Reviewer 2

This study proposes a CNN-based U-Net architecture named TS-cast for reconstructing subsurface temperature and salinity data. By integrating remote sensing products with in-situ observations to fill data gaps, the research topic holds certain value. However, the current manuscript requires substantial improvements in details such as data preprocessing and vertical interpolation. Specific comments are as follows:

Major Comments:

1) Many studies have established relationships between surface remote sensing data products and subsurface temperature-salinity profiles, enabling subsurface temperature-salinity reconstruction. While the abstract states that ADT represents integrated information for subsurface water properties, it remains challenging to correlate SST, SSS, and ADT with subsurface water profiles due to their complex spatial and temporal variations. The research challenges are not sufficiently clarified. Authors are advised to further clarify the research questions, innovative aspects of the study, and the problems addressed.

Response: We have revised the abstract to explicitly distinguish our approach from existing methods. We clarified that, unlike conventional models, TS-Cast introduces a physics-informed architecture and uncertainty quantification.

Revised text (Lines 4–11): While ADT reflects total ocean dynamics, its steric component represents the integrated information for subsurface water properties. However, relating surface variables to subsurface profiles remains challenging because surface signatures are often non-linearly related to interior structures, and satellite data contain inherent non-steric signals. To address these limitations, we introduce TS-Cast, a novel uncertainty-aware deep neural network. Unlike direct regression models, TS-Cast is designed to adjust monthly climatological profiles as a physical prior and learns to dynamically adjust them. By using a 31-day sequence of satellite inputs (SST, SSS, and ADT) and quantifying prediction uncertainty, the model effectively captures the temporal variation of mesoscale dynamics. It was trained on approximately 150,000 Argo and ship-based thermohaline profiles in the northwestern Pacific.

2) The author's X, Y, Z coordinate definitions are incorrect: Since cosine and sine transformations of longitude and latitude require consideration of periodicity, their current

transformations are entirely erroneous and lack any geographic meaning. Consequently, the model inputs encoded geographic information incorrectly.

The correct transformation should be: $X = \sin\left(lat \times \frac{\pi}{180}\right)$, $Y = \sin\left(lon \times \frac{\pi}{180}\right) \times \cos\left(lat \times \frac{\pi}{180}\right)$, and $Z = -\cos\,(lon \times \frac{\pi}{180}) \times \cos\,(lat \times \frac{\pi}{180})$

Response: We have revised the equation in Section 2.1 to explicitly include the conversion factor (180/π) as suggested by the reviewer.

3) High-resolution data can be interpolated to lower resolutions, but is it correct to interpolate 1/4° low-resolution climatological data to 1/8° high resolution? The text states: "All in-situ and climatological profiles were linearly interpolated onto 128 evenly spaced vertical layers between 10 and 700 dbar." This is a significant issue. Due to vertically discrete sampling, direct simple linear interpolation introduces substantial errors. On one hand, the interpolated results fail to match actual vertical variations—an inherent error of the method itself. On the other hand, if some profiles have only sparse observations (perhaps just one or two sampling points) between 10 and 700 dbar, interpolating these directly into 128 layers yields completely erroneous results.

Response: Regarding the horizontal regridding of the WOA23 climatology (1/4° to 1/8°), this was a necessary step to align the dimensions of the background climatological prior with the high-resolution satellite input data (1/8°) within the neural network architecture. We do not claim this process increases the information content of the climatology itself; rather, it allows the TS-Cast model to use the coarse climatology as a baseline and refine the spatial details using the high-resolution information provided by the satellite observations (SST, ADT).

We agree that applying linear interpolation to sparse data is problematic, and apologize for the textual error. The sentence "Profiles containing data gaps were retained by applying masks to use only valid-level observations for model training and error estimation" (Lines 72–73) should be corrected as "… for model test and error estimation". However, we would like to clarify our data selection process, which was designed to prevent this issue.

We used the CORA 5.2-Easy CORA dataset provided by CMEMS (Lines 65–67). This product is a processed version of the global in situ dataset, with rigorous quality controls already applied. We also applied a quality-control criterion to the training and test datasets. To avoid the linear interpolation artifacts, we selected only profiles containing more than 50 valid observation points within the sampled depth range (typically 10-1500 dbar). Given this sampling density, linear interpolation was used as a vertical resampling step. Profiles with gaps were excluded from training. We have revised Section 2.1 to explicitly state this selection criterion.

Revised Text (Section 2.1, Lines 65–67): For the training dataset, thermohaline profiles from CTD and Argo floats were sourced from the Coriolis Ocean dataset for Reanalysis version 5.2 (CORA5) Easy CORA product provided by CMEMS (ID: INSITU_GLO_PHY_TS_DISCRETE_MY_013_001). This is a delayed-mode dataset with preprocessed and quality-controlled profiles from CTD, Argo floats, and other platforms.

Revised Text (Section 2.1, Lines 74–77): "Only profiles containing more than 50 valid observation points within the sampled depth range (typically 10-1500 dbar) were used. For these profiles, linear interpolation was applied as a vertical resampling step to align the data onto the 128 standard vertical levels. Profiles with gaps were excluded from training. For the test dataset, profiles with gaps were retained by applying masks to use only valid-level observations for error estimation."

4) Due to the significant vertical interpolation bias in this study, their training and test datasets underwent identical interpolation operations. Since both datasets share the same bias issue, their results are relatively close. However, by using this biased test dataset as a benchmark to compare against GLORYS, HYCOM, and ARMOR3D datasets, they inevitably present the erroneous conclusion that TS-Cast is more accurate. If this comparison method is applied, then all results presented in Fig. 5 and subsequent figures are invalid.

Response: We respectfully disagree with the claim that our comparative analysis is invalid. This concern appears to stem from a misunderstanding of how the datasets were used in different figures.

1. Figure 4 (Internal Evaluation): The "identically processed" test dataset (interpolated profiles) was used solely for Figure 4. Crucially, Figure 4 does not present any comparison with other models (GLORYS, HYCOM, or ARMOR3D). It only illustrates the internal error distribution of our model. Therefore, the claim that we used a biased dataset to benchmark against other models is factually incorrect.

2. Figure 5 and Subsequent Figures (Model Comparison): The comparisons with other models (Figure 5 onwards) use independent in-situ mooring observations (KEO, EC1, PIES), which are not vertically interpolated. These are time-series measurements at specific

depths or vertically integrated values. If our model learned "linear interpolation artifacts" as suggested, it would fail to capture the complex variations observed in these in-situ datasets. The fact that TS-Cast outperforms or matches other models on these independent observations confirms that it has learned realistic states rather than artifacts.

Minor comments:

Page 2, Lines 25–26: This creates a critical “observational void” between the broad coverage of satellites and in-situ measurements, hindering a complete understanding of the ocean's 3D dynamics. This sentence is incorrect and should be rewritten. The disparity between abundant surface data and sparse internal data reflects differences in observational techniques and data acquisition. Describing this as an “observational void” is inappropriate. Furthermore, the obstacle to understanding the ocean's three-dimensional dynamic processes stems from the sparsity of three-dimensional environmental data, not from the aforementioned differences.

Response (Line 28): We agree with your comment and have rewritten the sentence, “This sparsity of subsurface observations limits our ability to fully resolve the ocean's 3-D dynamics.”

Pages 2–3, Lines 54–60: The four summaries of the objectives of this paper contain some redundancies and should be further refined and enhanced.

Response (Lines 58–59): We have revised the objectives to eliminate redundancies, particularly by integrating the validation steps. The second and third objectives are integrated and rewritten as “Assess the model's capability to reproduce continuous ocean dynamics by validating against multi-year, high-frequency timeseries mooring data, using spectral coherence analysis to quantify performance across different frequency bands.”

Page 7, Lines 112–114: Here, linear interpolation is used to interpolate the profiles into 128 uniformly spaced vertical layers. This processing yields erroneous results, with significant interpolation errors causing the outcomes to deviate substantially from the true temperature-salinity distribution.

Response: By interpolating the 10–700 dbar range into 128 layers, we achieve a vertical spacing of approximately 5.4 dbar. This resolution is sufficiently fine to capture sharp vertical gradients, including the main thermocline and halocline, with negligible interpolation error. This preprocessing step was necessary to generate fixed-dimension input tensors for the U-Net architecture.