. Dataset-specific mean multiplicity is computed as

$$\overline{m}^{(s)} = \frac{\sum_{i\in\Omega} V_i |\mathcal{P}_i^{(s)}|}{\sum_{i\in\Omega} V_i}. \tag{43}$$

Within a fixed inference context, let P_k denote the kth patch and let $\hat{x}_k(p)$ be its prediction at pixel p. Here k indexes patches only within this fusion definition; the inference context is held fixed. The unified fusion rule is

$$\hat{x}_{\mathrm{fused}}(p) = \frac{\displaystyle\sum_{k:p\in P_k} w_k(p)\hat{x}_k(p)}{\max\left(\displaystyle\sum_{k:p\in P_k} w_k(p),\, \varepsilon_f\right)}, \qquad \varepsilon_f = 10^{-6}. \tag{44}$$

The weights are nonnegative and computed from the corresponding observable state. Uniform fusion uses $w_k(p) = 1$; the resulting field is $\mathbf{\Pi}_{t,p}^{(s)} = \hat{x}_{\mathrm{fused}}(p)$.

Confidence weighting. Let $m_k(q) \in \{0, 1\}$ indicate whether pixel q in patch P_k is a retained context observation. We weight each patch by its retained context fraction:

$$c_k = \frac{1}{|P_k|} \sum_{q\in P_k} m_k(q), \qquad w_k(p) = \max(c_k^\gamma,\, w_{\min})\,. \tag{45}$$

Here $\gamma = \texttt{power}$ and $w_{\min} = \texttt{min_weight}$. The power is applied before the lower bound. The resulting weight is constant within each patch and increases with its retained context fraction.

Context-error weighting. For the single-channel ocean data, the uncertainty proxy is the reconstruction MSE on retained context observations:

$$u_k = \frac{\displaystyle\sum_{q\in P_k} m_k(q)\big(\hat{x}_k(q) - x_{\mathrm{obs}}(q)\big)^2}{\displaystyle\sum_{q\in P_k} m_k(q)}.$$

We then use

$$w_k(p) = \mathrm{clip}\big((u_k + \varepsilon_u)^{-\gamma},\, w_{\min},\, w_{\max}\big)\,. \tag{46}$$

Here $\varepsilon_u = \texttt{eps}$; γ, $w_{\min}$, and $w_{\max}$ correspond to `power`, `min_weight`, and `max_weight`, respectively. Only retained context entries enter u_k; ground-truth values at coordinates to be imputed are not used. Patches without context observations are assigned $u_k = +\infty$ and receive the minimum weight. The weight is constant within each patch. The table row "Uncertainty-weighted fusion uses this retained-context error proxy.

Gaussian weighting. Let $\xi_k(p), \eta_k(p) \in [-1, 1]$ denote normalized coordinates within patch P_k, and define

$$\sigma = \max(\texttt{sigma_scale}, 10^{-6}), \qquad g_k(p) = \exp\left(-\frac{\xi_k(p)^2 + \eta_k(p)^2}{2\sigma^2}\right).$$

The window is lower-bounded before normalization by its maximum:

$$\tilde{g}_k(p) = \max(g_k(p), w_{\min})\,, \qquad w_k(p) = \frac{\tilde{g}_k(p)}{\max(\max_{q\in P_k} \tilde{g}_k(q), 10^{-12})}. \tag{47}$$

Here $w_{\min} = \texttt{min_weight}$. The weights vary within a patch, emphasizing its center over its edges; smaller `sigma_scale` concentrates weights more strongly at the center. The unified fusion operator then normalizes these weights across patches covering each pixel.

Table 13: Local-prior reconstruction before global correction using full supplied masks on 64×64 ocean fields. Bold marks the lowest MSE within each dataset and panel.

View family / fusion	Black Sea CHL MSE↓ ($\times 10^{-3}$)	Baltic Sea NANO MSE↓ ($\times 10^{-3}$)	Global Ocean SSS MSE↓ ($\times 10^{-3}$)
A: identical overlap views; fusion weights vary			
Confidence-weighted mean	15.525	31.106	32.185
Gaussian-weighted mean	15.569	31.177	32.203
Uncertainty-weighted fusion	15.481	30.922	32.102
Uniform mean (ours)	**15.227**	**30.576**	**32.053**
B: covering and support union vary			
Uniform non-overlap	15.660	31.532	32.174
Half-stride overlap + uniform mean	**15.227**	**30.576**	**32.053**

Uniform fusion gives the lowest local-prior MSE among the tested weighting rules on each product (Table 13, Panel A). Its margin over the best alternative is small for Global Ocean SSS (32.053 versus 32.102 in the table's units), so the evidence supports the chosen simple rule within the tested family rather than a broad claim that adaptive weighting is harmful. Half-stride overlap also gives lower prior MSE than non-overlap on each product (Table 13, Panel B). This difference combines multiple changes: overlap supplies several predictions for some coordinates and alters the union of observations that can affect their estimates. It cannot be attributed to averaging alone.

Patch-scale sensitivity. The patch size controls how much context each local prediction can see, while the stride and number of views affect coverage and computational work. We compare 32×32/stride-16 and 16×16/stride-8 coverings, corresponding to nine and 49 in-bounds views per 64×64 field. Both use uniform overlap consensus and are evaluated through full-mask residual reconstruction. Validation selects 32×32/stride-16 for Black Sea CHL and Baltic Sea NANO, and 16×16/stride-8 for Global Ocean SSS.

Table 14: Patch-scale comparison for final full-mask residual reconstruction on 64×64 ocean fields. Bold marks the better result in each column.

Patch / stride	Views per field	Black Sea CHL		Baltic Sea NANO		Global Ocean SSS	
		MSE↓ ($\times 10^{-3}$)	PSNR↑ (dB)	MSE↓ ($\times 10^{-3}$)	PSNR↑ (dB)	MSE↓ ($\times 10^{-3}$)	PSNR↑ (dB)
32×32 / 16	9	**7.164**	**21.79**	**19.345**	**17.37**	28.297	15.76
16×16 / 8	49	7.542	21.45	21.405	16.84	**26.798**	**15.93**

The smaller covering increases final MSE on Black Sea CHL (7.164 to 7.542) and Baltic Sea NANO (19.345 to 21.405), but decreases it on Global Ocean SSS (28.297 to 26.798; Table 14). PSNR follows the same ordering. Thus the best tested scale depends on the product; the comparison changes patch support, covering geometry, and view count together, so it does not isolate any one of those factors.

F.4 Remote-Evidence Interventions on Full-Mask Reconstructions

The local-support argument in Appendix D.3 predicts that changing context outside $U(B)$ cannot change the prior on block B. The experiment asks whether the *corrected* prediction nevertheless changes, and whether removing such context raises or lowers error. We first use a complete remote-band deletion to test the structural prediction and measure output displacement. We then use

equal-sized deletions across distance bands to compare error effects. Both diagnostics intervene on the final reconstruction produced by the default ocean full-mask protocol.

Paired inputs and diagnostic support. For each evaluation item and training-seed checkpoint, we compare the reconstruction from the full supplied mask $\boldsymbol{M}_{\mathrm{in}}$ with reconstructions obtained after prescribed observation deletions. The unmodified and intervened inputs share the original near-zero perturbation. Each input is evaluated through local-prior construction, feature computation, and full-domain residual correction, using all observations remaining after its specified deletion without an additional context split. Scoring remains restricted to the original cached outer holdout constructed in Appendix E.2. Table 15 summarizes the diagnostic windows, local coverings, eligibility criterion, and pairing protocol.

For each diagnostic window W, define

$$B(W) = W \cap \boldsymbol{M}_{\mathrm{eval}}. \tag{48}$$

When B is understood, $W(B)$ denotes its associated diagnostic window. The contributing-patch support union $U(B)$ for the scored set $B(W)$ is defined in Appendix D.3. For one evaluation item, we suppress its sample index and let s index the training-seed checkpoint. The supplied observations outside the local-prior support form

$$R(B) = \{j : \boldsymbol{M}_{\mathrm{in},j} = 1\} \setminus U(B). \tag{49}$$

For $j \in R(B)$, its distance to the scored block is

$$\rho(j, B) = \min_{i \in B} \|j - i\|_2. \tag{50}$$

Sorting $R(B)$ by $\rho(j, B)$ and applying the mid-rank equal-frequency rule gives $R_N(B)$, $R_M(B)$, and $R_F(B)$. Remote Near remains outside $U(B)$; "near" is relative only to the other remote observations.

Table 15: Diagnostic settings for remote-evidence interventions on full-mask ocean reconstructions.

Item	Setting
Resolution	64×64
Black Sea / Baltic bank	W_{32}: 32×32, stride 32; four windows per field
Global SSS bank	W_{16}: 16×16, stride 16; 16 windows per field
Eligibility	At least eight scored coordinates, $\lvert B(W)\rvert \geq 8$
Black Sea / Baltic local views	32×32 patches, stride 16; nine views
Global SSS local views	16×16 patches, stride 8; 49 views
Distance bands	Mid-rank equal-frequency remote terciles
Evaluation unit	Paired reconstructions for one evaluation item, scored block, and checkpoint
Pairing	Shared checkpoint, original supplied input, and near-zero perturbation; arms differ only by the prescribed deletion

Support invariance and prediction displacement. The structural diagnostic removes the complete Remote Far tercile $R_F(B)$ from the full supplied mask. Here "full" denotes the unmodified input $\boldsymbol{M}_{\mathrm{in}}$. For deletion set $A_a(B) \subseteq \{j : \boldsymbol{M}_{\mathrm{in},j} = 1\}$, construct

$$\begin{aligned}
\boldsymbol{M}^a_{\mathrm{ctx}} &= \boldsymbol{M}_{\mathrm{in}} \odot \big(1 - \mathbf{1}_{A_a(B)}\big), \\
\boldsymbol{x}^a_{\mathrm{ctx},\delta,s} &= \boldsymbol{M}^a_{\mathrm{ctx}} \odot \boldsymbol{x}_{\mathrm{in},\delta,s}, \\
O^a_{\delta,s} &= (\boldsymbol{x}^a_{\mathrm{ctx},\delta,s}, \boldsymbol{M}^a_{\mathrm{ctx}}), \\
\boldsymbol{x}_{\mathrm{in},\delta,s} &= \alpha_\delta \boldsymbol{x}_{\mathrm{in}} + \sigma_\delta \boldsymbol{\epsilon}_s.
\end{aligned}$$

The unmodified state uses $A_{\mathrm{full}}(B) = \varnothing$, and the displacement diagnostic uses $A_{\mathrm{far}}(B) = R_F(B)$. Each state is evaluated through prior construction, feature computation, and full-domain residual correction with B fixed. Let $\widehat{\boldsymbol{x}}^a_{\mathrm{out},s}$ denote its final reconstruction. Because $B \subseteq \boldsymbol{M}_{\mathrm{eval}}$ is disjoint from $\boldsymbol{M}_{\mathrm{in}}$, restoring supplied observations does not change the prediction on B. We measure the prior and final-output displacements

$$d^{\mathrm{far}}_{\boldsymbol{\Pi},s}(B) = \frac{\left\| \boldsymbol{\Pi}^{\mathrm{far}}_{\delta,B,s} - \boldsymbol{\Pi}^{\mathrm{full}}_{\delta,B,s} \right\|_2^2}{d_x |B|}, \tag{51}$$

and

$$d^{\mathrm{far}}_{\widehat{\boldsymbol{x}},s}(B) = \frac{\left\|\widehat{\boldsymbol{x}}^{\mathrm{far}}_{\mathrm{out},B,s} - \widehat{\boldsymbol{x}}^{\mathrm{full}}_{\mathrm{out},B,s}\right\|_2^2}{d_x|B|}. \tag{52}$$

Table 16: Local-prior and final-reconstruction displacement after deleting the complete Remote Far tercile from the full supplied mask.

Dataset	$\max d^{\mathrm{far}}_{\boldsymbol{\Pi}}$	Mean $d^{\mathrm{far}}_{\widehat{\boldsymbol{x}}}$ [95% CI] ($\times 10^{-5}$)	Median $d^{\mathrm{far}}_{\widehat{\boldsymbol{x}}}$ ($\times 10^{-6}$)	Max $d^{\mathrm{far}}_{\widehat{\boldsymbol{x}}}$ ($\times 10^{-2}$)
Baltic Sea NANO	0	6.46 [4.71, 8.61]	3.35	1.617
Black Sea CHL	0	5.80 [4.71, 7.09]	2.67	1.156
Global Ocean SSS	0	62.30 [56.96, 68.02]	28.00	6.348

The local-prior displacement is zero for every eligible paired block evaluation (Table 16). Mean, median, and maximum final-output displacements are positive on all three datasets, with Global Ocean SSS having the largest reported mean. The full-domain estimator therefore responds to observations outside the support that determines the local prior on the scored block.

Matched-dose error changes. The local-near candidate set contains supplied observations inside the local-prior support union $U(B)$ but outside the current diagnostic window:

$$L_N(B) = (\{j : \boldsymbol{M}_{\mathrm{in},j} = 1\} \cap U(B)) \setminus W(B). \tag{53}$$

The matched deletion dose for each block is

$$m_B = \min\left\{|R_N(B)|,\, |R_M(B)|,\, |R_F(B)|,\, |L_N(B)|\right\}. \tag{54}$$

For each scored block, the deterministic compact selection rule removes m_B observations from each candidate set, producing five paired input states for each checkpoint:

$$O^{\mathrm{full}}_{\delta,s}, \quad O^{\mathrm{RN}}_{\delta,s}, \quad O^{\mathrm{RM}}_{\delta,s}, \quad O^{\mathrm{RF}}_{\delta,s}, \quad O^{\mathrm{LN}}_{\delta,s}. \tag{55}$$

The matched-dose experiment uses equally sized deletions from the full supplied mask for each scored block, whereas the displacement diagnostic removes the complete Remote Far tercile.

For removal arm $a \in \{\mathrm{RN}, \mathrm{RM}, \mathrm{RF}, \mathrm{LN}\}$, define the paired block-MSE change of the final reconstruction:

$$\Delta_{a,s}(B) = \mathrm{MSE}_B\left(\widehat{\boldsymbol{x}}^{a}_{\mathrm{out},s}\right) - \mathrm{MSE}_B\left(\widehat{\boldsymbol{x}}^{\mathrm{full}}_{\mathrm{out},s}\right). \tag{56}$$

Measurements are computed for each eligible scored block in each evaluation item and for each training-seed checkpoint, then summarized across these paired evaluations.

Table 17: Final-reconstruction block-MSE changes under matched-dose deletion from the full supplied mask. Entries are means with 95% confidence intervals; positive values indicate increased error after deletion. Bold marks means whose intervals exclude zero.

	Δ block MSE ($\times 10^{-5}$)			
Dataset	Remote Near	Remote Mid	Remote Far	Local Near
Baltic Sea NANO	0.95 [−3.32, 5.23]	**4.66** [1.08, 8.56]	**5.25** [1.14, 9.56]	**50.11** [32.63, 69.56]
Black Sea CHL	0.53 [−1.57, 2.77]	**2.53** [0.43, 4.84]	**3.72** [0.65, 6.75]	5.87 [−1.73, 13.10]
Global Ocean SSS	**6.76** [2.99, 10.54]	0.98 [−1.53, 3.49]	**−12.85** [−18.64, −7.09]	**64.12** [55.49, 73.16]

For Baltic Sea NANO, the intervals for Remote Mid, Remote Far, and Local Near error changes lie above zero, while the Remote Near interval crosses zero (Table 17). For Black Sea CHL, the

Table 18: Accuracy and request latency under full-mask and context-ensemble inference at 64×64. Entries are means $\pm$ standard deviations across independent training seeds. For latency, each seed contributes its median or 95th-percentile request time. Bold marks the best mean accuracy within each dataset.

Model	Inference	MSE↓ ($\times 10^{-3}$)	PSNR↑ (dB)	Median latency (ms/field)	P95 latency (ms/field)
A: Black Sea CHL					
Full Direct	Full-mask	9.350 $\pm$0.306	20.32 $\pm$0.12	69.46 $\pm$0.20	69.78 $\pm$0.31
Full Direct	Context $K = 1$	13.922 $\pm$0.055	18.61 $\pm$0.07	69.41 $\pm$0.10	69.76 $\pm$0.12
Full Direct	Context $K = 2$	13.823 $\pm$0.065	18.64 $\pm$0.07	80.05 $\pm$0.19	80.29 $\pm$0.26
Full Direct	Context $K = 4$	13.762 $\pm$0.072	18.66 $\pm$0.08	105.30 $\pm$0.20	105.51 $\pm$0.15
Full Direct	Context $K = 8$	13.721 $\pm$0.083	18.66 $\pm$0.08	210.30 $\pm$0.41	210.60 $\pm$0.50
Ours	Full-mask	**7.164** $\pm$0.060	**21.79** $\pm$0.04	191.82 $\pm$1.64	192.85 $\pm$2.84
Ours	Context $K = 1$	12.016 $\pm$0.110	19.37 $\pm$0.03	192.06 $\pm$1.11	196.95 $\pm$5.18
Ours	Context $K = 4$	11.510 $\pm$0.015	19.57 $\pm$0.01	345.38 $\pm$0.01	345.96 $\pm$0.07
B: Shallow Water Pixel 30%					
Full Direct	Full-mask	0.79503 $\pm$0.16693	37.77 $\pm$0.72	69.10 $\pm$0.03	69.35 $\pm$0.01
Full Direct	Context $K = 1$	0.70450 $\pm$0.02675	38.36 $\pm$0.15	69.20 $\pm$0.14	69.52 $\pm$0.08
Full Direct	Context $K = 2$	0.68605 $\pm$0.02891	38.46 $\pm$0.16	88.06 $\pm$0.12	88.36 $\pm$0.06
Full Direct	Context $K = 4$	0.67749 $\pm$0.02930	38.51 $\pm$0.16	110.50 $\pm$0.06	110.92 $\pm$0.10
Full Direct	Context $K = 8$	0.67295 $\pm$0.03067	38.53 $\pm$0.16	210.54 $\pm$0.05	210.83 $\pm$0.04
Full Direct	Context $K = 16$	0.67009 $\pm$0.03109	38.55 $\pm$0.16	399.74 $\pm$0.29	400.68 $\pm$0.15
Full Direct	Context $K = 32$	0.66856 $\pm$0.03128	38.55 $\pm$0.16	434.50 $\pm$0.19	434.86 $\pm$0.17
Ours	Full-mask	0.05246 $\pm$0.01542	50.19 $\pm$1.16	190.39 $\pm$0.08	190.99 $\pm$0.12
Ours	Context $K = 1$	0.04998 $\pm$0.01149	50.41 $\pm$0.98	190.86 $\pm$0.07	191.31 $\pm$0.08
Ours	Context $K = 4$	0.03771 $\pm$0.00819	51.40 $\pm$0.98	349.09 $\pm$0.66	349.53 $\pm$0.96
Ours	Context $K = 8$	**0.03521** $\pm$0.00738	**51.63** $\pm$0.96	668.28 $\pm$1.28	669.76 $\pm$1.19

Remote Mid and Remote Far intervals lie above zero, while Remote Near and Local Near cross zero. For Global Ocean SSS, Remote Near and Local Near deletion increase mean block MSE, Remote Far deletion decreases it, and the Remote Mid interval crosses zero. These full-mask interventions distinguish responsiveness beyond local support from the dataset-dependent accuracy effects of the removed observations. The remote bands contain different spatial locations even when their deletion counts match, and the mixed signs on Global Ocean SSS rule out a universal claim that farther context always improves this estimator's prediction.

F.5 Inference Cost and Observation Retention

Comparison protocol. We compare our estimator with parameter-matched Full Direct on Black Sea CHL and Shallow Water Pixel 30% at 64×64. Full Direct is based on OAMP and IDCD, respectively, as defined in Appendix E.4. For each estimator, all inference configurations reuse the same set of independently trained checkpoints. For each sample, the supplied support $\boldsymbol{M}_{\text{in}}$ and outer scoring support remain fixed across inference configurations. Context-ensemble comparisons additionally reuse one cached subcontext sequence, with an ensemble of size K using its first K entries. Full-mask and context-ensemble inference reuse the same near-zero perturbation and are scored on the same outer evaluation support. Full-mask uses all supplied observations, whereas Context $K = 1$ uses one subsampled context. The default modes and ensemble size are selected on validation data as specified in Appendix B.4.

Reconstruction latency is measured in milliseconds per request for one 64×64 field on a single NVIDIA A40 GPU in FP32. The measurement includes all local patch predictions, overlap fusion, feature construction, and full-domain correction. Inference subcontext masks, including BFN-generated masks, are cached, so mask generation is excluded from the timing interval. Ensemble contexts are batched for parallel prediction on the same GPU.

Accuracy under common latency ceilings. The ceilings below apply to the across-seed mean of per-seed median reconstruction latency. For each ceiling, we compare the lowest mean MSE among the measured configurations in Table 18 that satisfy it. On Black Sea, full-mask inference gives the lowest MSE for both estimators. Under a 211 ms ceiling, our full-mask estimator achieves 7.164×10^{-3} MSE at 191.82 ms, compared with 9.350×10^{-3} at 69.46 ms for the best eligible Full Direct configuration, a 23.4% MSE reduction. Full Direct's full-mask configuration is faster, but increasing its context ensemble to $K = 8$ raises latency to 210.30 ms without improving on its full-mask accuracy.

On Shallow Water Pixel 30%, our estimator retains its advantage at lower ensemble sizes. Under a 211 ms ceiling, our $K = 1$ configuration achieves 0.04998×10^{-3} MSE at 190.86 ms, compared with 0.67295×10^{-3} at 210.54 ms for Full Direct with $K = 8$, a 92.6% MSE reduction. Under a 400 ms ceiling, our $K = 4$ configuration achieves 0.03771×10^{-3} at 349.09 ms, compared with 0.67009×10^{-3} at 399.74 ms for Full Direct with $K = 16$, a 94.4% reduction. Increasing our ensemble from $K = 4$ to $K = 8$ provides a further 6.6% MSE reduction while increasing median latency from 349.09 to 668.28 ms.

Training and inference observation supports. The two settings differ in both acquisition masks and internal partitions (Appendices C.2, C.3, and E.2). Under Shallow Water Pixel 30%, the stored acquisition mask retains approximately 30% of the valid field, and training retains approximately 70% of those observations as context. The nominal training context and query therefore occupy 21% and 9% of the valid field, respectively. At test time, full-mask inference supplies approximately 30%, whereas each ensemble subcontext supplies approximately 21%. The subcontexts are thus closer to the training inputs in observation density and pixel-mask structure.

Black Sea training instead partitions each sample's source-supported observations directly. Its retained context density varies across samples and guided proposals rather than being fixed by the BFN anchor-retention parameter (Appendix C.3). At validation and test time, the empirical real-mask overlay protocol in Appendix E.2 constructs the cached supplied support $\boldsymbol{M}_{\text{in}}$ and outer scoring support. Full-mask inference retains $\boldsymbol{M}_{\text{in}}$, whereas context ensembling applies an additional guided-mask intersection within it.

Observation retention and spatial coverage. We compute mask statistics over all test samples using their first eight cached subcontexts. Observation density uses the valid domain as its denominator, whereas subcontext retention and context-union coverage use the supplied support $\boldsymbol{M}_{\text{in}}$. Ratios are computed within each sample before equal-weight averaging across samples; subcontext statistics additionally average over the eight contexts within each sample.

For sample n, let Q_n denote its fixed outer scoring coordinates. The mean distance to an observation mask A is

$$d_n(A) = \frac{1}{|Q_n|} \sum_{i \in Q_n} \min_{j : A_j = 1} \|i - j\|_2.$$

We compute this distance within each sample and then average equally across samples, averaging over the eight contexts for the subcontext statistic. Distances are Euclidean distances on the 64×64 grid, measured in pixels rather than physical units.

Table 19: Test-time observation retention and spatial coverage on Black Sea CHL and Shallow Water Pixel 30%. Entries are equally weighted means over test samples.

Mask statistic	Black Sea CHL	SW Pixel 30%
Full-mask observations / valid domain	50.72%	29.96%
Subcontext observations / valid domain	27.70%	20.97%
Supplied observations retained per subcontext	51.23%	70.00%
Supplied observations covered by the first 4 contexts	72.93%	99.19%
Supplied observations covered by the first 8 contexts	78.05%	99.994%
Mean nearest-observation distance: full-mask	3.09 pixels	1.15 pixels
Mean nearest-observation distance: subcontext	7.54 pixels	1.29 pixels

Black Sea subcontexts retain substantially less of the supplied evidence and create larger gaps around the scoring locations (Table 19). Even the first eight-context union covers only 78.05% of the original input observations, while the mean nearest-observation distance increases from 3.09 to 7.54 pixels. In Shallow Water Pixel 30%, the corresponding union covers 99.994%, and the distance increases only from 1.15 to 1.29 pixels. Union coverage does not imply that any individual member conditions jointly on all supplied observations. Restoring observations at the output enforces their values but does not recover their omitted conditioning influence on predictions at unsupplied locations.

Interpreting the different ensemble effects. Field structure may amplify these geometric differences. Black Sea chlorophyll observations have irregular spatial support and localized variation, for which removed nearby measurements may be difficult to replace with distant evidence. The Shallow Water subset shares radial dam-break structure, with variation in initial dam radius and evolution time. Spatially dispersed observations can repeatedly constrain this shared structure, potentially making moderate pixel subsampling less costly. Together with the closer match to training-context density, this provides an explanation consistent with the Shallow Water results.

Averaging helps relative to a single subcontext in both settings: increasing our ensemble from $K = 1$ to $K = 4$ reduces mean MSE by 4.2% on Black Sea and 24.6% on Shallow Water Pixel 30% (Table 18). On Black Sea, this gain does not compensate for observation removal, and full-mask inference still has 37.8% lower MSE than the $K = 4$ ensemble. On Shallow Water, the $K = 4$ ensemble instead improves over full-mask inference by 28.1%. These reductions are consistent with averaging context-sensitive errors, while errors shared across contexts can persist when relevant observations are repeatedly omitted. Because all members share the same near-zero perturbation, the gains are not attributable to averaging independent diffusion-noise draws.

The measured retention and distance statistics establish the different observation costs of subsampling; training-input similarity and error averaging provide complementary interpretations of the resulting accuracy trends.

G Limitations

Some block-missing reconstructions exhibit mild seam-like artifacts near observed–missing interfaces. These local transitions can remain imperfect even when fine-scale structures within the missing region are recovered. Improving cross-boundary consistency without suppressing local detail is a useful direction for future work.

H Reconstruction Visualizations

The ocean visualizations use full-mask inference. In Figures 5–11, black and red outlines mark supplied observations and withheld source-supported scoring locations, respectively; gray denotes outside-domain locations, and white in Reference denotes source-missing values.

H.1 Ocean Reconstruction at 64×64

Figure 5 decomposes the model predictions used in Figure 3, for the same samples and masks, and also includes the external OAMP baseline from Table 1. The local priors already contain large-scale field structure, while the learned residuals vary across the domain rather than applying a constant offset.

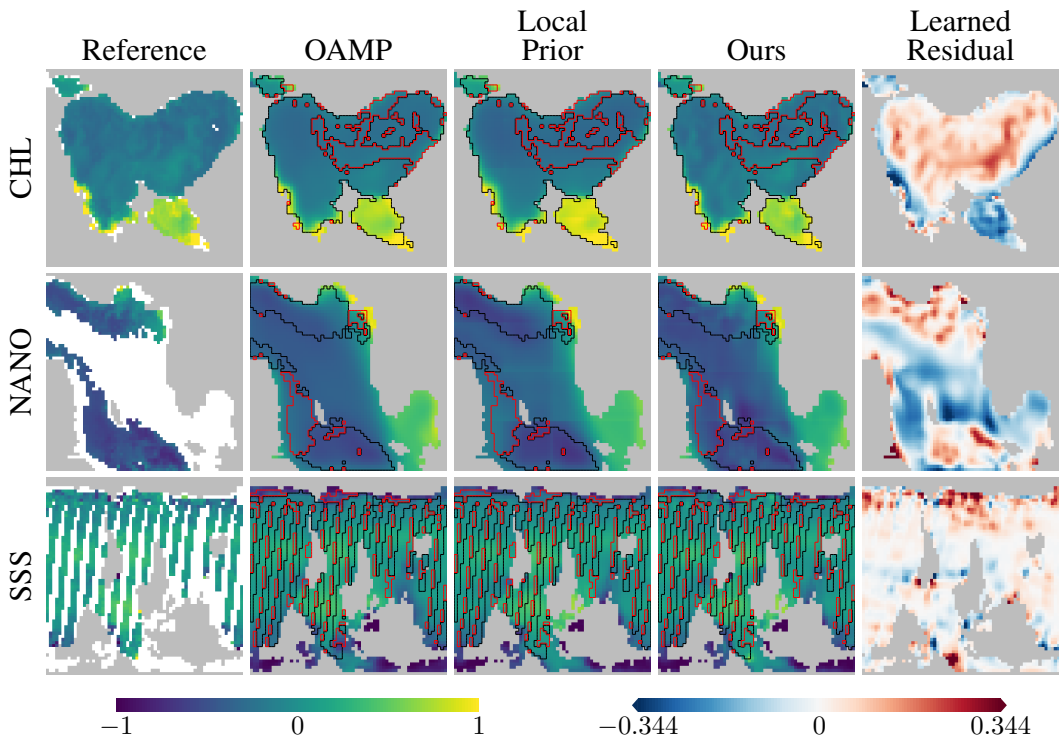

Figure 5: Local-to-global decomposition for full-mask ocean reconstruction at 64×64. Model predictions are shown before restoring supplied observations, preserving the pointwise relation Ours = Local Prior + Learned Residual.

Figures 6–8 compare Ours with parameter-matched Full Direct at 64×64, with supplied observations restored for display.

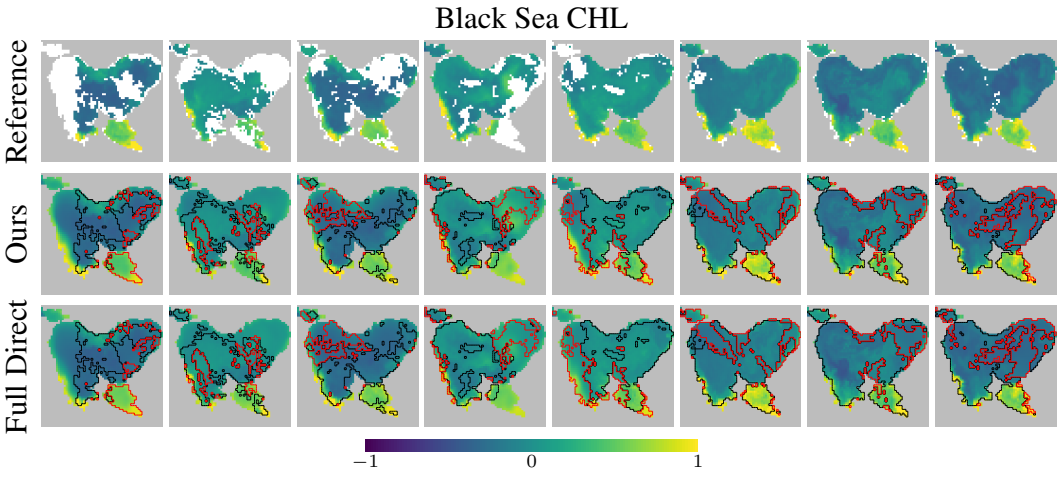

Figure 6: Full-mask reconstructions on Black Sea CHL at 64×64.

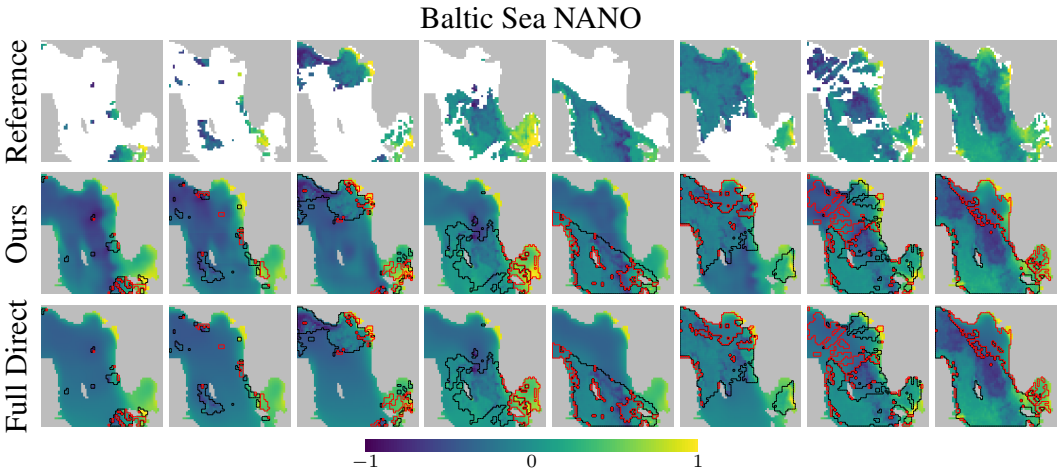

Figure 7: Full-mask reconstructions on Baltic Sea NANO at 64×64.

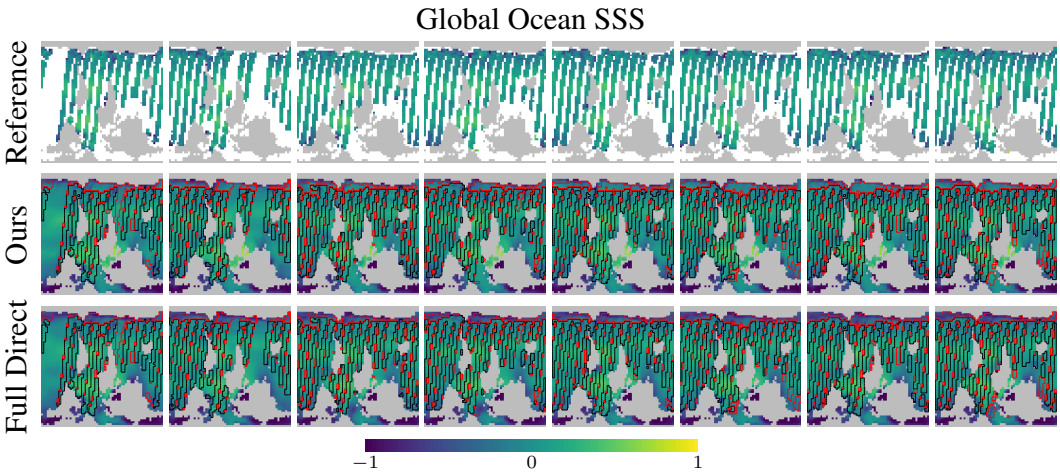

Figure 8: Full-mask reconstructions on Global Ocean SSS at 64×64.

H.2 Ocean Reconstruction at 128×128

Figures 9–11 compare Reference and Ours at 128×128, with supplied observations restored for display.

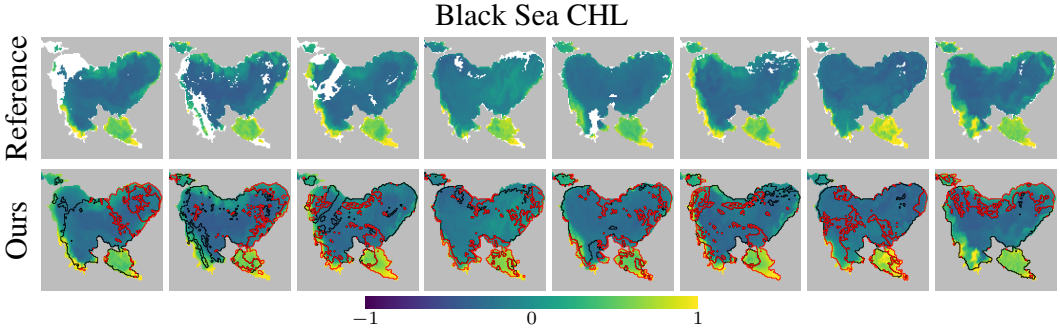

Figure 9: Full-mask reconstructions on Black Sea CHL at 128×128.

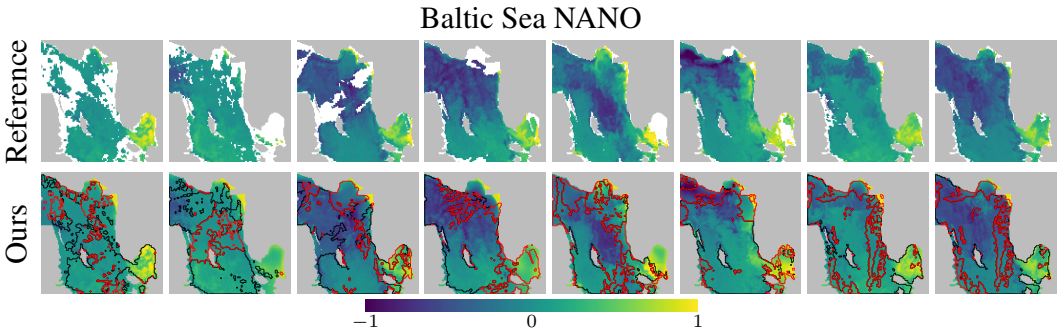

Figure 10: Full-mask reconstructions on Baltic Sea NANO at 128×128.

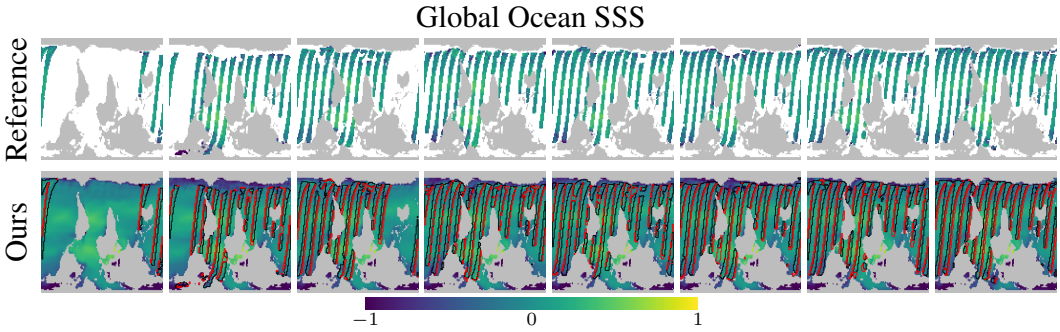

Figure 11: Full-mask reconstructions on Global Ocean SSS at 128×128.

H.3 PDE Reconstruction under Pixel Observations

Figures 12 and 13 show Navier–Stokes and Shallow Water Pixel 30% reconstructions, respectively. In Figures 12–16, Ours uses $K = 8$ context ensembling. Supplied observations are restored for display, and each column compares the same sample across methods.

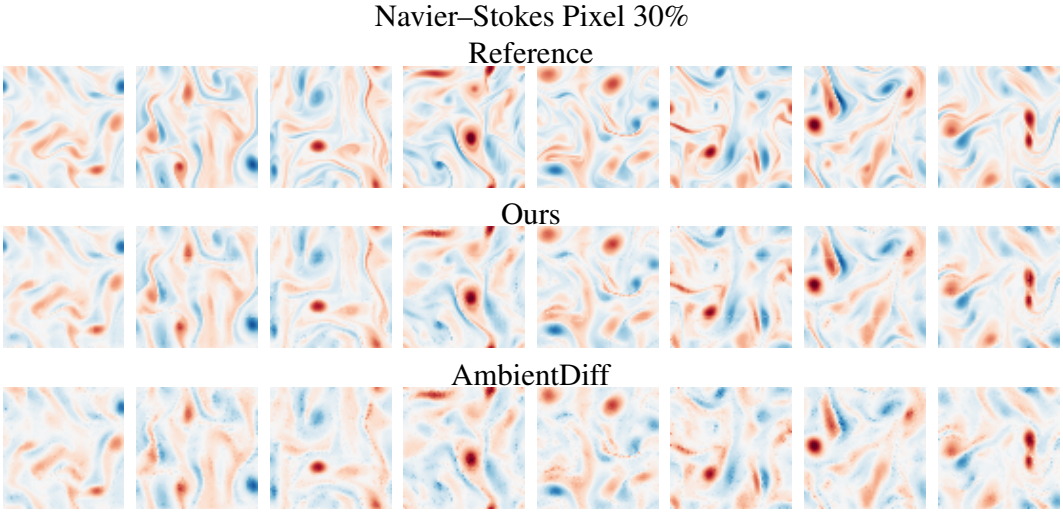

Figure 12: Navier–Stokes Pixel 30% reconstructions at 64×64, comparing Ours with AmbientDiff.

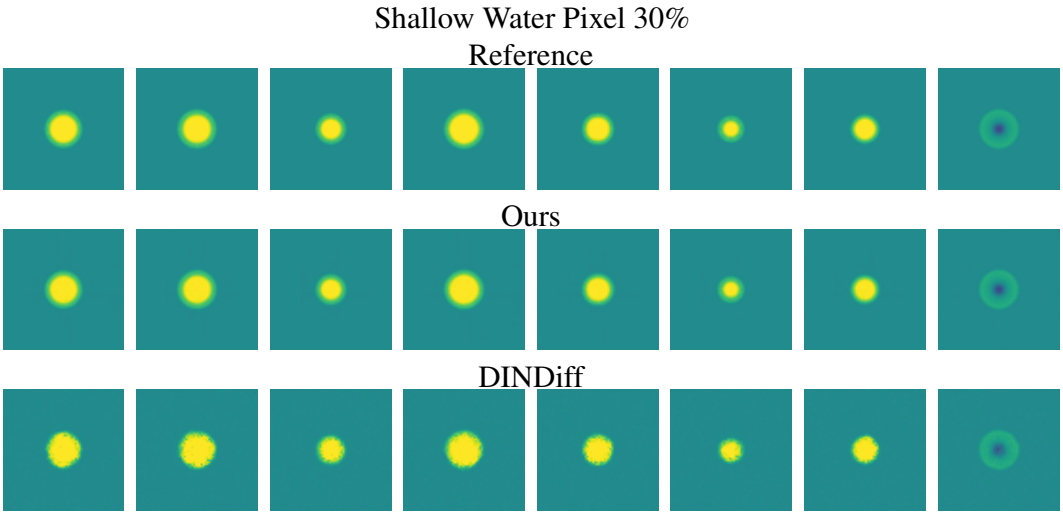

Figure 13: Shallow Water Pixel 30% reconstructions at 64×64, comparing Ours with DINDiff.

H.4 PDE Reconstruction under Block Observations

Figures 14–16 show Shallow Water and Navier–Stokes reconstructions under block observations. Solid lines mark the 3×3 acquisition partition, and red dashed outlines identify unsupplied blocks.

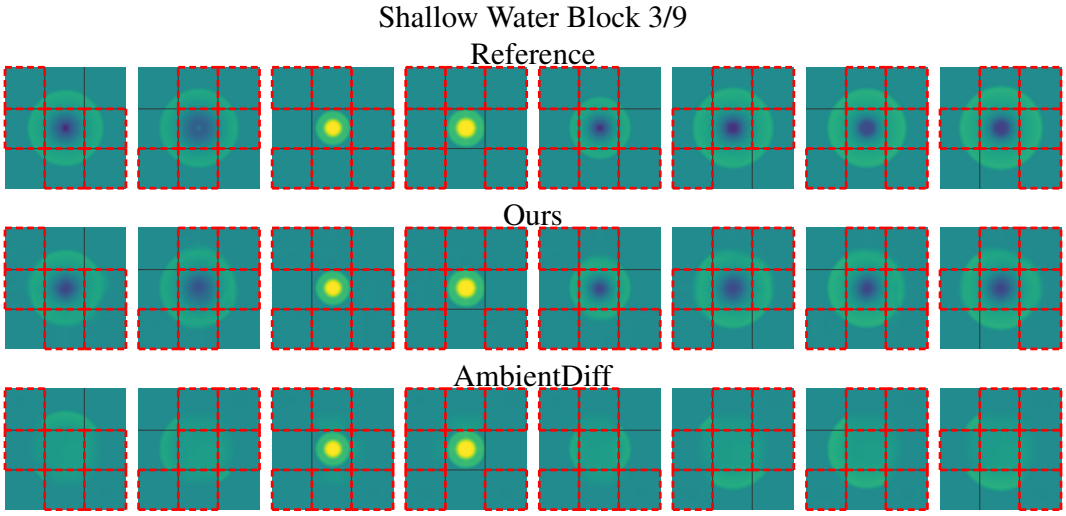

Figure 14: Shallow Water Block 3/9 reconstructions at 64×64, comparing Ours with AmbientDiff.

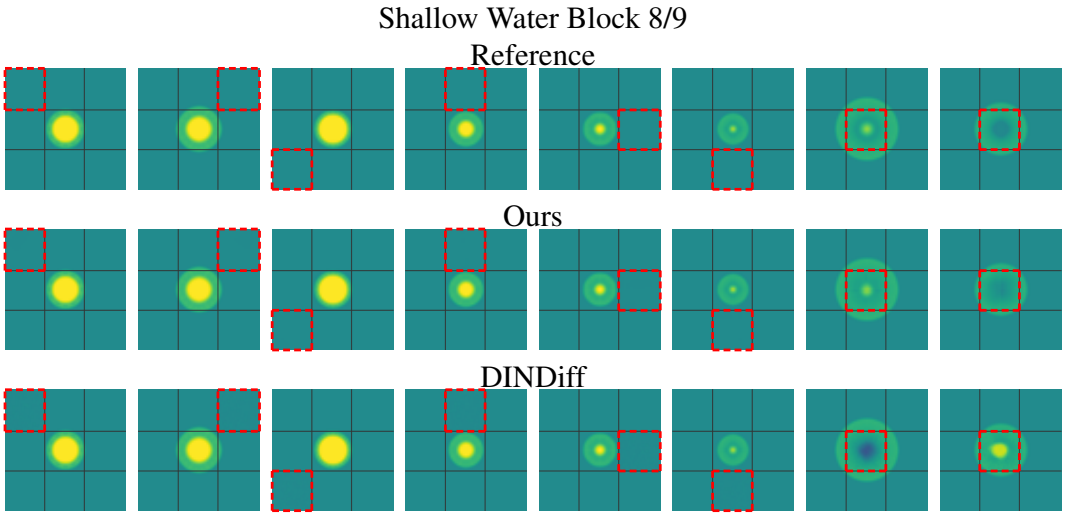

Figure 15: Shallow Water Block 8/9 reconstructions at 64×64, comparing Ours with DINDiff.

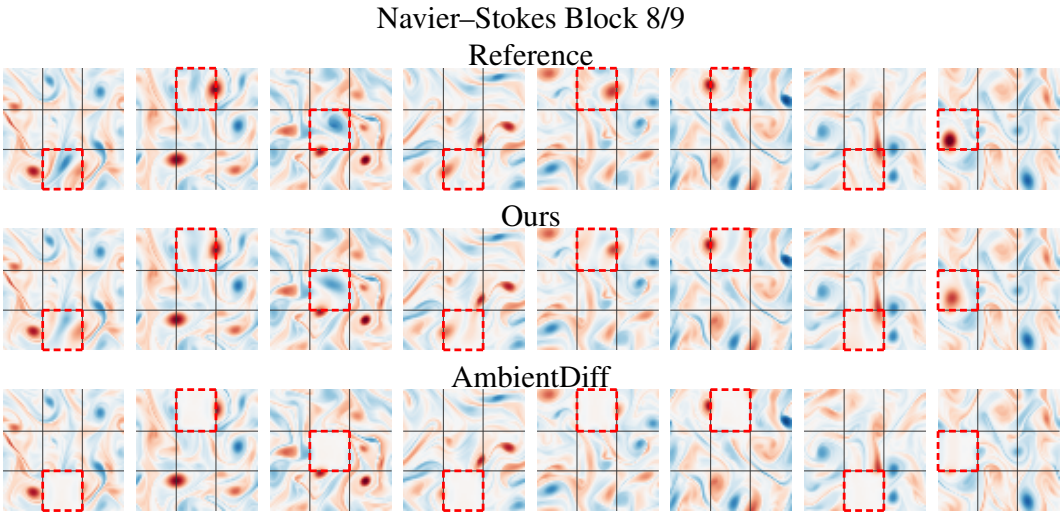

Figure 16: Navier–Stokes Block 8/9 reconstructions at 64×64, comparing Ours with AmbientDiff.