The authors present a deep neural network architecture able to reconstruct Temperature and Salinity profiles from satellite and climatology inputs. This architecture assimilates observations in delayed time. The applications

We thank the reviewer for the thoughtful and constructive comments, which have substantially improved the manuscript. We have carefully addressed each point, and our responses are provided below in blue, with the corresponding changes in the revised manuscript. Line numbers refer to the revised version.

Major comment

Validation method

The experiments reported show some methodological problems making the conclusions of the authors uncertain. The main problem comes from the fact that the authors compare their model on time period that were represented in their training data. Even if these data specifically where not used in training, the examples that they represent might be very corelated to training data. From Figure 1 we understand that EC1 and KEO sites are located in areas that are sampled by Argos. This means that the results that the author show in sections 3.2 can not be truly seen as "test" dataset but are highly correlated to training dataset. In section 3.1 the authors present results on Argos that are excluded from their dataset, but no comparison to the other models, which could be useful as it's the only truly test dataset.

Also, the authors mention that the other model assimilates the target data, so all the scores that we see in this analysis do not represent the true performance of the 3D methods, (the models because they assimilate evaluation profiles, and TS-cast because it is trained on observations that are close to the validation one). A solution could be to take the operational datasets from the compared products (GLORYS for example) and TS-cast estimation and compare the models in an operational manner. From what this paper describes, it's hard to know if the method brings a significant improvement compared to physical models.

As requested, we have added comparisons against GLORYS, ARMOR3D, and WOA on the test set based on the easyCORA5.2 dataset (Section 3.1, Figure 5). HYCOM is excluded from the basin-scale comparison because its reanalysis output does not extend to the 2021–2023 test period. We note that this comparison inherently favors the reanalysis products, as they directly assimilate these profiles, especially for the GLORYS and

ARMOR3D models. Nevertheless, TS-Cast achieves comparable performance with the GLORYS (Lines 207–213). Regarding the concern about data leakage at mooring sites, we respectfully clarify that the presence of nearby Argo profiles in the training set does not constitute data leakage (Lines 280–286. Discussion). TS-Cast learns a general satellite-to-subsurface mapping from scattered, instantaneous profiles, whereas the mooring validation tests a fundamentally different capability, reproducing continuous temporal variability over multiple years, information absent from the training data. We further emphasize that the PIES comparison provides the benchmark where all models are evaluated under equal conditions. PIES data are neither assimilated into any reanalysis product nor used in TS-Cast training, and TS-Cast consistently outperforms all reanalysis products in this test. Finally, while comparing against operational near-real-time reanalysis versions would be informative, the delayed-mode products used here represent the best available performance of these models, making TS-Cast's comparable results all the more meaningful.

Architecture

The work presented in this study mostly present a new architecture, trained specifically for TS profile inference. However, the authors don't show if the innovative parts of their architecture bring an improvement for their application. For instance, the core of the innovation being to compute the profile using climatology data conditioned by satellite variables, we should at least see an ablation study on the climatology data input. Is it possible to have similar performances with a more standard approach that outputs the profiles from satellite data only as already multiple study highlight?

Many features of network lack empirical justification such as: Including satellite objective analysis errors, including ADT error vector compared to climatology…

The description of the network is also incomplete and prevent reproducibility of the results (I summarized some points in the minor comments).

We appreciate the reviewer's suggestion regarding ablation studies. To address this, we conducted an ablation experiment in which the input related to the climatology was replaced with zeros, keeping all other architectural components identical (Figure 5, Section 3.1, Lines 200–206). The results show that removing the climatological prior has minimal impact on basin-scale RMSE. We acknowledge that the climatological prior provides modest gains in final prediction accuracy. However, the training dynamics show that it substantially stabilizes the learning process. We therefore retain the climatological input as a physically meaningful architectural component while being transparent about its limited contribution to prediction accuracy.

Regarding the other input features, each serves a physically distinct and well-motivated role. The satellite error fields, provided as formal mapping errors from the optimal interpolation process, allow the model to assess the spatially varying reliability of the gridded satellite inputs. The ADT–dynamic height anomaly vector informs the model how the current ocean state deviates from the climatological seasonal cycle, which is critical for both accurate profile estimation and depth-dependent uncertainty quantification. While a comprehensive ablation of every input feature is beyond the scope of this study, we believe the climatology ablation, which directly tests the core innovation, combined with the physical justification for the remaining features, adequately addresses this concern.

Minor comment

Figure 3 and 2.3 Thermohaline profile estimating network (TS-Cast)

This diagram in Figure 3 and the text presenting TS-Cast is sometime confusing.

Figure3:

1) What are the operations used to transform the Satellite images into a vector, I guess its 2D conv +activation +pooling?

2) Same question for the ADT climatology errors

3) From the red arrow I understand that the error variance is computed from primary input only is it correct? If it is why not computing using the rest of the network? Also I understand that you predict 2 means and 3 variances? Why don't you compute the density variance from the temperature and salinity variance?

The structure of the text describing the network could be improved.

- You start by redescribing the satellite data (this could be moved to the input data section) but you don't describe the module that you use to do so,
- Then you describe the climatology data with a description of the module.
- Then you get back to sateliite conditioning, again explaining the input data(why?)

- Then you mention the “temporal context” that you did not mention earlier, and how you combine it with satellite vector (still no explanation of how you processed satellite vector)
- You describe Film (this part is clear)
- But no description of the encoding and decoding upsampling/ downsampling?

Overall this description lacks many details and its structure should be improved, describing one module at a time, and to avoid repetitions move inputs description to previous section.

Thank you for your feedback. We have modified the section 2.3 overall (Lines 114–171).
- The satellite images are transformed using 3D convolutions followed by Mish activation and average pooling (Section 2.3.1).
- The ADT climatology anomalies are processed similarly using 2D convolutional blocks.
- The error variance is computed from the primary input alone to give seasonal, depth-dependent uncertainty information. (Section 2.3.4).
- We have added explicit descriptions of the downsampling (average pooling) and upsampling (linear upsampling) processes in the encoder and decoder (Section 2.3.2).

Line 120

Why giving 12 months in input as with 31 days only two months should be required? I think this could be confusing for the neural network and leading to overfitting.

We intended to give the seasonal information to network from the 12-climatological profile prior, and the ADT-DH_{clim} vector gives it how the current ADT is deviated from the monthly climatologies (Section 2.3.1). Furthermore, regarding the overfitting, our ablation study (Figure 5) empirically demonstrates that including this 12-month climatology prior actually stabilizes the validation loss and prevents unpredictable learning dynamics, indicating that it is better for generalization rather than causing overfitting.

Line 122

“*To integrate this annual climatological context, initial 2D convolutional layers are applied, using kernels span the time dimension, to collapse the temporal axis.*”

2D convolution with kernel_size=dimension correspond to 1D convolutions with 12x2 channels

We have clarified in the text that this is mathematically equivalent to a 1D convolution with 12×2 channels, while retaining the 2D formulation for clarity of tensor shape transformation (Lines 136–138).

Line 128 to 130:

"*To provide temporal context, we also compute anomalies by taking the difference between the ADT and the dynamic height (DH) from corresponding monthly climatological data for each month from January to December at the profile location.*"

I don't understand why this part is necessary. The authors mention that its for including temporal context, but I don't see why this operation does it. First its only ADT climatological anomalies, which should be lower than temperature anomalies for instance. Then the errors compared to monthly climatologies might not be sufficient to determine precisely the month of the year from the estimation. For instance, a front or an eddy could create the same error compared to the climatologies of two months which prevent the network to determine the target month. Finally, this ADT climatology error must be adapted by the network to learn temperature and salinity climatologies errors, which is not necessary the case. I suspect that during inference the network uses more the SST values than this vector in order to know the month of the year.

If the authors want to inform the network with the month of the estimation, why not giving it this directly, as a temporal positional encoding for instance?

We clarify that the purpose of this vector is not to inform the network of the target month. Rather, it characterizes how the current ADT deviates from each month's climatological dynamic height, providing a baseline for the steric signal across the full seasonal cycle (Section 3.1, Lines 123–128) . This allows the network to contextualize the current dynamical state relative to the annual cycle, including deviations caused by eddies or fronts. We acknowledge that our ablation experiment, in which this vector was replaced with zeros alongside the climatological prior, shows only a modest contribution to final accuracy. We have clarified the stated purpose of this vector in the revised text.

Section 2.4

Equations of Lt, Ls and Lp. I think that you should put a hat over what is estimated by the network (including sigmas)

Thank you for your observation. We have correct the equation (Section 2.4).

Lines 152--

"*In addition to the prediction accuracy of TS profiles, we enforce a physical constraint based on the equation of state for seawater (Fofonoff and Millard, 1983), adopting the physicsinformed approach (Sammartino et al., 2025). While the network does not directly output density, we compute the density profile (ρˆ) from the predicted TS profiles (T ˆ , Sˆ) and compare it to the ground-truth density (ρ) derived from the label profiles (T,S). This term penalizes the model for generating physically implausible combinations of temperature and salinity where σ 2 ρ is the predicted error variance for density.* "

From what I understand the density loss is supposed to reduce bad combination of Temperature/ salinity. This Is what is done if you put the density MSE term. But what happens when you learn the density error variance? To me the network can learn where the temperature and salinity wont match and increase its variance there to get less penalized. To be sure that this consistency terms does what you want you should either remove the uncertainty loss and use only the MSE, or derive the density variance from temperature and salinity ones.

We respectfully disagree that the density variance should be derived from T and S variances, as this would require ignoring the covariance between T and S errors, a term that cannot be neglected given the strong T-S correlation in the ocean. Furthermore, the learned σ^2_ρ cannot trivially inflate to avoid penalization, because the (1/2)log(σ^2) regularization term in the loss function explicitly penalizes large variance predictions (Lines 175–180), following the framework of Kendall et al. (2018). This formulation serves a dual role: enforcing physical consistency through the density constraint while providing adaptive normalization across the different scales of temperature, salinity, and density.

Section 3.1

Why not including other physical models inside the comparison? Glorys? Maybe also a comparison to the error of the Climatology that you transform using TS cast. It will help understand what TS cast adds to climatology performance. As mentioned in Major comments, it's really important that you validate on ARGO against at least GLORYS and maybe climatology, as the KEO and EC1 scores are hard to interpret due to data leakage.

Thank you for your comment and we have compared the WOA23, GLORYS, and ARMOR3D in Figure 5.

Figure 5, 6, 7

Its hard to know which model is better from these plots, especially TS-cast and ARMOR3D. Can you increase the shape of the lateral RMSE and Correlation plots?

As your suggestion, we rearranged the shape of plots (Figures 6, 7, and 8).

Line 271

At this site, TS-Cast, which uses no data assimilation, achieves a higher mean coherence than the ARMOR3D product.

Training on Argo data in similar areas and temporal extent can be seen as a form of Data assimilation for TS-cast, in any case claiming that the Ts cast uses no data assimilation is incorrect, as it was trained on satellite inputs and Argo data that intersect with you validation data.

We acknowledge that the term 'no data assimilation' was imprecise and have revised the manuscript accordingly (Lines 322–323). However, we note a fundamental distinction: data assimilation directly ingests observations to constrain the model state at a specific time and location, whereas TS-Cast learns a general surface-to-subsurface mapping from scattered Argo snapshots. The model has never seen the continuous temporal variability and must infer subsurface structure purely from the satellite signature at each time step. The revised text now states that "TS-Cast, which does not directly assimilate subsurface observations at the prediction time, achieves a higher mean coherence than the ARMOR3D product."

Although it is uncertain whether EC1 data was assimilated into the ARMOR3D, TS-Cast's performance near 0.5 is significant.

What is uncertain? Are you unsure that ARMOR3D assimilates EC1 data? In that case its performance should be rediscussed.

We have clarified this ambiguity, "While we could not definitively verify whether EC1 data were included in the CORA dataset used by ARMOR3D and GLORYS, the fact that TS-Cast achieves coherence near 0.5 at periods longer than ~30 days is significant." (Lines 323–325). We verified that KEO (WMO 28401) is included in the CORA5.2 dataset, but could not definitively confirm whether EC1 data are included, as EC1 is primarily distributed through SEANOE rather than the OceanSITES GDAC. The revised text reflects this distinction and no longer relies on the uncertain status of EC1 assimilation to make its argument.