

Note: The original referee comments are in black, and the authors' responses are in blue.

Response to Reviewer #1:

This paper presents a 3D reconstruction method of ocean temperature and salinity based on Attention Enhanced 3D-U-Net++ and Transfer Learning for the Northwest Pacific region, using real-time Sea Surface Temperature (SST) and Sea Surface Height (SSH) data to generate a daily high-resolution ($1/4^\circ$, 5-2,000 m depth) temperature-salinity field. The methods, despite incorporating designs such as attention and transfer learning, are those that have been widely used by previous authors and lack substantial innovation. Overall, this study requires significant refinement in terms of methodological detail, generalizability, and interpretability. I recommend a Major Revision.

Response: We sincerely thank the Editor and the Reviewers for their critical assessment and comprehensive feedback on our manuscript. We found the comments to be extremely helpful and have taken them very seriously. The rigorous review process has guided us to refine our methodology and improve the presentation of the results.

We have carefully addressed all the concerns raised and have revised the manuscript accordingly. The parts that have been revised have been marked in yellow in the revised manuscript. Our point-by-point responses are listed below.

Comment 1:

In line 60 of the article, it is mentioned that the data assimilation has the problem of “there remain significant challenges in accurately reproducing the vertical structures of mesoscale eddies”, but in this study, the reanalysis products based on the numerical model and data assimilation are used as the labels for training in fine-tuning stage, so is it possible that the problem of “inaccuracy of the vertical structure of mesoscale eddies” also exists in the present dataset? Please add extensive experimental analyses

to explain how this study used “inaccurate” reanalysis products as labels to train the model to obtain “accurate” 3D thermohaline fields?

Response:

We thank the reviewer for this critical and insightful comment. The reviewer raises a logical concern: how a model trained on reanalysis data (which may contain vertical structure biases) can produce results that are potentially more accurate than the training labels themselves.

We have provided a detailed response to this point, backed by extensive comparative analyses, in **Sections 3.3.2 and 3.4**. The results indicate that our proposed method yields subsurface thermohaline data that is highly consistent with GLORYS, while significantly exhibiting better agreement with WOD in situ observations.

Actually, this good performance is attributed to the transfer learning paradigm (from IPRC Argo to GLORYS2V4), which enables a transition from learning background signals to capturing perturbations. In the pre-training stage, the network captures real observational data and large-scale background signals from IPRC Argo. Using this as a baseline, the network effectively focuses on fitting the small-scale signals presented in GLORYS2V4. The architecture supports this process: 3D convolutions effectively extract volumetric features across spatial and depth dimensions, while dense skip connections enhance the modeling of fine-scale signals. Additionally, the inclusion of long time-series sea surface data provides robust physical constraints. These combined factors ensure the reconstructed data maintains high consistency with GLORYS while outperforming it in alignment with WOD field observations.

Evidence from Comparison with WOD Profiles:

Section 3.3.2 (Model vs. WOD / GLORYS vs. WOD): In this subsection, we compared the errors of the reconstructed data and GLORYS2V4 against the WOD observations using the validation set. The results indicate that the reconstructed data are generally

closer to the WOD in-situ observations in both vertical and spatial dimensions (Figures 12 and 13 in the revised manuscript). Furthermore, density scatter plots at representative depths demonstrate that the reconstructed data exhibit higher R^2 values and lower RMSEs compared to GLORYS2V4 when validated against WOD observations (Figures S1 and S2 in the Supplementary Materials).

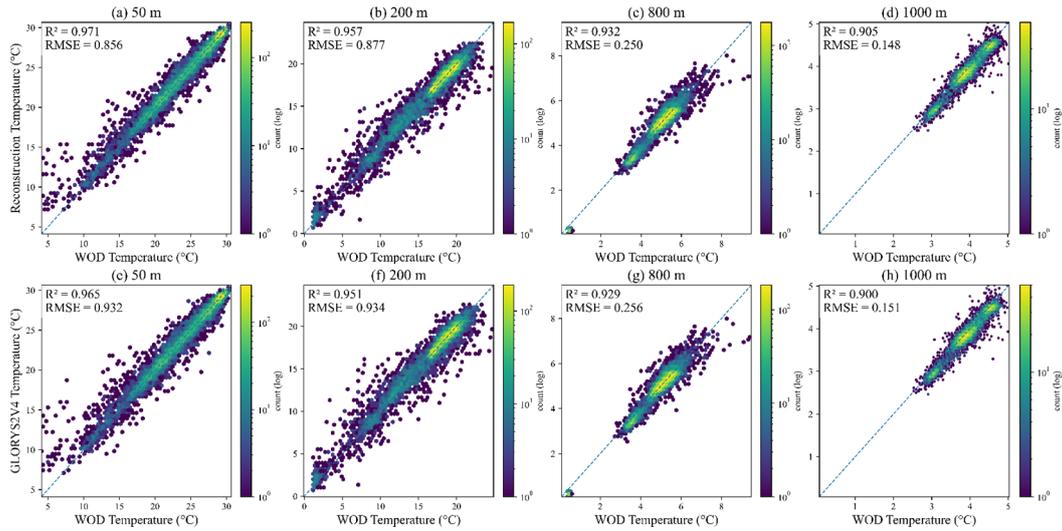


Figure S1. Density scatter plots of reconstructed temperature versus WOD observations at different depths in 2023: (a) 50 m, (b) 200 m, (c) 800 m, and (d) 1000 m; and density scatter plots of GLORYS2V4 temperature versus WOD observations at different depths in 2023: (e) 50 m, (f) 200 m, (g) 800 m, and (h) 1000 m.

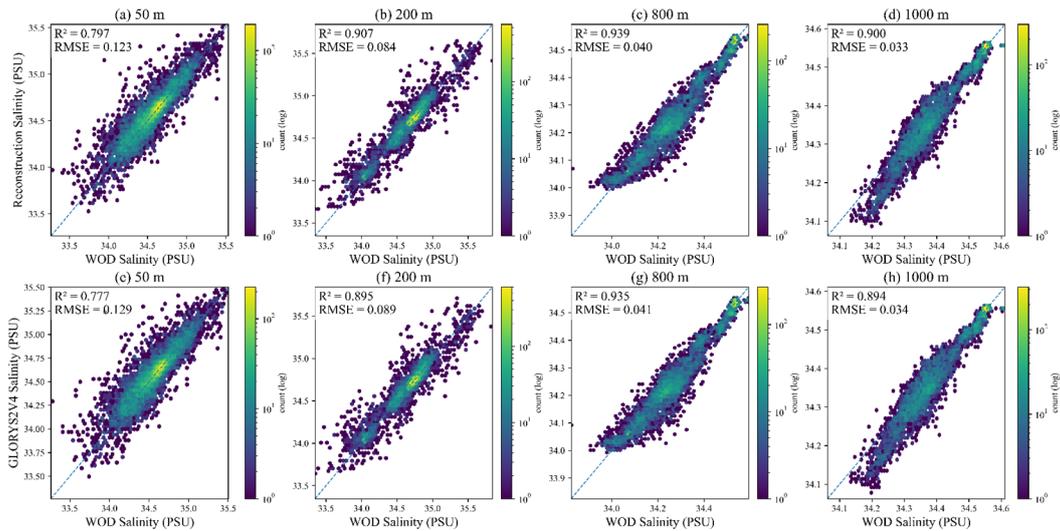


Figure S2. Density scatter plots of reconstructed salinity versus WOD observations at different depths in 2023: (a) 50 m, (b) 200 m, (c) 800 m, and (d) 1000 m; and density scatter plots of GLORYS2V4 salinity versus WOD observations at different depths in 2023: (e) 50 m, (f) 200 m, (g) 800 m, and (h) 1000 m.

The long-term analysis (1993–2023) further supports this (Section 3.4 in the revised manuscript). The accompanying statistical analysis (Figs. 14-16) and the density scatter plots (Figs. 17-18) confirm that the reconstructed data have higher correlation and lower errors relative to the WOD observations than the GLORYS2V4 dataset.

In summary, the “accuracy” of our 3D thermohaline fields is verified not by the training labels, but by the superior agreement with independent WOD observations. The model effectively corrects the deviations presented in the reanalysis data, providing a more reliable subsurface dataset. This highlights the value of the proposed transfer learning approach.

Comment 2:

Table 1 appears to have a non-English “、”

Response:

We appreciate the reviewer pointing out this formatting error. In the revised manuscript,

we have reorganized Table 1.

Table 1. Data processing and normalization methods.

Training Stage	Training Set	Validation Set	Data	Data Processing Method	Normalization Method
Pre-training	2005-2019	2020	AVISO SSH	Downsampling to 0.25° Monthly averaging	Standardization followed by Min-Max normalization
			OISST	Monthly averaging	normalization
			RPRC Argo	Linear interpolation to 0.25°	Min-Max normalization only
Fine-tuning	1993-2022	2023	AVISO SSH	Downsampling to 0.25°	Standardization followed by Min-Max normalization
			OISST	None	Min-Max normalization
			GLORYS2V	None	normalization

4

We have proofread the rest of the manuscript to ensure that all punctuation marks are in the correct English format.

Comment 3:

The combination of UNet and CBAM does not have novelty; many studies have been carried out by previous researchers [1], [2], [3], and this paper does not have a substantial improvement and is not innovative enough.

Response:

We sincerely thank the reviewer for pointing out these important references ([1] Xie et al., 2023; [2] Xie et al., 2022; [3] Qi et al., 2023), which we have carefully studied. It is true that U-Net-based architectures have been applied in oceanography. However, our proposed attention-enhanced 3D-U-Net++ differs significantly from the previous works in terms of model architecture, input logic, and training strategy. The specific contributions and distinctions are summarized below:

1) Advancement from 2D/Standard U-Net to 3D-U-Net++:

Dimensionality and Temporal Dynamics: The previous studies ([1], [2], [3]) primarily utilize 2D Convolutional Neural Networks (CNNs) or Attention U-Net architectures, treating the reconstruction task largely as a static spatial mapping problem. In contrast, our model adopts the 3D-U-Net++ architecture, where all operations—including the CBAM attention mechanism incorporated in this paper—are 3-dimensional.

2) Convolutional Kernel Differences:

There is a fundamental difference between 2D and 3D convolutions. A 2D convolution kernel (typically size $n \times n$) is suitable for processing two-dimensional static images. Conversely, a 3D convolution kernel (size $n \times n \times n$) possesses a natural advantage in extracting features from three-dimensional data blocks. This “3D” characteristic allows our model to perform 3D convolutions on 26 days of continuous input data, explicitly capturing temporal evolution and dynamic dependencies alongside spatial features. This capability is crucial for reflecting the time-lagged response of the subsurface to surface forcing.

We have added the relevant description to the 3rd paragraph of Section 2.2.1 in the revised manuscript:

‘The 3D-U-Net++ network employed in this study is derived from the conventional U-Net++ architecture by replacing all 2D operations—such as convolution, pooling, and up-sampling—with their 3D counterparts. Notably, 3D convolution

is capable of simultaneously processing data across four dimensions: depth, channel, height, and width (Tran et al., 2015). This capability renders the network highly suitable for physical oceanographic datasets, which typically encompass spatial extent, depth, and temporal information.’

Similarly, relevant descriptions have been added to the 5th paragraph of Section 2.2.1 in the revised manuscript:

‘As illustrated in Fig. 1, the neural network is designed to directly accept a 4D tensor with a shape of $(C \times D \times H \times W)$ as input. In this study, C is set to 2, representing the two input channels: SSH and SST. D is set to 26, denoting the continuous time series of sea surface data from the past 26 days (details regarding D are provided in Section 2.2.3). H and W are both set to 160, corresponding to the spatial dimensions of the study area. The output dimensions of the network are (2, 26, 160, 160). Specifically, the first dimension 2 represents the two target variables: seawater temperature and salinity; 26 corresponds to 26 depth levels spanning from 0 to 2000 m; and 160 represents the spatial dimensions of the study area. Within the network architecture, down-sampling is performed using 3D max pooling, while up-sampling is achieved via 3D transposed convolution.’

3) Nested Skip Pathways:

Furthermore, the “++” structure (nested dense skip pathways) effectively reduces the semantic gap between the encoder and the decoder. Unlike the standard U-Net used in previous researches, our densely connected design allows for more efficient aggregation of multi-scale features. This is valuable for preserving fine-scale ocean structures, such as fronts and eddies, which are often smoothed out by standard methods.

We have added the key differences between the standard U-Net and U-Net++ in in **the 3rd paragraph of Section 2.2.1:**

‘The dense skip connections play a pivotal role in bridging the semantic gap between the encoder and decoder feature maps. Unlike the simple long skip

connections in the standard U-Net, these nested, dense pathways pass feature maps through a series of convolutional blocks before fusing them with the decoder. This process gradually aggregates features at varying semantic levels, ensuring that the decoder receives more semantically rich and consistent information. Consequently, this mechanism allows the network to capture complex multi-scale features more effectively—acting essentially as an ensemble of U-Nets—which significantly improves gradient flow during training and enhances segmentation accuracy for objects of diverse sizes.’

Transfer Learning Strategy: Our study innovatively integrates a transfer learning framework. We first pre-train the model on IPRC Argo data and subsequently fine-tune it using GLORYS2V4 data. Our results demonstrate that this transfer learning approach yields temperature and salinity fields that are not only consistent with reanalysis data but also exhibit superior alignment with WOD observational data compared to direct training methods.

Comment 4:

The two-stage transfer learning is equally uninspiring. Combined with Fig. 4, the reconstruction accuracy is improved by less than 10% after transfer learning, and the result is not listed in the table with detailed values; is it intentionally avoided? Meanwhile, as shown in Figure 9, the reconstruction results are almost no different from GLORYS, especially the salinity reconstruction results, and the improvement of reconstruction accuracy is extremely limited. Based on this, is it necessary to carry out the process of such a complex reconstruction? Is it possible to achieve better results with more detailed model tuning? Or is it possible to train to a higher accuracy by replacing the training labels with reanalysis products that have a higher accuracy than GLORYS2V4?

Response:

We sincerely thank the reviewer for these critical questions regarding the transfer

learning strategy and the validation of our model results. We apologize if the presentation of the data in Fig. 4 gave the impression that detailed values were being withheld. This was certainly not our intention. To fully address your concerns, we have annotated this figure with specific numerical values in the revised manuscript (please see Fig. 5).

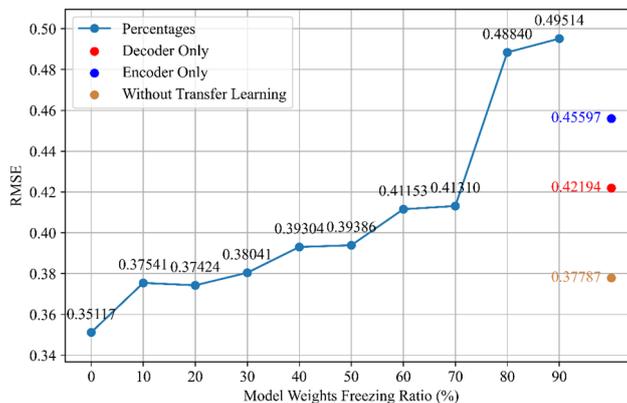


Figure 5. Average RMSE between model outputs and WOD T-S profiles data during 2023 under different transfer learning strategies

In a standard supervised learning setting (without transfer learning), the training labels (derived from GLORYS or other high-precision products) serve as the “ground truth” or the performance upper bound. Regardless of the extent of hyperparameter tuning, the model’s error relative to WOD will asymptotically approach the intrinsic error of GLORYS itself. Fundamentally, the model cannot “surpass” its teacher. The results presented in Fig. 9 are visually similar to GLORYS, a spatial consistency that we consider both crucial and advantageous. This demonstrates that the model has successfully learned the complex dynamic physical processes and high-resolution spatial patterns captured by GLORYS.

Furthermore, the results in Fig. 5 indicate that without transfer learning, regardless of parameter adjustments, the deviation between the network’s output and WOD observations ultimately aligns with the inherent deviation between GLORYS and WOD. It is only by employing transfer learning that we can achieve reconstruction results that maintain consistency with GLORYS while exhibiting superior alignment with in-situ

WOD observations.

In summary, regarding the absence of transfer learning: the model merely mimics GLORYS, inheriting its errors relative to WOD. With transfer learning: the model preserves the high-resolution structures of GLORYS but refines the numerical values to closer approximations to the in-situ WOD observations. Thus, the transfer learning strategy serves not merely to “fit” the data, but to synergize the strengths of two distinct domains: the spatial completeness of the reanalysis data and the high accuracy of the in-situ observations.

We have added relevant explanations in **the 5th paragraph of Section 3.1** of the revised manuscript:

‘In summary, since the neural network is trained exclusively on GLORYS2V4 data, the error between the reconstructed data and the WOD observational profiles can only asymptotically approach the error between the label data and the WOD profiles. Initializing the network weights using the IPRC Argo dataset allows the model to capture authentic observational information while establishing a background of the ocean environment at a monthly scale. Building upon this foundation, the fine-tuning phase enables the network to learn the complex dynamic mapping rules and smaller time-scale signals inherent in the GLORYS2V4 data. Validation against WOD profile data demonstrates that this transfer learning strategy maintains high physical consistency with GLORYS2V4 while achieving closer agreement with the WOD observations.’

Comment 5:

Why is it straightforward to say that inputting 26 days is optimal without any ablation experiments for other time periods, such as 2, 4, 6, 15, etc., up to 100 days, and is it not necessary to take into account the temporal correlation of the thermohaline high elements? Please analyze the temporal correlation of temperature and salt elements in this sea area with historical data, and also add ablation experiments for multiple days,

and analyze the results of the experiments against the temporal correlation, so as to make a strong case that 26 days is the optimal option.

Response:

We are grateful for the constructive suggestion, following which we have added the ablation experiments and discussions in the revised version. Due to computational overhead, 8 time intervals (1, 4, 8, 10, 20, 26, 30, and 40) are tested. The experimental results are illustrated in Fig. 6 and discussed in Section 3.2 in the revised manuscript.

Please refer to **Section 3.2** of the revised manuscript:

‘To determine the optimal length of the input time series for the network, an ablation study was conducted. Considering computational costs, time series lengths of 1, 4, 8, 10, 20, 26, 30, and 40 days were selected. The experimental results are illustrated in Fig. 6, which demonstrate a clear negative correlation between the input sequence length and the reconstruction error. When the time step increases from 1 to 20, the RMSE for both variables decreases significantly—Temperature RMSE drops from 0.63563°C to 0.61222°C, and Salinity RMSE from 0.09874 PSU to 0.09475 PSU. This indicates that the incorporation of historical surface data helps to mitigate the ambiguity associated with super-resolution tasks. Furthermore, continuous time-series of sea surface data provide robust physical constraints for the reconstruction of underwater 3D T-S fields, thereby enhancing the accuracy of the reconstructed data.

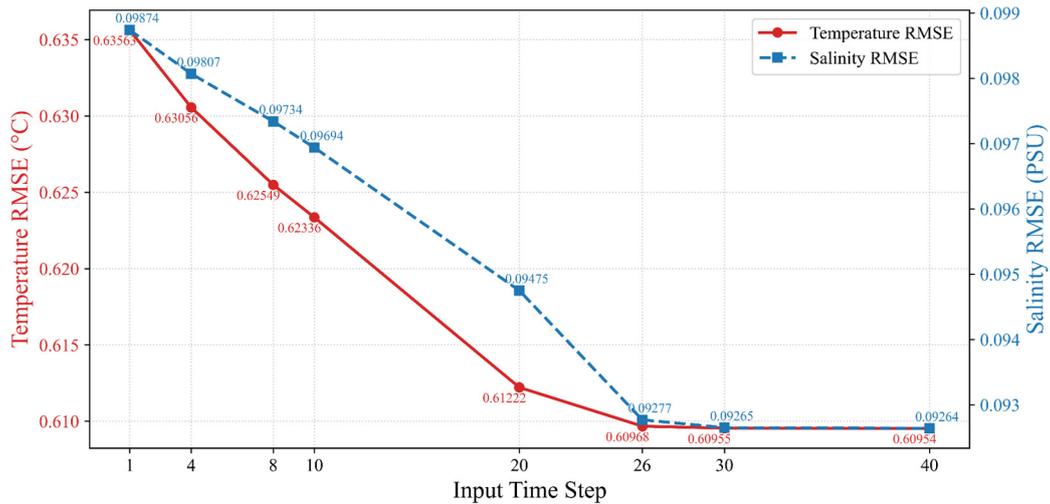


Figure 6. RMSE between the reconstructed data and WOD profiles under different input time series lengths in 2023.

Notably, the performance gain begins to saturate beyond an input length of 26 days. Between time steps 30 and 40, the RMSE curves for both temperature and salinity plateau (stabilizing around 0.6095°C and 0.0926 PSU, respectively). To balance computational cost and accuracy, the time series length of the input data was set to 26 in this study.'

Comment 6:

What is the quality control method for the profiles described in line 330? The number of profiles selected, 7833, is much less than the number of original profiles. Was it an intentional effort to select profiles that favored this study? Also, in terms of spatial distribution, the profiles are not uniform, so large blank areas of the profiles are not assessable for reconstruction accuracy, so the dataset is not entirely credible. How can we verify the reconstruction accuracy of the model in regions with no or sparse profiles?

Response:

We thank the reviewer for this critical question regarding the data quality control (QC) process and the spatial distribution of validation data. We would like to address the two concerns as follows:

1) Clarification on Data Selection and Quality Control:

We respectfully wish to clarify that there was no intentional effort to select profiles that favored our study. The WOD dataset contains many raw profiles with sensor errors, missing values, or single-variable records (e.g., XBTs measuring only temperature). Validating against erroneous data would lead to incorrect assessment of our model. The reduction in the number of profiles is strictly due to a rigorous, automated quality control pipeline designed to ensure that we are validating against reliable “ground truth.” Specifically, our automated QC procedure, which led to the final selection, includes the following five steps:

Step 1. Variable Completeness Check: The NetCDF file must contain both Temperature and Salinity variables. Profiles lacking one of these (indicating potential sensor damage or single-variable probes) are excluded.

Step 2. Profile-level Flag Check: Both `Temperature_WODprofileflag` and `Salinity_WODprofileflag` must be 0 (accepted WOD standard for good data).

Step 3. Depth-wise Filtering: We remove specific depth points where `Temperature_WODflag` or `Salinity_WODflag` are non-zero.

Step 4. Gross Error Check: If a profile contains excessive anomalies (values equal to 0, temperature $< -0^{\circ}\text{C}$, or salinity > 35 psu [specific to low-salinity contexts if applicable, or general outliers]), the entire profile is discarded.

Step 5. Regional Statistical consistency (3-Sigma Rule): For each profile, we search for neighboring profiles within a specific spatiotemporal window. Current data is checked against the mean and standard deviation of its neighbors. Data falling outside three standard deviations (3σ) are identified as outliers and removed.

In 2023, a total of 8,963 PFL profiles were originally available within the study area. After applying Steps 1 through 5 of the quality control procedure, 7,833 PFL profiles were ultimately retained. The strict application of Step 1 (requiring paired T-S) and Step 5 (statistical consistency) significantly reduced the raw count but ensured high

reliability.

Such descriptions as mentioned above have been added in supplementary material.

2) Verification in Sparse Regions:

We acknowledge the reviewer's valid point regarding the non-uniform spatial distribution. It is indeed an inherent limitation of in situ ocean observation networks (like WOD and Argo) that large gaps exist, making direct point-to-point accuracy verification impossible in those specific blank areas. Therefore, more in situ observations are needed in the future. In spite of this fact, the selected 7,833 profiles cover various dynamic environments, including the Kuroshio extension, eddies, and open ocean areas. Thus, the results are representative to a certain extent.

Regarding the limitation of verification in data-sparse regions, we have already explicitly discussed this in **Section 5** (Discussion and Conclusion):

'Despite the promising results, several limitations and uncertainties must be acknowledged. First, the computational cost of training is substantial due to the employment of operations such as 3D convolutions. Second, despite validation against in situ observations, the spatial sparsity of these profiles implies inherent uncertainty regarding the model's reliability in unsampled regions.'

Comment 7:

This paper has repeatedly emphasized that this dataset has the advantage of "real-time", so please add detailed information on the update cycles of various products, the hardware environment for model training and inference, and the time spent. At the same time, please list the update cycles of several mainstream ocean reanalysis products and real-time objective analysis products. By comparison, please illustrate the advantages of this dataset in "real-time".

Response:

We thank the reviewer for this valuable suggestion, following which we have added the

detailed hardware specifications, inference time, and a comparative analysis of data latency in the revised manuscript (please see the **1st paragraph of Section 3**) as follows:

‘Model training and inference were conducted on a supercomputing cluster equipped with an Intel® Xeon® Gold 5218R CPU and an NVIDIA A100-PCIE-40GB GPU. The average inference time required to generate daily 3D T-S fields for the entire region is approximately 5 seconds. In contrast to traditional reanalysis products (e.g., GLORYS and EN4) which typically suffer from latencies ranging from weeks to months due to the assimilation of sparse in-situ data (see Table 2), our method relies exclusively on real-time satellite observations, enabling real-time reconstruction using data from the current day.

Table 2. Comparison of update cycles and approximate time lag (latency) between the proposed method and mainstream products.

Dataset / Product	Type	Approximate Update Cycle / Latency	Dependence on In-situ Data
Proposed Method	Deep Learning Reconstruction	Daily / 0 day lag	No (Only relies on NRT Satellite SST & SSH)
GLORYS12V1 (CMEMS)	Reanalysis	Monthly / ~2-3 years lag (Delayed Mode)	Yes (High dependence)
PSY4V3R1 (CMEMS Analysis)	Operational Analysis	Weekly / ~7 days lag (Best estimate)	Yes
EN4 (Met Office)	Objective Analysis	Monthly / ~1-2 months lag	Yes
RG-Argo (Scripps)	Gridded Argo Product	Monthly / ~1-2 months lag	Yes
SODA3	Reanalysis	Monthly / Several months lag	Yes

,

Comment 8:

This study lacks comparisons with other mainstream marine reanalysis products, such as HYCOM, ECCO2, ORA5, CORA2, SODA3, and so on.

Response:

We sincerely thank the reviewer for this constructive suggestion. We agree that benchmarking against mainstream reanalysis products is essential to verify the robustness and practical value of our proposed method.

Following your recommendation, we have introduced two widely used ocean reanalysis

datasets, HYCOM and ORAS5, for comparative analysis in the revised manuscript (please see **Sections 2.2.2**).

Furthermore, in **Section 3.4** (Figures 13–17), we have included the RMSE and scatter density plots comparing these two datasets with the WOD observational data. The results indicate that, compared with the newly introduced HYCOM and ORAS5 datasets, the reconstructed data exhibits a higher R^2 and a lower RMSE with respect to WOD.

Comment 9:

Some sentences are too long and could be split to improve readability.

Response:

We sincerely thank the reviewer for this constructive suggestion. We have carefully examined the entire manuscript and modified those very long sentences (please see Line 234, 523). We believe these changes have made the manuscript much clearer and more concise.

Comment 10:

As shown in Figures 7 and 8, the reconstruction results are almost identical to the GLORYS reanalysis. However, the model inputs are sea surface temperature and height information, and not even sea surface salinity information. How to restore so many small- and medium-scale details of the approximate reanalysis products with so little information? Please give a more detailed description of the training process.

Response:

We appreciate the reviewer's insightful comment. We acknowledge the concern regarding the model's ability to retrieve detailed subsurface structures—particularly salinity—relying solely on Sea Surface Temperature (SST) and Sea Surface Height (SSH) inputs. This capability arises from the strong physical coupling between surface

dynamics and the ocean's interior state and also between temperature and salinity, as well as our specific deep learning architecture. This is achievable for three primary reasons:

1) Transfer Learning and Architecture:

The model initially undergoes pre-training on observational gridded data (IPRC-Argo) to capture fundamental large-scale T-S relationships. Subsequently, it is fine-tuned using GLORYS reanalysis data. Given that GLORYS is a physics-based product, the neural network effectively learns the complex, non-linear “mapping rules” (covariances) linking surface and subsurface variables specific to the physics within GLORYS. Furthermore, our attention-augmented U-Net++ architecture, utilizing skip connections, is explicitly designed to preserve small- to meso-scale features during this mapping process.

The relevant analysis and description are presented in **the 4th and 5th paragraph of Section 3.1:**

‘The experimental results demonstrate that the optimal transfer learning strategy is global fine-tuning (0% frozen weights), rather than acting as a rigid feature extractor (which would favor partial freezing). This indicates that the fundamental mechanism of transfer learning in this study is providing a physically constrained parameter initialization rather than directly reusable features. Oceanographically, the transfer process mimics a “background-to-perturbation” learning paradigm. During the pre-training phase with monthly, coarse-resolution IPRC Argo data, the network learns the large-scale climatological background, encompassing basic stratification, seasonal cycles, and global vertical covariance structures. By establishing this robust physical framework, the pre-trained weights place the model in a physically plausible region within the high-dimensional optimization landscape. During the fine-tuning phase with daily, high-resolution GLORYS data, global fine-tuning allows the network to bypass the struggle of learning fundamental ocean physics from scratch. Instead, it fully

allocates its learning capacity to resolving high-frequency, synoptic-scale dynamics—such as how mesoscale eddies and fronts perturb the pre-established climatological background.

In summary, since the neural network is trained exclusively on GLORYS2V4 data, the error between the reconstructed data and the WOD observational profiles can only asymptotically approach the error between the label data and the WOD profiles. Initializing the network weights using the IPRC Argo dataset allows the model to capture authentic observational information while establishing a background of the ocean environment at a monthly scale. Building upon this foundation, the fine-tuning phase enables the network to learn the complex dynamic mapping rules and smaller time-scale signals inherent in the GLORYS2V4 data. Validation against WOD profile data demonstrates that this transfer learning strategy maintains high physical consistency with GLORYS2V4 while achieving closer agreement with the WOD observations.’

2) Long-term Time Series Input: As detailed in the manuscript, we avoid relying on single snapshots. Instead, we employ a 26-day time series of SST and SSH as input. This temporal dimension enables the 3D-U-Net++ to capture the evolution and propagation of surface features. The dynamic history of the surface state provides additional constraints, allowing for a more accurate resolution of subsurface states than would be possible relying solely on spatial correlations.

The relevant analysis and description are presented in **the 2nd paragraph of Section 3.2:**

‘The experimental results are illustrated in Fig. 6, which demonstrate a clear negative correlation between the input sequence length and the reconstruction error. When the time step increases from 1 to 20, the RMSE for both variables decreases significantly—Temperature RMSE drops from 0.63563°C to 0.61222°C, and Salinity RMSE from 0.09874 PSU to 0.09475 PSU. This indicates that the incorporation of historical surface data helps to mitigate the ambiguity associated

with super-resolution tasks. Furthermore, continuous time-series of sea surface data provide robust physical constraints for the reconstruction of underwater 3D T-S fields, thereby enhancing the accuracy of the reconstructed data.'

3) Physical Correlation (SSH as a Proxy): Although Sea Surface Salinity (SSS) is not utilized, SSH serves as a robust integral constraint. Variations in SSH are primarily driven by steric height changes (thermosteric expansion and halosteric contraction) integrated over the entire water column. Consequently, SSH contains comprehensive information regarding the subsurface density structure, enabling the network to infer vertical temperature and salinity distributions.

References

- [1] H. Xie, Q. Xu, Y. Cheng, X. Yin, and K. Fan, "Reconstructing three-dimensional salinity field of the South China Sea from satellite observations," *Front. Mar. Sci.*, vol. 10, p. 1168486, /05/08 2023, doi: 10.3389/fmars.2023.1168486.
- [2] H. Xie, Q. Xu, Y. Cheng, X. Yin, and Y. Jia, "Reconstruction of Subsurface Temperature Field in the South China Sea From Satellite Observations Based on an Attention U-Net Model," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–19, 2022, doi: 10.1109/TGRS.2022.3200545.
- [3] J. Qi, B. Xie, D. Li, J. Chi, B. Yin, and G. Sun, "Estimating thermohaline structures in the tropical Indian Ocean from surface parameters using an improved CNN model," *Front. Mar. Sci.*, vol. 10, Apr. 2023, doi: 10.3389/fmars.2023.1181182.