2nd Review of Long-term Prediction of the Gulf Stream Meander Using OceanNet: a Principled Neural Operator-based Digital Twin

NOTE: We have noticed an error in the numbering of sections. This has been updated. Please refer to the newest iteration of the manuscript for any mention of lines and sections by the authors.

Many of the comments from my original review have been suitably addressed, thank you. However, I feel there are still some elements which need to be addressed. Apologies for the length, I've tried to be very thorough so it's clear where I feel there are still issues and what these are.

For me it's still very unclear how the model works, and the response to my questions around the wording in section 2.4.1, figure 3, and the use of $\mathcal{N}$ and H, still leave me quite confused. I feel figure 3 and section 2.4.1 need to be substantially updated to clarify the model architecture.
The description in section 2.4.1 doesn't seem to match with figure 3, and this, along with equations 1 and 2, mean it's unclear from the paper how the model works. In particular I have the following questions around this section/figure/equations:

- The only info in section 2.4.1 about the way the network itself works seems to be the sentence 'a Fourier transform is performed on the input data, the highest Fourier modes are reduced to zero, then an inverse Fourier transform brings the data back to a real space where it is concatenated with the input image'. To me, this implies the data is moved to Fourier space, reduced, and re-projected back to real space. I.e. a reduced representation of the data (still at time t) is the output of the network. This leaves the question as to where any prediction is happening (to either predict the increment to X(t), or to predict X(t+1)). Figure 3 however appears far more complicated, and includes far more than just FFT and IFFT (i.e. the schematic of an MLP, the box labelled R, and the concatenation of items) It seems the write up misses a lot of steps, and to me would greatly benefit from a broader explanation of the steps in the network in section 2.4.1.

We thank the reviewer for taking the time to dissect this section so thoroughly and provide us with areas of improvement.

- The purpose of section 2.3.1 is to describe why the FNO is chosen and how it is employed theoretically for our problem. We leave the specifics of the FNO to the reference of Li et al. 2020 since we believe its complexity would distract the reader from the intent of the study. Thanks to the feedback from the reviewer, we have noticed the figure reference is incorrect on line 160. We added the following to section 2.3.1 to provide clarification on the FNO and have altered the original statement for continuity:
    - Line 158-160:"Following the methods of Li et al. (2020), the Fourier layer takes a high-dimensional representation of the input field, applies a Fourier transform, reduces the highest Fourier modes to zero, and applies an inverse Fourier transform to bring the data back to its original space. The resulting tensor is then concatenated with the input to the Fourier layer, to which a 2D convolution has been applied to account for aperiodicity in the data (Fig.3b).
- We are averse in changing the actual figure if it can be avoided since the figure is a reproduction of Figure 6 in our partner publication "OceanNet: A principled neural operator-based digital twin for regional oceans" (Chattopadhyay et al. 2023), which is conveyed to the reader (lines 81-84); however, we have updated the caption for Fig.3 to be clearer for the reader and better reflect the edits mentioned above:
    - (a) A schematic of the OceanNet model with input image X(t). Prior to entering the Fourier layers, the input field is lifted to a higher dimensional space by means of two convolutional

layers. The data then flows through four Fourier layers. The output of each of the Fourier layers is activated with the Gaussian Error Linear Units function. Following the last Fourier layer, the data is fed through two more convolutions to preserve the dimensions of the final output. (b) The Fourier Neural Operator, depicted as N. A Fourier transform is performed on v(t), the higher-dimensional representation of the input image, followed by a linear operation, R, to reduce the highest Fourier modes to zero, resulting in v~(t). An inverse Fourier transform brings v~(t) back to its original space. The resulting tensor is then concatenated with the input to the Fourier layer, to which a 2D convolution has been applied.
    - (c) and (d) are addressed below

- Figure 3a schematic implies a multi-layer perceptron (MLP) is used (the orange dots, connected with black lines), prior to the fourier layer, after the fourier layer, and within the Fourier layer. But nothing in section 2.4.1 mentions this. What is this part of the schematic and what is it doing? What are the input nodes in this MLP, and what is output (i.e. the dimensions of the output, and what does it represent, if anything)?
    - Thank you for catching this inconsistency between the figure and the text. Our proposed resolution for this comment can be found above.
- Figure 3b is defined as the 'Fourier Layer', but the sideways curly brace under $\boldsymbol{N}$, moving from part a to part b implies that the grey box in part b covers the entire process shown in part a. This is really confusing to me. If schematic 3b is a subset of schematic 3a, please can you clarify which of 3a is covered. I suspect that what's intended is that all of 3a combines give $\boldsymbol{N}$, if this is the case, the curly brace needs to clearly start and end either side of the schematic (currently it starts and ends in the middle of the MLP), and most importantly, the brace needs to be above $\boldsymbol{N}$, pointing at $\boldsymbol{N}$. It would also help to then have a clear space between figure 3a and figure 3b. I would also suggest having $\boldsymbol{N}$(x(t) at the right hand side of fig. 3a, to make it clear what the output is here.
    - (thanks, etc). We are averse in changing the actual figure if it can be avoided since the figure is a reproduction of Figure 6 in our partner publication "OceanNet: A principled neural operator-based digital twin for regional oceans" (Chattopadhyay et al. 2024), which is conveyed to the reader (lines 81-84). We ask that this suggestion be disregarded so continuity between the two papers may be maintained for readers.
- It's not clear what is being predicted by $\boldsymbol{N}$. Is this simply giving a reduced representation of X(t), is it predicting the increment to X(t), or is this predicting X(t+1) in a way which is then adapted by H? Please can this be clearly stated in the text, and in figure 3 (i.e. $\boldsymbol{N}(x(t) = \Delta x(t)$)
    - We appreciate the reviewer's attention to the intricacies of our model's theory and acknowledge that our original test was unclear in this regard. What is being predicted by N[X(t), \theta] is dependent on the integration scheme selected by the user. For example, if one elects to use the implicit Euler integration scheme, then the neural network, N, would be predicting the increment of X(t). In another example, if one was to choose no integration scheme and would instead train the neural network N to directly predict the next timestep, then N[X(t), \theta] would output X(t + \delta t). To clarify this in the paper, we have changed lines 181-182 to read the following: "In practical terms, a future timestep X(t + \delta t) is predicted by feeding the initial image X(t) into our neural network N with parameters \theta. The numerical integration scheme H is then applied to the outputs as discussed in Chattopadhyay & Hassanzadeh (2023)"
    - To further address this point, we have provided another example of an integration scheme in the section immediately following the above, section 2.3.2.

- The use of $\boldsymbol{\nu}$ in figure 3b is confusing, as the only explanation given is '$\boldsymbol{\nu}$(t) represents some 2d field'. I think more clarification is needed as to what $\boldsymbol{\nu}$ is - is this the input fields, or inputs processed in some way (the schematic implies $\boldsymbol{\nu}$ is the output of the MLP, is this correct?). Is each single 2d field processed independently as implied at present, with no awareness of the others? Given this case deals with single 2d SSH field inputs, the schematic feels very complex, implying combinations of inputs create $\boldsymbol{\nu}$.
    - Thank you for catching this inconsistency between the figure and the text. Our proposed resolution for this comment can be found above.
- What is R in figure 3b? This doesn't seem to be mentioned anywhere.
    - Thank you for catching this inconsistency between the figure and the text. Our proposed resolution for this comment can be found above.
- Equation 2 feels a bit confusing in its set up. Again, careful consideration needs to be given to the maths and descriptions here. In particular, there's an integral applied to F(X(t)), but then H is referred to as the integrator and covers the whole of the right hand side, not just the integral. It's not clear to me where the boundaries of $\boldsymbol{N}$ should sit here - the curly brace starts after the integral, but includes dt. I would have thought either the integral (both the integral sign and dt) are included in $\boldsymbol{N}$, or just the inside (F(X(t)) is included in $\boldsymbol{N}$ (not just dt as is shown here). The use of $\boldsymbol{N}$ later, especially the discussion of the PEC scheme in section 2.4.2 implies $\boldsymbol{N}$ is a prediction of the increment, i.e. it *includes* the integral sign in equation 2. And then H deals with how you add that increment on, whereas the current schematic doesn't give this. If this is the case, I think calling H the integrator needs to be carefully described to distinguish the way H is integrating from the integral sign used, perhaps referring to the time-stepping when explaining H.

We appreciate the thoroughness of the reviewer in her following of the mathematics of our paper and have added clarifications to the text to improve the consistency of the terms used throughout:

- We have carefully reviewed the mathematics throughout the submission and have found no computational errors. However, we do acknowledge that equation 2 has a typo that has been addressed: the bracket used for visually defining N[o,\theta] no longer includes dt.
- We have only found one instance of H being referred to as an "integrator". In the caption of Figure 3c, we have updated the clause in question to read: "the two-time-step scheme with the numerical integration operator, H". H is originally defined in lines 178-179 as "some implicit integration scheme", which is not contradicted in the visual depiction in equation 2. Other references to H used are "integration scheme" and "operator", both of which we believe to be appropriate. To be more explicit in its definition, we have amended the introduction H to be the following: "H[o] represents an operator encompassing the numerical technique used to evaluate the right-hand side of Eq.2 and will henceforth be referred to as an the 'numerical integration scheme'."
- We have found instances where the "PEC integration scheme" has been referred to as an "integrator." We have replaced such instances with "integration scheme" to remain consistent with the references to H

- If leaving the loss function in figure 3, then the caption should explain what N_0 is please.
    - Thank you for pointing out the lack of completeness in the caption of our figure. The caption of figure 3d has been updated to the following: "The point-wise loss function used, constructed by the spectral regularizer \mu and MSE L1 for M samples and applied to all N_o ocean points. The loss function is discussed in greater detail in section 2.4.1"

I commented previously on the mathematical use of H and N, and the need for mathematical rigour - The paper still states that H is a function of $\boldsymbol{N}$ only. However this isn't the case, as X is a direct input to H (the first term on the right hand side in eq 4b).
I think this needs updating throughout the paper to replace H($\boldsymbol{N}$(X(t),θ)) with H(X(t), $\boldsymbol{N}$(X(t),θ)).

- (thanks, etc) We respectfully disagree with the reviewer's assessment of the uses of H and N throughout the article. H is explicitly referred to as an "operator" and an "integration scheme," not a function. 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 section 3 we reference model configurations with "no integration scheme", which means the neural network is directly predicting the next timestep. In such a case, the output is only estimated by N[X(t), \theta]. The decision for our notation allows for consistency across all possible numerical integration schema. Please refer to our responses above for the improvements we have made to increase clarification throughout the text in this regard.

Figure 7 has an erroneous " at the end.

- Thank you for correcting this error. The quotation mark in question has been removed.

The paper still refers to 'no integration scheme', and 'the absence of integration'. I cannot understand how this could be the case. If there is no integration how is the value changing? The model is timestepping somehow, as the blue line in figure 8 varies with time. It may be a very simple integrator is used, but the model is being integrated over time. This still needs correcting.
This perhaps speaks to a wider confusion over the use of the term 'integrator' through the paper - if this is being used to refer to the way the model moves from time step t, to time step t+1, given some output from the network, then perhaps this could be clarified when first used, to distinguish this from mathematical integration as used in the equations.
If, as I think is the case, 'integrator' is referring to how the model moves from one time step to the next, then there must be something used to do this in all versions of the model runs - any reference to 'no integrator' or similar needs to be updated throughout the paper, including in figure captions and tables etc. If I've misunderstood, then there needs to be some clarification in the paper as to what the 'integrator' means, as well as a description of how different versions of the model are being time stepped.

- We thank the reviewer for her thorough investigation of our mathematics. We have addressed all instances of the use of the word "integrator" (see above). Our definition of N on lines 176-177 identifies N[o, \theta] as the neural network parameterizing F, regarding equation 2, while H is defined on line 178-179 as "an operator encompassing the numerical technique used to evaluate the right-hand side of Eq.2" (amended, see above response). H can be *any* numerical integration scheme (e.g. implicit Euler, Runge-Kuta 2nd order, Runge-Kuta 4th order, PEC, etc). To clarify the phrases "no integration scheme" and "the absence of integration", we have added the following after line 199: "As mentioned above, experiments were performed on a variety of models, some of which did not employ a numerical integration scheme in their methods. In such cases, the neural network N is directly predicting the next timestep, as is commonly seen in CNN and U-Net models such as those discussed in section 2.2.1. The equation representing such these models can be given as: X(t + \delta t) = H[N[X(t), \theta]] = N[X(t), \theta]"

Line 350-355 currently reads: *"The comparisons between OceanNet and ROMS can also be considered to have a major caveat: ROMS as a regional ocean model depends on* ***persistent forcing conditions*** *at open boundaries, for which persistent boundaries were provided in this study.* ***While this method of modelling is operational in the sense*** *that the absence of subseasonal-to-seasonal prediction often lacks forcing information on similar timescales thus persistence must be used, it may be more fair to compare the performance of OceanNet to a model which does not require boundary conditions, such as a global ocean model."*
The first use of persistence here is incorrect - ROMS does not depend on *persistent* forcing. It depends on boundary and atmospheric forcing, which can be provided from a variety of sources, not simply from persistence (this forcing rarely comes from persistence, most commonly it comes from large models, or from climatological forecasts)
Secondly, as I mentioned in my first review, I feel it is very misleading to refer to the long term use of persistence forcing as 'operational' in any sense here. Having worked in operational regional oceanography I don't think ROMS would ever be used in this way in an operational setting, certainly I have never heard of anything like this. In almost all operational settings, a forecast from a broader model would be available and the model would be forced with this. I don't feel the comparison needs to be re-done, but this caveat needs to properly clarify that it is not being compared to anything resembling a regularly used configuration, and certainly not to anything operational. 'Operational' has a specific meaning in this context and does not seem to be correctly used here - the comments made in the paper imply a very poor approach is taken in operational ocean forecasting, which is not the case. Perhaps there is some confusion over the terminology here. In operational cases (and I would imagine almost all research cases) ROMS would not be used over these timescales with persistence forcing, far better options exist and would be used.
I would like to see this updated, with any reference to operational applications using persistence removed. I suggest something along the lines of the following:
'*The comparisons between OceanNet and ROMS can also be considered to have a major caveat: ROMS as a regional ocean model depends on* ***providing*** *forcing conditions at open boundaries, for which persistent boundaries were provided in this study.* ***This method would not be used in meaningful prediction scenarios over the timescales considered here, and as such*** *it may be more fair to compare the performance of OceanNet to a model that does not require boundary conditions, such as a global ocean model****, or to a configuration of ROMS forced with boundary conditions taken from predictions produced by a global ocean model."***

The authors also do not seem to have made any reference in these caveats to the use of persistence forcing for the atmospheric forcing conditions as I flagged as a concern in my first review. Again this element of the set up severely degrades the quality of the predictions from ROMS, and again ROMS would not be used to provide forecasts over these timescales with persistent atmospheric forcing. I think this caveat must be noted in the paper.

We appreciate the reviewer for the thoughtful comments and assistance in revision the texts. We have adopted the suggested revision to amend the statement about persistence surface and boundary forcing:

"*The comparisons between OceanNet and ROMS can also be considered to have a major caveat: ROMS as a regional ocean model depends on providing forcing conditions on the ocean surface and at open boundaries, for which persistence was provided in this study. This method would not be used in conventional prediction scenarios over the timescales considered here, and as such it may be more fair to compare the performance of OceanNet to a model that does not require boundary conditions, such as a* *global ocean model, or to a configuration of ROMS forced with forcing and boundary conditions taken from predictions produced by a global model***.”**

Following my review, the authors updated the paper to state: *“Currently, almost all global ocean and atmosphere forecasts extend only 7-10 days”.* This is not correct, many operational forecast centres run forecasts for almost all lead times. Short term systems give forecasts for 7-10 days, but many forecast centres run operational sub-seasonal to seasonal (S2S) forecast systems (generally providing ocean and atmospheric forecasts within a coupled system) which predict out to a few months, and many other systems provide even longer forecasts, all the way out to mutli-decadal climate forecasts and beyond. The paper they cite (Jacox et al., 2020) seems to be a *research* application of marine biology predictions, not an operational system, and not forecasting the ocean itself. The incorrect statements over *operational* forecasts must be amended.

Similarly, the paper now states *“in practice. Persistence is commonly used in short- to medium-term ocean forecasting due to its simplicity (e.g., Jacox et (2020)), but it does not account for changes in climatic conditions such as those driven by El Niño or other large-scale climate.”* Persistence is not ‘commonly’ used, and the issues with persistence are far greater than this - persistence forecasts do not account for *any* variability, over any timescales. For applications such as those in the paper it's not just the lack of climatic variability, but any weather impacts, such as winds, storms/hurricanes, warming or cooling over a few days/weeks etc. The caveat here does not accurately capture the significant limitations of the use of persistence.

We thank the reviewer for the thoughtful comments. Following your guidance, we have amended the paragraph as the following:

“*In regional ocean forecasting, defining surface and boundary forcing is a significant challenge, particularly when accurate and continuous global ocean and atmosphere forecasting data for extended periods are unavailable. In this study, persistence refers to the assumption that future conditions will resemble past conditions. 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. While persistence can provide a baseline, it is not expected to capture the full variability or trends in long-term forecasts. We acknowledge the limitations of using persistent forcing to drive ROMS forecasts in this study*.”

In general, I think that paper has some really interesting science in it, but still needs notable work in describing the methods, and in being clear about the caveats of the comparison given. Along with clarification over the alternative methods to iterate/time-step the models used.

We thank the reviewer again for her very thoughtful and constructive comments and suggestions, which have greatly improved the quality of this manuscript.