4DVarGen: A 4D Variational-Inspired Generative Model for Eddy-Resolving Surface Ocean Reconstruction

Junpeng Huang[1] , **Wuxin Wang**[1,*] , **Xiaoyong Li**[1,*] , **Juan Zhao**[1,*] , **Senliang Bao**[1] , **Di Zhang**[1] and **Difu Sun**[1]

[1]National University of Defense Technology, Changsha, China
wuxinwang@nudt.edu.cn, sayingxmu@nudt.edu.cn, zhaojuan@nudt.edu.cn

Abstract

Sea surface variable reconstruction from sparse observations is a key ocean-science challenge. Traditional methods, such as the four-dimensional variational (4DVar) approach, rely on numerical models for background information, leading to high computational costs. Deep learning methods are more efficient but often fail to capture eddy dynamics, resulting in limited effective resolution. We propose 4DVarGen, a 4DVar-inspired generative framework for reconstructing sea surface variable fields at eddy-resolving scales from sparse remote-sensing observations. 4DVarGen establishes a mathematical equivalence between 4DVar and an observation-guided denoising process. Its key innovation is injecting the observation-likelihood gradient into denoising iterations, driving the generated trajectories to evolve in the direction of minimizing the 4DVar objective function toward a maximum a posteriori solution. Spatiotemporal priors learned by a diffusion model serve as background information, reducing computational costs and mitigating the adverse effects of Gaussian assumptions. Experiments show that 4DVarGen effectively leverages the temporal evolution patterns of sea surface temperature (SST) and sea surface height (SSH), as well as their dynamical mappings learned by the diffusion model, leading to improved reconstruction accuracy and effective resolution. Our model, pretrained on GLORYS12V1 reanalysis data, generates sea surface variable fields guided by real observations, achieving accuracy and effective resolution improvements of 18% and 58%, respectively, compared to GLORYS12V1. This study offers a novel framework for reconstructing Earth system states from sparse observations.

1 Introduction

Reconstructing SSH fields at eddy-resolving scales is crucial for understanding ocean dynamics and supporting related human activities [Ding *et al.*, 2022]. Currently, global SSH observations primarily rely on satellite-borne nadir radar altimeters and the wide-swath satellite altimeter, SWOT. The former provides only one-dimensional observations along the orbital path, with significant gaps between orbits. While the latter offers two-dimensional observations, its long revisit cycle leads to substantial temporal and spatial gaps in the observations [Archer *et al.*, 2025]. Such sparse and irregular observations make it difficult to effectively support the study of mesoscale ocean processes, making the sparse reconstruction of SSH fields one of the key issues in ocean science research.

Early work on ocean variable reconstruction was primarily based on statistical or dynamical formulation modeling [Hong *et al.*, 2025]. Such methods typically rely on numerical models to provide spatiotemporal background information, and the complexity of these models leads to high computational costs [Lellouche *et al.*, 2021]. Furthermore, when background information is utilized for reconstruction, background errors are commonly assumed to follow a Gaussian distribution to simplify computation. Although this assumption offers mathematical convenience, it may fail to accurately characterize actual background errors, potentially compromising reconstruction accuracy [Fablet *et al.*, 2023]. With the advancement of artificial intelligence technologies and the expansion of ocean data scales, deep learning-based reconstruction methods are increasingly adopted for ocean variable reconstruction [Yu *et al.*, 2025b]. These methods often operate within a supervised learning framework, training neural networks to learn nonlinear mappings between ocean variables from historical data, thereby enabling modeling of complex ocean systems [Yu *et al.*, 2025a]. Although this approach replaces certain modules (e.g., observation operators) in traditional methods with neural networks, reducing computational costs and significantly improving reconstruction quality, it still struggles to accurately represent eddy dynamics, resulting in limited effective resolution of reconstruction results [Martin *et al.*, 2024].

To address the aforementioned issues, this paper proposes a 4DVar-Inspired Generative Model (4DVarGen) for Eddy-Resolving Surface Ocean Reconstruction. This method establishes a mathematical equivalence between 4DVar and observation-guided inverse denoising processes, explicitly injecting the likelihood gradient computed from observations within a time window into denoising iterations. This transforms the dominant direction of trajectory generation into the gradient direction of the log posterior, aligning it with the minimization direction of the 4DVar objective function. At the same time, the denoising process uses the spatiotemporal prior learned by the diffusion model as background information, which avoids the high computational cost associated with traditional methods relying on numerical models and also mitigates the impact on reconstruction quality caused by assuming Gaussian-distributed background errors when computing the log-posterior probability function.

This study uses 4DVarGen to reconstruct ocean surface variable fields at eddy-resolving scales. Our contributions are summarized as follows:

- This study clarifies the equivalence between the minimization direction of the 4DVar objective function and the primary direction of observation-guided denoising, showing that 4DVar can be viewed as a Gaussian-prior special case of the proposed learned-prior framework.
- We propose 4DVarGen, a generative framework for reconstructing sea surface states from sparse satellite observations. By leveraging prior information learned from diffusion model, 4DVarGen addresses two key limitations of conventional 4DVar methods: first, it avoids the Gaussian assumption of background errors, which improves reconstruction quality; second, it avoids reliance on numerical models for providing background information, which reduces computational cost.
- Experiments demonstrate that the proposed method achieves reconstruction of sea surface variable fields at eddy-resolving scales. It effectively leverages temporal evolution patterns of SST and SSH learned by the diffusion model, as well as the dynamical mappings between SST and SSH, thereby enhancing reconstruction quality.

2 Related Works

Traditional ocean variable reconstruction methods primarily rely on dynamical constraints or statistical models to realize reconstruction. With the accumulation of ocean data and advancements in artificial intelligence, machine learning and deep learning methods are increasingly used for ocean variable reconstruction because of their ability to model complex nonlinear mappings [Wang *et al.*, 2025b].

2.1 Traditional Reconstruction Methods

Traditional reconstruction methods can be primarily categorized into statistical analysis methods and dynamical constraint methods. Statistical analysis methods establish mappings between different ocean variables through statistical or empirical formulas without introducing dynamical equations, and integrate satellite and in situ observations to complete reconstruction [Taburet *et al.*, 2019]. These methods feature low computational cost and high speed, but due to the lack of explicit dynamical constraints, the reconstructed results often exhibit deficiencies in physical consistency [Amores *et al.*, 2018]. Dynamically constrainted methods, such as 4DVar, establish mappings between ocean variables based on dynamic equations and utilize numerical models to assimilate and fuse multi-source observations across different time points [Chen *et al.*, 2023]. These methods possess explicit physical constraints, and their reconstructions generally satisfy fundamental physical laws such as geostrophic balance, hydrostatic equilibrium, and mass conservation [Liu *et al.*, 2021]. However, the numerical models they rely on are generally complex, resulting in high computational costs [Chen *et al.*, 2020]. Furthermore, traditional methods typically assume background errors follow a Gaussian distribution when utilizing background information provided by numerical models. While this assumption simplifies mathematical processing, it struggles to accurately describe real background errors, thereby affecting reconstruction quality [Zhou *et al.*, 2024].

To balance computational efficiency and physical consistency, researchers have attempted to combine statistical methods with dynamic constraint approaches. However, the logic of such hybrid methods often relies on manual design, making it difficult to fully extract complex information from high-dimensional, nonlinear ocean data [Yan *et al.*, 2021].

2.2 Supervised Learning-Based Reconstruction Methods

With the expansion of ocean data scale and continuous advancements in artificial intelligence, machine learning and deep learning methods have become important tools for ocean variable reconstruction [Liu and Wang, 2025]. Instead of explicitly solving physical equations, these methods learn nonlinear mappings between ocean variables from historical data, enabling efficient and flexible modeling of complex ocean systems [Cutolo *et al.*, 2024]. Early supervised learning approaches primarily relied on machine learning techniques, commonly employing methods such as Support Vector Machines (SVM), Random Forests (RF), and Self-Organizing Maps (SOM) to establish mappings between variables for reconstruction [Wang *et al.*, 2025a]. Although these methods offer significant improvements in accuracy and computational efficiency over traditional approaches, they typically rely on feature engineering and exhibit limited expressive power, making it challenging to fully characterize high-dimensional, multimodal, and strongly nonlinear ocean variable fields [Bao *et al.*, 2019]. With the advancement of artificial intelligence and the continuous accumulation of ocean observation and reanalysis data, deep learning methods have gained widespread attention in ocean variable reconstruction due to their automatic feature extraction and robust nonlinear modeling capabilities [Zhu *et al.*, 2025]. These methods have further enhanced reconstruction quality, yet most still rely on regression models that produce a single optimal solution, making it challenging to provide reliable uncertainty quantification for reconstruction results. This limitation is particularly evident under conditions of sparse ocean observations and complex error sources [QI *et al.*, 2024].

2.3 Generative Reconstruction Methods

Diffusion model, as generative deep learning models, can learn complex probability distributions of data through pretraining, making them ideal for representing highly nonlinear dynamical systems such as oceans and atmospheres [Chao *et al.*, 2025]. Many studies leverage this property by incorporating sparse observation constraints within a Bayesian framework for reconstruction under sparse observation conditions [Li *et al.*, 2024]. For instance, Martin et al. applied diffusion model to sparse reconstruction of sea surface variables within this framework, incorporating randomness during generation to quantify uncertainty for reconstructed structures [Martin *et al.*, 2025]. Currently, research on diffusion model for ocean variable reconstruction remains nascent, with significant exploration needed on how to leverage the temporal evolution patterns and dynamical mappings among sea surface variables for joint reconstruction. Mesoscale oceanic processes (such as mesoscale eddies and oceanic fronts) typically exhibit time scales ranging from days to months, implying relatively stable evolution of mesoscale features in surface ocean variables over short periods [Seo *et al.*, 2023]. Therefore, incorporating neighboring-day spatiotemporal data may aid surface-variable reconstruction at specific times, especially for sparsely observed variables such as SSH. In summary, investigating whether diffusion model can accurately learn the temporal evolution patterns of surface ocean variables and the dynamical mappings among multiple variables, and thereby effectively fuse observations to achieve high-quality reconstruction, holds significant importance.

3 Preliminary

3.1 Diffusion Model

Diffusion model comprise a forward noise addition process and a backward denoising process [Song *et al.*, 2021]. The forward process progressively adds noise to samples over time t via a linear stochastic differential equation (SDE), as shown in Figure 1(a). The backward denoising process is expressed by inverting the forward SDE:

$$dx(t) = [f(t)x(t) - \frac{1}{2}g^2(t)\nabla_{x(t)} \log p(x(t))]dt + g(t)dw(t), \tag{1}$$

where $f(t) \in R$ is the drift coefficient; $g(t) \in R$ is the diffusion coefficient; $w(t) \in R^D$ represents the Wiener process; the sample $x \in R^D$ follows the distribution $p(x)$; $x(t) \in R^D$ is the perturbed sample at time $t \in [0, 1]$. Since the forward SDE is linear in $x(t)$, the perturbation kernel from x to $x(t)$ is Gaussian. Thus, we readily obtain $\nabla_{x(t)} \log p(x(t)) = \frac{\epsilon}{\sigma(t)}$. In practice, the "score" $\nabla_{x(t)} \log p(x(t))$ is approximated by the score network $S_\theta(x(t), t)$, whose loss function is expressed as

$$\arg\min_\theta E_{p(x)p(t)p(\epsilon)} \left[\left\| S_\theta(x(t), t) - \frac{\epsilon}{\sigma(t)} \right\|_2^2 \right], \tag{2}$$

where $p(t) = U(0, 1)$; $p(\epsilon) = N(0, I)$; and $\sigma(t)$ denotes the noise intensity corresponding to time step t. When the score network is sufficiently trained, $S_\theta(x(t), t) \approx \nabla_{x(t)} \log p(x(t))$.

Finally, by employing appropriate discretization schemes for the inverse denoising process, such as variance exploding (VE) SDE and variance preserving (VP) SDE [Song and Ermon, 2019], and combining them with pretrained score networks, denoised generation can be achieved.

4 Method

Diffusion model can generate realistic ocean states when applied to ocean state estimation. However, the generated samples are unrelated to any observations and result from random sampling from the ocean data distribution. Therefore, we aim to introduce observational constraints so that the generated samples are no longer random but instead approximate observations at a specific time. 4DVar provides the theoretical basis and implementation framework for this purpose.

4.1 Mathematical Equivalence

The principle of 4DVar is as follows: under the constraint of dynamic trajectories constrained by observation within a time window, the state estimate corresponding to the maximum value of the posterior probability density function of the state vector is defined as the analysis field. The posterior can be written as:

$$p(x_{t_0} \mid y_{0:L}) \propto p(x_{t_0}) \prod_{k=0}^{L} p(y_{t_k} \mid H_k(M_{t_0 \to t_k}(x_{t_0}))), \tag{3}$$

where x_{t_0} is the state of variables at t_0, y_{t_k} is the observation at time t_k within the time window, L is the time window length, $M_{t_0 \to t_k}$ is the model propagator from t_0 to t_k, and $H_k(\cdot)$ is the observation operator at time t_k. The 4DVar objective function is as follows:

$$\begin{aligned} J(x_{t_0}) &\equiv -\log p(x_{t_0} \mid y_{0:L}) \\ &= -\log p(x_{t_0}) - \sum_{k=0}^{L} \log p(y_{t_k} \mid H_k(M_{t_0 \to t_k}(x_{t_0}))) \\ &\quad + C, \end{aligned} \tag{4}$$

where J is the 4DVar cost function and C is a constant independent of x_{t_0}. Taking the derivative yields the gradient direction of the objective function:

$$\begin{aligned} \nabla_{x_{t_0}} J(x_{t_0}) &= -\nabla_{x_{t_0}} \log p(x_{t_0}) \\ &\quad - \sum_{k=0}^{L} \nabla_{x_{t_0}} \log p(y_{t_k} \mid H_k(M_{t_0 \to t_k}(x_{t_0}))) . \end{aligned} \tag{5}$$

At the same time, after incorporating observational constraints, the dominant direction of the denoising process in diffusion model can be expressed as:

$$\begin{aligned} \nabla_X \log p(X \mid y_{0:L}) &= \nabla_X \log p(X) \\ &\quad + \sum_{k=0}^{L} \nabla_X \log p(y_{t_k} \mid H_k(X)) , \end{aligned} \tag{6}$$

where $X = x_{t_0:t_L}$. Since the score network learns spatiotemporal trajectory priors, it implicitly plays a role analogous to dynamical propagation in 4DVar in the update direction, avoiding reliance on explicit dynamical models. Therefore, when focusing on the variable x_{t_0} at time t_0, we establish the equivalence between the 4DVar objective-function minimization process and the observation-constrained denoising iterative process in the update direction, namely:

$$\nabla_{x_{t_0}} \log p(x_{t_0} \mid y_{0:L}) = -\nabla_{x_{t_0}} J(x_{t_0}). \tag{7}$$

Moreover, traditional 4DVar often assumes Gaussian background errors in practice, so that $\nabla_{x_{t_0}} \log p(x_{t_0})$ is easy to compute, namely:

$$\nabla_{x_{t_0}} \log p(x_{t_0}) = -B^{-1}(x_{t_0} - x_b(t_0)) , \tag{8}$$

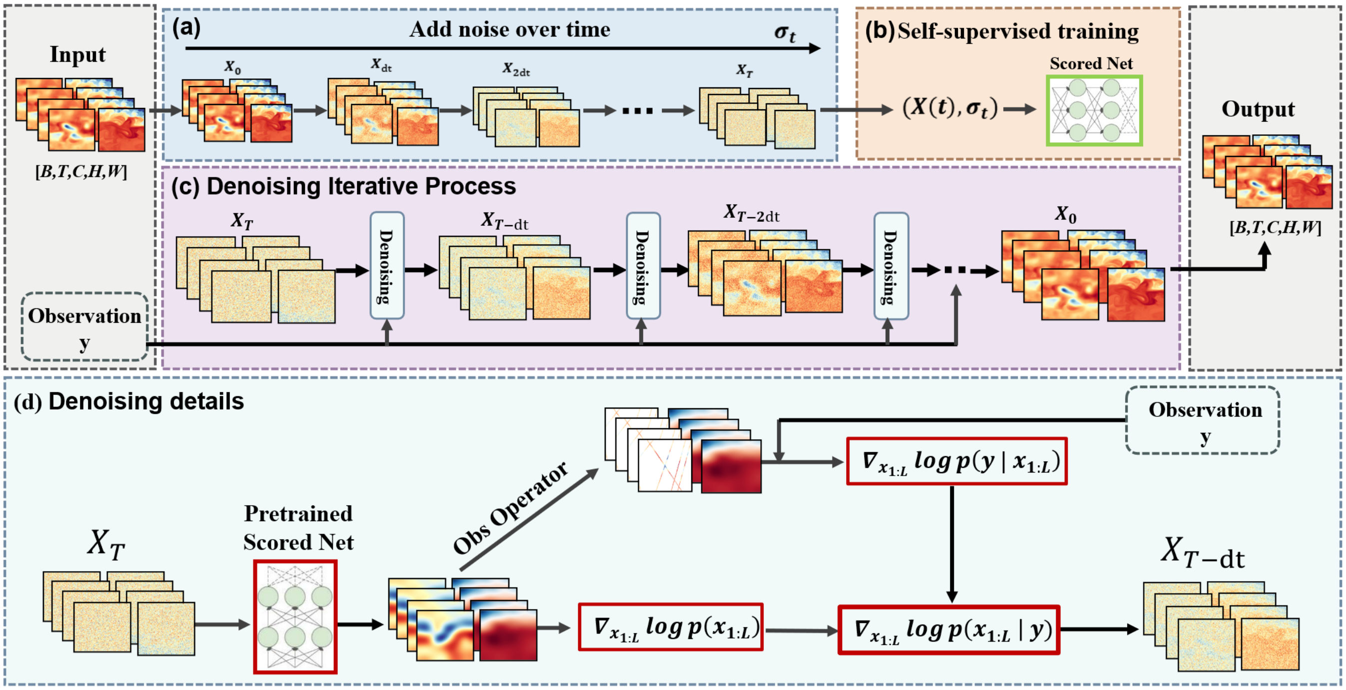

Figure 1: 4DVarGen Framework. (a) The process of progressively adding noise to the sample; (b) The score network undergoes self-supervised training using the perturbed samples after each noise addition step in (a) and their corresponding noise intensities as inputs; (c) The process of progressively denoising the noisy samples to generate the sea surface variable fields under observational constraints; (d) Details of a single denoising step in process (c).

where $x_b(t_0)$ is the background information and B is the background error covariance matrix. However, diffusion model estimate $\nabla_{x_{t_0}} \log p(x_{t_0})$ directly via the score network, thereby avoiding the potential impact of the Gaussian assumption on reconstruction quality. If the learned prior is Gaussian with covariance B, this score reduces to the 4DVar background gradient above; otherwise, it naturally represents non-Gaussian priors.

4.2 4DVarGen

4DVarGen first sequentially adds noise to spatiotemporal sea surface variables data to generate perturbed samples and their corresponding noise intensity data pairs. Then, using the data pairs obtained in the first step as input, it trains a score network through self-supervised learning to capture the temporal evolution patterns and dynamical mappings of sea surface variables. Subsequently, we introduce observational constraints based on 4DVar and perform denoising by integrating the background information, temporal evolution patterns, and dynamical mappings of sea surface variables provided by the pretrained score network. Finally, we iterate the third step to generate ocean surface fields.

Gradual Noise Addition

This study employs the VESDE strategy as the discretization scheme for the diffusion process, discretizing the stochastic differential equation into 2000 steps. Noise levels ranging from 0.1 to 20.0 are applied to incrementally add noise to spatiotemporal samples of sea surface variables. The perturbed samples and their corresponding noise intensities are retained for subsequent model training, as illustrated in Figure 1(a).

Self-Supervised Training

Unet-video [Li *et al.*, 2024] was selected as the backbone network for the score network, with its architecture shown in Figure 2. The score network takes the perturbed samples generated by the noise addition process and their noise intensities as input. Through self-supervised training, it learns background information, temporal evolution patterns, and dynamical mappings from the spatiotemporal data of sea surface variables, as illustrated in Figure 1(b). The pretrained score network provides background information and temporal evolution patterns for the subsequent denoising process, thereby avoiding the background error assumptions and high computational costs associated with complex numerical models and adjoint operators in traditional methods.

One-Step Denoising

At any step in the denoising process, the pretrained score network directly outputs the gradient of the background term and samples containing temporal information, using the perturbed sample and its corresponding noise intensity as input. Simultaneously, the observation term gradient is computed using observational data and the sample containing temporal information. Under the combined influence of the observation term gradient and background term gradient, the perturbed sample undergoes denoising updates along the direction of maximum a posteriori probability. The specific process is illustrated in Figure 1(c). During inference, the observation-gradient weight is scheduled relative to the prior score as 5%, 20%, and 55% for denoising steps 0–500, 500–800, and 800–1000, respectively, balancing early large-scale prior-driven generation with late-stage observation fitting.

Segmented Reconstruction Strategy

To enhance the utilization of temporal information and ensure temporal continuity of reconstructed results, this study employs a Segmented reconstruction strategy. The Segmented reconstruction strategy divides long temporal-sequence data into multiple overlapping temporal blocks for parallel reconstruction. This approach reduces the length of the time sequence that the score network needs to output, alleviating the high demands on data volume and computational resources when training long-sequence score networks. The overlapping regions between adjacent reconstruction blocks are utilized to compute the continuity loss. This loss term ensures smooth temporal transitions between blocks while enabling observational information to implicitly propagate backward through overlapping regions. With the introduction of the continuity loss, the backward denoising process is as follows:

$$
\begin{aligned}
dX_t = &-\frac{1}{2}\frac{d[\sigma_t^2]}{dt}\, S_\theta(X_t,t)\, dt + \sqrt{\frac{d[\sigma_t^2]}{dt}}\, dW_t \\
&- \alpha_t \nabla_{X_t} \|y - H(\hat{X}_t)\|_2^2 \, dt \\
&- \beta_t \nabla_{X_t} \sum_{i=1}^{B-1} \left(\sum_{l=i(L-N)+1}^{i(L-N)+N} \left\| \hat{x}^{\,l,(i+1)} - \mathrm{sg}(\hat{x}^{\,l,(i)}) \right\|_2^2 \right) dt
\end{aligned},
\tag{9}
$$

where $X_t = \left[X_t^{(1)}, \ldots, X_t^{(B)}\right]^\top$ and $X_t^{(B)}$ denotes the Bth sample block; $S_\theta(X_t,t) = \left[S_\theta(X_t^{(1)},t), \ldots, S_\theta(X_t^{(B)},t)\right]^\top$, with $S_\theta(X_t^{(B)},t)$ the score of the Bth block; $dW_t = \left[dW_t^{(1)}, \ldots, dW_t^{(B)}\right]^\top$, where $\{W_t^{(i)}\}_{i=1}^{B}$ are B independent standard Wiener processes; $\hat{X}_t$ is the long-sequence reconstructed sample obtained by stacking B reconstruction blocks; N is the overlap length; $\hat{x}^{\,l,(i)} \in \hat{X}_t^{(i)}$ is the state at the lth time step within the ith sequence; $\mathrm{sg}(\cdot)$ indicates no gradient is computed; and α_t and β_t are the observation and continuity constraint weights, respectively.

5 Experiments

The Observing System Simulation Experiment (OSSE) demonstrates how the model learns the temporal evolution patterns of sea surface variables and thereby contributes to reconstruction quality through the propagation of observations over time. The Observing System Experiment (OSE) compares the proposed method with state-of-the-art approaches to validate the effectiveness of 4DVarGen. The ablation experiment further analyzes whether the temporal evolution patterns of sea surface variables learned by the model from reanalysis data are applicable to the real world.

5.1 Experimental Setup

Data used in this study include SSH and SST products from the GLORYS12V1 reanalysis data (2010–2020); Sea Level Anomaly (SLA) from Global Ocean Along Track L3 satellite observations (2017) and SST from ODYSSEA L3 satellite observations (2017). It should be clarified that this study reconstructs SLA and then converts it back to SSH during evaluation. The study region was selected as the Gulf Stream area (65°W–55°W, 33°N–43°N), one of the most energetically intense and mesoscale eddy-active zones in the global ocean. The year 2017 was designated as the model evaluation period, with data from other years used for model training. Due to the irregular nature of satellite data, preprocessing is required to interpolate the data onto a regular latitude-longitude grid before inputting it into the model.

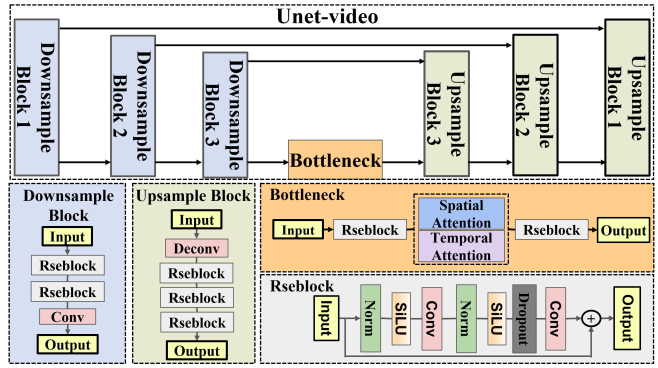

Figure 2: Architecture of the score network. The score network adopts a U-Net-video backbone composed of downsampling blocks, a bottleneck module, and upsampling blocks to learn spatiotemporal priors of sea surface variables.

Unless otherwise specified, the time window was set to 10 days, with a 2-day overlap. Training samples were extracted from the training sets based on the time window length and divided into training and validation sets in an 8:2 ratio according to time sequence. Model training comprised 300 epochs using the Adam optimizer [Kingma and Ba, 2015]. Validation followed each epoch, and the model with the lowest validation loss was saved for subsequent experiments. Further details on dataset construction and model training are provided in the Supplementary Materials. Finally, this study employs the following metrics [Martin *et al.*, 2023]:

- Accuracy Evaluation: $RMSE_{score}$ and its standard deviation quantify the deviation between the reconstructed field and the reference ground truth. $RMSE_{score}$ closer to 1 indicates higher accuracy;

$$RMSE_{score} = 1 - \frac{RMSE_{recon}}{RMS_{obs}} \tag{10}$$

- Dynamic Feature Assessment: Spatial effective resolution quantifies the spatial scale at which reconstructed results can be accurately resolved; Spectral analysis evaluates their kinetic consistency.

5.2 Observing System Simulation Experiment

We constructed pseudo-observation-truth data pairs based on satellite observations and the GLORYS12V1 reanalysis dataset. Using pseudo-observations as input, we visu-

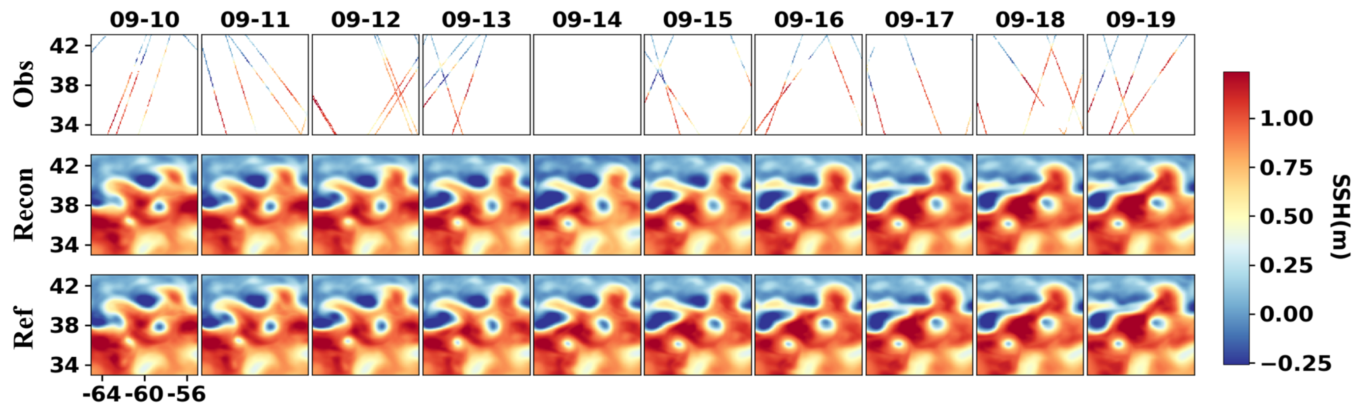

Figure 3: Visualization of OSSE results (2017.09.10-2017.09.19). The first row shows the observation field, where blank areas indicate no observations; 09-14 was a non-observation day. The second row displays the reconstruction results from 4DVarGen. The third row presents the true values (GLORYS12V1)

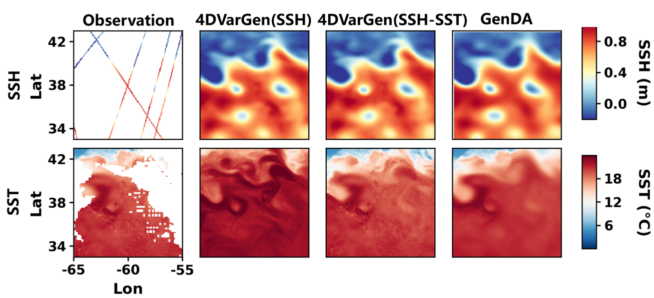

Figure 4: Visualization of OSE results (2017-01-03). Observations used by the proposed method, the corresponding reconstructions, and reconstructions from GenDA.

Method	$\mu\uparrow$	$\sigma\downarrow$	λ_x (km) $\downarrow$
DUACS	0.876	0.066	151
MOIST	0.886	0.086	139
DYMOST	0.888	0.065	130
4DVarNet	0.888	0.091	108
NeurOST(2024)	<u>0.899</u>	**0.063**	114
GenDA(2025)	**0.900**	<u>0.064</u>	109
GLORYS12V1	0.751	0.163	236
4DVarGen (SSH)	0.882	0.074	<u>105</u>
4DVarGen (SSH-SST)	0.889	0.067	**100**

Table 1: Quantitative comparison of reconstruction performance. μ denotes $RMSE_{score}$ and σ denotes the standard deviation of $RMSE_{score}$, and λ_x indicates the spatial effective resolution. A value of μ closer to 1 indicates better reconstruction quality, whereas smaller values of σ and λ_x indicate better performance. Bold values indicate the best performance, and underlined values indicate the second-best performance.

ally compared reconstruction results with the truth (GLORYS12V1). We analyzed results from September 10 to 19, 2017, with September 14 being an observation-free day.

As shown in Figure 3, the reconstructed SSH field on the non-observation day exhibits high spatial consistency with the truth. This indicates that the model effectively learned and utilized the temporal evolution patterns of SSH, enabling the transfer of observational information from other time periods to the non-observation day, thereby achieving SSH field reconstruction for non-observation days.

5.3 Observing System Experiment

This experiment evaluates 4DVarGen under real-observation scenarios. We use preprocessed SLA and SST observations as inputs to generate 50 ensemble members. The ensemble mean is used for quantitative comparison with deterministic products, while the ensemble spread serves as an uncertainty indicator. Since the true ocean state is unknown, CryoSat-2 SLA observations are retained for independent validation.

This study preliminarily assessed the effectiveness of the 4DVarGen method by comparing its visualization results with those of the advanced method (GenDA). As shown in Figure 4, the SSH fields reconstructed by 4DVarGen using only SSH observations, as well as those using both SSH and SST observations, were highly consistent with GenDA results. Furthermore, the comparison reveals that the SST reconstructed solely from SSH observations exhibits systematic

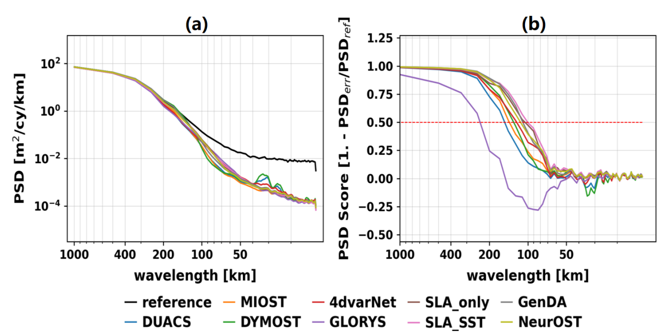

Figure 5: (a) SSH wavenumber spectrum plotted for each method on the retained CryoSat-2 observation track, along with the observed wavenumber spectrum (b) Spectral coherence for each method (Values closer to 1 indicate more accurate SSH signals at the corresponding scale).

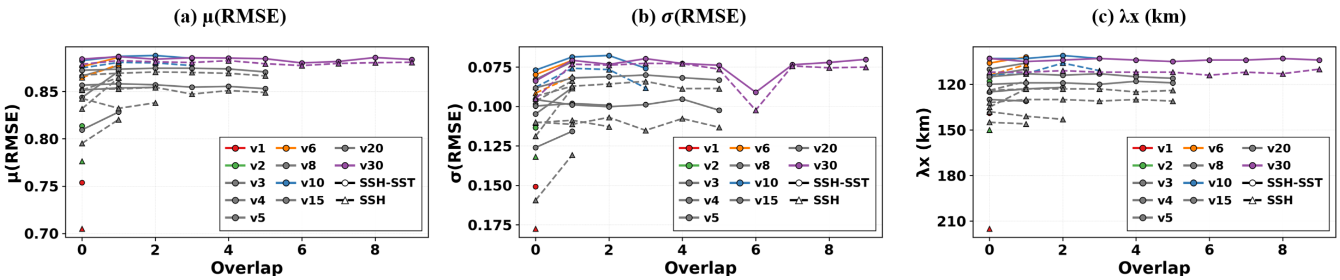

Figure 6: Ablation Experiment Results: (a) $RMSE_{score}$, (b) $RMSE_{score}$ Standard Deviation, (c) Spatial Effective Resolution. The higher the curve, the better the reconstructed results perform on that metric.

biases, though its spatial distribution characteristics resemble those of other reconstruction results. This indicates that the model has learned the mapping relationship between SSH and SST from reanalysis data. By leveraging this mapping relationship, the SSH observations serve as a valuable supplement to SST observations, providing crucial information for reconstructing SST in the absence of observation.

The 50-member ensemble mean outperforms representative single members on all reported metrics; therefore, we compare it with existing reconstruction methods, including dynamical data assimilation systems (e.g., GLORYS12V1 [Lellouche *et al.*, 2021]), traditional interpolation products (e.g., DUACS [Taburet *et al.*, 2019]), and deep learning-based reconstruction methods (e.g., NeurOST [Martin *et al.*, 2024] and GenDA [Martin *et al.*, 2025]). In addition, the ensemble spread provides an uncertainty estimate consistent with observational coverage: the mean standard deviation is 0.118 cm in observed areas and 2.364 cm in unobserved areas, indicating higher uncertainty where observations are absent.

As shown in Table 1, the SSH reconstruction results of 4DVarGen under both observation scenarios demonstrate performance comparable to state-of-the-art methods across relevant metrics. Compared to the reanalysis product GLORYS12V1, which provides pretraining data, 4DVarGen exhibits significant advantages. When using SLA observations alone, $RMSE_{score}$ improved by 17% and the spatial effective resolution improved by 56%. When combining SLA and SST observations, the improvements further increased to 18% and 58%, respectively.

To assess the physical consistency, we compare the spectral characteristics of the reconstructed results with other advanced methods. As shown in Figure 5(a), all methods underestimate the SSH signal strength at small scales. However, 4DVarGen maintains a spectral slope similar to observations, indicating that it effectively captures the energy transfer mechanisms between different scales. Figure 5(b) shows that 4DVarGen(SSH-SST) performs better within the observation-reliable scale range. This indicates that it more accurately restores the spatial structure of the SSH field and achieves stronger dynamical consistency.

5.4 Ablation Study

This experiment further analyzes the effectiveness of the spatio-temporal evolution patterns and dynamical mappings of sea surface variables learned by the model from reanalysis data in the real world by comparing reconstruction quality under different observation types (SSH, SSH-SST), time-window lengths (v), and overlap durations. The same metrics as in the OSE are used, and the results are shown in Figure 6.

Figure 6 indicates that the temporal evolution patterns of SSH significantly aid in SSH reconstruction. When the time window is set to 1, the model fails to learn the temporal evolution patterns of SSH and cannot acquire observational information from other time periods through Segmented reconstruction strategy. Under this condition, the reconstruction results significantly underperform across all metrics compared to other parameter settings. Furthermore, all experimental results demonstrate that incorporating SST observations effectively enhances the reconstruction quality of SSH, indicating that the model can effectively learn and utilize the dynamical mapping between SSH and SST.

6 Conclusion

This study establishes a theoretical connection between diffusion model and sea surface variable reconstruction under 4DVar, showing that conventional 4DVar can be regarded as a special case of an observation-guided diffusion model under a Gaussian prior. By leveraging the spatiotemporal prior information provided by diffusion model, the proposed framework avoids two key limitations of conventional 4DVar methods: the degradation in reconstruction quality caused by the Gaussian assumption of background errors and the high computational cost from dependence on numerical models.

The OSSE demonstrates that 4DVarGen effectively utilizes the spatiotemporal evolution patterns of sea surface variables provided by the diffusion model to propagate observations from other time points to a specific time point, thereby aiding sea surface variable field reconstruction at that time point. The OSE demonstrates that 4DVarGen achieves reconstruction outcomes comparable to state-of-the-art methods for $RMSE_{score}$ and effective resolution. Moreover, the reconstruction results significantly outperform GLORYS12V1, the dynamical data assimilation product used for model pretraining. The ablation experiment results confirm the importance of the model learning the temporal evolution patterns of sea surface variables for reconstruction accuracy. In summary, this study provides a learned-prior assimilation framework for high-dimensional spatiotemporal inverse problems, with potential extensions to atmospheric, land-surface, hydrological, and broader Earth-system reconstruction tasks.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant Nos. 42275170, 42305166, and 42305170), the Science and Technology Innovation Program of Hunan Province (2022RC3070), and the Natural Science Foundation of Hunan Province (Grant No. 2023JJ40678).

References

[Amores *et al.*, 2018] Angel Amores, Gabriel Jordà, Thomas Arsouze, and Julien Le Sommer. Up to what extent can we characterize ocean eddies using present-day gridded altimetric products? *Journal of Geophysical Research: Oceans*, 123(10):7220–7236, 2018.

[Archer *et al.*, 2025] Matthew Archer, Jinbo Wang, Patrice Klein, Gerald Dibarboure, and Lee-Lueng Fu. Wide-swath satellite altimetry unveils global submesoscale ocean dynamics. *Nature*, 640(8059):691–696, 2025.

[Bao *et al.*, 2019] Senliang Bao, Ren Zhang, Huizan Wang, Hengqian Yan, Yang Yu, and Jian Chen. Salinity profile estimation in the pacific ocean from satellite surface salinity observations. *Journal of Atmospheric and Oceanic Technology*, 36(1):53–68, 2019.

[Chao *et al.*, 2025] Jie Chao, Baoxiang Pan, Quanliang Chen, Shangshang Yang, Jingnan Wang, Congyi Nai, Yue Zheng, Xichen Li, Huiling Yuan, Xi Chen, Bo Lu, and Ziniu Xiao. Learning to infer weather states using partial observations. *Journal of Geophysical Research: Machine Learning and Computation*, 2(1):1–18, 2025.

[Chen *et al.*, 2020] Zhiqiang Chen, Xidong Wang, and Lei Liu. Reconstruction of three-dimensional ocean structure from sea surface data: An application of isqg method in the southwest indian ocean. *Journal of Geophysical Research: Oceans*, 125(6):1–13, 2020.

[Chen *et al.*, 2023] Zhiqiang Chen, Xidong Wang, Lei Liu, and Xiaoting Wang. Estimating three-dimensional structures of eddy in the south indian ocean from the satellite observations based on the isqg method. *Earth and Space Science*, 10(10):1–21, 2023.

[Cutolo *et al.*, 2024] 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.

[Ding *et al.*, 2022] Mengrong Ding, Hailong Liu, Pengfei Lin, Yao Meng, Weipeng Zheng, Bo An, Yihua Luan, Yongqiang Yu, Zipeng Yu, Yiwen Li, Jinfeng Ma, Jian Chen, and Kangjun Chen. A century-long eddy-resolving simulation of global oceanic large- and mesoscale state. *Scientific Data*, 1(1):691, 2022.

[Fablet *et al.*, 2023] Ronan Fablet, Quentin Febvre, and Bertrand Chapron. Multimodal 4dvarnets for the reconstruction of sea surface dynamics from sst-ssh synergies. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1, 2023.

[Hong *et al.*, 2025] Yingxiang Hong, Xuan Wang, Bin Wang, Wei Li, and Guijun Han. A fusion method based on physical modes and satellite remote sensing for 3d ocean state reconstruction. *Remote Sensing*, 17(8):1468, 2025.

[Kingma and Ba, 2015] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*, 2015.

[Lellouche *et al.*, 2021] Jean-Michel Lellouche, Eric Greiner, Romain Bourdallé-Badie, Gilles Garric, Angélique Melet, Marie Drévillon, Clément Bricaud, Mathieu Hamon, Olivier Le Galloudec, Charly Regnier, Tony Candela, Charles-Emmanuel Testut, Florent Gasparin, Giovanni Ruggiero, Mounir Benkiran, Yann Drillet, and Pierre-Yves Le Traon. The copernicus global 1/12° oceanic and sea ice glorys12 reanalysis. *Frontiers in Earth Science*, 9:1–27, 2021.

[Li *et al.*, 2024] Zeyu Li, Wang Han, Yue Zhang, Qingfei Fu, Jingxuan Li, Lizi Qin, Ruoyu Dong, Hao Sun, Yue Deng, and Lijun Yang. Learning spatiotemporal dynamics with a pretrained generative model. *Nature Machine Intelligence*, 6(12):1566–1579, 2024.

[Liu and Wang, 2025] Xiaoming Liu and Menghua Wang. Detection of ocean eddies from satellite ocean color and sst measurements using a deep learning approach. *International Journal of Applied Earth Observation and Geoinformation*, 144:104929, 2025.

[Liu *et al.*, 2021] Lei Liu, Huijie Xue, and Hideharu Sasaki. Diagnosing subsurface vertical velocities from high-resolution sea surface fields. *Journal of Physical Oceanography*, 51(5):1353–1373, 2021.

[Martin *et al.*, 2023] Scott A. Martin, Georgy E. Manucharyan, and Patrice Klein. Synthesizing sea surface temperature and satellite altimetry observations using deep learning improves the accuracy and resolution of gridded sea surface height anomalies. *Journal of Advances in Modeling Earth Systems*, 15(5):1–26, 2023.

[Martin *et al.*, 2024] Scott A. Martin, Georgy E. Manucharyan, and Patrice Klein. Deep learning improves global satellite observations of ocean eddy dynamics. *Geophysical Research Letters*, 51(17):1–13, 2024.

[Martin *et al.*, 2025] Scott A. Martin, Georgy E. Manucharyan, and Patrice Klein. Generative data assimilation for surface ocean state estimation from multi-modal satellite observations. *Journal of Advances in Modeling Earth Systems*, 17(8):1–33, 2025.

[QI *et al.*, 2024] Jifeng QI, Guimin SUN, Bowen XIE, Delei LI, and Baoshu YIN. Deep learning to estimate ocean subsurface salinity structure in the indian ocean using satellite observations. *Journal of Oceanology and Limnology*, 42(2):377–389, 2024.

[Seo *et al.*, 2023] Hyodae Seo, Larry W. O'Neill, Mark A. Bourassa, Arnaud Czaja, Kyla Drushka, James B. Edson, Baylor Fox-Kemper, Ivy Frenger, Sarah T. Gille, Benjamin P. Kirtman, Shoshiro Minobe, Angeline G. Pendergrass, Lionel Renault, Malcolm J. Roberts, Niklas

Schneider, R. Justin Small, Ad Stoffelen, and Qing Wang. Ocean mesoscale and frontal-scale ocean–atmosphere interactions and influence on large-scale climate: A review. *Journal of Climate*, 36(7):1981–2013, 2023.

[Song and Ermon, 2019] Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. In *Advances in Neural Information Processing Systems*, 2019.

[Song *et al.*, 2021] 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.

[Taburet *et al.*, 2019] Guillaume Taburet, Antonio Sanchez-Roman, Maxime Ballarotta, Marie-Isabelle Pujol, Jean-François Legeais, Florent Fournier, Yannice Faugere, and Gerald Dibarboure. Duacs dt2018: 25 years of reprocessed sea level altimetry products. *Ocean Science*, 15(5):1207–1224, 2019.

[Wang *et al.*, 2025a] Penghao Wang, Kefeng Mao, Xi Chen, Ming Li, Yuhang Zhu, Hongchen Li, Jiahao Wang, Kefeng Liu, and Yangjun Wang. Three-dimensional structure reconstruction of ocean mesoscale eddies based on physical process modeling and data-driven machine learning. *Ocean Modelling*, 196:102558, 2025.

[Wang *et al.*, 2025b] Wuxin Wang, Boheng Duan, Weicheng Ni, Jingze Lu, Taikang Yuan, Dawei Li, Juan Zhao, and Kaijun Ren. A four-dimensional variational informed generative adversarial network for data assimilation. *Journal of Advances in Modeling Earth Systems*, 17(6):1–27, 2025.

[Yan *et al.*, 2021] Hengqian Yan, Ren Zhang, Huizan Wang, Senliang Bao, and Chengzu Bai. Practical dynamical-statistical reconstruction of ocean's interior from satellite observations. *Remote Sensing*, 13(24):5085, 2021.

[Yu *et al.*, 2025a] Xiaoyu Yu, Daling Li Yi, and Peng Wang. Enhancing ocean temperature and salinity reconstruction with deep learning: The role of surface waves. *Journal of Marine Science and Engineering*, 13(5):910, 2025.

[Yu *et al.*, 2025b] Xiaoyu Yu, Daling Li Yi, and Peng Wang. Reconstruction of three-dimensional temperature and salinity in the equatorial ocean with deep-learning. *Remote Sensing*, 17(12):2005, 2025.

[Zhou *et al.*, 2024] Chaojie Zhou, Wei Cui, Ruili Sun, Ying Huang, and Zhanpeng Zhuang. Enhancing the assimilation of swot simulated observations using a multi-scale 4dvar method in regional ocean modeling system. *Remote Sensing*, 16(5):778, 2024.

[Zhu *et al.*, 2025] Qiaoshi Zhu, Hongping Li, Haochen Sun, Tianyu Xia, Xiaoman Wang, and Zijun Han. 3dv-unet: Eddy-resolving reconstruction of three-dimensional upper-ocean physical fields from satellite observations. *Remote Sensing*, 17(19):3394, 2025.