

The authors have improved the manuscript and addressed my initial comments. The addition of coastal regions in their neural network (NN) model evaluation is very helpful, and it provides better support for the reported results of the NN estimates. The revised manuscript is also easier to follow, and only minor revisions are needed to improve the writing in places.

We sincerely thank the reviewer for their time. Their insights have helped us strengthen the manuscript and adding the coastal regions was a valuable addition that better supports our analysis. We are particularly grateful to them for dedicating their time to review our manuscript a second time, especially considering the significant interval since the previous submission. We have performed a final editorial review based on their comments to ensure that the manuscript is clear and concise throughout.

In parts of the manuscript, one of my earlier criticisms still applies; some sentences remain difficult to understand. I especially had trouble with the two NN prediction approaches that are not clearly distinguished from one another in Section 2.4. Having read through the full manuscript, I think I now understand the two approaches, but when I initially read Section 2.4, I was confused by the description. There are a few more instances where sentences are confusing or misleading for a first-time reader; please see my specific comments for details. We thank the reviewer for their feedback. To address the confusion, we have revised the first paragraph of section 2.4 to explicitly clarify the rationale for both approaches. We provide more details on this revision below.

The revised manuscript contains many large figures; some of them could be moved to a supporting information document. For example, the open ocean and coastal region raw-value chlorophyll-a figures (11 and 13) are shown alongside anomaly figures (12 and 14), each taking up a page. Here the authors could decide to focus on showing either raw values or the anomalies, or not displaying all regions.

We thank the reviewer for the suggestion to streamline the presentation. After careful consideration, and noting the Editor's general preference for maintaining the data within the main text, we have decided to retain Figures 11–14. While we acknowledge the length of the manuscript, we would prefer it to remain self-contained.

specific comments (line numbers are based on the revised manuscript, not the tracked changes version)

L 4: "a neural architecture based on the U-Net that reconstructs surface, near-global chlorophyll-a based on observations and four physical predictors": This sentence can easily be interpreted to say that observations are used as inputs to the neural network, when actually model (reanalysis or forecast) fields are used. The next sentence is a bit more helpful, but the reader may think that the listed inputs "mixed layer depth, sea surface height, salinity, and temperature" are observations. Please rephrase to avoid confusion. We thank the reviewer for this observation. We have removed the word "observations" from that sentence to avoid confusion. The source of the data remains described in the following paragraph (L9).

L 71: "Because the target exhibits strong seasonality and our focus was not necessarily maximal architectural expressiveness, we did not adopt a single monolithic model covering

the full range of variability." What exactly does this mean? The following sentences make the reader understand the methodology better but this sentences is too general and not helpful. Please rephrase.

We thank the reviewer for this suggestion. We have rewritten this sentence to more clearly explain the modeling choice.

L 75: "... each network was trained on six-month time series starting from its initialization month (m_1 - m_6): The use of " m_1 - m_6 " is confusing here: the symbols are not used elsewhere in the section and the reader is lead to believe forecasts start in months 1 through 6 when I presume the m_1 through m_6 are meant to refer to the "six-month time series" and not the "initialization month". Please rephrase and maybe just do not use the symbols yet.

We agree that the notation (m_1)–(m_6) was unnecessary and potentially confusing. In the revised text we have removed these symbols and instead describe the training data in plain language.

L 75: "For example, the network initialized in January was trained on January-June data from 1998-2016, while the December network was trained on December 1998-May 2017, and so on.": The date range "December 1998-May 2017" may make most readers believe all months in that time period were used when (I assume) this is not the case. I suggest using the same formulation from earlier in the sentence: "December-May data (December 1998 to May 2017)". For some readers it may further be helpful to mention the word "climatology" in this section.

We appreciate this suggestion. We have revised the description of the training regime to clarify that only periods of six months are used.

L 110: "As mentioned above, the network operates on log-transformed chl-a values, where $p_i = \log(\hat{y}_i)$...": because it was mentioned above that the NN produces log-transformed chl-a values, I would have expected \hat{y} to be in log-space already. We thank the reviewer for this observation. To ensure absolute clarity between the physical units (y^i) and the optimization space (p_i), we have opted to keep the explicit notation. This distinction is emphasized to differentiate from the training objective in log-space back to the evaluation of model performance in linear-space

L 147: "we used 6-month forecasts from SEAS5": The previous sentence stated that the neural network is "using the reanalysis (GLORYS12) as input". Are both SEAS5 and GLORYS12 being used here, and how? This is a bit confusing. Currently it reads like the first sentences of Section 2.4 are describing one single approach, please be more specific and mention the two approaches explicitly.

We thank the reviewer for pointing out the potential confusion. To address the confusion, we have revised the first paragraph of section 2.4 to explicitly clarify the rationale for both approaches. In the revised manuscript we have clarified that our evaluation uses two distinct setups with different temporal resolutions:

1. **Validation with reanalysis at 5-day resolution**: we use GLORYS12 as input to generate 5-day chl-a predictions over 2018–2023, which allows us to assess the

model's ability to map physical fields to chl-a at higher temporal resolution when the physics are well constrained by observations.

2. **Seasonal forecast generation with SEAS5 at monthly resolution**: we use 6-month SEAS5 forecasts as input to generate monthly chl-a predictions. This is consistent with the temporal resolution of SEAS5 and the intended seasonal forecasting application.

L 160: "monthly resolution": Here, everything is expressed in months, for both the temporal resolution and the lead time. What happened to the 5-day predictions mentioned in the first sentence of the section? Are these 5-day predictions? If so, please add that information somewhere, if not, modify the first sentence. -- This comment is related to the last one, please describe the forecast generation approaches better.

We thank the reviewer for their comment. We have revised section 2.4 to clarify this point, we refer the reviewer to the previous comment on L147.

L 199: "Physics=G12": Is G12 the same as GLORYS12, GLO12 used in line 179, or something else? Without additional information this statement is not very helpful. If it is just a reference to Fig. 5, I don't think it needs to be included here. We thank the reviewer for their comment. We have removed this reference from the text, as the relationship is defined in the caption and legend of Figure 5.

Fig. 5 and 7: Why not include the BIO4 results here?

We thank the reviewer for this suggestion. We have updated Figures 5 and 7 to include the BIO4 data as requested.

Fig. 8: The color scale is broken: white color does not indicate 0. When recreating the figure, I would suggest using the same color scale as in Fig. 6 for easy comparison. We are especially grateful to the reviewer for identifying this inconsistency. The color scale in Figure 8 was indeed shifted. We have recreated the figure ensuring that it correctly maps to [-1, 1] and this revealed that the degradation in skill with respect to the GLORYS12 version of the model was less important than previously thought. We have also updated the manuscript to reflect this (L219-223).

Fig. 11 and following: I would suggest using the same color for the same NN output, i.e., "Init. JAN" should have the same color in different figures. We thank the reviewer for this suggestion. We have updated the figures to ensure a more consistent color identity across the figures. However, to maintain the distinction between anomalies and raw fields, we have utilized varying tints of the base colors.

L 308: "and so this framework would fall short whenever three-dimensional biogeochemical consistency is required.": There are many cases where surface chlorophyll-a is not representative of phytoplankton biomass, for example in the presence of a deep chlorophyll maximum. So, I agree that the framework as presented falls short in some cases, but the expression "three-dimensional biogeochemical consistency" is awkward and not helpful to the reader. We thank the reviewer for their suggestion. We agree that the term "three-dimensional biogeochemical consistency" was overly abstract. Following the

reviewer's suggestion, we have revised this sentence to explicitly mention the vertical distribution of biomass and the specific example of a DCM.

L 309: "lead time 1" I think it would be useful to mention the units here and say that this is a 1-month lead time. I would prefer if the months were added throughout the manuscript, but they are important to mention here, as some readers may just skip to the Discussion. We agree that specifying this is essential for clarity. We have updated the Discussion and have ensured that time units (months) are included throughout the manuscript.

L 329: "... can be penalizing of modest timing and spatial shifts in bloom features.": What is meant by "modest timing" here? I suggest changing to something like "can penalize even modest changes in timing or small spatial shifts of chl-a blooms." We thank the reviewer for this suggestion. We have incorporated it into the manuscript.

L 330: "the monthly evaluation record over is relatively short": something is missing here. We thank the reviewer for catching this typo. We have revised the unclear phrasing.