## 3rd Review of Long-term Prediction of the Gulf Stream Meander Using OceanNet: a Principled Neural Operator-based Digital Twin

Many thanks for the clarifications, particularly around the model. I want to reiterate that I think this is interesting work, and would love to see it published, however I maintain my view that some amendments are needed first.

We sincerely appreciate the time and effort both the editor and the reviewer have dedicated to reviewing our manuscript and providing your valuable feedback. We are grateful for your insightful comments and valuable suggestions, all of which have been carefully incorporated in our revision.

The improvements to the model description have hugely improved readability. Thank you for this. However, the use of the convolutional layers are still not mentioned here, please could they be added (the figure 3 caption now includes reference to these, a similar sentence in the model text would help considerably). I note the authors comment that the specifics of the FNO are described elsewhere, and agree that a full and thorough model design paper should not form part of this paper. However, I feel as it stands the text gives a brief description of some parts of the model, ignoring others… a full (but very brief) description of the model is needed – as a description of only part leaves the reader assuming that the model only consists of the parts mentioned, and is misleading. To be clear, I am requesting only a brief overview, but which includes all elements of the model to a simple degree. Currently the convolutional layers are entirely missed in the text.

The Figure 3 caption has been updated and the paragraph starting at line 158 has been changed to reflect the requested brief description of the model.

Re figure 3, I note the authors reluctance to make changes to the figure due to a sister paper. However, in my view, its current form is misleading, and incredibly unhelpful to the reader. The location of the start and end of the curly braces between part a and b of the figure very much seem to be incorrect. Leaving this misleading element (even while acknowledging this is present in the sister paper) is, to me, incredibly unhelpful to readers. The curly braces currently sit part way through the convolution layers, implying that some of the convolutions belong to the Fourier layer, and some don't. My understanding is that the curly braces should begin and end at the edges of the yellow 'Fourier layer' box. I would like to see the curly braces correctly located, so the start and end clearly define what is included in part b of the figure.

Further to this, it is still unclear to me which part of the figure is included in the definition of *N*. If this is the output from part a, can this be clarified. To me, this is now even more confusing than before --- the caption implies that N refers just to the Fourier part ("(b) The Fourier Neural Operator, depicted as N."), but part c of the figure 3 shows that N is applied to x... if both these are correct, then when/how are the convolutions in fig3a applied to x?

I think the authors would benefit from having a discussion with someone who was not involved in the creation of the figure, to try and ensure the figure, caption and text make sense to someone who has no prior knowledge of the figure. I am able to offer my time if the authors would like to discuss this by v/c, but it may be just a productive to speak to a colleague in their group/deparment who was not involved with the original figure creation.

Figure 3 and its caption have been amended. $\mathcal{N}$ is correctly referred to as the Fourier Neural Operator. The OceanNet model is comprised of a Fourier Neural Operator, a numerical integration scheme, and a loss function. As conveyed by the mathematics, the Fourier Neural Operator, $\mathcal{N}$, predicts the increment of the system while the numerical integration scheme, **H**, uses the initial field **X** and the increment to give the predicted field. In Figure 3, (a) shows a high-level overview of the Fourier Neural Operator, (b) shows a high-level overview of the Fourier Layers contained in the Fourier Neural Operator. $\mathcal{N}$ is also referred to as the neural network in the sections containing the mathematical theory of the model since we are generalizing (e.g. it could be any neural network).

For me the added lines 181-182 "In practical terms, a future timestep X(t + Δt) is predicted by feeding the initial image X(t) into our neural network N with parameters θ. The numerical integration scheme H is then applied to the outputs as discussed in Chattopadhyay and Hassanzadeh (2023)." do not make sense with other comments in the paper and in response to the review. In the majority of cases (and that discussed here), the tendency, Δt, is predicted by N. Then H is applied to give X(t + Δt). This would read better/make more sense with something like:

"In practical terms, the increment, Δt, is obtained by feeding the initial image X(t) into our neural network N with parameters θ. The numerical integration scheme H is then applied to the this as discussed in Chattopadhyay and Hassanzadeh (2023) to give the future timestep X(t + Δt)"

Thank you for the close attention to detail. This mistake has been corrected as recommended.

The added description in section 2.3.2 hugely improves my understanding of the methods, many thanks. Its particularly useful for enlightening me on what is meant by 'no integration scheme'. I still however feel that the phrase 'no integration scheme' is a bit confusing. Please could this be changed to "directly predicting future fields". I feel it's important to clarify that these experiments have not simply removed the way the increment is applied, but instead retrained a new model, N, to do a totally different thing. 'no integration scheme' does not capture this message. Please could this be changed throughout the text and figure captions.

This information on (what are now) lines 204-207 has been amended:
"In Sect. \ref{Results}, experiments are described for a variety of models, some of which did not employ a numerical integration scheme in their methods. In such cases, the integration scheme is not present during the training of the neural network, $\mathcal{N}$, thus the model is directly predicting the next timestep as is commonly seen in CNN and U-NET models such as those discussed in Sect. \ref{sect:DLOP}. The equation representing such models can be given as:..."

Furthermore, a recall to this method has been added to lines 306 and 307, where the references to models without an integration scheme begin:
"The RMSE, anomaly correlation coefficient (ACC), and MHD are compared across different iterations of the DLOP and FNO models, focusing on integration schemes and loss function terms. The two integration schemes compared were the absence of integration (Eq. \ref{noint}) and PEC. Recall that a difference in integration schemes corresponds to an entire retraining of the model and thus results in a different model."

The description of v(t) in figure 3 caption has hugely improved readability, thanks. However, it seems the tilda is lost on some of the references – some should refer to v_tilda, not v.

Thank you for catching this error. The typo has been corrected.

I note the authors response to my comments re the inputs of H. I maintain my viewpoint that H should take both x and N as inputs. The authors note 'While our selection of integration scheme is implicit and thus depends on both X(t) and N[X(t), \theta], this does not have to be the case in general.' In my view, the *general approach* would be to have both variables available as inputs, in some cases both are used, in other cases only one. I.e. F(x,y), sometimes F(x,y)=x+y, sometimes F(x,y)=x+2. It's not ideal to have a function which takes different inputs, but certainly including all potential inputs provides this generality. By listing h as a function *only* of N, it is then incorrect to give any inputs other than N, and the generality is lost.

This suggestion has been implemented. Where appropriate, H has been given inputs of X(t) and N[X(t), \theta]

The caveats in the use of persistence etc are much better stated now, thank you. However, I still have issue with the line “Persistence is sometimes used in short-to-medium term ocean forecasting due to its simplicity (e.g. Jacox et al. (2020)), but it does not account for changes in weather and climate conditions.” I have not thoroughly read the Jacox paper, but firstly, and most importantly it refers to prediction of marine ecosystems. Ocean forecasting (including in the applications of this paper) is very distinct biogeochemical/ecosystem forecasting. Further, this paper seems to be a review paper discussing the predictability of systems, and commenting on how *anomaly* persistence *contibutes* to predictability. Again this is very different to stating persistence can be suitable used for forcing. I would very much like to see this single sentence removed.

The sentence has been removed as requested.

Thank you again for the great attention to detail from both the reviewer and the editor - this has greatly improved our submission and we look forward to any further discussions.