arXiv:2506.10391v1 [cs.CV] 12 Jun 2025

RECONMOST: MULTI-LAYER SEA TEMPERATURE RECONSTRUCTION WITH OBSERVATIONS-GUIDED DIFFUSION

A PREPRINT

Yuanyi Song[1], **Pumeng Lyu**[2], **Ben Fei**[3,*], **Fenghua Ling**[2], **Wanli Ouyang**[3], **Lei Bai**[2]
[1]Shanghai Jiao Tong University, [2]Shanghai AI Laboratory, [3]The Chinese University of Hong Kong
norsheep919@sjtu.edu.cn, benfei@cuhk.edu.hk
[*] Corresponding Author

ABSTRACT

Accurate reconstruction of ocean is essential for reflecting global climate dynamics and supporting marine meteorological research. Conventional methods face challenges due to sparse data, algorithmic complexity, and high computational costs, while increasing usage of machine learning (ML) method remains limited to reconstruction problems at the sea surface and local regions, struggling with issues like cloud occlusion. To address these limitations, this paper proposes ReconMOST, a data-driven guided diffusion model framework for multi-layer sea temperature reconstruction. Specifically, we first pre-train an unconditional diffusion model using a large collection of historical numerical simulation data, enabling the model to attain physically consistent distribution patterns of ocean temperature fields. During the generation phase, sparse yet high-accuracy in-situ observational data are utilized as guidance points for the reverse diffusion process, generating accurate reconstruction results. Importantly, in regions lacking direct observational data, the physically consistent spatial distribution patterns learned during pre-training enable implicitly guided and physically plausible reconstructions. Our method extends ML-based SST reconstruction to a global, multi-layer setting, handling over 92.5% missing data while maintaining reconstruction accuracy, spatial resolution, and superior generalization capability. We pre-train our model on CMIP6 numerical simulation data and conduct guided reconstruction experiments on CMIP6 and EN4 analysis data. The results of mean squared error (MSE) values achieve 0.049 on guidance, 0.680 on reconstruction, and 0.633 on total, respectively, demonstrating the effectiveness and robustness of the proposed framework. Our source code is available at https://github.com/norsheep/ReconMOST.

1 Introduction

The ocean, covering around 70% of Earth's surface, can absorb and release large amounts of heat and thus stabilize Earth's climate [Johnson et al., 2016, von Schuckmann et al., 2016]. Tremendous efforts have been made to monitor ocean temperature, but still with irregular and sparsely distributed observations from Argo floats, ships and satellites [Roemmich et al., 2009, Abraham et al., 2013, Von Schuckmann et al., 2013]. Sea temperature (ST) reconstruction is thus needed for a more complete picture of ocean temperature and different depths and locations. It has a significant impact to help us understand global warming [Mendelssohn and Bessey, 2012], discover factors driving extreme weather [Jewson et al., 2007, Prochaska et al., 2021], and understand climate dynamics [Cutolo et al., 2024].

Conventional approaches for climate field reconstruction mostly include Optimal Interpolation (OI) [Smith and Reynolds, 2003], Kriging [Ding et al., 2008], and Empirical Orthogonal Functions (EOF) [Alvera-Azcárate et al., 2005]. However, these data science methods exhibit limitations, for example, limited spatial coverage, lower reconstruction accuracy in large and irregular areas of missing data, high computational costs, and limited generalization capability across the various ocean domains [Smith and Murphy, 2007, Zhang et al., 2022].

Recent development of AI has shown promising solutions for ocean reconstruction. Neural networks and generative AI models are used to reconstruct sea surface temperature (SST) [Ouala et al., 2018] and understand ocean dynamics [Zheng

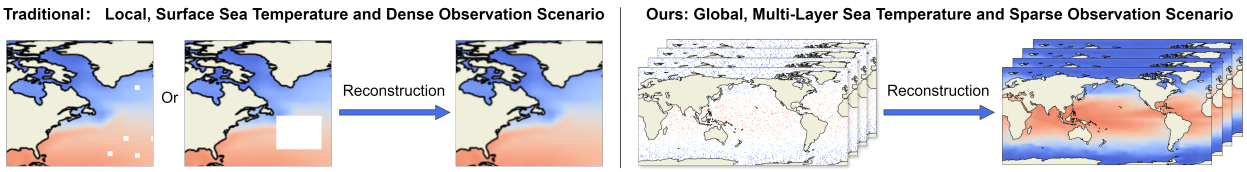


Figure 1: Difference between traditional methods and our ReconMOST.

et al., 2024]. These studies indicate that ML-based methods have the potential to outperform traditional statistical approaches in ocean reconstruction. However, these models are still used for dense observation or cloud occulasion scenarios in the local sea, shown as the left part of Fig. 1, and often fail in scenarios with extreme data sparsity, such as global sea temperature reconstruction, where observations (e.g., from Argo floats or satellites) are significantly fewer than required grid points. This sparsity complicates learning the underlying physical patterns, particularly for multi-layer ocean temperature fields. To our knowledge, no existing work has addressed the reconstruction of global, deep ocean temperature fields under such sparse conditions, a critical task for advancing climate modeling and oceanographic research.

To address these challenges, we propose ReconMOST, a novel model based on the Guided Diffusion Model [Dhariwal and Nichol, 2021] for reconstructing near-globally distributed, multi-layered sea temperature fields from extremely sparse observational data shown as the right part of Fig. 1, demonstrating strong generalization capability to real-world observations. Alessio et al. [1999] and Kawamura [Kawamura, 1994] have indicated that the spatial distribution of ocean temperature fields exhibits consistent principal patterns across varying temporal scales. Inspired by these findings, we first trained an unconditional diffusion model using a large collection of historical ocean temperature data from numerical model simulations. Using the pre-trained diffusion model, we guide the reverse generation process to reconstruct target multi-layer sea temperature fields under real-world sparse observation conditions. While ensuring accurate reconstruction at the observational guidance points, the physically consistent distribution patterns learned by the pre-trained diffusion model implicitly enhance the reconstruction quality of unobserved regions as well. Experimentally, a diffusion model was trained on seven numerical model's historical simulation data, which effectively learns a physically consistent spatial prior for 3D ocean temperature fields. We then conduct reconstruction experiments on both remaining CMIP6 and EN4 analysis data, and ensure that the number of guidance points matches the buoy data density. Our method achieves a mean squared error (MSE) of 0.633. Additionally, we conduct ablation experiments to evaluate the effects of different pre-training modes, dataset sizes, and guidance parameters, providing robust evidence of our method's effectiveness in various aspects. Our main contributions are threefold:

- **We introduce a data-driven guided diffusion framework for ocean reconstruction for the first time.** Our ReconMOST captures physically consistent spatial patterns through large-scale numerical simulations prior, enabling the reliable generation of ocean states aligned with fundamental physical properties.
- **Our framework reconstructs global, multi-layer sea temperature fields under extreme sparsity, with up to 92.5% missing data.** Even in such scenarios, the method maintains a high reconstruction accuracy, achieving an MSE as low as 0.633, demonstrating strong robustness in sparse-data regimes.
- **We simultaneously achieve high reconstruction accuracy and generalization performance.** Our ReconMOST adapts to various densities of sparse observational points according to real-world observations. As the number of guided observations increases, reconstruction accuracy improves correspondingly, while retaining reliable performance under minimal supervision.

2 Related Work

Guided Diffusion. Guided diffusion is a conditional sampling technique for diffusion models, where external information is incorporated by computing the gradient between the generated sample and the conditioning data, and using this gradient to steer the sampling trajectory [Tu et al., 2024, Lyu et al., 2024]. This approach has been widely applied across various domains. Wang et al. [2022] mitigates adversarial perturbations in image classification, while SGDiff and GDUI [Sun et al., 2023, Xie and Zhao, 2024] focus on style-controlled image synthesis and unsupervised image generation, respectively. These methods leverage essential latent features of the distribution while matching relevant conditions. Beyond image, guided diffusion has been extended to other tasks by reframing them as a structured image processing problem. For instance, DMDiff [Zhang et al., 2025] addresses cloud removal by fusing remote sensing data, and Klarner et al. [2024] handle out-of-distribution molecular and protein generation. Inspired by these advances, sea temperature reconstruction from sparse observations can similarly be formulated as a conditional large-scale image inpainting task, where the goal is to recover a physically consistent, multi-layered ocean field conditioned on limited point-wise measurements.

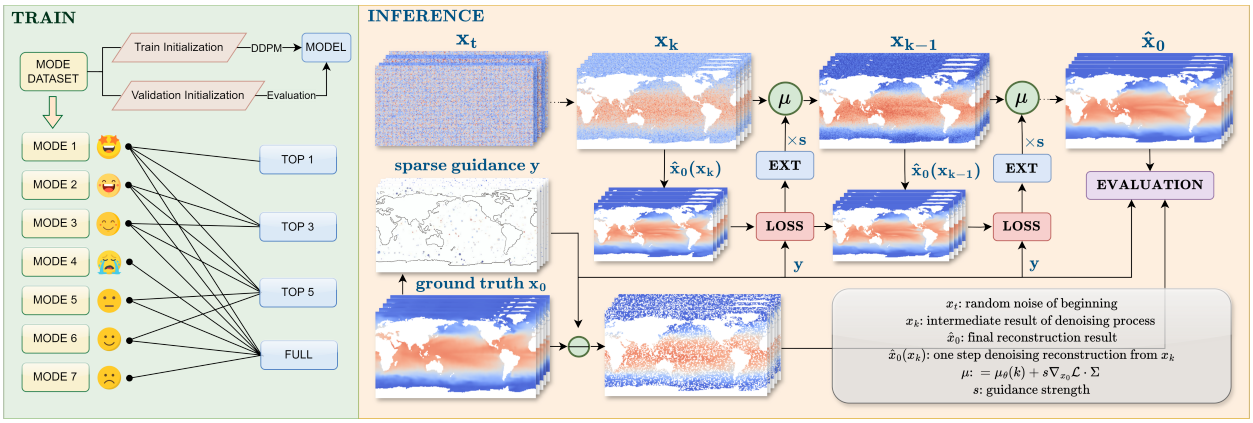


Figure 2: The framework of ReconMOST: In the training phase, we pre-train an unconditional diffusion model using a large collection of historical numerical simulation data chosen by effectiveness experiment in Sec. 4.3, enabling the model to learn physically consistent distribution patterns of ocean temperature fields. In the inference phase, we split the test dataset into sparse guidance part (7.5%), which is utilized to guide the reverse process to generate accurate reconstruction results, and evaluation part, which is utilized for the evaluation the accuracy of reconstruction results in Sec. 4.4.

Ocean Reconstruction. Ocean field reconstruction has been constrained by the lack of accurate ground truth data, primarily due to observational errors and the inherent discrepancies between physical simulations and real-world ocean conditions. Existing methods predominantly rely on numerical outputs or reanalysis products from authoritative institutions, such as the ECMWF or the WCRP, as reference datasets [Hersbach, 2016, Gutowski Jr et al., 2016]. Consequently, most data-driven ocean and atmospheric reconstruction approaches are trained on these numerical model outputs. Recent advances in AI, particularly neural networks, offer promising solutions. Ouala et al. [2018] first used deep learning to reconstruct sea surface temperature (SST) by introducing bilinear residual neural network representations, which mimic numerical integration schemes such as Runge-Kutta [Cartwright and Piro, 1992], for the forecasting and assimilation of geophysical fields from satellite-derived remote sensing data. Zheng et al. [2024] redesigned a purely data-driven latent space DA framework (DeepDA) that employs a generative AI model to capture the nonlinear evolution in sea surface temperature. Despite their potential to outperform traditional statistical methods, these ML approaches do not work well in scenarios with extreme data sparsity, where limited observations hinder the learning of underlying physical patterns.

Guided Diffusion in Reconstruction Issues. In atmospheric sciences, forecasting and field reconstruction problems typically produce spatially discrete results, which can naturally be reframed as conditional image generation or restoration tasks. Guided diffusion models, with their inherent capability for conditional sampling, have demonstrated strong adaptability in this context. SGD [Tu et al., 2025] integrates conditional diffusion models with attention mechanisms to reconstruct high-resolution meteorological states from low-resolution inputs. Hua et al. [2024] incorporates numerical weather prediction outputs during sampling, enabling flexibility across varying forecast lead times and improving long-range forecast stability. Similarly, DGDM [Yoon et al., 2023] uses deterministic forecasts to guide probabilistic diffusion models, achieving high-precision probabilistic predictions. These advances collectively demonstrate the effectiveness of guided diffusion models in ocean reconstruction tasks, highlighting their potential for addressing the highly sparse and multi-layer global sea temperature field reconstruction.

3 Method

Our ReconMOST framework follows a train-inference process as Fig. 2. First, we train a diffusion model on a large dataset of historical ocean temperature profiles derived from numerical simulations. This process allows the model to learn the underlying spatial distribution patterns of sea temperature fields across both horizontal and vertical dimensions. Second, during the reconstruction phase, sparse observational data of both surface and subsurface measurements are incorporated as guidance. These observations dynamically guide the reverse sampling trajectory of the diffusion process, enabling the model to reconstruct multi-layer sea temperature fields that not only match the provided observations but also preserve the learned global statistical distribution with physical consistency of the ocean temperature field.

3.1 Pretrain Physically Consistent Diffusion Model

The physically consistent diffusion model is obtained by using the classical denoising diffusion probability model (DDPM) [Ho et al., 2020]. Specifically, we pre-train the model using a large corpus of historical ocean temperature profile data generated by numerical simulations. It is well-known that ocean temperature fields exhibit fundamental and stable physical distribution patterns on a global scale. For example, sea surface temperatures are typically higher near the equator and decrease progressively toward the polar regions [Oceanic and Administration, 2022]. Additionally, due to the effects of oceanic circulation, different ocean basins exhibit distinct temperature characteristics [Oceanic and Administration, 2024a]. For instance, the surface temperatures of the Pacific Ocean are generally higher than those of the Atlantic, and the deep-water masses in various basins also follow corresponding distribution patterns. These statistical characteristics, derived from extensive historical observations and numerical simulations, imply the underlying physical laws governing the distribution of ocean temperatures. Such basic distribution can be modeled by DDPM, effectively capturing the spatial distribution patterns and physical consistency of ocean temperature fields. This ensures that the subsequent reconstruction process produces physically plausible results that conform to both observational constraints and large-scale statistical regularities.

Algorithm 1 Observation Guided Reconstruction

Input: Sparse observation data y, gradient scale s, distance measure $\mathcal{L}$, Gaussian Kernal μ, σ.
Output: Output image x_0 conditioned on y
Sample x_T from $\mathcal{N}(0, \mathbf{I})$
for t from T to 1 **do**
  $\hat{x}_0 = \frac{x_t - \sqrt{1-\alpha_t}\epsilon_\theta(x_t, t)}{\sqrt{\alpha_t}}$
  $g = -s \cdot \nabla_{x_0} L(\hat{x}_0, y)$
  $g = \text{SoftExtension}(g, \mu, \sigma)$
  Sample $x_{t-1} \sim \mathcal{N}(\mu + \Sigma g, \Sigma)$
end for
return x_0

The method for pre-training our base DDPM includes two stages: the **forward diffusion process** and the **reverse denoising process** [Ho et al., 2020]. In the forward process, a Markov chain is constructed to progressively add noise to the original data sample $x_0 \sim p_{\text{data}}$ by $x_t = \sqrt{1-\beta_t}x_{t-1} + \sqrt{\beta_t}\boldsymbol{\epsilon}_t$ to get $x_1, x_2, \cdots, x_T$. The mathematical formulation of the forward diffusion process is given by:

$$q(x_1, \cdots, x_T|x_0) = \prod_{t=1}^{T} q(x_t|x_{t-1}), \tag{1}$$

where t denotes as diffusion step, $q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I)$, and β_t are variance schedule. According to the additivity property of Gaussian distributions, the distribution of x_t in the forward diffusion process at step t is formulated by $x_t = \sqrt{\bar{\alpha}}x_0 + \sqrt{1-\bar{\alpha}}\boldsymbol{\epsilon}_t$, where $\epsilon \sim \mathcal{N}(0, \mathbf{I})$,$\alpha_t = 1 - \beta_t$ and $\bar{\alpha} = \prod_{i=1}^{t} \alpha_i$. Therefore, $q(x_t|x_0) = \mathcal{N}(x_t; \sqrt{\bar{\alpha}}x_0, (1-\bar{\alpha})I)$. $\bar{\alpha}$ goes to 0 when t goes to large T, and eventually $x_T \sim \mathcal{N}(0, 1)$.

In the reverse process, we sample reverse steps $q(x_{t-1}|x_t)$ until reaching x_0, which can also be a Markov Process defined as:

$$p_\theta(x_0, \cdots, x_{T-1}|x_T) = \prod_{t=1}^{T} p_\theta(x_{t-1}|x_t). \tag{2}$$

To make the reverse progress learnable, Sohl-Dickstein et al. [2015] proposed to train a neural network to predict a mean μ_θ and a diagonal covariance matrix Σ_θ:

$$q(x_{t-1}|x_t, x_0) = \mathcal{N}(x_{t-1}; \tilde{\mu}(x_t, x_0), \tilde{\beta}_t I), \tag{3}$$

$$\tilde{\mu}(x_t, x_0) = \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}x_0 + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}x_t. \tag{4}$$

The ReconMOST effectively learns the latent horizontal and vertical distribution patterns of ocean temperature fields and their inter-layer interactions. Random samples generated from the model naturally exhibit spatial structures consistent with fundamental oceanographic principles.

3.2 Observation Guided Reconstruction

We perform conditional sampling reconstruction based on the pretrained DDPM model, using a set of known observations, such as buoy data, as constraints. Specifically, the sampling process is guided by gradients derived from these observations, enabling the reconstruction of accurate multi-layer global sea temperature fields from sparse observation points. The reconstruction procedure is detailed in Algo. 1.

3.2.1 Gradient Guidance

In this conditional diffusion setting, the condition is defined by the observed sea temperature data x_0, represented as a sparsely masked y where only a few observation points remain, consistent with buoy-type data. The main objective is to reconstruct an accurate global sea temperature x_0 conditional on sparse observations y and intermediate sampling result x_t.

Sohl-Dickstein et al. [2015] has proven that:

$$\log p_\theta(x_{t-1}|x_t, y) = \log\left(p_\theta(x_{t-1}|x_t)p(y|x_t)\right) + C_1 \approx \log p(z) + C_2, \tag{5}$$

where $z \sim \mathcal{N}(\mu(x_t, t) + \Sigma g, \Sigma)$ and $g = \nabla_{x_t} \log p(y|x_t)|_{x_t=\mu(x_t,t)}$, $\Sigma = \Sigma_\theta(x_t)$. Dhariwal and Nichol [2021], and Ho and Salimans [2022] have provided concrete implementations.

For $p(y|x_t)$, we make an intuitive assumption that the probability of generating observations sea temperature data y from x_t is equivalent to that from the reconstructed $\hat{x}_0$ obtained via directly reverse diffusion from x_t. Accordingly, this conditional probability can be evaluated by the difference between reconstruction sea temperature $\hat{x}_0$ and observation sea temperature y:

$$p(y|x_t) \approx \exp\left(-s \cdot L(\hat{x}_0(x_t), y)\right), \tag{6}$$

$$\nabla_{x_t} \log p(y|x_t) \approx -s \cdot \nabla_{x_t} L(\hat{x}_0(x_t), y). \tag{7}$$

At each reconstructing(denoising) step from x_t to x_{t-1}, we steer the distribution of x_{t-1} towards the actual sea temperature distribution associated with the observed values y. This is achieved by computing the discrepancy between the reconstructed $\hat{x}_0$ from x_t and the observations y, and using the resulting gradient to adjust the reconstructing direction of x_t [Fei et al., 2023]. In this work, we approximate the gradient with respect to x_t by that of x_0, namely $g = \nabla_{x_0} \log p(y|x_t)|_{x_t=\mu(x_t,t)}$, which preserves the effectiveness of the heuristic formulation for $p(y|x_t)$ and achieves satisfactory reconstruction accuracy. By incorporating this gradient-based guidance into the sampling process, the method not only improves reconstruction accuracy at the observation points but also leverages the physical consistency embedded in the pretrained model to produce plausible and coherent reconstructions in unobserved regions.

3.2.2 Extension of Observation Guidance

Given the extreme sparsity of current buoy observations, directly applying observation-derived gradients results in insufficient constraints for unobserved regions. Inspired by Tobler's first law of geography, which states that all spatial variables are related, with nearby values being more strongly correlated [Tobler, 1970, Huang et al., 2024], we introduce a Soft-Extension strategy. Without altering the sparsity pattern of observation points, this method spatially diffuses the gradients g of observed points to their neighbors via a Gaussian kernel as Fig. 3. Since sea temperature fields typically exhibit continuous and smooth spatial variations, nearby points are expected to have similar values. The Soft-Extension effectively enhances local reconstruction accuracy while preserving the sparse observational constraints.

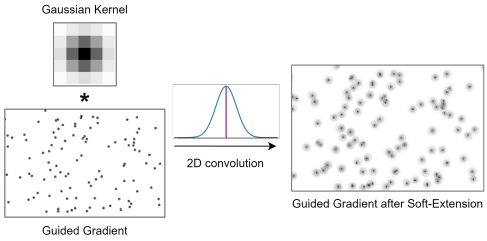


Figure 3: Demonstration of Soft-Extension

Before Extension: Only observed points carry gradients, leaving surrounding regions unconstrained.

After Extension: Gradients at observed points are spatially diffused to neighboring regions with a Gaussian weighting, enabling nearby points to acquire indirect constraints and improving local reconstruction quality.

We do not perform additional extensions in the vertical dimension for two main reasons. First, the dataset exhibits heterogeneous depth intervals and representational scales across different vertical layers, which may compromise spatial consistency if directly extended. Second, the vertical distribution of sea temperature typically follows a relatively stable pattern compared to the horizontal direction, making additional model extensions in the vertical dimension unnecessary. Thereby, we avoid redundant complexity and improving computational efficiency.

4 Experiments and Analysis

4.1 Datasets

CMIP6. Due to the scarcity of monthly ocean data, we utilize historical simulations from seven models that participated in the Coupled Model Intercomparison Project Phase 6 (CMIP6) [Eyring et al., 2016] to increase the

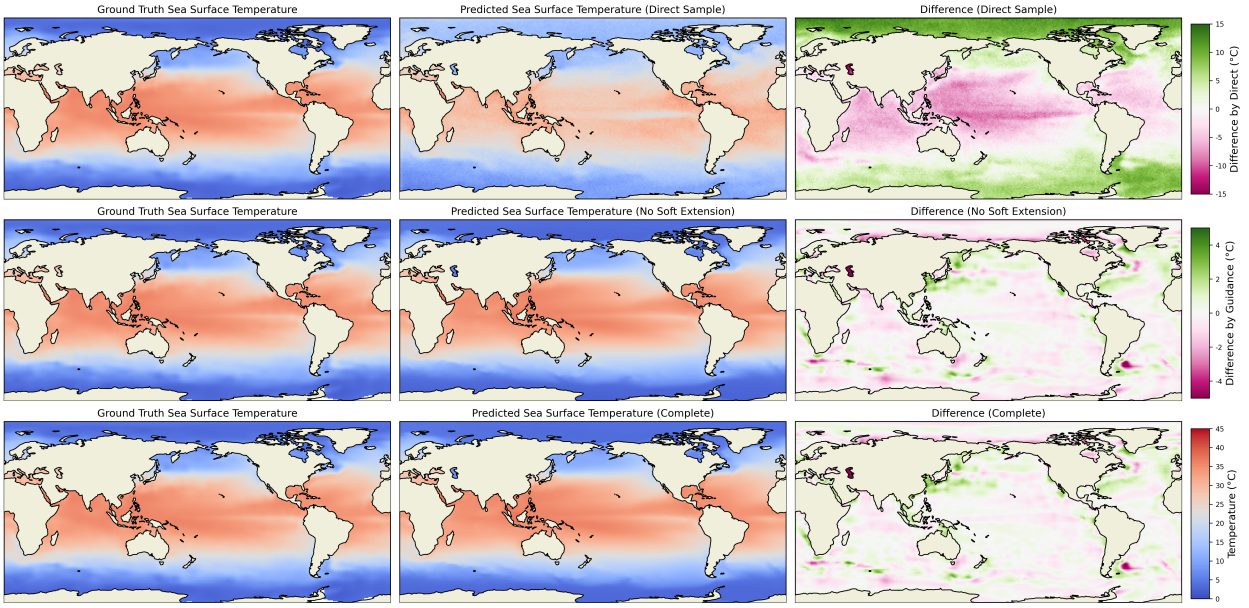


Figure 4: Comparison of SST reconstruction of EN4 dataset in ground truth, reconstruction results, and difference: for each role, the figure shows ground truth, reconstruction result and difference of SST between ground truth and reconstruction result with no guidance, guidance with no soft extension, and complete guidance.

amount of training data and also reduce systematic biases among dynamical models. CMIP6 comprises a suite of comprehensive Earth system models developed by leading climate research centers to simulate past, present, and future climate conditions under standardized experimental protocols. Here, historical simulations from these Earth system models can help our model obtain more robust predictions that capture the interconnected dynamics across ocean basins, thus enhancing model generalization and predictive capability.

EN4 is a global ocean temperature and salinity dataset developed by the UK Met Office [Good et al., 2013], providing quality-controlled subsurface ocean profile observations from various sources, including Argo floats [Roemmich et al., 2009], expendable bathythermographs (XBTs) [Goni et al., 2019], and ship-based measurements. Available from 1900 to the present, EN4 offers monthly objective analyses of temperature and salinity at standard depth levels, making it a valuable resource for climate studies and ocean model validation. In this study, we utilize upper ocean temperature data from EN4 during the 21st century (i.e., after 2000) as ground truth to guide and evaluate our proposed method for complete ocean temperature reconstruction.

4.2 Experimental Setups

The DDPM model is configured as follows: 'in_channels' is 42, corresponding to 42 ocean temperature layers. The Unet backbone adopts channel-mult (1, 2, 2, 2, 4, 4), progressively increasing feature channels to capture multi-scale spatial patterns. The base number of channels is set to 128, with 3 residual blocks per layer. The number of diffusion steps is 1000, with a linear noise schedule. In the **training phase**, the learning rate is 1e-4, and the batch size is 4. We trained each model for 200,000 steps, requiring approximately 24 hours on a single A100 GPU. In the **inference phase**, we independently reconstructed each observation sparsity month four times. Observation points are randomly sampled from the global sea temperature field according to the specified 'guided_rate', with consistent ratios across all layers but varied locations in each trial. The final performance is evaluated by averaging the MSE over four reconstruction results with ground truth temperature. The default setting of 'guided_rate' is 0.075, which means selecting 7.5% of grid points as known observations, matching real-world buoy distribution densities. During reconstruction, the Euclidean distance between $\hat{x}_0(x_t)$ and observation y is used as the guided distance function to compute the gradient-based guidance term. Additionally, $\mathbf{\Sigma}$ can optionally be set to 0 (controlling noise influence), while the **guidance strength** s and the **Gaussian kernel size for soft-extension** are treated as hyperparameters, alterable according to observation sparsity and spatial distribution characteristics.

Table 1: Guided reconstruction result of CMIP6: train and validation datasets are from the same mode, and the last three columns are respectively average MSE of upper ten levels on guidance points (MSE-g), reconstruction points (MSE-r) and all complete global points (total MSE).

mode	s	sigma	guided rate(#)	soft-ext size	MSE-g	MSE-r	total MSE
FGOALS-f3-L	4	0	7.5%(3188)	5	0.046	0.607	0.565
CanESM5	4	0	7.5%(3188)	5	0.043	0.559	0.521
IPSL-CM6A-LR	4	0	7.5%(3188)	5	0.044	0.585	0.544
FGOALS-g3	4	0	7.5%(3188)	5	0.045	0.585	0.545
BCC-CSM2-MR	4	0	7.5%(3188)	5	0.033	0.428	0.398
FIO-ESM-2-0	4	0	7.5%(3188)	5	0.005	0.068	0.063
MRI-ESM2-0	4	0	7.5%(3188)	5	0.025	0.330	0.308

4.3 Effectiveness Across Various Data Modalities

To validate the effectiveness of the proposed method, we selected seven modes in CMIP6 with large data volume to conduct experiments. Our ReconMOST works across all seven modes. Specifically, for each model with 3 or 5 initializations spanning 165 years (1980 samples), independent training and testing were conducted by reserving the last initialization as the test set and using the others for training. Only data from 2000 to 2014 (180 samples) were used for testing, as the number of buoy observations before 2000 was too limited for practical application. As described in the settings, each case was reconstructed four times and computed the mean MSE. The results in Table 1 demonstrate that the pretrained model effectively captures the implicit physical patterns of sea temperature distributions when the training and testing models are consistent. The best-performing model achieved the reconstruction with the lowest MSE of 0.063.

Table 2: Guided reconstruction result of EN4 analysis: train dataset is CMIP6 historical numerical simulation profile, whose total MSE of guided reconstruction is top k least on EN4 analysis, and test dataset is EN4 analysis. The last three columns are respectively average MSE of upper ten levels on guidance points (MSE-g), reconstruction points (MSE-r) and all complete global points (total MSE).

mode	s	sigma	guided rate(#)	soft-ext size	MSE-g	MSE-r	total MSE
Top1	4	0	7.5%(3188)	5	0.055	0.737	0.686
Top3	4	0	7.5%(3188)	5	0.091	1.229	1.144
Top5	4	0	7.5%(3188)	5	0.051	0.696	0.648
Top7	4	0	7.5%(3188)	5	0.053	0.711	0.662

4.4 Generalization Capabilities Across Diverse Scenarios

This experiment aims to evaluate the generalization capability of our ReconMOST. We investigate whether a pretrained model can achieve high-precision multi-layer sea temperature reconstruction on datasets derived from unseen simulation modes. This setting closely reflects real-world application scenarios, where temporal, climatic, and noise variations pose significant challenges to model generalization. We applied the seven models pretrained during the effectiveness verification phase to the EN4 analysis dataset, using global average MSE as the evaluation metric, where a lower MSE indicates stronger generalization. Additionally, we selected the Top1, Top3, and Top5 modes (ranked by MSE performance), along with the complete seven-mode dataset, to retrain DDPM models. This strategy balances generalization capacity with training data volume to capture more diverse distribution patterns. The results in Table 2 show that the Top5 model achieves the best performance of 0.648, while the Top3 performs worst, likely due to increased distributional variance arising from a smaller mixture of modes. Meanwhile, Top1 model also performed well, with only 0.038 difference from the best model, which means that we can achieve good generalization performance by learning from just one single best mode. So as long as the pre-trained analytical data physicality is good enough, one mode is sufficient. Moreover, we visualize our reconstruction results as shown in Fig. 4, where the first row is the unconditional generation result using the pretrained model, and the other rows are conditional generation results guided by observations. Although there is a noticeable deviation between the unconditional generation result and the ground truth, the spatial patterns of temperature variation remain physically consistent.

Table 3: Guided reconstruction result of EN4 analysis with difference guided rate (number of guided points): use best model Top5 in Table 2 and other parameters s=4, Sigma=0, soft-ext size is 5.The last three columns are, respectively, average MSE of upper ten levels on guidance points (MSE-g), reconstruction points (MSE-r), and all complete global points (total MSE).

guided rate	# guided points(round)	MSE-g	MSE-r	total MSE
7.5%	3188	0.051	0.696	0.648
10%	4251	0.068	0.672	0.612
20%	8502	0.132	0.563	0.476
30%	12754	0.151	0.370	0.305
40%	17005	0.165	0.259	0.222

Table 4: MSE result of ablation on *sigma*, *s*, *softmask scale*. The last three columns are respectively average MSE of upper ten levels on guidance points (MSE-g), reconstruction points (MSE-r) and all complete global points (total MSE).

Config	s	sigma	guided rate(#)	soft-ext size	MSE-g	MSE-r	total MSE
Top5	4	0	7.5%(3188)	5	0.051	0.696	0.648
S-3	3	0	7.5%(3188)	5	0.055	0.739	0.688
S-5	5	0	7.5%(3188)	5	**0.049**	**0.680**	**0.633**
S-6	6	0	7.5%(3188)	5	0.049	0.689	0.641
EXT-3	4	0	7.5%(3188)	3	0.055	0.754	0.702
EXT-7	4	0	7.5%(3188)	7	0.052	0.717	0.667
EXT-9	4	0	7.5%(3188)	9	0.051	0.701	0.653
W.O Sigma=0	4	DDPM	7.5%(3188)	5	0.052	0.721	0.671
W.O. Extension	4	0	7.5%(3188)	0	0.112	1.479	1.376
W.O. Sigma & Ext	4	DDPM	7.5%(3188)	0	0.113	1.487	1.384

4.5 Scalability Under Increasing Observation Density

To evaluate the scalability of the proposed method under varying observational guidance ratios, we conducted reconstruction experiments with different proportions of observational points, considering potential future increases in buoy coverage. The results in Table 3 indicate a consistent and significant improvement in the accuracy of reconstruction as the guidance ratio increases. This observation is intuitive, as a higher density of observational points provides richer spatial information for reconstructing the temperature fields, thereby enhancing model performance. We hope that in the future, more buoy observations combined with our method will contribute to a better understanding of the ocean.

4.6 Ablation Studies

Guidance Strength *s*. In the reconstruction process, s controls the constraint strength imposed by the distance function on the observational guidance points. Theoretically, a smaller s results in a higher heuristic probability $p(y|x_t)$ for a given distance, allowing the reconstruction to tolerate greater deviations from the observed values while still achieving satisfactory global performance. To investigate the impact of s, we conducted an ablation study using the optimal Top5 model under varying integer values of s. The results in Table 4 reveal that the accuracy of reconstruction initially improves and subsequently declines as s increases. This suggests that excessively strict constraints enforcing exact matching with observations may impair overall reconstruction performance. Therefore, selecting an appropriate s requires balancing constraint intensity and global accuracy of reconstruction, potentially in conjunction with other hyperparameters and the spatial distribution of observation points.

W.O. *Sigma*. Experimental results indicate that during the gradient-guided sampling process, retaining the original diffusion variance Σ when computing the guiding term g for updating μ, while setting the diffusion variance Σ to zero in the sampling step from x_t to x_{t-1} with new μ, significantly improves the accuracy of reconstruction. This configuration avoids unnecessary noise accumulation in the diffusion process, stabilizing the sampling trajectory and enhancing reconstruction performance under sparse observational guidance (Table 4).

Extension Kernel Size In the soft-extension module, the Gaussian kernel size determines the horizontal influence range of temperature observations. A larger kernel implies a stronger guiding effect of individual observation points on their surrounding areas. Experimental results from Table 4 indicate that increasing the kernel size consistently improves

the accuracy of reconstruction. Without applying the extension operation, the reconstruction performance becomes significantly worse. This can be attributed to the highly sparse distribution of observations, where local guidance alone is insufficient, and spatial diffusion of observation influence effectively enhances reconstruction performance. The second row in Fig. 4 illustrates the reconstruction results guided by sparse observations without applying soft-extension. Compared to the third row, which uses the complete ReconMOST framework, it is evident that ReconMOST achieves superior reconstruction accuracy in both open ocean and coastal regions. This highlights the critical role of soft-extension in enhancing reconstruction performance. These findings further validate the effectiveness of our ReconMOST.

5 Conclusion and Discussion

This paper proposes ReconMOST, a guided diffusion-based framework for ocean reconstruction tasks, achieving for the first time high-accuracy global multi-layer sea temperature field reconstruction. The method demonstrates excellent performance even under extremely sparse observation scenarios, with up to 92.5% of missing data, effectively addressing the limitations of the current global buoy observation network.

Despite its promising performance, several aspects of our ReconMOST might be further improved. First, we think the relationship between the guidance strength s, the soft-extension kernel size, and both the number and spatial distribution of observation data remains underexplored. Establishing a functional relationship or adaptive adjustment strategy is expected to enhance reconstruction accuracy. Second, the current framework is evaluated to static three-dimensional spatial reconstructions without incorporating temporal dynamics. Future work will extend our framework to spatiotemporal reconstructions.

Most importantly, ReconMOST provides a novel methodology and perspective for ocean reconstruction problems, with excellent generalization and scalability. It holds the promising potential for extension to other oceanic variables and diverse observation conditions, contributing to improved ocean monitoring, climate prediction, and sustainable marine resource development.

References

Gregory C Johnson, John M Lyman, and Norman G Loeb. Improving estimates of earth's energy imbalance. *Nature Climate Change*, 6(7):639–640, 2016.

Karina von Schuckmann, MD Palmer, Kevin E Trenberth, Anny Cazenave, Don Chambers, Nicolas Champollion, James Hansen, SA Josey, Norman Loeb, P-P Mathieu, et al. An imperative to monitor earth's energy imbalance. *Nature Climate Change*, 6(2):138–144, 2016.

Dean Roemmich, Gregory C Johnson, Stephen Riser, Russ Davis, John Gilson, W Brechner Owens, Silvia L Garzoli, Claudia Schmid, and Mark Ignaszewski. The argo program: Observing the global ocean with profiling floats. *Oceanography*, 22(2):34–43, 2009.

John P Abraham, Molly Baringer, Nathaniel L Bindoff, Tim Boyer, LJ Cheng, John A Church, Jessica L Conroy, Catia M Domingues, John T Fasullo, John Gilson, et al. A review of global ocean temperature observations: Implications for ocean heat content estimates and climate change. *Reviews of Geophysics*, 51(3):450–483, 2013.

Karina Von Schuckmann, Jean-Baptiste Sallée, D Chambers, Pierre-Yves Le Traon, C Cabanes, Fabienne Gaillard, Sabrina Speich, and M Hamon. Monitoring ocean heat content from the current generation of global ocean observing systems. *Ocean Science Discussions*, 10:923–949, 2013.

Roy Mendelssohn and CIndy Bessey. A reanalysis of north pacific sea surface temperatures using state-space techniques: The pdo in a new light. *arXiv preprint arXiv:1204.2486*, 2012.

Stephen Jewson, Roman Binter, Shree Khare, Kechi Nzerem, and Adam O'Shay. Predicting basin and landfalling hurricane numbers from sea surface temperature. *arXiv preprint physics/0701170*, 2007.

J Xavier Prochaska, Peter C Cornillon, and David M Reiman. Deep learning of sea surface temperature patterns to identify ocean extremes. *Remote Sensing*, 13(4):744, 2021.

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

Thomas M Smith and Richard W Reynolds. Extended reconstruction of global sea surface temperatures based on coads data (1854–1997). *Journal of Climate*, 16(10):1495–1510, 2003.

Youzhuan Ding, Dongyang Fu, Zhihui Wei, Zhihua Mao, and Juhong Zou. Reconstruction of incomplete satellite oceanographic data sets based on eof and kriging methods. In *Image and Signal Processing for Remote Sensing XIV*, volume 7109, pages 321–329. SPIE, 2008.

Aïda Alvera-Azcárate, Alexander Barth, Michel Rixen, and Jean-Marie 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.

Doug M Smith and James M Murphy. An objective ocean temperature and salinity analysis using covariances from a global climate model. *Journal of Geophysical Research: Oceans*, 112(C2), 2007.

Chunling Zhang, Danyang Wang, Zenghong Liu, Shaolei Lu, Chaohui Sun, Yongliang Wei, and Mingxing Zhang. Global gridded argo dataset based on gradient-dependent optimal interpolation. *Journal of Marine Science and Engineering*, 10(5):650, 2022.

Said Ouala, Cédric Herzet, and Ronan Fablet. Sea surface temperature prediction and reconstruction using patch-level neural network representations. In *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, pages 5628–5631. IEEE, 2018.

Qingyu Zheng, Guijun Han, Wei Li, Lige Cao, Gongfu Zhou, Haowen Wu, Qi Shao, Ru Wang, Xiaobo Wu, Xudong Cui, et al. Generating unseen nonlinear evolution in sea surface temperature using a deep learning-based latent space data assimilation framework. *arXiv preprint arXiv:2412.13477*, 2024.

Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34:8780–8794, 2021.

Silvia Alessio, Arnaldo Longhetto, and Luo Meixia. The space and time features of global sst anomalies studied by complex principal component analysis. *Advances in Atmospheric Sciences*, 16:1–23, 1999.

Ryuichi Kawamura. A rotated eof analysis of global sea surface temperature variability with interannual and interdecadal scales. *Journal of physical oceanography*, 24(3):707–715, 1994.

Siwei Tu, Weidong Yang, and Ben Fei. Taming generative diffusion prior for universal blind image restoration. *Advances in Neural Information Processing Systems*, 37:21172–21206, 2024.

Zhaoyang Lyu, Ben Fei, Jinyi Wang, Xudong Xu, Ya Zhang, Weidong Yang, and Bo Dai. Getmesh: A controllable model for high-quality mesh generation and manipulation. *arXiv preprint arXiv:2403.11990*, 2024.

Jinyi Wang, Zhaoyang Lyu, Dahua Lin, Bo Dai, and Hongfei Fu. Guided diffusion model for adversarial purification. *arXiv preprint arXiv:2205.14969*, 2022.

Zhengwentai Sun, Yanghong Zhou, Honghong He, and PY Mok. Sgdiff: A style guided diffusion model for fashion synthesis. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 8433–8442, 2023.

Xuanyuan Xie and Jieyu Zhao. Gdui: Guided diffusion model for unlabeled images. *Algorithms*, 17(3):125, 2024.

Wenjuan Zhang, Junlin Mei, and Yuxi Wang. Dmdiff: A dual-branch multimodal conditional guided diffusion model for cloud removal through sar-optical data fusion. *Remote Sensing*, 17(6):965, 2025.

Leo Klarner, Tim GJ Rudner, Garrett M Morris, Charlotte M Deane, and Yee Whye Teh. Context-guided diffusion for out-of-distribution molecular and protein design. *arXiv preprint arXiv:2407.11942*, 2024.

Hans Hersbach. The era5 atmospheric reanalysis. In *AGU fall meeting abstracts*, volume 2016, pages NG33D–01, 2016.

William J Gutowski Jr, Filippo Giorgi, Bertrand Timbal, Anne Frigon, Daniela Jacob, Hyun-Suk Kang, R Krishnan, Boram Lee, Christopher Lennard, Grigory Nikulin, et al. Wcrp coordinated regional downscaling experiment (cordex): a diagnostic mip for cmip6. *Copernicus GmbH*, 2016.

Julyan HE Cartwright and Oreste Piro. The dynamics of runge–kutta methods. *International Journal of Bifurcation and Chaos*, 2(03):427–449, 1992.

Siwei Tu, Ben Fei, Weidong Yang, Fenghua Ling, Hao Chen, Zili Liu, Kun Chen, Hang Fan, Wanli Ouyang, and Lei Bai. Satellite observations guided diffusion model for accurate meteorological states at arbitrary resolution. *arXiv preprint arXiv:2502.07814*, 2025.

Zhanxiang Hua, Yutong He, Chengqian Ma, and Alexandra Anderson-Frey. Weather prediction with diffusion guided by realistic forecast processes. *arXiv preprint arXiv:2402.06666*, 2024.

Donggeun Yoon, Minseok Seo, Doyi Kim, Yeji Choi, and Donghyeon Cho. Deterministic guidance diffusion model for probabilistic weather forecasting. *arXiv preprint arXiv:2312.02819*, 2023.

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

National Oceanic and Atmospheric Administration. How does the temperature of ocean water vary?, 2022. URL https://oceanexplorer.noaa.gov/facts/temp-vary.html. Accessed: 2025-04-21.

National Oceanic and Atmospheric Administration. Ocean temperature patterns are strongly controlled by ocean forces on a decadal timescale, 2024a. URL https://www.climate.gov/news-features/feed/ocean-temperature-patterns-are-strongly-controlled-ocean-forces-decadal. Accessed: 2025-04-21.

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International conference on machine learning*, pages 2256–2265. pmlr, 2015.

Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598*, 2022.

Ben Fei, Zhaoyang Lyu, Liang Pan, Junzhe Zhang, Weidong Yang, Tianyue Luo, Bo Zhang, and Bo Dai. Generative diffusion prior for unified image restoration and enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9935–9946, 2023.

Waldo R Tobler. A computer movie simulating urban growth in the detroit region. *Economic geography*, 46(sup1): 234–240, 1970.

Langwen Huang, Lukas Gianinazzi, Yuejiang Yu, Peter D Dueben, and Torsten Hoefler. Diffda: a diffusion model for weather-scale data assimilation. *arXiv preprint arXiv:2401.05932*, 2024.

Veronika Eyring, Sandrine Bony, Gerald A Meehl, Catherine A Senior, Bjorn Stevens, Ronald J Stouffer, and Karl E Taylor. Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization. *Geoscientific Model Development*, 9(5):1937–1958, 2016.

Simon A. Good, Matthew J. Martin, and Nick A. Rayner. En4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *Journal of Geophysical Research: Oceans*, 118(12):6704–6716, 2013. doi:https://doi.org/10.1002/2013JC009067. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2013JC009067.

Gustavo J Goni, Janet Sprintall, Francis Bringas, Lijing Cheng, Mauro Cirano, Shenfu Dong, Ricardo Domingues, Marlos Goes, Hosmay Lopez, Rosemary Morrow, et al. More than 50 years of successful continuous temperature section measurements by the global expendable bathythermograph network, its integrability, societal benefits, and future. *Frontiers in Marine Science*, 6:452, 2019.

Garrett Wilson and Diane J. Cook. A survey of unsupervised deep domain adaptation. *ACM Trans. Intell. Syst. Technol.*, 11(5), July 2020. ISSN 2157-6904. doi:10.1145/3400066. URL https://doi.org/10.1145/3400066.

Rachel Furner, Peter Haynes, Dave Munday, Brooks Paige, Daniel C Jones, and Emily Shuckburgh. A sensitivity analysis of a regression model of ocean temperature. *Environmental Data Science*, 1:e11, 2022.

A Derivation of Denoising Diffusion Probabilistic Models

Consider a data distribution $q(x_0)$ over continuous data $x_0 \in \mathbb{R}^d$. DDPM introduces a latent variable sequence $\{x_1, \ldots, x_T\}$ by defining a Markov chain that gradually adds Gaussian noise to the data [Ho et al., 2020]:

$$q(x_{1:T}|x_0) = \prod_{t=1}^{T} q(x_t|x_{t-1}) \tag{8}$$

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1-\beta_t}\, x_{t-1}, \beta_t \mathbf{I}), \tag{9}$$

where

$$\{\beta_t\}_{t=1}^{T}$$

is a predefined variance schedule satisfying $0 < \beta_t < 1$. Using the properties of Gaussians, the marginal distribution at any time t can be derived as:

$$q(x_t|x_0) = \mathcal{N}(x_t; \sqrt{\bar{\alpha}_t}\, x_0, (1-\bar{\alpha}_t)\mathbf{I}). \tag{10}$$

where $\bar{\alpha}_t = \prod_{s=1}^{t}(1-\beta_s)$. The generative process aims to learn the reverse Markov chain [Ho et al., 2020]:

$$p_\theta(x_{0:T}) = p(x_T) \prod_{t=1}^{T} p_\theta(x_{t-1}|x_t) \tag{11}$$

with prior $p(x_T) = \mathcal{N}(x_T; 0, \mathbf{I})$ and parameterized reverse transitions:

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t)). \tag{12}$$

By minimizing the variational bound on the negative log-likelihood:

$$\mathbb{E}_q \left[-\log p_\theta(x_0|x_1) + \sum_{t=2}^{T} D_{KL}\left(q(x_{t-1}|x_t, x_0) \| p_\theta(x_{t-1}|x_t)\right) + D_{KL}\left(q(x_T|x_0) \| p(x_T)\right) \right], \tag{13}$$

the model can be trained efficiently. It is further simplified by noting that the true posterior $q(x_{t-1}|x_t, x_0)$ is Gaussian with known closed-form expressions for its mean and variance:

$$q(x_{t-1}|x_t, x_0) = \mathcal{N}(x_{t-1}; \tilde{\mu}(x_t, x_0), \tilde{\beta}_t \mathbf{I}), \tag{14}$$

where

$$\tilde{\mu}(x_t, x_0) = \frac{\sqrt{\bar{\alpha}_{t-1}}\,\beta_t}{1-\bar{\alpha}_t} x_0 + \frac{\sqrt{1-\beta_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t} x_t, \tag{15}$$

and $\tilde{\beta}_t = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t$. However, in practice, we do not have access to the true x_0 during sampling, we use a rearranged forward process formula to express x_0 as:

$$\hat{\mathbf{x}}_0 = \frac{\mathbf{x}_t - \sqrt{1-\bar{\alpha}_t}\epsilon_\theta}{\cdot} \sqrt{\bar{\alpha}_t} \tag{16}$$

By substituting this estimated x_0 into the expression for $\tilde{\mu}_t$, we obtain an approximate mean for the reverse process transition:

$$\mu_\theta(x_t, t) = \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\hat{x}_0 + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t} x_t. \tag{17}$$

This formulation allows us to iteratively sample x_{t-1} from x_t using:

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}\left(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \Sigma_\theta(t)\right), \tag{18}$$

until reaching x_0 at the final step. In practice, the network is trained to predict the noise term ϵ using a U-Net network whose loss function is as follows:

$$L := E_{t\sim[1,T], x_0\sim q(x_0), \epsilon\sim N(0,\mathbf{I})}[\|\epsilon - \epsilon_\theta(x_t, t)\|^2]. \tag{19}$$

Note that L does not provide any learning signal for $\Sigma_\theta(xt, t)$. Ho et al. [2020] find that instead of learning $\Sigma_\theta(xt, t)$, they can fix it to a constant, choosing either $\beta_t\mathbf{I}$ or $\tilde{\beta}_t\mathbf{I}$. These values correspond to upper and lower bounds for the true reverse step variance.

B Observation Guidance

Here, we supplement the exploitation of observation data as guidance to improve the reconstruction. Specifically, we utilize a conditioner $p_\phi(y|x_t, t)$ and use $\nabla_{x_t} p_\phi(y|x_t, t)$ to guide the diffusion sampling process towards given observation y. We will describe how to use such conditioners to improve the accuracy of the reconstruction result. The notation is chosen as $p\phi(y|x_t, t) = p_\phi(y|x_t)$ and $\epsilon_\theta(x_t, t) = \epsilon_\theta(x_t)$ for brevity, referring to separate functions for each time step t.

B.1 Diffusion Process

According to Bayes' theorem:

$$p(x_t|x_{t+1}, y) = \frac{p(x_t, x_{t+1}, y)}{p(x_{t+1}, y)} = \frac{p(x_t|x_{t+1})p(x_{t+1})p(y|x_t, x_{t+1})}{p(y|x_{t+1})p(x_{t+1})} \tag{20}$$

$$= \frac{p(x_t|x_{t+1})p(y|x_t, x_{t+1})}{p(y|x_{t+1})} = \frac{p(x_t|x_{t+1})p(y|x_t)}{p(y|x_{t+1})} \tag{21}$$

$$\propto p(x_t|x_{t+1})p(y|x_t), \tag{22}$$

where the penultimate step is derived from :

$$p(y|x_t, x_{t+1}) = p(x_{t+1}|x_t, y)\frac{p(y|x_t)}{p(x_{t+1}|x_t)} = p(x_{t+1}|x_t)\frac{p(y|x_t)}{p(x_{t+1}|x_t)} = p(y|x_t). \tag{23}$$

The $\hat{p}(y|x_{t+1})$ can be treated as a constant because y does not depend on x_{t+1}. Therefore, we want to sample from the distribution $C \cdot p(x_t|x_{t+1})\hat{p}(y|x_t)$ where C denotes the normalization constant. $p(x_t|x_{t+1})$ is just the $p_\theta(x_t|x_{t+1})$ approximated by a neural network, so the rest is $\hat{p}(y|x_t)$ that can be obtained by computing a conditioner (evaluated by the distance between y and $\hat{x}_0(x_t)$) on noised images x_t.

B.2 Guided Reverse Process

Assume a diffusion model with an unconditional reverse noising process is $p_\theta(x_t|x_{t+1})$. In sea temperature reconstruction, we will obtain some sparse observation data and aim to impute the whole temperature field. Therefore, we regarded y as sparse observation data, x_t as the generated reconstruction in time step t, $\hat{x}_0(x_t)$ as the final reconstruction result from x_t. Then the conditioner is formulated as follows:

$$p(y|x_t) = exp(-s \cdot \mathcal{L}(\hat{x}_0(x_t), y)), \tag{24}$$

where $\mathcal{L}$ stands for Euclidean distance between $\hat{x}_0(x_t)$ and y. We have known that

$$p_\theta(x_t|x_{t+1}, y) = C \cdot p_\theta(x_t|x_{t+1})p(y|x_t), \tag{25}$$

where $p_\theta(x_t|x_{t+1})$ is the unconditional reverse process in denoising diffusion probabilistic models, which can be approximated as a perturbed Gaussian distribution $p_\theta(x_t|x_{t+1}) = \mathcal{N}(\mu, \Sigma)$:

$$\log p_\theta(x_t|x_{t+1}) = \mathcal{N}(\mu, \Sigma) = -\frac{1}{2}(x_t - \mu)^T\Sigma^{-1}(x_t - \mu) + C. \tag{26}$$

We can assume that $\log p(y|x_t)$ owns low curvature when compared with Σ^{-1}. This assumption is reasonable under the constraint that the infinite diffusion step, where $\|\Sigma\| \to 0$. Under the circumstances, $\log p(y|x_t)$ can be approximated via a Taylor expansion around $x_t = \mu$ as:

$$\log p(y|xt) \approx \log p(y|x_t)|_{x_t=\mu} + (x_t - \mu)\nabla x_t \log p(y|x_t)|_{x_t=\mu} \tag{27}$$

$$= (x_t - \mu)g + C_1, \tag{28}$$

where $g = \nabla x_t \log p(y|x_t)|_{x_t=\mu}$, and C_1 is a constant. This gives [Dhariwal and Nichol, 2021]:

$$\log(p_\theta(x_t|x_{t+1})p(y|xt)) \approx -\frac{1}{2}(x_t - \mu)^T\Sigma^{-1}(x_t - \mu) + (x_t - \mu)g + C_2 \tag{29}$$

$$\approx -\frac{1}{2}(x_t - \mu - \Sigma g)^T\Sigma^{-1}(x_t - \mu - \Sigma g) + \frac{1}{2}g^T\Sigma g + C_2 \tag{30}$$

$$\approx -\frac{1}{2}(x_t - \mu - \Sigma g)^T\Sigma^{-1}(x_t - \mu - \Sigma g) + C_3 \tag{31}$$

$$= \log p(z) + C_4, z \sim \mathcal{N}(\mu + \Sigma g, \Sigma), \tag{32}$$

where the constant term C_4 could be ignored. In practice, we use $g = \nabla x_0 \log p(y|x_t)|_{x_t=\mu} = \nabla x_0 \log \mathcal{L}(\hat{x}_0(x_{t+1}))$ instead, which simplifies the calculation and maintains great performance.

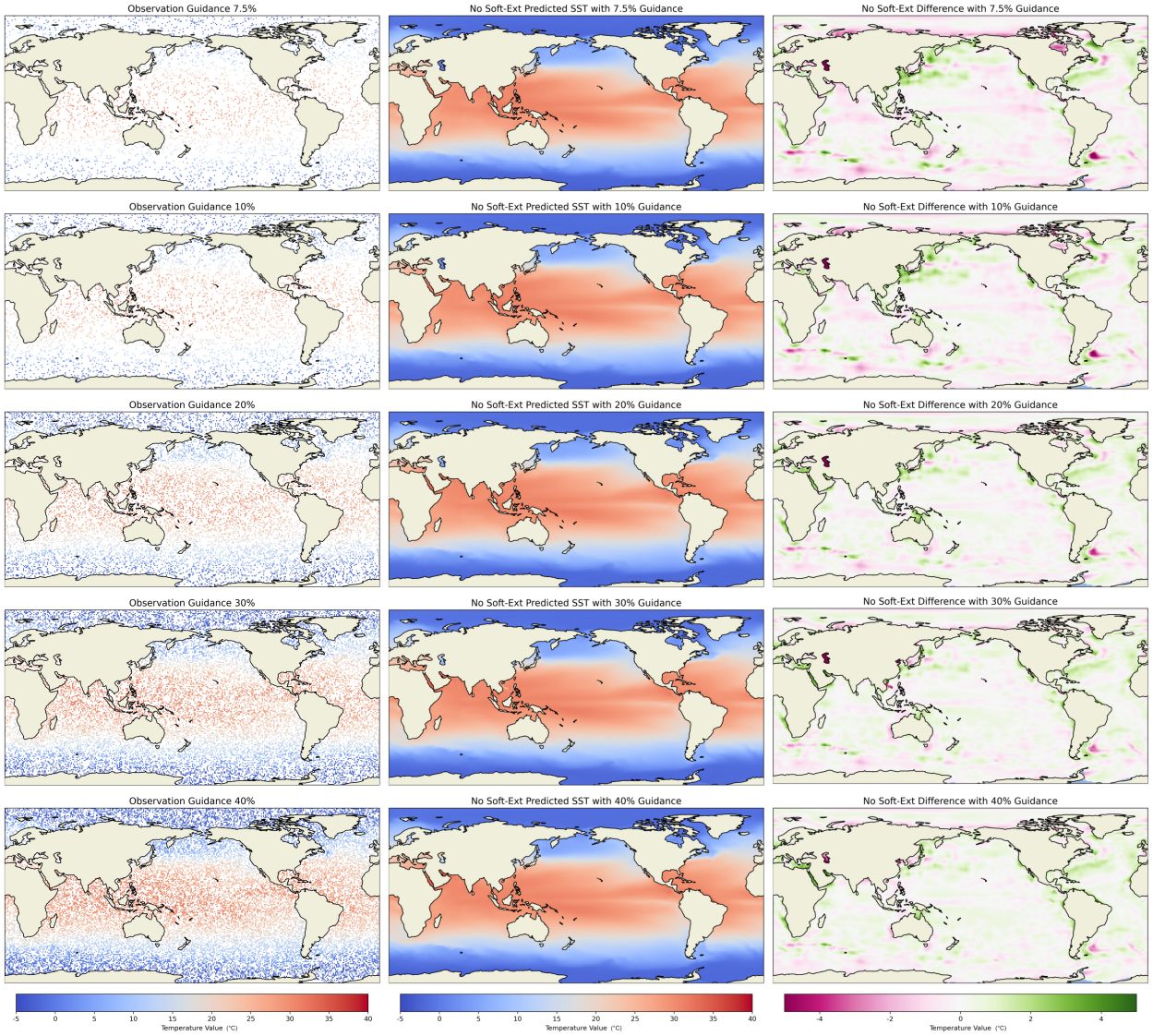


Figure 5: Comparison of increasing observation density with no soft extension, separately 7.5%, 10%, 20%, 30%, 40% from top to bottom.

C Visualization of Scalability Under Increasing Observation Density

We also visualize the reconstruction result of different guided rates to show the scalability under increasing observation density. We evaluate two variants: ReconMOST with and without the soft-extension module. Fig. 5 demonstrates the reconstruction result without soft-extension, demonstrating that higher guidance density leads to improved accuracy in reconstruction. However, the soft-extension may cause variance in the equatorial region and loss of otherwise good physical consistency when guidance density increases, as shown in Fig. 6. This occurs because sufficiently dense observations enable accurate temperature field reconstruction through guided reverse sampling alone, without requiring soft-extension. Adding extra information via a Gaussian kernel will destroy the learned local physical distribution underlying the pre-trained model, but most areas performed better than that without soft-extension.

D Visualization Across Temporal Dimension

We chose different timestamps (every two months in 2023) to indicate the temporal compatibility of our ReconMOST. As shown in Fig. 7, our method achieved excellent reconstruction performance at different times, only April and August

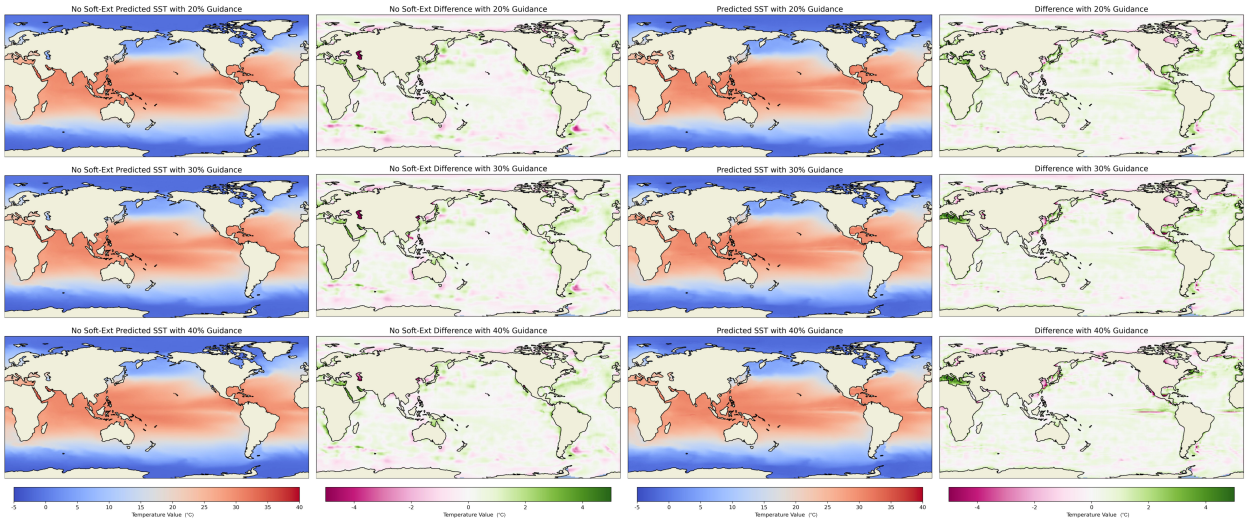


Figure 6: Comparison of complete method(right) and method with no soft extension at dense observation density(left), separately 20%, 30%, 40% from top to bottom.

Table 5: UNet reconstruction result of EN4 analysis with difference guided rate (number of guided points): use best model Top5 in Table 5. Guided rate during pre-training and number of training steps are displayed. The last three columns are, respectively, average MSE of upper ten levels on guidance points (MSE-g), reconstruction points (MSE-r), and all complete global points (total MSE).

pre-train guided rate	**guided rate**	**# training steps**	**MSE-g**	**MSE-r**	**total MSE**
7.5%	10%	200,000	7.975	0.870	7.264
7.5%	10%	800,000	4.946	0.530	4.504
7.5%	40%	200,000	11.247	7.414	9.714
7.5%	40%	800,000	3.683	2.408	3.173
40%	10%	200,000	9.102	1.223	8.314
40%	10%	800,000	9.816	1.362	8.970
40%	40%	200,000	6.022	4.011	5.218
40%	40%	800,000	6.305	4.198	5.462

have relatively obvious local errors. These results confirm the temporal stability of our method and support its potential for temporal extension in future work.

E Compared with Vanilla U-Net Method

To evaluate the performance, robustness and adaptability of our ReconMOST for ocean temperature reconstruction, we compared it with a vanilla U-Net model. The architecture of U-Net here is identical to that we used for guided diffusion, ensuring the same number of model parameters. We pre-trained U-Net models with Top5 CMIP6 simulations ranked by MSE performance as defined in our Experiments and Analysis section. We evaluated model performance on the EN4 dataset from the Met Office. Results here in Table 5 show that our ReconMOST outperforms U-Net, with lower reconstruction loss and requiring fewer training steps. The performance gap here can be attributed to a distribution mismatch that exists between the CMIP6 model simulations and real-world EN4 analysis data, a common challenge in applying machine learning to physical systems [Wilson and Cook, 2020, Furner et al., 2022]. Unlike U-Net, our ReconMOST takes partial real-world observations to adjust its distributions, improving its accuracy and adaptability for sea temperature reconstruction. The percentage of observation points during pre-training that affects reconstruction quality is another point worth considering, and will be left for future work to discuss.

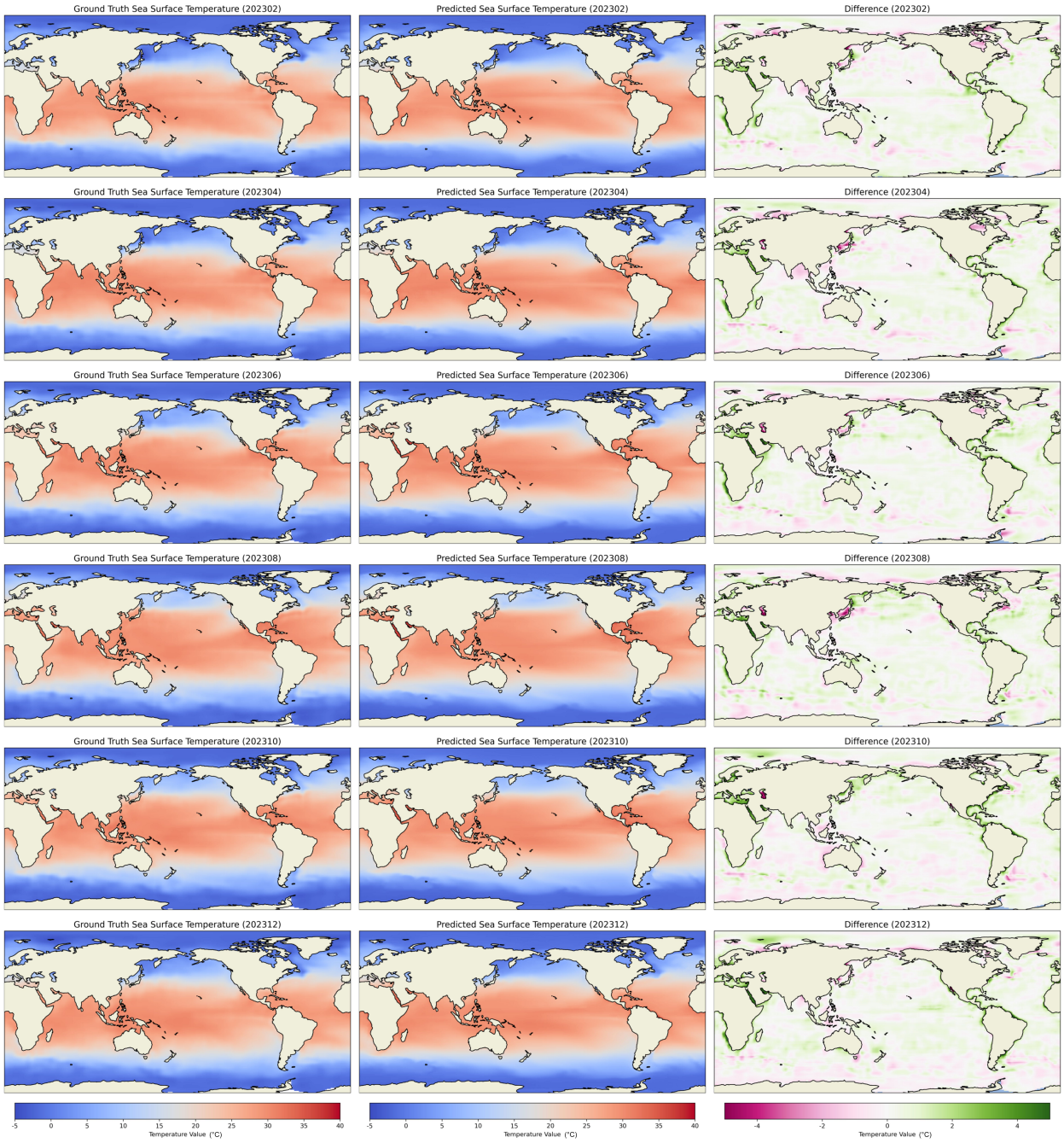


Figure 7: Enter Caption