Response to Anonymous Reviewer #1

We thank reviewer 1 for your thorough and thoughtful review. We greatly appreciate your positive feedback on our work and your helpful suggestions for improving the clarity and organization of our manuscript. Our point-by-point responses to your comments are detailed below.

Specific comments:

- Q.1: l.65. Even if this is more explained in the following sections, can you precise from here what you intend by "predictability" of the SSH, which can be a very general formulation?

To improve the clarity, we have revised this sentence as follow:

"This method achieved valuable results, showing improved prediction accuracy of SSH fields in the Gulf of Mexico for up to 12 weeks compared to persistence, which was used as the baseline".

- Q.2: l.96. In the paper, you state that you used EnKF for producing the reanalysis, which has no knowledge of future observations. I am not sure that it would be a problem to have reanalysis with knowledge from both past and future observations. In the end, what you need for training OceanNet is a sequence of Ground Truth (GT) that you will use to compare OceanNet predictions vs GT. So having the best reanalysis available would be compatible with your framework no?

We acknowledge that these sentences are confusing. We have revised them as follows:

"Unlike the 4D-var method, the EnKFDA method does not rely on future timestep observations or require forward and adjoint model iterations during data assimilation. This approach enables the efficient creation of a data-assimilative ocean reanalysis, allowing OceanNet to be trained on a time-space continuous reanalysis dataset."

- Q.3: l.96. I would clearly state this as an Equation in the text, something like: x*_{t+4} = DLOP(x*_t)

For the sake of clarity, we have updated this sentence as suggested:

"If $X(t)$ is the initial five-day mean field of SSH at timestep t, then $X(t + \Delta t) = DLOP(X(t))$,

where Δt was determined during training to be four days."

- Q4. In the legend of Fig.3, the training losses are not yet introduced. You may want to let the readers know that all the related notations are defined in Section 2.5.

We have revised the caption of Figure 3 to include a statement recommended by the reviewer:

"The loss function is discussed in Sect. 2.5.1"

- Q5. l.172. You precise that the model has 80*10e6 trainable parameters which is rather big. Can you give some information here or in Section 2.5 about the current computational capabilities you needed to train the model and how long it is (not only for inference which is very fast)?

Following the reviewer's suggestion, we have included the training time information, but have positioned it at the end of the introduction.

"OceanNet, on average, outperforms SSH predictions made by the state-of-the-art Regional Ocean Modeling System (ROMS) across a 120-day period while maintaining a computational cost that is 4,000,000x cheaper (ROMS:10 hours across 144 CPUs; OceanNet: 1.18 seconds on a single NVIDIA A100 GPU) following a training period of approximately 12 hours (on an NVIDIA A100 GPU with 40GB of memory).".

- Q6. Fig.4. Can you provide some complementary analyses regarding the plateau reached at day 40 by residual geostrophy? You mentioned it for PV but not for residual geostrophy. I guess there is a clear explanation why the plateau is reached by geostrophy at the same day that the model stops agreeing with reanalysis PV (since they are connected). Maybe for the reader not a specialist in oceanography, it would be great to precise how these variables are connected.

As also pointed out by Reviewer 2, the framing of Figure 4 and its discussion is not as clear, and more importantly, it is not the focus of this study. We have decided to removed Figure 4 and its related discussion in the revision.

- Q7. l199. After reading the entire Section 2, I got the feeling that it would be valuable to add a Table here to summarize all the 4 architectures (*2 with the loss functions after but not yet presented at this point of the paper) tested among DLOP and OceanNet with specific architectures, pros and cons. This would also help the reader having an overview of the methods tested here.

Following the reviewer's suggestion, we have include a table of architecture configuration in the revision.

- Q8. Eqs 5a and Eqs 5b. You did not precise what are the notations \hat{F}x and \hat{F}y, even if we easily understand this corresponds to the Fourier decomposition. The same for |\hat{F}x|_k for the kth mode.

Following the reviewer's suggestion, we have updated the notation in the revision

- Q9. Eq.6. It is not clear to me how a and b are defined. You state that they are defined to agree with the MSE order of magnitude, which is clear but is it defined once and for all during training while the MSE magnitude may vary from batch to batch no? It is always tricky to define some weighting between multiple losses for sure but some comments added on that point would be great. What would happen if MSE prevails on spectral regularization? And the opposite? Would these parameters could have been trainable?

We have clarified this in the revision:

"Coefficients a and b are scaling factors used to ensure the order of magnitude of μx agrees with the order of magnitude of μy as well as the magnitude of the MSE loss (Eq. 7). Both a and b were determined via hyperparameter optimization to be 0.25"

- Q10. l.290. Similarly to Q.7, adding a Table to intercompare the 8 models, together with ROMS and persistence, with the metrics presented would be a good way to summarize everything, not only with Figs.7-9.

Following the reviewer's suggestion, we have included a table in the revision.

- Q11. Conclusion. You identified the generalization of your work as an issue. Can you provide some ideas to overcome this problem? Maybe a decomposition of the global domain with specific training for each area? Even better, do you think there is a way to inject more information as inputs. For the specific SSH prediction, what would be great to add for instance to ease transfer learning on other domains (atmospheric forcings?, addition of physical constraints in the losses?)

An approach for OceanNet's generalization is to implement it for the global ocean. Our team has been working on this and will report the results in a sperate correspondence. We have included the following sentence in the conclusion as a next step:

"Efforts to apply OceanNet to the global ocean are currently underway by our research team and will be reported in a future correspondence".

3. Technical corrections

Please find below a list of grammatical or typing errors to consider before publication:

l.27. reattached the to GSM -> to the

l.175. What is the arrow in front of Eq.3?

l.251. an example of it can… -> One example can

l.254. validations images -> may be replaced by ground truth reanalyses?

l.276. that that l.278. If the means of two objects of comparisons… -> replace by: if the average metric of two models…?

l.282. can still outperforms -> outperform

We thank the reviewer for this list. All have been corrected and incorporated in our revision.