Response to Reviewer 2

We would like to express our sincere gratitude for the time and effort you have dedicated to reviewing our manuscript. Your insightful comments and constructive feedback have been invaluable in enhancing the clarity, rigor, and overall quality of our work. Each of your suggestions has provided us with the opportunity to refine our methodologies, clarify our explanations, and ensure that our findings are presented in the most comprehensive and accessible manner possible.

We deeply appreciate your attention to detail and your commitment to advancing research in this field. Your expertise has significantly contributed to the improvement of our study, and we are confident that the revisions made in response to your feedback have strengthened our manuscript. Thank you once again for your thoughtful and thorough reviews.

Best regards,

Olmo Zavala-Romero

This work explores the ability of Convolutional Neural Networks (CNNs) to serve as a surrogate to the Tendral Statistical Interpolation System (TSIS) method of data assimilation between the Hybrid Coordinate Ocean Model (HYCOM) and observations. In addition to evaluating the performance of their models in the Gulf of Mexico, the authors also quantify the difference in skill of various hyperparameters and model architectures in this highly dynamic region. The results of this paper show the technique is sound and has potential to be used operationally.

While the study is well conducted and the results are relevant to the community, the report will benefit from more specificity in the model architectures and data preprocessing. All such information is discernible in the code repository provided but should be expressed in the report. This includes:

- Preprocessing
- The handling of land points (e.g., masked as 0)
- Structure of tensors used as input
- Use of batch normalization
- Point where normalization/denormalization occurs
- Structure of output tensors
- Which parameters were tuned via hyperparameter optimization

Thank you for your thoughtful comments and for recognizing the relevance of our study. We appreciate your suggestion to provide more detailed information on the model architectures and data preprocessing in the manuscript. We agree that including these specifics will enhance the clarity and reproducibility of our work.

In response to your feedback, we have thoroughly addressed each of your points below and have made corresponding revisions to the manuscript.

We have added a new subsection titled "Data Preprocessing" in Section 3.2 of the manuscript, where we detail the entire preprocessing steps applied to our data before training the CNN models.

Summary of the preprocessing steps:

- *We use daily outputs from the assimilative HYCOM model for the years 2009 and 2010. The data includes background state variables, observation fields, and increment fields (the corrections applied by T-SIS). Each input variable is normalized to have zero mean and unit standard deviation (mean 0, standard deviation 1). The mean and standard deviation are computed from the training set (2009 and 2010). This normalization is applied consistently to the test sets (last 10% of 2010 and the years 2002 and 2006).*

- *Missing values in the input data (e.g., due to land areas or lack of observations) are filled with zeros after normalization. The loss function is computed only over ocean grid points. By masking out land points during loss computation, the network focuses on learning the ocean dynamics and does not penalize predictions over land.*

- *We include an input layer mask indicating areas with depths greater than 200 meters. Since T-SIS does not provide SSH increments for shallow areas (<200m), we provide this mask as an input to the CNN to help it learn this restriction. After the CNN makes predictions, we apply the land mask to set the values at land grid points to NaN.*

- *The input tensors to the CNN models are four-dimensional arrays with the following dimensions: [Batch Size, Height, Width, Channels]. The numbers for Height, Width, and Channels vary depending on the experiment. All input tensors are of type float32.*

We have added this description of the input tensor structures in the "Data Preprocessing" section.

Batch Normalization is used in our U-Net architecture to improve training stability and performance. Batch Normalization layers are applied after each convolutional layer and ReLU activation in the U-Net architecture, except for the final output layer. This is reflected in the updated Figure 2, which now includes Batch Normalization layers. We have updated the description of the U-Net architecture in Section 3.3 to include this information.

We did not perform extensive hyperparameter optimization. The hyperparameters used were chosen based on standard practices and preliminary testing: We used the Adam optimizer with a learning rate of 0.001 (1e-3). This learning rate provided stable convergence across our experiments. A batch size of 32 was used for all models. Training was monitored using validation loss. We implemented early stopping if the validation loss did not improve for 20 consecutive epochs. The model with the lowest validation loss was selected for evaluation.

Finally, there should be some comparison to other models/techniques used for similar purposes in geosciences. Much has been written on this topic in the atmospheric sciences. How could this study be extended to newer, more sophisticated model architectures?

Thank you for your comments and for highlighting the importance of situating our study within the broader context of geoscientific modeling techniques. We agree that comparing our approach with other models and techniques, particularly those prevalent in ocean sciences, would enhance the comprehensiveness and relevance of our work.

We have included a discussion on the potential integration of advanced architectures such as Convolutional Block Attention Modules (CBAM) in SmaAt U-Nets, Masked Autoencoder Vision Transformers (ViT), and Denoising Diffusion Inpainting models. Given that CNNs remain the backbone for some of these more advanced models, our findings regarding CNN performance and data assimilation are expected to extend to these.

Specifically, the principles of spatial feature extraction, suggesting that our insights on window size and input configurations would still be applicable.

We have also included in the conclusions the scope of our work and potential future research directions with newer techniques.

Line 80: The "Markov process" mentioned is not described or referenced. Presumably, modelers and computational scientists will be familiar with the meaning, but it wouldn't hurt to briefly describe this meaning.

Thank you for your comment. We agree that providing a brief description of the Markov process will enhance the clarity of our manuscript. We have revised the relevant section to include a concise explanation and an appropriate reference.

Revised Text:

"In this approach, it is assumed that the model forecast follows a Markov process, which means that the future state of the system depends only on its current state and not on any previous states (Davis, 2013). Observations can improve the estimate of the model state in a least squares sense, taking into account the modeled and observed error covariances as follows."

Line 98: Similar to above; Gaussian Markov Random Field not explained or referenced.

Thank you for your valuable feedback. We agree that providing a brief explanation of the Gaussian Markov Random Field (GMRF) will enhance the clarity and comprehensiveness of our manuscript. We have revised the relevant section to include a concise description of GMRF and an appropriate reference.

Revised Text:

"The information matrix is modeled using a Gaussian Markov Random Field (GMRF), which is a probabilistic model consisting of a set of random variables having a multivariate Gaussian distribution, with the Markov property that each variable is conditionally independent of all others given its immediate neighbors (Rue and Held, 2005)."

Sect. 2.2: If CNNs are going to be explained in this detail, it would help to show a figure differentiating traditional CNNs from UNets/encoder-decoder networks since the difference is hard to visualize. UNets are referenced, CNNs are not. There should be consistency in the degree of explanations in this section.

Thank you for your feedback regarding Section 2.2 of our manuscript. We agree that providing a visual comparison between traditional Convolutional Neural Networks (CNNs) and U-Net/encoder-decoder architectures would enhance the reader's understanding of the structural differences and functional advantages of these models.

To address your suggestion, we have incorporated the classical LeCun et al. (1998) reference for traditional CNNs to acknowledge their foundational role in deep learning. Additionally, we have included a new figure that clearly differentiates traditional CNN architectures from U-Net/encoder-decoder networks, highlighting key components such as the encoder, decoder, and skip connections inherent to U-Nets.

Furthermore, we have expanded the description of U-Net architectures to ensure consistency in the level of detail provided for both CNNs and U-Nets, thereby offering a more balanced and comprehensive explanation of each model type.

Line 156: “hindcast” is used to describe the training data (I assume). There is no mention of performing the TSIS technique with awareness of time, so can this be called a hindcast?

Thank you for your observation regarding the use of the term "hindcast" in our manuscript. You are correct that "hindcast" may not be the most appropriate term in this context. While we are indeed predicting past states, the TSIS technique employed does not utilize observations beyond the current time step, which is part of hindcasting.

To address this, we have revised the manuscript to replace "hindcast" with "analysis."

Line 150: I am not familiar with the use of the word “innovations” here and in the following line. Could you rephrase?

Thank you for your feedback. We understand that the term "innovations" may not be universally familiar outside the data assimilation community. To enhance clarity, we have revised the manuscript to define "innovations" explicitly.

Line 291: Why this day specifically? Whether it was randomly selected or chosen based on best/worst performance should be noted.

Thank you for pointing this out. The day was selected randomly to represent a typical example of our model's performance. We will update the manuscript to clarify that this day was chosen at random to showcase the model behavior.

Revised Text:

"This day was randomly selected to provide a representative example of the model's performance in a typical scenario."

Line 347: Units should be placed on “58” and “22”. It is unclear that these are factors until reading the table.

Thank you for your feedback. The speedup factors "58" and "22" are unitless ratios representing the performance improvements of our CNN model compared to the T-SIS method. To enhance clarity, we have added a descriptive explanation in the manuscript. This clarification ensures that the context and meaning of the speedup factors are immediately apparent.

Added Text:

"The speedup or speedup factor is a unitless measure defined as the ratio of the time taken by the traditional T-SIS method to the time taken by our proposed CNN model for the same assimilation task. For example, a speed-up factor of 58 indicates that the CNN model performs the assimilation 58 times faster than the T-SIS method running on a 32-processor cluster, while a speed-up factor of 22 signifies a 22-fold increase in performance compared to the 96-processor T-SIS configuration."

Technical Comments

Throughout the paper, citations are not delineated from the clause that precedes them (i.e., separated with a comma or wrapped in parentheses). Some examples: Line 48-49, Line 117-118, Line 126, Line 144, Line 153, Line 219.

Thank you for your attention to the citation formatting in our manuscript. We understand the importance of clear and consistent citation practices for enhancing the readability and professionalism of our work. In response to your comment, we have thoroughly reviewed the manuscript and revised all instances where citations were not properly delineated from the preceding clauses. Specifically, we have ensured that all citations are appropriately wrapped in parentheses and, where necessary, separated by commas to maintain clarity.

Line 114: “coming” → “coming”

We reviewed this comment, but could not find the error in the document. If there is a specific instance that still needs correction, we would appreciate any further clarification.

Line 125: “U-net” should be plural.

Thank you for this suggestion. We have revised the text to use the plural form "U-Nets" where appropriate to maintain consistency.

Line 156: “test” → “testing”

Thank you for pointing this out. We have corrected the text to use "testing" instead of "test" to maintain proper terminology.

Line 197: “1[h]”?

Thank you for bringing this to our attention. We have fixed this issue by providing the correct time unit. The text has been updated for clarity.

We appreciate your thorough review and constructive feedback, which have helped us improve the quality and clarity of our manuscript. We hope the revised version addresses all of your concerns and contributes meaningfully to the field.