https://doi.org/10.5194/egusphere-2026-5227
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

Physics-guided Spatial Downscaling of Three-dimensional Oceanic Temperature in the South China Sea via the DIFFDS3v1 Generative Diffusion Model

Xueming Zhu, Weiyang Fan, Shuo Wang, Xuri Zhang, Yingqiang He

[1]Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai) & School of Marine Sciences, Sun Yat-sen University, Zhuhai, 519082, China

Correspondence to: Xueming Zhu (zhuxueming@sml-zhuhai.cn)

Abstract. High-resolution three-dimensional (3D) ocean temperature fields are crucial for advancing the understanding of ocean dynamics in the South China Sea (SCS). However, their acquisition remains constrained by the inherent scarcity of in-situ subsurface observations and the computational demands of physics-based numerical models. While the deep learning technique has been applied in ocean reconstruction, existing approaches often fail to preserve high-frequency structures and vertical physical consistency. To address these limitations, this paper proposes DIFFDS3v1, the first version of our generative diffusion model for 3D spatial downscaling. DIFFDS3v1 reconstructs high-fidelity 3D temperature fields from coarse-resolution inputs using a dual-network architecture that integrates atmospheric forcing and ocean background fields for physical guidance. Validated against the GLORYS12V1 dataset in the SCS, DIFFDS3v1 achieved root mean square error values of 0.06, 0.07, and 0.10 °C for 3×, 4×, and 6× downscaling, respectively, significantly outperforming CNN-, generative adversarial network-, and transformer-based benchmark models. Critically, DIFFDS3v1 demonstrates exceptional capability in resolving complex mesoscale and submesoscale phenomena while maintaining physical fidelity, a capability that prior models lacked. Overall, DIFFDS3v1 provides a robust, physics-guided, and generalizable methodology for enhancing sparse oceanographic observations.

1 Introduction

As the largest marginal sea in the northwestern Pacific Ocean, the South China Sea (SCS) exhibits a highly complex bottom topography, encompassing a broad continental shelf, a steep continental slope, and deep-sea basins. Driven by strong air–sea interactions, alternating East Asian monsoons, and the intrusion of the Kuroshio current, the SCS sustains exceptionally rich multiscale dynamical processes, including mesoscale eddies, coastal and island upwellings, and active submesoscale frontal activities. Three-dimensional (3D) oceanic temperature fields are of paramount importance for understanding these dynamics and for synoptic oceanographic forecasting, because their vertical structures govern the spatial organization of key dynamical processes, including the oceanic mixing layer, thermocline, barrier layer, upwelling, and oceanic fronts (Zhao et al., 2024; Pastor, 2021; Zhu et al., 2025).

https://doi.org/10.5194/egusphere-2026-5227
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

The accurate characterization of 3D oceanic temperature structures is fundamental for advancing oceanographic research. The recent emergence of high-resolution digital twin platforms and AI-based meteorological and regional oceanographic forecast large models in the SCS has sharply increased the demand for high-resolution 3D subsurface ocean data. The performance of these data-driven models depends critically on the quality of their training datasets: they require refined, high-resolution fields to effectively learn and reproduce complex mesoscale or submesoscale dynamics that are otherwise lost in coarse-resolution inputs. Therefore, generating continuous 3D ocean data with enhanced spatiotemporal resolution has become an urgent priority for improving the training and initialization of these advanced systems (S. Wang et al., 2024).

In-situ observations, primarily obtained through scientific research vessels, mooring arrays, and Argo floats, remain the most effective means of acquiring high-quality 3D oceanic data, particularly for the subsurface layer (Wu et al., 2024). However, the discrete, sparse, and costly nature of in-situ measurements constrains real-time monitoring and limits the ability to capture continuous internal structures (Shi et al., 2025). To bridge the gap between sparse observations and continuous fields, data assimilation techniques typically combine in-situ measurements with numerical ocean models to generate reanalysis datasets, which currently serve as standards for ocean status estimation. While numerical simulations and DA systems can produce high-resolution oceanographic environmental reanalysis data, they remain inherently constrained by model limitations and empirical parameterization schemes. Additionally, their high computational costs limit adaptability for real-time 3D oceanic forecasting and reanalysis (de Burgh-Day & Leeuwenburg, 2023).

The volume of sea-surface observation data has increased exponentially with the development of ocean satellite remote sensing technologies. Nevertheless, satellite remote-sensing observations face inherent limitations in penetrating below the sea surface, restricting their ability to resolve dynamic subsurface structures (Su et al., 2021). Most ocean dynamic processes extend far beyond the surface layer (Xing et al., 2024). Therefore, satellite data alone cannot provide comprehensive oceanic information. Efficient acquisition of high-resolution oceanic data and the accurate resolution of mesocale to submesoscale processes therefore remain persistent challenges in physical oceanography (Liu et al., 2024), particularly for operational oceanography forecasting requiring low latency.

Deep learning-based downscaling originates primarily from super-resolution (SR) technology in computer vision. The first SR convolutional neural network (SRCNN) (Dong et al., 2016) was proposed for image SR. Subsequently, numerous CNN-based SR models have been developed, and their performances have been further improved. For instance, bicubic interpolation has been utilized to generate low-resolution inputs from the operational sea surface temperature and sea ice analysis (OSTIA) dataset, followed by SRCNN training to recover high-resolution targets (Ducournau & Fablet, 2016). The emergence of SR generative adversarial models, such as the enhanced SR generative adversarial network (ESRGAN) (Ledig et al., 2017), SR transformer models (Lu et al., 2022; Conde et al., 2023), and SR diffusion models (DMs) (Xia et al., 2023; H. Li et al., 2022) also represent significant advances in the SR field.

In oceanography, some studies have begun to adopt deep learning-based downscaling methods to generate higher-resolution oceanic datasets, demonstrating their effectiveness in spatial downscaling for oceanic temperature or wave data. For example,

https://doi.org/10.5194/egusphere-2026-5227
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

a spectral normalization (SN) ESRGAN neural network was developed to downscale low-resolution SST to high-resolution SST (Khoo et al., 2022). Similarly, a diffusion-based downscaling model has been proposed to transform 1/3° low-resolution input data into 1/12° high-resolution sea surface temperatures (SSTs) (S. Wang et al., 2024).

The above methodologies collectively underscore the potential of deep learning approaches to overcome seawater temperature downscaling challenges. Most previous downscaling methods are two-dimensional (2D) downscaling neural network models. However, research on downscaled 3D models is scarce in historical literatures. 3D downscaling models can simultaneously output high-resolution temperature fields for multiple depth layers (Feng et al., 2025). To achieve this, 2D downscaling models must build and train separate models for temperature at different depth layers individually, incurring substantial computational and time costs (Zhang et al., 2022). Additionally, 2D downscaling models must consider issues such as whether to use the same parameters across different layers or tune parameters separately for each layer and whether different models may cause inconsistencies in temperature across different vertical layers. In contrast, 3D downscaled models eliminate these overheads, and their advantages become increasingly prominent as the number of vertical layers increases. Nevertheless, applying 3D models to oceanic data downscaling presents significant challenges beyond computational costs and graphics processing unit (GPU) memory constraints. For instance, the difficulty lies in the strong anisotropy of ocean dynamics: horizontal correlation scales are orders of magnitude larger than the vertical scales, making it difficult for standard isotropic 3D convolutional kernels to extract features effectively.

In this study, we propose a regional, physics-guided 3D spatial downscaling method, the DM for spatial downscaling 3D (DIFFDS3v1), built upon recent advances in generative diffusion modeling. To enhance the accuracy of reconstructed temperature fields, DIFFDS3v1 incorporates wind fields, ocean currents, and salinity as additional physical factors to guide the temperature downscaling process. Compared with benchmark models, consistently generates high-resolution temperature fields with superior accuracy across standard evaluation metrics, while also capturing mesoscale and submesoscale oceanic phenomena with greater fidelity. These results demonstrate both the effectiveness and feasibility of DIFFDS3v1 as a robust approach for 3D oceanic temperature downscaling.

This paper is structured as a description of the DIFFDS3v1 model and an evaluation of its performance. The data sources, downscaling methodology and the model architecture used in this study are presented in Section 2. The design, results and analysis of experiments are included in Section 3. In Section 4 for discussion, we show the specific samples from experiments, reconstruction of seasonal MLD, anticyclonic eddies, and ablation experiments on the guidance effects of physical factors. Finally, the main conclusions are provided.

2 Data and Methodology

2.1 Data

In this study, 3D temperature, salinity, and ocean current data derived from GLORYS12V1, together with the U- and V-components of wind at 10 m above the sea surface derived from the ERA5 atmospheric reanalysis dataset—were used to

https://doi.org/10.5194/egusphere-2026-5227
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

conduct 3D oceanic temperature spatial downscaling experiments. GLORYS12V1 data were provided by the Copernicus Marine Environment Monitoring Service as high-resolution reference ground truth. Its spatial resolution reached 1/12° horizontally, with 50 standard vertical levels and a daily temporal resolution spanning from 1993 to the present. GLORYS12V1 is a global eddy-resolving ocean reanalysis dataset using the NEMO numerical model platform, incorporating a reduced-order Kalman filter assimilation technique with large-scale bias correction implemented via a 3D-Var scheme. This dataset deeply integrates multi-source observation and simulation data, including along-track satellite sea level anomalies, satellite SST, sea ice concentration observations, and in-situ observation data such as Argo float temperature and salinity vertical profiles and drifter velocity data (Lellouche et al., 2021). ERA5, the fifth-generation European Centre for Medium-Range Weather Forecasts reanalysis, provides global climate and weather data spanning the past eight decades, with a daily temporal resolution and a horizontal resolution of 1/4° on a regular latitude–longitude grid (Hersbach et al., 2020).

The study area spans 106°E–122°E and 8°N–24°N, encompassing the entire SCS basin (Figure 1). This region is subject to a tropical maritime monsoon climate driven by the East Asian Monsoon, with substantial rainfall and is characterized by complex multiscale dynamical processes, such as coastal upwelling, mesoscale eddies, internal waves, Kuroshio intrusion through the Luzon Strait, and oceanic fronts. Downscaling the 3D oceanic temperature from a large scale to meso- or sub-mesoscale is a big challenge. Considering the typical vertical ocean structure characteristics of the SCS, as well as constraints related to GPU memory and computational efficiency, the following depth layers were selected: 0.49, 11.40, 29.44, 55.76, 92.33, 155.85, 222.48, 318.13, 643.57, and 1062.44 m. These depth layers fully cover the key dynamic process zones from the surface to 1000 m mid-depth in the SCS, including vertical mixing in the mixed layer, temperature gradient transitions in the thermocline, and the vertical propagation characteristics of mesoscale eddies.

To enhance the physical plausibility of the 3D downscaling results, this study incorporated three other important physical variables closely related to seawater temperature spatiotemporal evolution as additional physical guidance conditions for the downscaling model. Seawater salinity, together with temperature at the same depth, it determines seawater density and is a core factor controlling ocean stratification stability and vertical mixing intensity, directly affecting the vertical distribution structure of seawater temperature. Seawater currents, zonal velocity (U) and meridional velocity (V), together with temperature at the same depth, control the horizontal spatial distribution of seawater temperature through horizontal advection and play a critical role in reconstructing key dynamic processes such as mesoscale eddies and coastal upwelling in the SCS. The U (U10) and V (V10) of wind at 10 m above the sea surface, as the primary drivers of air–sea interaction, determine the mixing intensity and heat exchange rate in the upper ocean, and are important controlling factors for seawater temperature variations.

The low-resolution 3D temperature field was obtained using Gaussian blur (kernel size = 0.5) (X. Li & Chen, 1995) combined with nearest-neighbor interpolation. This procedure simulates the nonlinear numerical dissipation associated with grid coarsening in low-resolution ocean numerical simulations, as well as the spatial averaging effect of microwave remote sensors under large observation footprints. Following Gaussian smoothing, high-gradient signals in the data (such as frontal transitions and submesoscale eddy filaments) were reasonably attenuated. Subsequently, based on the smoothed temperature field, nearest-neighbor interpolation was applied to perform regular grid point extraction, ultimately obtaining the low-resolution

https://doi.org/10.5194/egusphere-2026-5227
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

input field of the downscaling model. To evaluate the reconstruction performance of the generative DM at different downscaling ratios, three experimental groups were established with input low resolutions of 1/4°, 1/3°, and 1/2°, respectively. These resolutions represent the grid scales of the current mainstream medium-resolution ocean reanalysis datasets and satellite observation products.

The experimental period spanned 2005–2021, with the dataset divided chronologically as follows: 2005–2018 was the training set, 2019 was the validation set, and 2020–2021 was the test set. Prior to network input and training, normalization processing was applied to the different data using the min–max normalization method to map the data to the [-1, 1] interval to accelerate the convergence of the AI model training process.



Figure 1: Study area enclosed by red rectangle.

2.2 Generative DM

Generative models, including generative adversarial networks (GANs) and variational autoencoders (VAEs), have been widely adopted for producing highly realistic samples. However, both approaches face inherent limitations (Karras et al., 2021). GANs often suffer from challenging training dynamics, leading to instability and mode collapse—a phenomenon where the model generates only a limited variety of samples and fails to capture the full diversity of the data distribution (Goodfellow et al., 2014). Conversely, VAEs, while stable for training, often produce results with insufficient detail, leading to blurred or overly smooth outputs (Vahdat & Kautz, 2020).

https://doi.org/10.5194/egusphere-2026-5227
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

By contrast, DMs have emerged as principled and robust alternatives, achieving state-of-the-art generation quality by fundamentally reframing generative tasks (Rombach et al., 2022). The core insight behind DMs is elegant, yet powerful: they learn the reverse of a slow, gradual data corruption process (Saharia et al., 2022), transforming the intractable problem of direct generation into a series of well-defined, manageable denoising steps. This formulation effectively avoids the training instabilities and mode collapse issues associated with GANs, while producing samples with exceptional detail (Saharia et al., 2023). The theoretical foundations of this model, inspired by non-equilibrium thermodynamics (Sohl-Dickstein et al., 2015), have been progressively refined in influential studies, such as denoising diffusion probabilistic models (DDPMs) (Ho et al., 2020), improved DDPMs (Nichol & Dhariwal, 2021), and denoising diffusion implicit models (Song et al., 2020).

Taking DDPM as an example, DMs typically encompass two processes: forward and reverse denoising, both of which are characterized by a T-step Markov chain.

In the forward noise diffusion process, the initial data distribution is progressively transformed into a standard Gaussian distribution. First, the initial data x0 are defined, and then Gaussian noise is progressively added at each time step until it reaches pure Gaussian noise xT~ $\mathcal{N}$ (0, I) at time step T by T iterations. During each mid-time step t, noisy data xt are generated with the same shape as x0. The noise incorporated in the diffusion process is specified by a predefined sequence of $\beta_{1:T} \epsilon\ (0,1]^T$. Denote $\alpha_t = 1 - \beta_t$, $\overline{\alpha_t} = \prod_{n=1}^{T} \alpha_n$ and β1 < β2 < . . . < βt (t ∈ [1, T]). Each iteration of the forward noising process can be described as follows:

$$q(x_t|x_0) = \mathcal{N}\left(x_t;\ \sqrt{\overline{\alpha_t}}x_0, (1-\overline{\alpha_t})\boldsymbol{I}\right) \quad (1)$$

The reverse denoising process focuses on learning the inverse of the forward noise process and generate a distribution that resembles the real data. The reverse denoising process first samples a random noise xT ∈ $\mathcal{N}$ (0, I) and then gradually denoises until it reaches a high-quality output x0. pθ (xt) is defined as the data distribution at timestep t in the reverse process. A neural network $\epsilon_\theta(x_t, t)$ is involved in predicting the added noise component, where θ represents the network parameters. Each iteration of the reverse denoising process can be described as follows:

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1};\ \mu_\theta(x_t,\ t), \sigma_t^2 \boldsymbol{I}) \quad (2)$$

where $\mu_\theta(x_t,\ t)$ is the mean value calculated using Eq. (3):

$$\mu_\theta(x_t, t) = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \epsilon_\theta \frac{1-\alpha_t}{\sqrt{1-\overline{\alpha_t}}}\right) \quad (3)$$

and σ_t^2 is the variance value calculated using Eq. (4):

$$\sigma_t^2 = \frac{1-\overline{\alpha_{t-1}}}{1-\overline{\alpha_t}}\beta_t \quad (4)$$

In addition, only one unknown variable ϵ_θ exists that should be learned in the reverse denoising process, which is estimated by the DMs using a neural network $\epsilon_\theta(x_t, t)$. To train the network of $\epsilon_\theta(x_t, t)$, a time step t and a noise $\epsilon \in \mathcal{N}(0, \mathbf{I})$ were randomly sampled to generate noisy data xt with the given real data x0, according to Eq. (2). Subsequently, all network parameters were optimized using the following loss function:

$$Loss = \mathbb{E}_{x_0,\epsilon,t}[\|\epsilon - \epsilon_\theta(x_t, t)\|_2^2] \quad (5)$$

https://doi.org/10.5194/egusphere-2026-5227
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

2.3 3D Downscaling Method Based on Generative DM

Applying generative DMs to 3D oceanographic downscaling introduces unique challenges, particularly regarding the computational efficiency and effective fusion of multimodal physical data. A naive approach, common in 2D image SR (Saharia et al., 2023), involves upscaling all input variables, including multi-depth layers of seawater temperature, salinity, velocity, and wind fields to the target resolution and concatenating them as input channels of the denoising network. For 3D spatiotemporal data, this strategy is computationally prohibitive, leading in exorbitant GPU memory demands and unmanageable slow training cycles. To address both the computational and architectural challenges, we propose a novel dual-network architecture, DIFFDS3v1, which decouples the primary denoising task from the feature extraction of the guidance vector (winds, seawater salinity, and currents), thereby facilitating the downscaling process as physical guidance. DIFFDS3v1 consists of two main components: a core 3D denoising network and an auxiliary element fusion network (EFN), which work synergistically to generate high-resolution results (Figure 2).

2.3.1 Workflow: Training and Inference

The DIFFDS3v1 model operates in two stages, as shown in Figure 2.

Training Stage (forward noise process): The denoising network and the EFN are trained together. The model learns to predict noise added to a high-resolution temperature field. The loss is calculated based on the difference between the predicted noise and the actual noise, conditioned on both the interpolated low-resolution temperature and the guidance vectors from the EFN.

Inference Stage (reverse denoising process): Starting with random Gaussian noise, the trained model iteratively denoises data. At each time step of the reverse denoising process, the EFN processes the wind, salinity, and velocity inputs to generate a guidance vector, which is then used by the denoising network to refine the emerging 3D oceanic temperature field, ultimately yielding an objective high-resolution output.

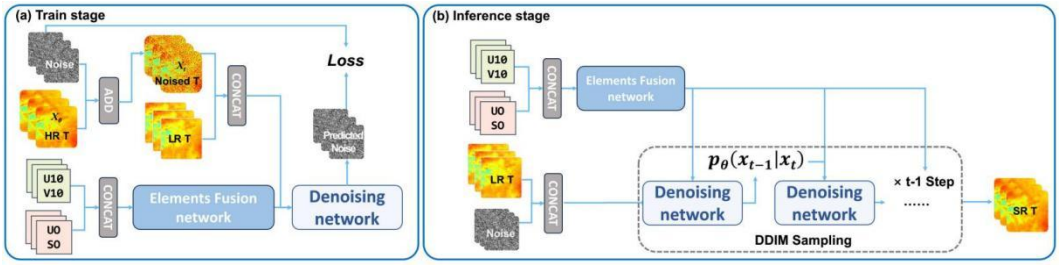


Figure 2: (a) Loss calculation workflow within a single training step, detailing the collaboration of the EFN and denoising network under the forward diffusion process. (b) Workflow of the inference stage of DIFFDS3v1. HR T: high-resolution oceanic temperature; LR T: low-resolution oceanic temperature; Noised T: oceanic temperature with noise added.

2.3.2 Architectural Components

EFN serves as the core module for conditional feature extraction. It consists of stacked successive ResnetBlock3D and TransformerBlock modules, followed by a FusionLayer to compress and reorganize the features, as shown in Figure 3a. The

https://doi.org/10.5194/egusphere-2026-5227
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

physical essence of this module is to extract prior information that characterizes the multi-scale spatial dynamic structures of the SCS from the input guidance vectors, thereby guiding the ocean temperature spatial downscaling process.

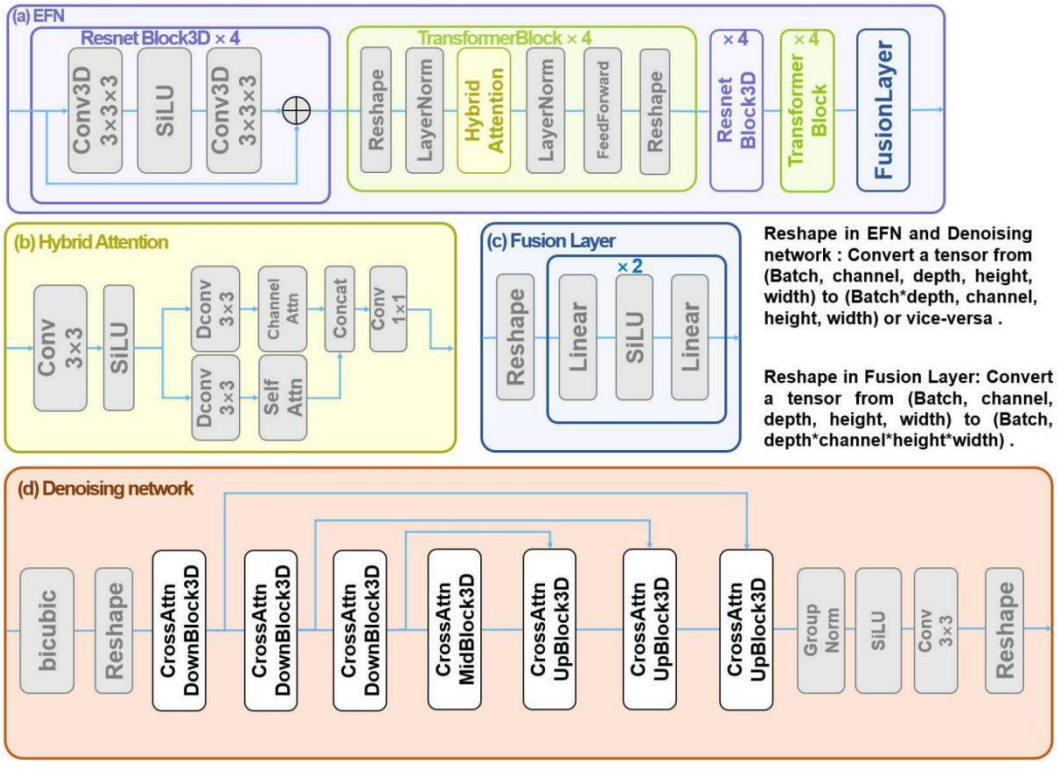
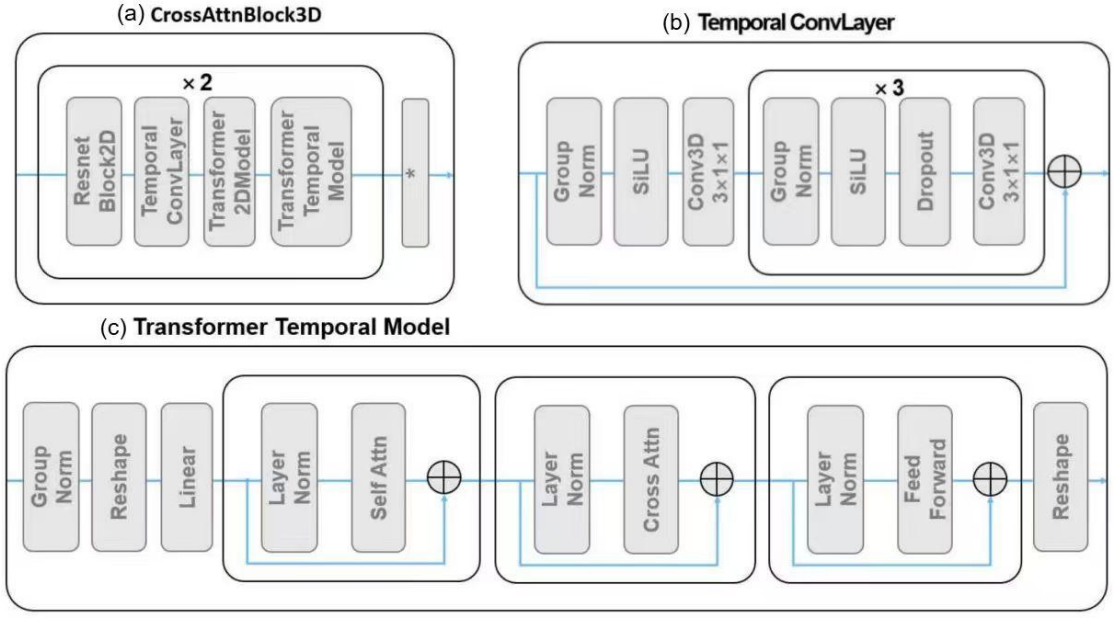


Figure 3: Detailed architecture for the EFN and denoising network.

The input tensor first passes through the stacked ResnetBlock3D modules. This module is composed of a 3×3×3 3D convolutional layer and an SiLU activation function, followed by another 3D convolutional layer. It first captures the input 3D oceanic temperature features through 3D convolution and then enhances the network's feature representation capability via nonlinear transformation. To facilitate gradient propagation and preserve the original information, a residual connection was introduced at this stage. To process the input data, the ResnetBlock3D was primarily responsible for capturing the local spatial gradients and short-range dependencies. Once local features were extracted using ResnetBlock3D, the data were passed into the TransformerBlock module. Because Transformers excel at modeling long-range dependencies, this module first flattens the tensor through a reshaping operation. After layer normalization (LayerNorm), the tensor was processed into the core hybrid attention module (Figure 3b), which extracts spatial and depth-level features. The output was then through a feed-forward network. This stage aims to establish global contextual associations among the features. Finally, the processed features were alternately refined through two sets of ResnetBlock3D and TransformerBlock modules before entering the FusionLayer (Figure 3c). This layer is responsible for mapping the high-dimensional features into a latent space suitable for injection into the denoising network.

https://doi.org/10.5194/egusphere-2026-5227
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.



225 **Figure 4: Detailed architecture for CrossAttn Block3D. * denotes downsample Layer for DownBlock, upsamle layer for UpBlock, and null for MidBlock.**

The denoising network shown in Figure 3d serves as the core backbone of the generative DM. It adopts a classic UNet architecture tailored to the modeling requirements of the horizontal, and vertical depth features of 3D oceanic temperature data in the SCS. From a physical oceanography perspective, this decoupled design aligns naturally with the strongly anisotropic 230 fluid characteristics of the SCS. The network consists of three parts: downsampling (DownBlock), middle bottleneck layer (MidBlock), and upsampling (UpBlock) (Figure 3d). CrossAttn DownBlock3D and its corresponding MidBlock3D/UpBlock3D modules (Figure 4) decompose 3D data processing into two independent branches: horizontal and vertical directions. In the horizontal spatial dimension, ResnetBlock2D performs local feature extraction on the 2D oceanic temperature field of each depth layer. Through residual connections and 2D convolutions, it captures the spatial gradients of 235 oceanic temperature in the latitudinal and longitudinal directions, representing fine-scale structures, such as active submesoscale fronts in the interior of the SCS and cold/warm eddy filaments induced by complex coastal topography, while mitigating the degradation problem of deep networks. The Transformer 2DModel applies a self-attention mechanism in the horizontal direction to model the long-range spatial dependencies of the SCS oceanic temperature, breaking the local receptive field limitations of traditional convolutions. This effectively captures the submesoscale modulations of the overall oceanic 240 temperature patterns exerted by large-scale circulation forcing from the East Asian monsoon. In the vertical direction, the temporal ConvLayer is utilized to capture dynamical associations among different depth layers (such as vertical convection and mixing processes). Concurrently, the transformer temporal model applies a self-attention mechanism in the vertical direction to model the vertical long-range dependencies of the SCS oceanic temperature, capturing the complex dynamical coupling between surface, subsurface, and deep-layer oceanic temperatures.

https://doi.org/10.5194/egusphere-2026-5227
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

Within each sub-module stage of DownBlock, data flows sequentially through the ResnetBlock2D, temporal ConvLayer, Transformer 2DModel, and the transformer temporal model (Figure 4a). Furthermore, the physical conditional features (such as atmospheric forcing from wind) extracted by the EFN were injected into the Transformer modules via a cross-attention mechanism, guiding the model to learn a denoising direction that conformed to the laws of ocean dynamics. The middle layer, MidBlock, acts as a network bottleneck to further integrate the global semantic information. The upsampling stage (UpBlock) gradually recovers the resolution through upsampling operations and concatenates with the corresponding features from DownBlock via long skip connections. Finally, through progressive denoising, the network outputs high-resolution 3D oceanic temperature data with statistical consistency and physical plausibility.

To extract the vertical feature associations of 3D oceanic temperature in a fine-grained manner, the denoising network employs the temporal ConvLayer and transformer temporal model modules to perform depth feature extraction. The temporal ConvLayer (Figure 4b) adopts stacked multilayer 3×1×1 asymmetric 3D convolutions, where the convolutional kernel slides locally only along the depth dimension. This design aims to extract local feature associations between adjacent depths, compelling the network to learn the physical continuity of the seawater vertical profile. As a global complement to the local features, the transformer temporal model (Figure 4c) utilizes a self-attention mechanism to establish connections across different depth layers. Through tensor reshaping operations, the input is converted from (Batch*depth, channel, height, width) to (Batch*height*width, depth, channel). This transformation reorganizes the 3D spatial features into shallow-to-deep one-dimensional sequences at each independent spatial grid point, enabling the model to capture long-range, non-local dynamical couplings such as downward surface heat transfer or deep-layer upwelling. It further leverages a cross-attention mechanism to inject external physical conditions such as atmospheric forcing and background fields from both salinity and currents extracted by the EFN into the model as physical guidance information. This ensures that the generated high-resolution 3D oceanic temperature field from downscaling not only exhibits fine textures on the horizontal plane but also displays vertical thermal evolution that conforms to submesoscale ocean dynamical laws.

2.3.3 Fusion Mechanism: Cross-Attention

A critical step in DIFFDS3v1 is the integration of physical guidance from the EFN into the denoising network. This was achieved by employing a cross-attention mechanism, a technique that enables one network to selectively focus on the most relevant information provided by the EFN (Lin et al., 2022). The guidance vector generated by the EFN is injected into multiple layers of the U-Net structure of the denoising network (specifically within the CrossAttnDownBlock3D, CrossAttnMidBlock3D, and CrossAttnUpBlock3D modules).

Let $\mathbf{X} \in R^{N \times d}$ denote the intermediate feature map of the denoising network at a specific layer, where N represents the flattened spatial (or volumetric) dimensions and d is the channel dimension. Let $\mathbf{G} \in R^{M \times dg}$ represents the guidance feature sequence produced by the EFN. To enable interaction between these two distinct domains, we project them into query (**Q**), key (**K**), and value (**V**) matrices using the learnable linear projections $\mathbf{W}_Q$, $\mathbf{W}_K$, and $\mathbf{W}_V$:

$$\mathbf{Q} = \mathbf{X}\mathbf{W}_Q, \quad \mathbf{K} = \mathbf{G}\mathbf{W}_K, \quad \mathbf{V} = \mathbf{G}\mathbf{W}_V \tag{6}$$

https://doi.org/10.5194/egusphere-2026-5227
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

where **Q** represents the target features from the noisy input and **K** and **V** encode the physical constraints and guidance information, respectively. The cross-attention output is then computed as

$$Attention(\mathbf{Q},\mathbf{K},\mathbf{V}) = softmax\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V} \quad (7)$$

where d_k is the dimension of the key vectors, and is used as a scaling factor to ensure numerical stability. In this process, the feature map within the denoising network acts as a query and the output of the EFN serves as the key and value. This formulation enables the denoising network, at each level of its feature hierarchy, to request and receive pertinent physical guidance based on the correlation between noisy features and physical embeddings, effectively steering the denoising process toward a more physically plausible and accurate reconstruction.

3 Experiments and Results

3.1 Experiments Design

Based on the dataset and downscaling methods described in Section 2, we conducted 3×, 4× and 6× scale downscaling experiments. For comparison, the selected downscaling methods include Bicubic, Enhanced deep residual networks for single-image SR (EDSR) (Lim et al., 2017), Real-ESRGAN (J. Wang et al., 2020), hybrid attention transformer (HAT) (Zheng et al., 2021), video DM, and DIFFDS3v1. All the models were trained based on the same low- and high-resolution datasets described in Section 2.1. For the DIFFDS3v1 model, the sampling steps of DDPM are set to 100, with the β scheduler configured to linear, β_1 and β_t set to 0.0001 and 0.02, respectively. The parameter values of each model are listed in Table 1. Among them, the parameter quantities of the CNN-, GAN-, and transformer-based models were relatively close, whereas those of the diffusion-based models were of the same order of magnitude.

Table 1: Comparison of model complexity in terms of learnable parameters.

Model	Methodology	Parameters (M)
Bicubic	Interpolation	0
EDSR	CNN	21.68
Real-ESRGAN	GAN	16.71
HAT	Transformer	41.60
VDM	Diffusion	134.46
DIFFDS3v1	Diffusion	69.80

To rigorously assess the quality and accuracy of the proposed method with respect to other benchmark models, we employed a comprehensive set of objective evaluation metrics, including the root mean square error (RMSE), mean absolute error (MAE), and peak signal-to-noise ratio (PSNR). These metrics facilitate the quantification of discrepancies between the enhanced spatial resolution temperature values and actual high-resolution temperature measurements.

https://doi.org/10.5194/egusphere-2026-5227
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

3.2 Results and Analysis

3.2.1 Metrics Evaluation

Table 2 summarizes the objective statistical metrics of the different comparison methods for three spatial downscaling factors: 3×, 4×, and 6×. In the 3× downscaling experiments, all deep learning models demonstrated clear reconstruction advantages over traditional bicubic interpolation. Bicubic interpolation produced relatively large reconstruction error with an RMSE of 0.18 ℃, MAE of 0.08 ℃, and a PSNR of 45.58 dB. By contrast, deep learning methods—EDSR, Real-ESRGAN, HAT, VDM, and DIFFDS3v1—successfully restricted the RMSE to the range of 0.06–0.08 ℃, with one order lower than that of bicubic interpolation, and improved the PSNR to 54.4–57.1 dB. These results illustrate the efficacy of deep neural networks in handling complex nonlinear mappings. Among them, DIFFDS3v1 achieved the best reconstruction performance, with a minimum RMSE of 0.06 ℃ and MAE of 0.02 ℃, and a maximum PSNR of 57.07 dB. Compared with the traditional bicubic interpolation, the RMSE of DIFFDS3v1 decreased by 66.7%, and its PSNR increased by 25.2%. Furthermore, when compared with the runner-up model HAT at the same scale, DIFFDS3v1 yielded a 0.66 dB improvement in PSNR while maintaining the same lowest error level (RMSE of 0.06 °C and MAE of 0.02 °C).

As the spatial downscaling factor increased to 4× and 6×, the effective information in the input data was significantly reduced and the ill-posedness of the task increased substantially, leading to varying degrees of performance degradation across all models. In the 4× experiments, the degradation of Bicubic interpolation was particularly evident, with the RMSE rising to 0.24 ℃ and the PSNR dropping to 45.08 dB. Among the deep learning models, HAT and VDM exhibited favorable anti-interference capabilities, both maintaining an RMSE of 0.09 ℃ and a PSNR of approximately 53.3 dB. Nevertheless, DIFFDS3v1 continues to lead the performance, achieving an RMSE of 0.07 ℃, a MAE of 0.03 ℃, and a PSNR of 54.85 dB. This represents a 22.2% reduction in the RMSE compared with the runner-up model HAT, indicating that DIFFDS3v1 possesses superior spatial reconstruction precision when facing moderate information loss.

Table 2: Average quantitative metrics of all methods under 3×, 4×, and 6× downscaling.

Model	3×			4×			6×		
	RMSE (℃)	MAE (℃)	PSNR (dB)	RMSE (℃)	MAE (℃)	PSNR (dB)	RMSE (℃)	MAE (℃)	PSNR (dB)
Bicubic	0.18	0.08	45.58	0.24	0.11	45.08	0.36	0.16	41.68
EDSR	0.08	0.03	54.41	0.11	0.05	50.83	0.15	0.07	46.68
Real-ESRGAN	0.07	0.03	55.42	0.11	0.05	51.75	0.14	0.07	49.33
HAT	0.06	0.02	56.41	0.09	0.04	53.39	0.17	0.10	47.71
VDM	0.07	0.03	55.64	0.09	0.04	53.31	0.12	0.06	51.19
DIFFDS3v1	0.06	0.02	57.07	0.07	0.03	54.85	0.10	0.04	52.84

In the highly challenging 6× large-scale downscaling experiments, the robustness differences among the various models were further amplified. The RMSE of Bicubic acid increased to 0.36 ℃. Notably, the transformer-based HAT model, which performed favorably at lower scaling factors, experienced severe performance degradation, with its RMSE rising to 0.17 ℃

https://doi.org/10.5194/egusphere-2026-5227
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

and MAE to 0.10 ℃, falling behind the classic CNN-based model EDSR (RMSE of 0.15 ℃). This indicates that traditional discriminative SR models are prone to generalization deficiencies in large-scale downscaling scenarios, where local feature guidance is extremely scarce. By contrast, generative diffusion-based models such as VDM and DIFFDS3v1 demonstrate stronger distribution modeling and anti-interference capabilities. VDM achieved an RMSE of 0.12 ℃ and a PSNR of 51.19 dB under the 6× scale. Meanwhile, DIFFDS3v1 maintained the best reconstruction performance, with an RMSE of 0.10 ℃, an MAE of 0.04 ℃, and a PSNR of 52.84 dB. Compared with the runner-up model VDM, the RMSE of DIFFDS3v1 was reduced by 16.7%, and the PSNR was improved by 1.65 dB. These results suggest that, in high-ratio downscaling scenarios with extremely limited input information, the DIFFDS3v1 model can more robustly restore complex detailed features, demonstrating favorable generalization performance and cross-scale robustness.

3.2.2 Analysis of Temporal Error Characteristics

As illustrated by the monthly mean RMSE time-series variations under 3×, 4×, and 6× spatial downscaling factors, the RMSE of the reconstructed oceanic temperature for all models exhibited clear depth dependency and prominent seasonal oscillation patterns (Figg. 5-7). As the downscaling factor increases, the overall RMSE baseline continuously increases, the seasonal polarization of errors intensifies, and the performance divergence among different models becomes increasingly pronounced. In the upper layers from 0.49 to 92.33 m, the RMSE curves at all depths demonstrate a distinct seasonal oscillation characterized by higher values in summer and lower values in winter. The peaks were concentrated from boreal summer to early autumn, while troughs occurred during boreal winter. This feature aligns closely with regional ocean dynamical processes. During summer, the prevailing southwest monsoon and intense solar radiation drive rapid warming of the upper ocean, forming a shallow seasonal thermocline with a steep gradient. Meanwhile, mesoscale eddies and sub-mesoscale frontal activities are frequent, substantially increasing the spatial heterogeneity of the seawater temperature field and exacerbating the difficulty of downscaling reconstruction (Hu et al., 2012; Kim & Hong, 2010). Conversely, during winter, the strong winter northeast monsoon forces intensive vertical mixing of the upper water column, deepening the mixed layer and making the temperature distribution relatively uniform, which leads to a significant decrease in the RMSE.

In terms of model performance, the DIFFDS3v1 model consistently maintained the lowest RMSE across all upper depths and for most months. Notably, its error peaks during the high-error summer periods were significantly lower than those of all other models. Under 3× and 4× downscaling experiments, the EDSR and Real-ESRGAN models show the largest RMSEs among all the methods. Under 6× downscaling, both the EDSR and HAT models performed poorly in the upper ocean, exhibiting the highest error peaks and the most severe fluctuations, with their RMSEs nearly double those of DIFFDS3v1. This indicates that these SR-based image models struggle to effectively extract and restore submesoscale and mesoscale thermal perturbations when dealing with oceanographic data characterized by strong spatiotemporal correlations and complex physical processes. The performance of VDM and Real-ESRGAN lies between that of DIFFDS3v1 and EDSR/HAT, with VDM showing slightly higher errors than DIFFDS3v1 in most instances, although this gap narrows during the high-error summer periods.

https://doi.org/10.5194/egusphere-2026-5227
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

As the depth extends to the middle ocean layers (155.85–318.13 m), corresponding to the upper section of the permanent thermocline, the water column stratification gradually stabilizes. The influences of the high-frequency air–sea forcing and surface dynamics weakened significantly, leading to a marked reduction in the seasonal oscillation amplitude of the RMSE and a generally flattened time series curve. For instance, at a depth of 318.13 m under 3× downscaling, the fluctuation range of the RMSE for all the models was smaller than that in the upper ocean. At this depth, the performance gaps among the models narrow; however, DIFFDS3v1 still maintains a stable lead, with its RMSE curve consistently positioned at the bottom and exhibiting the smoothest fluctuations (0.033–0.05 ℃).

In the deep layers from 643.57 to 1062.44 m, the seasonal oscillation of RMSE virtually disappears, and the curves become remarkably flat, showing only faint annual fluctuations. This behavior reflects the fact that the deep ocean is far from the air–sea interface, resulting in stable water mass properties, extremely slow temperature variations, and minimal spatial heterogeneity. Consequently, the difficulty of downscaling the reconstruction was substantially reduced. All the models exhibited the lowest RMSE values, and the performance differences between them were further minimized. Nonetheless, DIFFDS3v1 maintained the lowest RMSE and smoothest curve across most time periods among all methods, suggesting favorable robustness and minimal noise levels when reconstructing the temperature attributes of stable deep-water masses.

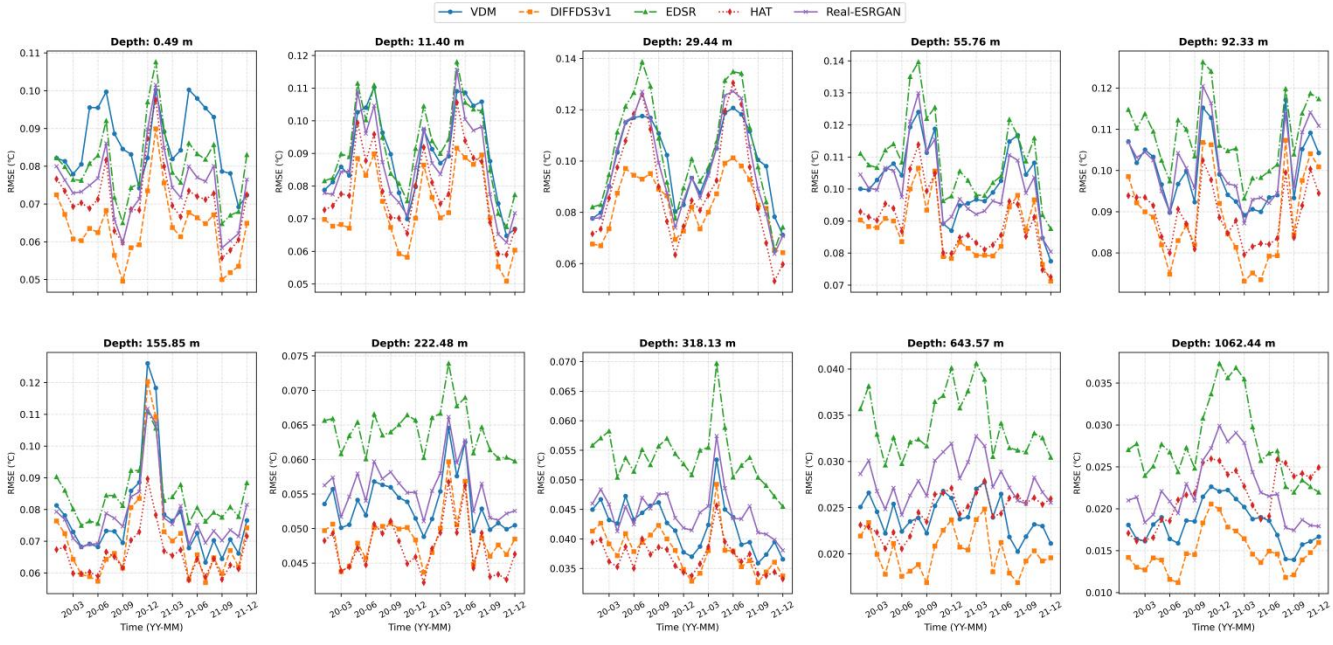


Figure 5: Variations of RMSE time series for each depth layer at 3×.

In conclusion, the reconstruction errors of seawater temperature downscaling primarily originate from high spatial heterogeneity regions in the upper ocean, which are controlled by intensive, nonlinear dynamic processes in the boreal summer. The magnitude of these errors was positively correlated with the downscaling factor. The DIFFDS3v1 model demonstrates optimal performance across all downscaling scales, depth layers, and time points. Its advantages are particularly prominent

https://doi.org/10.5194/egusphere-2026-5227
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

under a large 6× downscaling factor, and during high-error summer periods in the upper ocean. This suggests that by integrating spatiotemporal convolutions and attention mechanisms combined with the physical prior learning capability of the diffusion generative model, DIFFDS3v1 can effectively enhance both the accuracy and robustness of spatial downscaling for oceanic temperature.

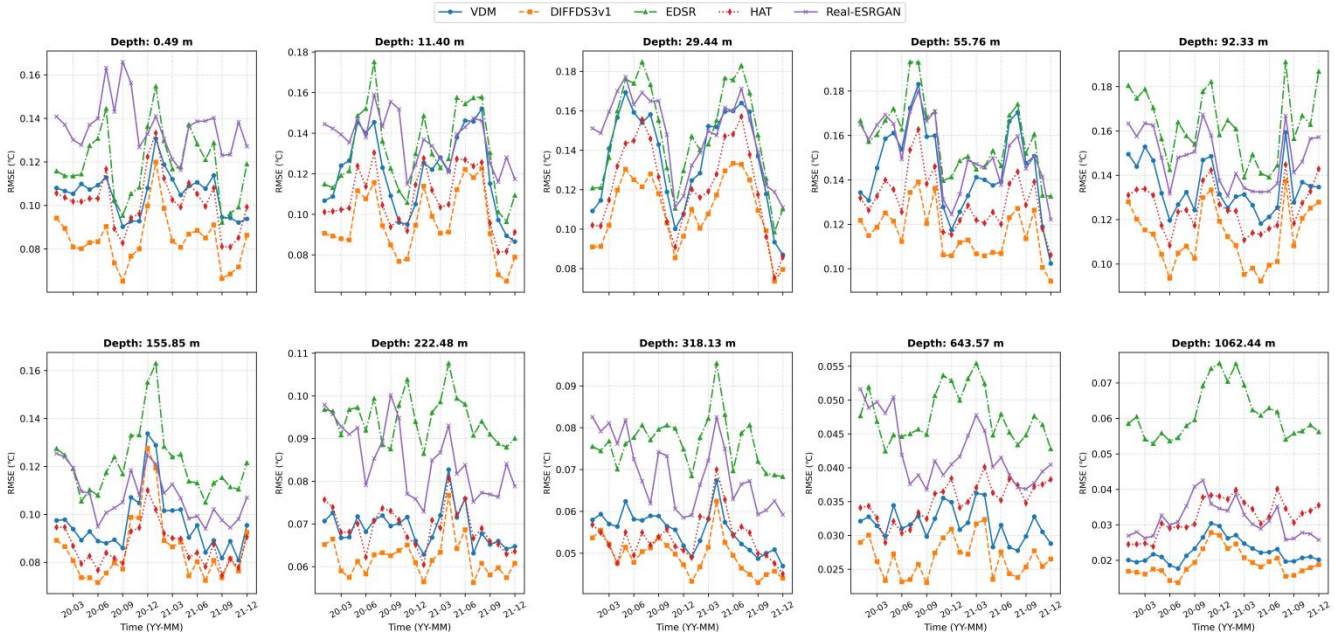


Figure 6: Variations of RMSE time series for each depth layer at 4×.

https://doi.org/10.5194/egusphere-2026-5227
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.



Figure 7: Variations of RMSE time series for each depth layer at 6×.

3.2.3 Vertical Error Distribution Analysis

Figure 8 shows the vertical distributions of the RMSE for all models under spatial downscaling factors of 3×, 4×, and 6×. The RMSE of each model exhibited a prominent vertical pattern: values were highest in the surface thermocline layer and lowest in the deep ocean layers. As the downscaling factor increased from 3× to 6×, the RMSE baseline systematically rose at all depths, and the performance divergence among models intensified. This variation pattern was highly consistent with the vertical heterogeneity of the thermal structure of the oceanic temperature field. Specifically, the upper ocean is characterized by active dynamic processes and steep temperature gradients, presenting the greatest difficulty for downscaling reconstruction, whereas the deep ocean features stable water masses and slow temperature changes, resulting in the lowest reconstruction difficulty.

https://doi.org/10.5194/egusphere-2026-5227
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

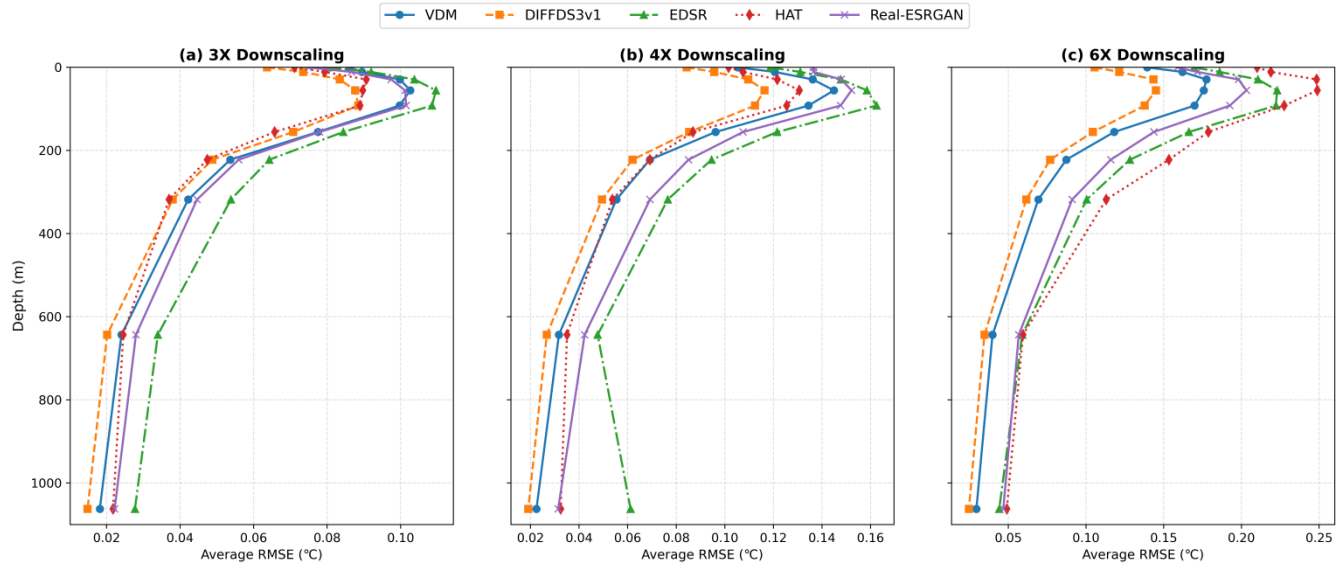


Figure 8: Vertical variations of RMSE.

The overall reconstruction error was lowest in the 3× downscaling factor experiment. The RMSE peaks (0.09–0.11 ℃) occur in the surface layer (0–100 m) and the seasonal thermocline (100–300 m), after which the values gradually decreased with increasing depth, reaching its minimum (0.01–0.03 ℃) at approximately 1000 m. DIFFDS3v1 maintains the lowest RMSE across all depths, with the maximum value 0.09 ℃ and the minimum value 0.01 ℃, and its error curve consistently positioned on the far left and exhibiting the most gradual vertical variation.

In the 4× downscaling experiment, the overall RMSE baseline rises noticeably compared with the 3× case, with the maximum RMSE of all models escalating to the 0.11–0.16 ℃ range. The vertical distribution characteristics remained consistent with those under 3× downscaling and the error peaks remained concentrated in the surface and thermocline layers. Although the relative performance ranking of the models remains unchanged, the gaps between them widen. Specifically, the RMSE gap between DIFFDS3v1 and models such as EDSR in the thermocline expands further, indicating that DIFFDS3v1 possesses stronger adaptability under more severe information loss.

Under a 6× downscaling factor, the overall baseline RMSE systemically increased further. Although the vertical distribution pattern persisted, the error peaks became more prominent and the performance divergence among the models reached its most significant level, with the leading advantage of DIFFDS3v1 expanding further. The RMSE values of EDSR and HAT in the surface and thermocline layers increased sharply, indicating that these image-SR-based models struggled to effectively reconstruct the fine structures of oceanic temperatures under extreme conditions of information deficiency.

Analyzing the error characteristics of different depth layers reveals distinct vertical regimes of reconstruction difficulty. In the upper ocean, all models exhibited the highest RMSE values and the most pronounced performance disparities. This result is consistent with that of the temporal analysis. Driven by intensive, nonlinear dynamic processes during summer, the temperature

https://doi.org/10.5194/egusphere-2026-5227
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

field in the upper ocean exhibits the highest spatial heterogeneity, presenting the greatest difficulty for downscaling reconstruction. The advantage of DIFFDS3v1 was most prominent in this region, with its error curve being the most gradual, demonstrating its capability to effectively capture complex vertical temperature gradient variations in the upper ocean. In the subsurface layer, RMSE values decreased compared with that in the upper layer, and the performance gaps among the models were narrow. This is attributed to the stabilizing stratification of the water column in the middle ocean, which weakens the influence of high-frequency air–sea interactions and surface dynamics, thereby reducing the spatial heterogeneity of the temperature field. In the deep ocean, the RMSE drops to its lowest level, with the error curves of all models nearly overlapping, and the performance disparities are minimized. Because the deep ocean is far from the air–sea interface, it features stable water mass properties, extremely slow temperature variations, and minimal spatial heterogeneity, which significantly reduces the difficulty of downscaling reconstruction. Even in these stable deep layers, DIFFDS3v1 maintained the lowest RMSE, indicating its robustness when reconstructing the temperature properties of stable deep-water masses.

Vertical error distribution analysis further validated that the reconstruction errors of oceanic temperature downscaling primarily originated from high spatial heterogeneity regions in the upper ocean, and the magnitude of these errors was positively correlated with the downscaling factor. The DIFFDS3v1 model demonstrated optimal performance across all downscaling factors and depth layers, with its error curves exhibiting the most gradual vertical variation. This suggests that combined with the physical prior learning capability of the diffusion generative model, DIFFDS3v1 can effectively capture the dynamical correlations among oceanic temperatures at different depth layers by integrating spatiotemporal convolutions and attention mechanisms. This ensures the vertical physical continuity of the 3D oceanic temperature field. Its advantages are particularly prominent under a large 6× downscaling factor and in high-error upper ocean regions.

3.2.4 Spatial Error Distribution Analysis

In this section, further analyzes the spatial distribution characteristics of the RMSE from the 6× spatial downscaling experiment. Notably, the RMSE values from the 3× and 4× experiments (not shown) were smaller than those from the 6× experiment but exhibited the same spatial pattern as the 6× experiment. Figure 9 illustrates the spatial distribution of the average RMSE for the six mainstream downscaling methods in the upper layer (11.40 m), whereas Figure 10 presents the spatial distribution of the average RMSE for the DIFFDS3v1 model across the 10 depth levels. The results indicated that the spatial downscaling errors of seawater temperature across all models exhibited high spatial heterogeneity, a distribution closely coupled with the complex upper-ocean dynamic environment of this sea area.

Pronounced high-error zones are primarily concentrated in active dynamic areas, including the Kuroshio main-axis region east of Taiwan island and the Philippines, the Luzon Strait, and the SCS western boundary current region east of Vietnam. The Kuroshio, as a strong western boundary current, is accompanied by intense baroclinic instability and substantial horizontal temperature gradients. Meanwhile, the Luzon Strait region is influenced by the Kuroshio intrusion by loop currents or eddy shedding processes, resulting in high mesoscale eddy kinetic energy and active submesoscale processes. The high-frequency spatial variability of temperature in these regions was absent in the 1/2° low-resolution input field, leading to a significant

https://doi.org/10.5194/egusphere-2026-5227
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

increase in the difficulty of reconstructing the thermal structure. Traditional bicubic interpolation, which relies solely on the linear smoothing of local grid points, struggles to recover the sharp fronts of current boundaries and nonlinear eddy structures, resulting in extensive areas of high errors along the Kuroshio path and Luzon Strait. In contrast, deep learning models significantly reduce the extent of these high-error zones through nonlinear feature extraction, although differences remain when resolving dynamic fronts with severe isotherm fluctuations. Among these, the DIFFDS3v1 model proposed in this study showed favorable performance in reconstructing spatial details, exhibiting narrower high-error bands in both the Kuroshio main axis and Luzon Strait. This suggests that DIFFDS3v1 can capture and map the physical constraints implicit in low-resolution flow fields and recover mesoscale and submesoscale dynamic processes from the background field to achieve competitive accuracy in 3D seawater temperature spatial downscaling under strongly dynamical conditions.

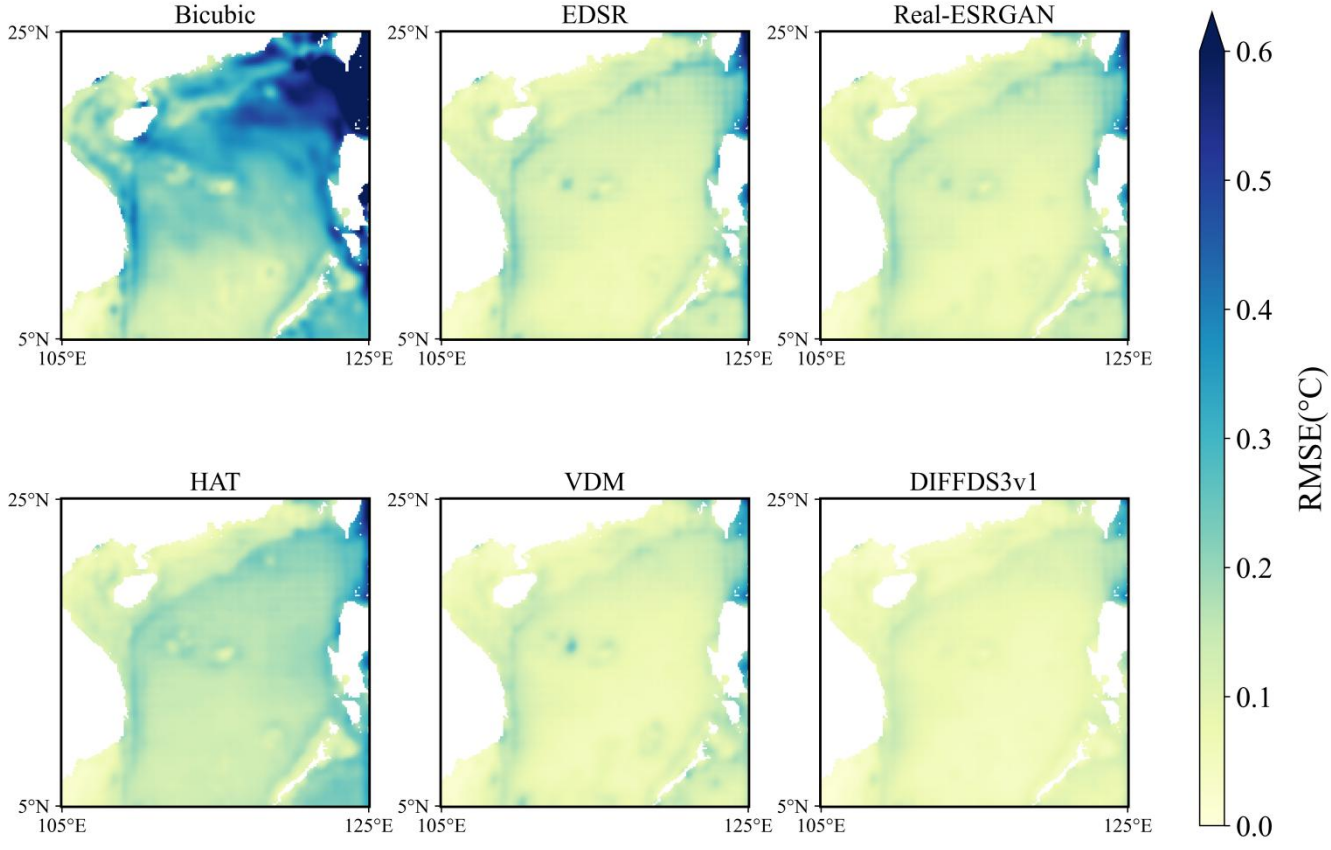


Figure 9: Spatial distribution of mean RMSE for downscaling results of 6× factor at 11.40 m.

https://doi.org/10.5194/egusphere-2026-5227
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

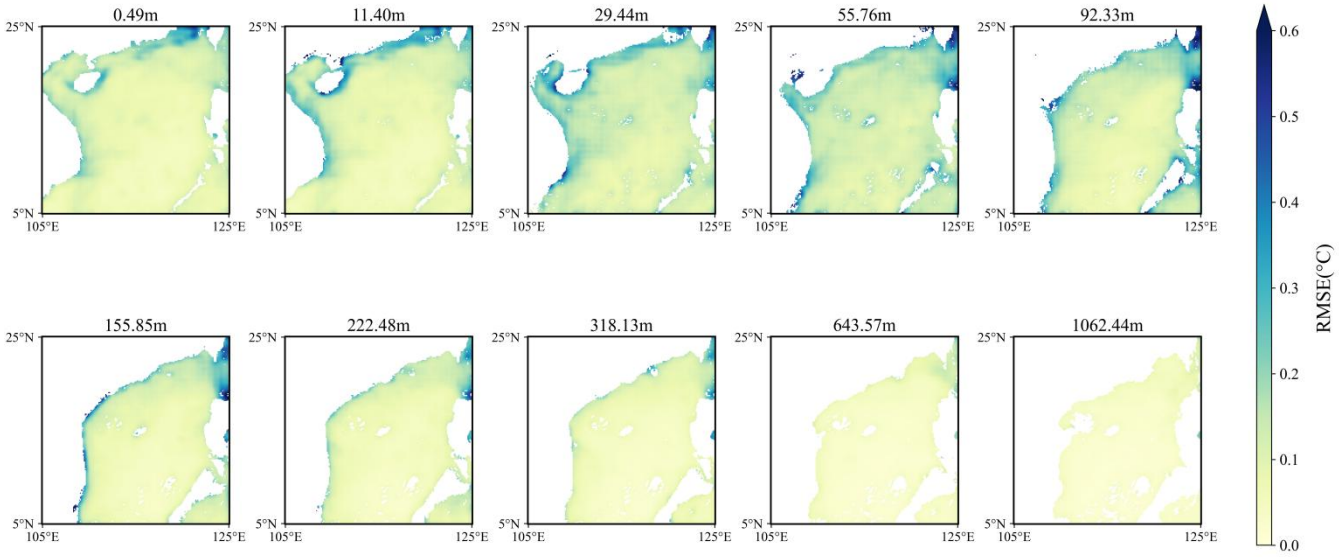


Figure 10: Spatial Distribution of mean RMSE for DIFFDS3v1 at 6× factor.

To further explore the spatial distribution of errors at different depths, the average RMSE spatial distribution of DIFFDS3v1 across the ten depth levels is shown in Figure 10. The RMSE spatial distribution of DIFFDS3v1 exhibited clear region-depth coupling characteristics, closely aligning with the spatial heterogeneity of oceanographic dynamical processes in the study area. Vertically, high RMSE values in the surface layer (11.40 m) are primarily concentrated in the core region of the Vietnamese coastal upwelling, the shelf break on the northeast of the SCS, and the upwelling zone eastern of Hainan Island, whereas the RMSE in the central deep-sea basin is generally below 0.1 ℃, presenting a "high nearshore, low offshore" pattern. In the mixed layer (55.76 m), the extent of the high-value area contracts compared with the surface layer, maintaining relatively high values only in the subsurface core of the Vietnamese coastal upwelling and slope zones, where the thermocline changes abruptly. Down to the subsurface layers (92.33 and 222.48 m), the spatial heterogeneity of RMSE further weakens, with high-value areas sporadically distributed only in abrupt topographic zones at the basin margins, and the RMSE in the open deep sea decreasing further.

The physical cause of this distribution pattern is that, under the 6× downscaling factor, the original data can only partially capture the core signals of large-scale ocean processes in the SCS. DIFFDS3v1 relies on these physical constraints to reconstruct fine temperature gradients, with minor errors occurring only in high-gradient regions (such as the upwelling and Kuroshio zones) owing to the resolution limitations of the original data. Shielded by the vertical barrier of the thermocline, spatial variations in seawater temperature in the subsurface are mainly controlled by horizontal advection and topography-induced circulation, where the gradient intensities are weaker than those at the surface, leading to a higher reconstruction accuracy. At deeper levels (such as 222.48 m), which approach the homogeneous layer of deep SCS water, spatial variations in sea temperature are dominated by weak topography-induced circulation and diffusion. With extremely small spatial gradients, the downscaling error tends to be uniform and remains low across the entire region.

https://doi.org/10.5194/egusphere-2026-5227
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

The spatial error distribution analysis indicated that the high-error regions of 3D seawater temperature downscaling coincided strongly with the intensely dynamic areas of the SCS. By incorporating key physical elements such as flow fields, the DIFFDS3v1 model effectively narrowed the high-error bands in the strong current and frontal regions. Even under the challenging 6× downscaling factor, DIFFDS3v1 prevented the widespread dispersion of errors and maintained clear marine thermal boundaries, validating the importance of multiphysical data-driven approaches in improving reconstruction accuracy.

4 Discussion

4.1 Specific Samples Examination

The preceding sections provide a comprehensive evaluation of the overall performance of the DIFFDS3v1 model using quantitative statistical metrics. To further illustrate the 3D reconstruction capability of the model across different seasons and depths, this section presents the seawater temperature reconstruction results of DIFFDS3v1. Two representative 4× downscaling cases were examined. These two cases corresponded to July 15, 2020 (summer) and January 15, 2021 (winter), and were evaluated at four depths (11.40, 55.76, 92.33, and 222.48 m).

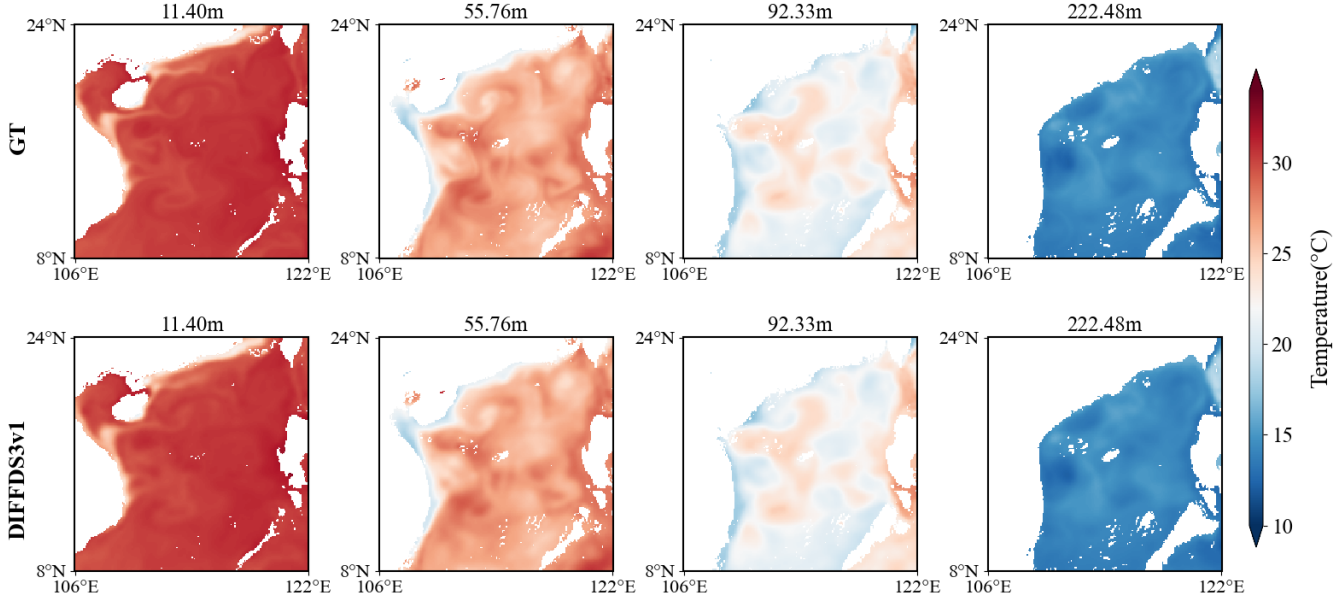


Figure 11: 4× spatial downscaling results of DIFFDS3v1 at the depths of 11.40, 55.76, 92.33, and 222.48 m on July 15, 2020.

July 15, 2020, falls within the peak period following the onset of the SCS summer monsoon, a season characterized by intense air–sea interactions and pronounced spatial heterogeneity and vertical stratification in the temperature field. The surface layer (11.40 m) generally showed elevated seawater temperatures, with a distinct cold-water tongue formed along the coast of central Vietnam. This is a hallmark summertime dynamic feature of the SCS, which is closely associated with Ekman transport and coastal upwelling driven by the southwesterly monsoon (Xie et al., 2003; Kuo et al., 2000). As depth increases, temperature

https://doi.org/10.5194/egusphere-2026-5227
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

progressively decreases, and the field becomes nearly homogeneous by 222.48 m, consistent with the understanding that water properties below the thermocline are less affected by seasonal forcing and tend toward stable water-mass characteristics (Shaw, 1991). As shown in Figure 11, DIFFDS3v1 satisfactorily reproduced the spatial distribution of the temperature field at all four depths. At the surface, the model accurately captured the position of Vietnam's cold core eddy and its coastal extension. At 55.76 and 92.33 m, it reasonably resolves the subsurface structure of the upwelling. Even at the deeper 222.48 m level, the reconstructed field remained in close agreement with the ground truth, without generating obvious spurious structures or non-physical artifacts.

January 15, 2021, corresponded to the peak period of the SCS winter monsoon. A strong northeasterly monsoon drives vigorous vertical mixing, resulting in a markedly deepened mixed layer (Wang et al., 2020; Wang et al., 2025). The upper-ocean temperature field displays a clear zonal distribution pattern. Temperatures on the northern continental shelf generally fall below 20 °C, whereas the southern basin remains largely above 25 °C. This north–south contrast is linked to winter monsoon-driven meridional circulation and the modulating influence of the Kuroshio intrusion through the Luzon Strait (Lin et al., 2022; Wang et al., 2020). Meanwhile, a pronounced temperature gradient developed along the coast of mainland China under the influence of coastal currents. As shown in Figure 12, DIFFDS3v1 effectively reproduced the overall winter distribution pattern. The boundary between the low-temperature zone on the northern shelf and the high-temperature southern basin was sharply and clearly delineated, and the cold-water tongue along the coast of mainland China was well represented. From the surface to 222.48 m, the spatial pattern of the reconstructed field remained in close agreement with the ground truth.

These two representative case studies corroborate the quantitative and statistical conclusions presented earlier from the perspective of 3D spatial and vertical structures. The results demonstrate that DIFFDS3v1 exhibits robust 3D temperature reconstruction capability under contrasting seasonal dynamical regimes. The model not only recovers the absolute values and spatial distribution of sea temperature, but also captures depth-dependent dynamical features associated with monsoon forcing, upwelling, and coastal currents. This is consistent with previous studies emphasizing the importance of multiphysics information for improving the reconstruction accuracy of ocean variables (Wang et al., 2021; Yang et al., 2025). These findings further validate the effectiveness of the proposed multiphysics guidance mechanism combined with the joint horizontal–vertical feature extraction strategy.

https://doi.org/10.5194/egusphere-2026-5227
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

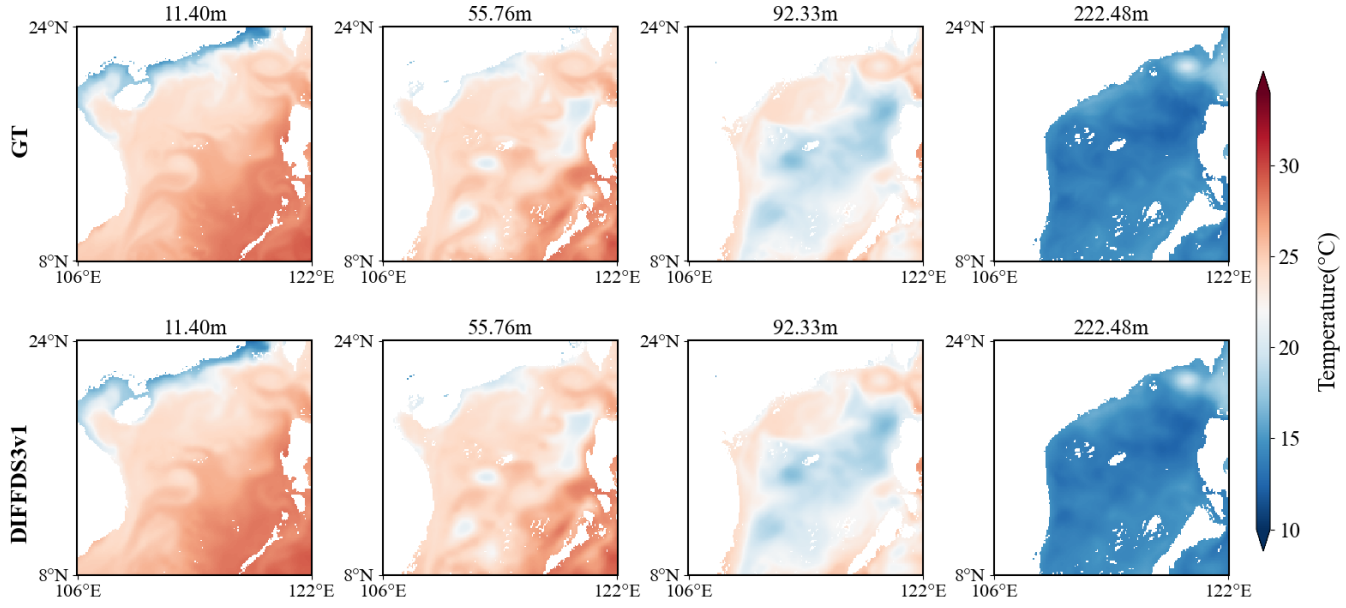


Figure 12: 4× spatial downscaling results of DIFFDS3v1 at the depths of 11.40, 55.76, 92.33, and 222.48 m on January 15, 2021.

4.2 Reconstruction of Seasonal Mixed Layer Depth

530 This section further examines the reconstruction performance of each model for mixed layer depth (MLD) under the 6× spatial downscaling, using the summer case of July 15, 2020, and the winter case of January 15, 2021, as illustrative examples (Figures 13 and 14).

https://doi.org/10.5194/egusphere-2026-5227
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

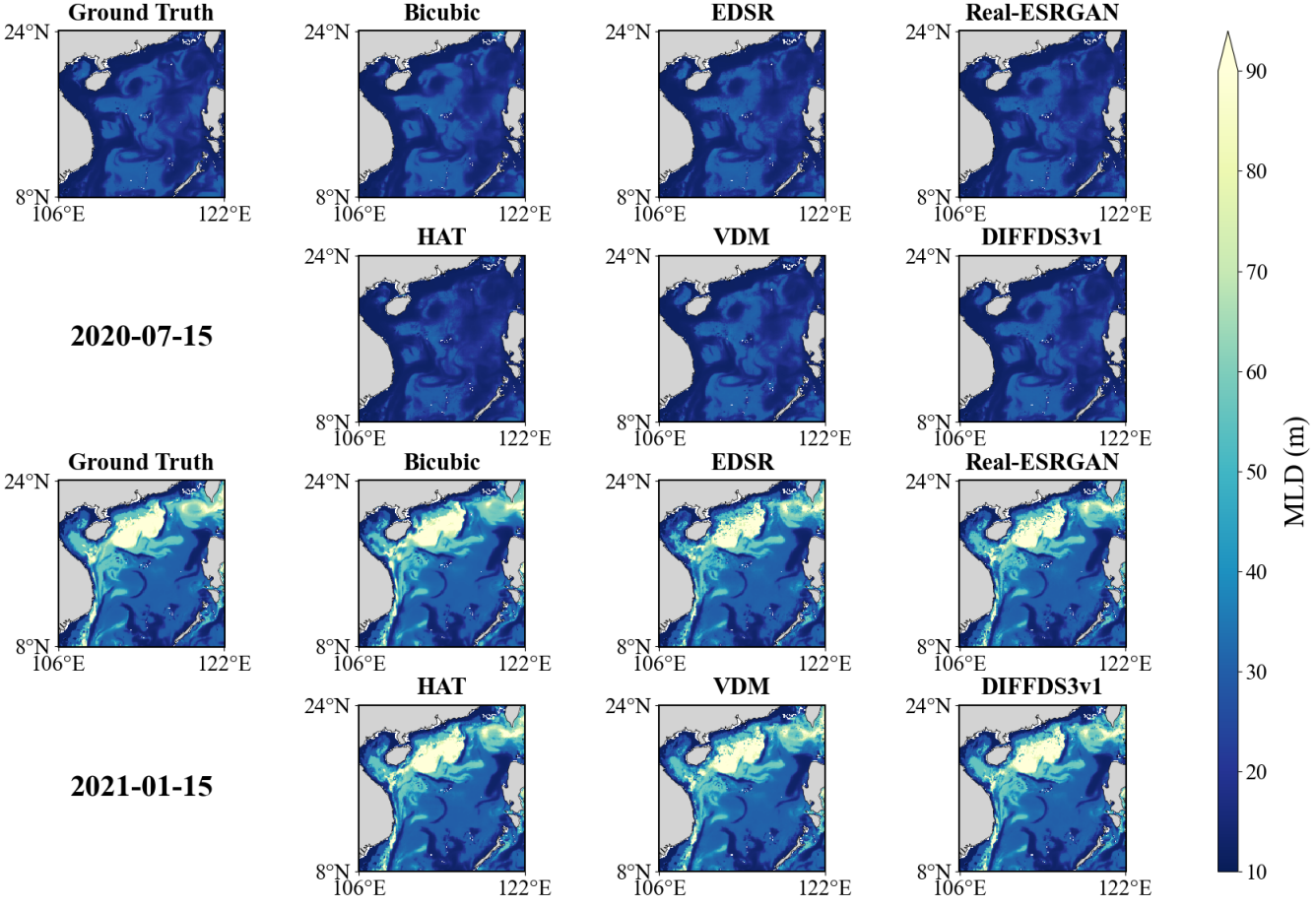


Figure 13: MLD distributions of ground truth and different methods under 6× spatial downscaling on July 15, 2020 and January 15, 2021.

The MLD field in the study region exhibited pronounced seasonal and regional coupling. During the summer southwest monsoon, the mixed layer across the northern SCS is generally shallow (10–30 m), reflecting weaker wind-driven mixing during this season (Qu, 2001). Within the core of the Vietnamese coastal upwelling, cold water advection driven by wind-induced Ekman transport further suppressed vertical mixing, shoaling the MLD to below 10 m (Xie et al., 2003; Wang et al., 2020). In contrast, to the west of Luzon Strait, horizontal advection associated with the Kuroshio intrusion moderately deepened the mixed layer to 20–30 m, consistent with previous observations of Kuroshio-induced modulation of upper ocean stratification in this region (Nan et al., 2015; Metzger & Hurlburt, 2001). During winter, the intensification of the northeast monsoon substantially strengthens surface wind stress and buoyancy loss (Qu et al., 2001), deepening the basin-wide MLD to 20–80 m. Along the coastal and shelf zones, topographically enhanced tidal and wind-driven mixing further deepens the layer to 60–80 m (Wang et al., 2025).

https://doi.org/10.5194/egusphere-2026-5227
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

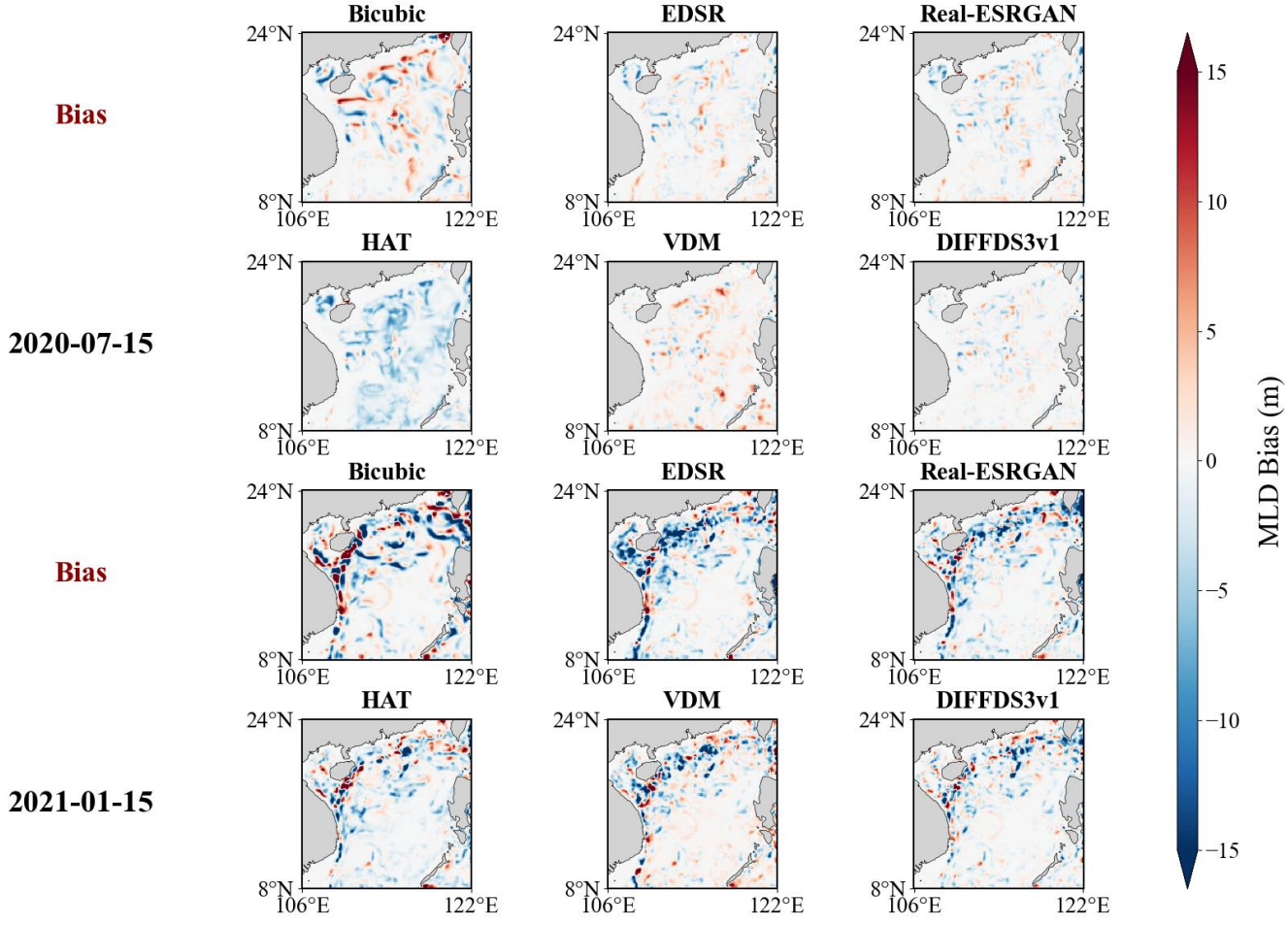


Figure 14: MLD bias distributions of different methods under 6× spatial downscaling on July 15, 2020 and January 15, 2021.

The bias maps (Figure 14) revealed that the bicubic interpolation exhibited the most pronounced errors along the coastal boundaries and regions of abrupt bathymetric change, with a clear seasonal dependence. Because bicubic interpolation relies solely on geometric interpolation, without incorporating any physical constraints related to upwelling or frontal dynamics, it fails to capture the sharp MLD gradients associated with the Vietnam upwelling core during summer (Xie et al., 2003). In winter, as topography-enhanced mixing deepens the MLD to 60–80 m over the shelf, bicubic interpolation cannot reproduce this spatially heterogeneous deepening, resulting in large localized biases. However, in an open basin, where the MLD field is comparatively smooth, its bias remains modest.

Data-driven SR models, such as EDSR and Real-ESRGAN, along with the diffusion-based VDM, learn the underlying spatial structure of the temperature field directly from training data. This reduced the overall MLD bias and reasonably reproduced the basin-scale seasonal cycle of a shallow summer mixed layer and a deepened winter mixed layer, which is consistent with the general seasonal MLD variability reported for the SCS (Qu, 2001; Wang et al., 2025). Nonetheless, a noticeable residual

https://doi.org/10.5194/egusphere-2026-5227
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

bias remains in dynamically active regions, such as the Vietnam upwelling zone and Kuroshio intrusion front near the Luzon Strait, as purely data-driven architectures lack explicit constraints on the underlying ocean dynamical processes governing these features (Dueben et al., 2018; Raissi et al., 2019). By contrast, DIFFDS3v1 incorporates 3D temperature-derived physical priors during the downscaling process, allowing it to implicitly represent monsoon-driven mixing and topographically modulated circulation (Nan et al., 2015). As a result, DIFFDS3v1 achieved a markedly lower bias in the key dynamical regions examined here, namely, the Vietnam upwelling core in summer, Kuroshio intrusion front in winter, and shelf-break zone, yielding a more reliable reconstruction of both the spatial heterogeneity and seasonal evolution of the mixed layer.

Overall, the MLD analysis indicates that by integrating dynamical prior information such as wind- and current-related fields, DIFFDS3v1 more faithfully reconstructs the spatial heterogeneity and seasonal evolution of the SCS mixed layer than the purely interpolation-based or data-driven baseline models. Its superior performance is particularly evident in dynamically critical regions, including the Vietnamese coastal upwelling zone and Kuroshio intrusion area near the Luzon Strait. This underscores the effectiveness of the proposed multiphysics-guided spatial downscaling strategy for reconstructing physically consistent ocean subsurface fields.

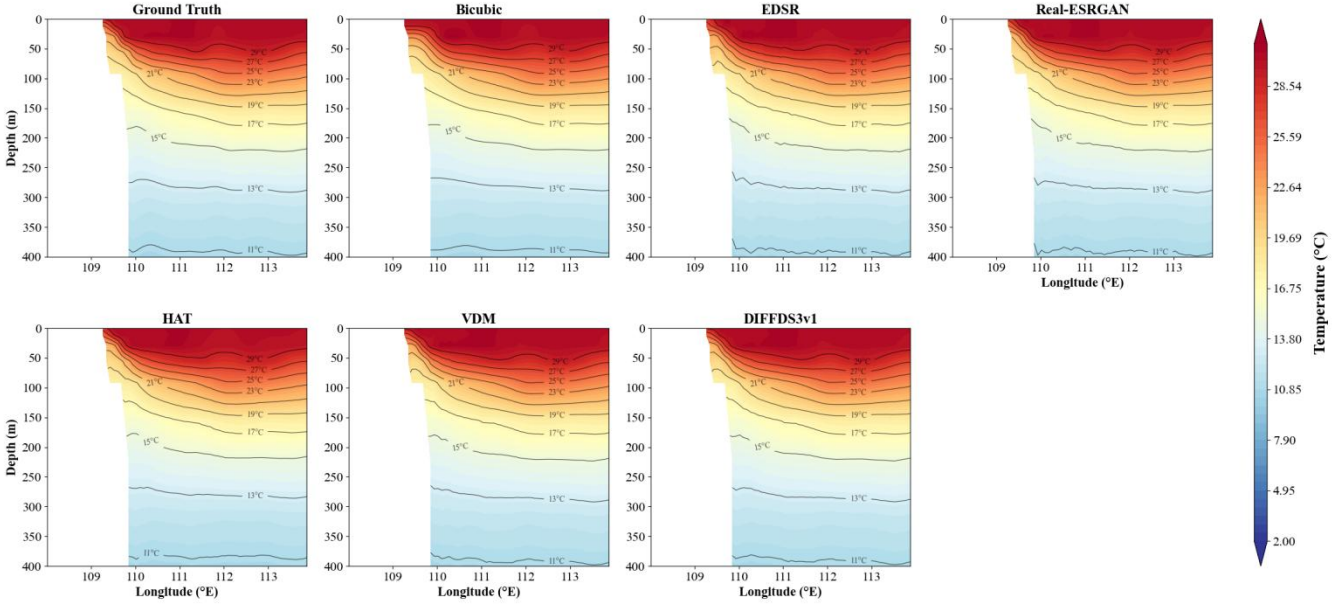
4.3 Sections across Anticyclonic Eddies

Mesoscale eddies constitute the primary pathway for the lateral transport of heat, mass, and biogeochemical tracers throughout the global oceans (Chelton et al., 2011; Zhang et al., 2014). To further assess the capability of the model to reconstruct subsurface thermal structures, we examined a representative anticyclonic warm eddy centered near 12.0°N within the study region (Figure 15). Its 3D temperature structure exhibited a characteristic bowl-shaped signature, with isotherms in the upper 50–200 m depressed markedly downward. This signature is consistent with the convergent downwelling of warm surface water associated with anticyclonic rotation, and the resulting deepening of the thermocline beneath such eddies (Chelton et al., 2011; Zhang et al., 2016).

When performing high-ratio spatial downscaling reconstruction of this dynamically complex thermal field, the mainstream deep-learning baselines, including EDSR, Real-ESRGAN, HAT, and VDM, all exhibit reasonable basic mapping capability. They successfully recovered the overall downward-bowing shape of the eddy isotherms from the low-resolution input, with only minor discrepancies among the baseline outputs at the mesoscale and submesoscale. This is broadly consistent with prior findings that data-driven SR networks are generally effective in reproducing large-scale coherent structures in oceanic fields (Wang et al., 2021; Yang et al., 2025). However, the DIFFDS3v1 model proposed in this study showed a markedly higher degree of consistency with the ground truth when reconstructing fine-scale isotherm structures. It achieves an accurate recovery of the temperature field while avoiding gradient blurring, an artifact commonly reported in interpolation-based and convolutional SR approaches when reconstructing sharp thermal fronts (Kadow et al., 2020; Su et al., 2021). Moreover, within the weakly stratified deep-water zone below 200 m, DIFFDS3v1 maintains a clean, artifact-free background field, whereas the baselines EDSR, Real-ESRGAN, and VDM all introduce a spurious 11 °C isotherm not present in the ground truth. This form

https://doi.org/10.5194/egusphere-2026-5227
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

of downscaling-induced noise has also been noted in purely data-driven reconstructions that lack explicit physical constraints in weakly forced ocean regions (Su et al., 2021).



595 **Figure 15: Temperature along the section cross anticyclonic eddy for ground truth and different models under 6× spatial downscaling.**

These results indicate that DIFFDS3v1 can accurately fit the absolute values of the 3D temperature field and faithfully capture high-frequency spatial gradients associated with eddy-induced thermal fronts. This dual capability reflects the high degree of physical fidelity in reconstructing the ocean's subsurface thermal structure, further corroborating the effectiveness of incorporating 3D physical priors into the spatial downscaling of mesoscale features (Chelton et al., 2011; Zhang et al., 2014).

600 **4.4 Ablation Experiments on the Guidance Effects of Multi-Physical Factors**

To quantify the individual contributions of each physical guiding variable, this section presents an ablation study of the auxiliary inputs of the EFN under a 6× spatial downscaling factor (Table 3). Specifically, five configurations of DIFFDS3v1 were compared: without any auxiliary physical field, without wind, without current, without salinity, and with the complete model incorporating wind, current, and salinity simultaneously.

605 The results indicated that multi-physical guidance systematically improved the accuracy of 3D temperature downscaling. With respect to the baseline configuration without any auxiliary input, the complete model incorporating all three physical fields reduces the RMSE by 9.6% (from 0.1056 to 0.0955 °C) and increases the PSNR from 52.01 to 52.84 dB. This confirms that external dynamical constraints can partially compensate for the information loss inherent in low-resolution temperature fields, which is consistent with previous studies emphasizing that auxiliary environmental variables help resolve fine-scale ocean 610 structures that temperature fields alone cannot fully constrain (Wang et al., 2021; Meng et al., 2021). Removing individual

https://doi.org/10.5194/egusphere-2026-5227
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

physical factors degrades the model performance to differing extents, revealing a clear hierarchy of importance among the three guiding variables. The absence of the current field produces the most severe degradation, with RMSE increasing from 0.0955 to 0.1034 °C, an 8.3% increase. This underscores the dominant role of ocean currents in shaping the 3D temperature structure of the SCS, which is consistent with the well-established view that geostrophic and eddy-driven currents govern subsurface heat redistribution and the formation of thermal fronts in this region (Chelton et al., 2011; Nan et al., 2015). The removal of salinity leads to a moderate degradation, with RMSE rising to 0.1011 °C, a 5.9% increase, reflecting the physical coupling between salinity and density-driven stratification that modulates subsurface temperature gradients (Qu, 2001; Shaw, 1991). By contrast, omitting the wind field resulted in the smallest performance loss, with RMSE increasing only to 0.0979 °C, a 2.5% increase. This suggests that wind forcing primarily influences the upper mixed layer and exerts a comparatively limited direct constraint on the broader 3D temperature field compared to current and salinity information (Herrmann et al., 2024).

Taken together, these results indicate that ocean currents constitute the most critical physical driver of 3D temperature downscaling in the SCS, followed by salinity, with wind forcing playing a secondary role. The joint extraction and fusion of these three physical fields through the EFN module provided implicit dynamic constraints during the denoising process of the DM, thereby enhancing the physical consistency of the reconstructed temperature field (Yang et al., 2020). This hierarchy offers practical guidance for future lightweight model designs. In operational settings with limited computational resources, prioritizing the current field as the core guiding variable may substantially reduce training costs while preserving the reconstruction accuracy.

Table 3. Ablation experiments on the effectiveness of different physical factors under 6× downscaling with DIFFDS3v1.

Auxiliary Inputs			Metrics		
Wind	Current	Salinity	RMSE (°C) ↓	MAE (°C) ↓	PSNR (dB) ↑
			0.1056	0.05	52.01
	✓	✓	0.0979	0.05	52.63
✓		✓	0.1034	0.05	52.20
✓	✓		0.1011	0.05	52.35
✓	✓	✓	**0.0955**	**0.04**	**52.84**

5 Conclusion

To address the growing demand for high-resolution 3D oceanic temperature data, this study proposed and validated DIFFDS3v1, a novel spatial downscaling framework based on a generative DM. This study explores the significant potential of the DIFFDS3v1 approach to bridge the gap between sparse, low-resolution observations and the data-intensive requirements of current oceanographic science.

The experimental results unequivocally demonstrate that DIFFDS3v1 establishes a state-of-the-art approach across multiple downscaling scales, significantly outperforming a suite of benchmark models including those based on CNN, GAN, and transformer architectures. Quantitatively, for 3×, 4×, and 6× downscaling tasks, DIFFDS3v1 achieved RMSE values of 0.06, 0.07, and 0.10 ℃, respectively, with PSNR reaching 57.07, 54.85, and 52.84 dB. This superior performance stems from a

https://doi.org/10.5194/egusphere-2026-5227
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

synergistic architectural design; the robust 3D conditional U-Net backbone proved to be highly effective for the core denoising task, whereas the innovative EFN successfully distilled the critical physical context from wind, salinity, and ocean velocity data. This dual-network approach, integrated via cross-attention, proved particularly effective in reconstructing complex ocean dynamic processes, such as MLD, mesoscale eddies, and frontal structures, with fidelity that other models could not match.

The success of DIFFDS3v1 has significant implications for the field. This demonstrates that physics-guided generative models can serve as powerful virtual sensors capable of generating multi-depth, dynamically consistent, and physically plausible datasets that are comparable in quality to raw high-resolution observations. This provides a viable pathway for enriching existing ocean datasets and enhancing the accuracy of downstream tasks such as operational forecasting.

Future work will focus on enhancing the physical realism and interpretability of the model. One critical direction involves leveraging widely available high-resolution satellite SST products. We plan to investigate how these fine-scale surface observations can serve as strong boundary constraints to guide the 3D reconstruction process, thereby improving the vertical propagation of surface information into subsurface layers. Additionally, we aim to integrate other oceanic or atmospheric variables (e.g., sea surface height anomalies and air–sea heat fluxes) and explore methods to embed explicit physical constraints, such as conservation laws, directly into the architecture of the model. These advancements will continue to refine the effectiveness of the DIFFDS3v1 framework, pushing the frontier of AI-driven oceanographic data enhancement, and enabling more accurate and reliable Earth system science.

Code and data availability

The seawater temperature, salinity, and velocity data used in this study are publicly available from the Copernicus Marine Data Store (https://data.marine.copernicus.eu/products/GLOBAL_MULTIYEAR_PHY_001_030/description, DOI: 10.48670/moi-00021, Jean-Michel et al., 2021), while the wind data are publicly available from the Copernicus Climate Data Store (https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels?tab=overview, DOI: 10.24381/cds.adbb2d47, Hersbach et al., 2020). The code used in this study can be obtained from https://zenodo.org/records/24214824. The data used in this study span 2005–2021, with 2005–2018 used for training, 2019 for validation, and 2020–2021 for testing. Both datasets were accessed on 13 June 2025.

Author contributions

SW, XMZ and XRZ, Writing original draft; XMZ, conceptualization, and writing review & editing; SW and WF, investigation, data curation and methodology; XRZ and YH, formal analysis.

https://doi.org/10.5194/egusphere-2026-5227
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

Competing interests

The authors declare no conflicts of interest relevant to this study.

Acknowledgements

The authors are grateful for the professional suggestions and comments of anonymous reviewers and the handling editors.

Financial support

This work was supported by the project of Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai) (No. SML2023SP202，SML2022SP401).

References

Conde, M. V., Choi, U.-J., Burchi, M., & Timofte, R.: Swin2sr: Swinv2 transformer for compressed image super-resolution and restoration. In Computer vision–eccv 2022 workshops (pp. 669–687). Berlin, Heidelberg: Springer-Verlag. https://doi.org/10.1007/978-3-031-25063-742, 2023.

de Burgh-Day, C. O., & Leeuwenburg, T.: Machine learning for numerical weather and climate modelling: a review. Geoscientific Model Development, 16 (22), 6433–6477. https://doi.org/10.5194/gmd-16-6433-2023, 2023.

Dong, C., Loy, C. C., He, K., & Tang, X.: Image super-resolution using deep convo- lutional networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(2), 295–307. https://doi.org/10.1109/TPAMI.2015.2439281, 2016.

Ducournau, A., & Fablet, R.: Deep learning for ocean remote sensing: an application of convolutional neural networks for super-resolution on satellite-derived sst data. In 2016 9th iapr workshop on pattern recognition in remote sensing (prrs). Cancun, Mexico: IEEE. https://doi.org/10.1109/prrs.2016.7867019, 2016.

Feng, Z., Qi, J., Li, D., Xie, B., Sun, G., Yin, B., & Yang, S.: Attention-enhanced deep learning model for reconstruction and downscaling of thermocline depth in the tropical indian ocean. Ocean Modelling, 196, 102537. https://doi.org/10.1016/j.ocemod.2025.102537, 2025.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al.: Generative adversarial nets. In Proceedings of the 28th international conference on neural information processing systems - volume 2 (pp. 2672–2680), 2014.

Herrmann, M., To Duy, T., & Marsaleix, P.: Mechanisms and intraseasonal variability in the south vietnam upwelling, South China Sea: the role of circulation, tides, and rivers. Ocean Science, 20(4), 1013–1033. https://doi.org/10.5194/os-20-1013-2024, 2024.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horanyi, A. J., et al.: The era5 global reanalysis. Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049. https://doi.org/10.1002/qj.3803, 2020.

https://doi.org/10.5194/egusphere-2026-5227
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

Ho, J., Jain, A., & Abbeel, P.: Denoising diffusion probabilistic models. In 34th international conference on neural information processing systems (neurips) (pp. 6840– 6851), 2020.

Hu, Y., Zhong, Z., Liu, X., & Zhu, Y.: Influence of air–sea interaction on the simulation of east asian summer monsoon: A case study. Dynamics of Atmospheres and Oceans, 53-54, 1–16. https://doi.org/10.1016/j.dynatmoce.2011.12.001, 2012.

Karras, T., Laine, S., & Aila, T.: A style-based generator architecture for gen- erative adversarial networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(12), 4217–4228. https://doi.org/10.1109/TPAMI.2020.2970919, 2021

Khoo, J. J. D., Lim, K. H., & Pang, P. K.: Deep learning super resolution of sea surface temperature on south china sea. In 2022 international conference on green energy, computing and sustainable technology (gecost). https://doi.org/10.1109/gecost55694.2022.10010371, 2022.

Kim, E.-J., & Hong, S.-Y.: Impact of air-sea interaction on east asian summer monsoon climate in wrf. Journal of Geophysical Research: Atmospheres, 115 (D19). https://doi.org/10.1029/2009JD013253, 2010.

Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., Acosta, A., et al.: Photo-realistic single image super-resolution using a generative adversarial network. In 2017 IEEE conference on computer vision and pattern recognition (cvpr), 2017.

Jean-Michel, L., Eric, G., Romain, B.-B., Gilles, G., Angélique, M., Marie, D., Clément, B., Mathieu, H., Olivier, L.G., Charly, R., Tony, C., Charles-Emmanuel, T., Florent, G., Giovanni, R., Mounir, B., Yann, D. and Pierre-Yves, L.-T.: The Copernicus Global 1/12° Oceanic and Sea Ice GLORYS12 Reanalysis. Front. Earth Sci. 9:698876. http://doi.or/10.3389/feart.2021.698876, 2021.

Li, H., Yang, Y., Chang, M., Chen, S., Feng, H., Xu, Z., Chen, Y.: Srdiff: Single image super-resolution with diffusion probabilistic models. Neurocomputing, 479 (4), 47–59. https://doi.org/10.1016/j.neucom.2022.01.029, 2022.

Li, X., & Chen, T.: Efficient synthesis of parameterized gaussian-like filters by approximation. Signal Processing, 41(2), 119–134, https://doi.org/10.1016/ 0165-1684(94)00095-H, 1995.

Lim, B., Son, S., Kim, H., Nah, S., & Lee, K. M. Enhanced deep residual networks for single image super-resolution. In 2017 ieee conference on computer vision and pattern recognition workshops (cvprw) (pp. 1132–1140). https://doi.org/10.1109/CVPRW.2017.151, 2017.

Lin, H., Cheng, X., Wu, X., & Shen, D.: Cat: Cross attention in vision transformer. In 2022 ieee international conference on multimedia and expo (icme) (pp. 1–6). https://doi.org/10.1109/ICME52920.2022.9859720, 2022.

Liu, S., Jia, W., Wang, Q., Zhang, W., & Wang, H.: Enhancing the resolution of satellite ocean data using discretized satellite gridding neural networks. Remote Sensing, 16(16), 3020. https://doi.org/10.3390/rs16163020, 2024.

Lu, Z., Li, J., Liu, H., Huang, C., Zhang, L., & Zeng, T.: Transformer for single image super-resolution. In 2022 ieee/cvf conference on computer vision and pattern recognition workshops (cvprw) (pp. 456–465). https://doi.org/10.1109/CVPRW56347.2022.00061, 2022.

Ngo, M.-H., & Hsin, Y.-C.: Impacts of wind and current on the interannual variation of the summertime upwelling off southern Vietnam in the South China Sea. Journal of Geophysical Research: Oceans. https://doi.org/10.1029/2020JC016892, 2021.

https://doi.org/10.5194/egusphere-2026-5227
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

Nichol, A. Q., & Dhariwal, P.: Improved denoising diffusion probabilistic models. In 38th international conference on machine learning (icml) (pp. 8162–8171), 2021.

Pastor, F.: Sea surface temperature: From observation to applications. Journal of Marine Science and Engineering, 9(11), 1284. https://doi.org/10.3390/ jmse9111284, 2021.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B.: High-resolution image synthesis with latent diffusion models. In 2022 ieee/cvf conference on computer vision and pattern recognition, 2022.

Saharia, C., Chan, W., Saxena, S., Lit, L., Whang, J., Denton, E., et al.: Photorealistic text-to-image diffusion models with deep language understanding. In 36th international conference on neural information processing systems (pp. 36479– 36494), 2022.

Saharia, C., Ho, J., Chan, W., Salimans, T., Fleet, D. J., & Norouzi, M.: Image super-resolution via iterative refinement. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(4), 4713–4726. https://doi.org/10.1109/tpami.2022 .3204461, 2023.

Shi, J., Zhao, Y., & Yu, F.: Synergistic remote sensing and in situ observations for rapid ocean temperature profile forecasting on edge devices. Applied Sciences, 15 (16), 9204. https://doi.org/10.3390/app15169204, 2025.

Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., & Ganguli, S.: Deep unsupervised learning using nonequilibrium thermodynamics. In 32nd international conference on machine learning (pp. 2256–2265), 2015.

Song, J., Meng, C., & Ermon, S.: Denoising diffusion implicit models. In 8th international conference on learning representations, 2020.

Su, H., Zhang, T., Lin, M., Lu, W., & Yan, X.-H.: Predicting subsurface thermohaline structure from remote sensing data based on long short-term memory neural networks. Remote Sensing of Environment, 260, 112465. https://doi.org/10.1016/j.rse.2021 .112465, 2021.

Vahdat, A., & Kautz, J.: Nvae: a deep hierarchical variational autoencoder. In Proceedings of the 34th international conference on neural information processing systems, 2020.

Wang, J., Chen, Y., Wu, Y., Shi, J., & Gee, J.: Enhanced generative adversarial network for 3d brain mri super-resolution. In 2020 ieee winter conference on applications of computer vision (pp. 3616–3625). https://doi.org/10.1109/WACV45572.2020.9093603, 2020.

Wang, S., Li, X., Zhu, X., Li, J., & Guo, S.: Spatial downscaling of sea surface temperature using diffusion model. Remote Sensing, 16 (20), 3843. https:// doi.org/10.3390/rs16203843, 2024.

Wu, X., Zhang, M., Wang, Q., Wang, X., Chen, J., & Qin, Y.: A deep learning method for inversing 3d temperature fields using sea surface data in offshore China and the northwest Pacific Ocean. Journal of Marine Science and Engineering, 12 (12), 2337. https://doi.org/10.3390/jmse12122337, 2024.

Xia, B., Zhang, Y., Wang, S., Wang, Y., Wu, X., Tian, Y., Van Gool, L.: Diffir: Efficient diffusion model for image restoration. In 2023 ieee/cvf international conference on computer vision. https://doi.org/10.1109/ICCV51070.2023 .01204, 2023.

https://doi.org/10.5194/egusphere-2026-5227
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

Xing, Q., Yu, H., & Wang, H.: Global mapping and evolution of persistent fronts in large marine ecosystems over the past 40 years. Nature Communications, 15, 4090. https://doi.org/10.1038/s41467-024-48566-w, 2024.

Zhang, Y., Liao, Q., Ding, L., & Zhang, J.: Bridging 2d and 3d segmentation networks for computation-efficient volumetric medical image segmentation: An empirical study of 2.5d solutions. Computerized Medical Imaging and Graphics, 99, 102088. https://doi.org/10.1016/j.compmedimag.2022.102088, 2022.

Zhao, Y., He, Z., & Jiang, Y.: The inversion of three-dimensional ocean temperature and salinity fields for the assimilation of satellite observations. Journal of Marine Science and Engineering, 12(4), 534. https://doi.org/10.3390/jmse12040534, 2024.

Zheng, M., Luo, J., & Cao, W.: Video super-resolution based on spatial-temporal transformer. In 2021 ieee 7th international conference on cloud computing and intelligent systems (pp. 403–407). https://doi.org/10.1109/CCIS53392.2021 .9754604, 2021.

Zhu, R., Yu, J., Zhang, X., Yang, H., & Ma, X.: Air–sea interaction during ocean frontal passage: A case study from the northern South China Sea. Remote Sensing, 17(17), 3024. https://doi.org/10.3390/rs17173024, 2025.