Review: Deep learning-based chlorophyll prediction: comparison with a dynamic model and applications to fish catch forecasting

The manuscript shows substantial improvement. The introduction is well structured and clearly outlines the research topic. The methods section presents the essential details in a more coherent manner and provides a clearer description of the architectural framework. Similarly, the results section is now more clearly presented, with accurate description of the conducted experiments and more detailed analysis of the presented results. Despite these advances, a few issues remain that still require attention.

METHODS:

- Section 2.1 outlines the architecture design and provides several key details necessary for reproducibility. Table B1, referenced in the review, offers a clear and concise summary of these architectural specifications; however, it is not included in the manuscript, and no supplementary material section is currently available. Its inclusion within the manuscript is therefore recommended. Although this information is presented in the results section, the description of the model outputs would benefit from this clarification. In particular, it remains unclear whether the output consists of a single value or a time series. Providing this information more explicitly in the methods section would improve the reader's understanding of the architecture design.

- In the Introduction, the authors state that "*the network produces monthly or annual chlorophyll forecasts at the LME scale with lead times of 1–24 months*" (line 64). However, in the Methods section it is reported that "*each model directly predicts a 3-month mean chlorophyll anomaly centered on the targeted month from a three-month window of preceding input variables*" (line 93). Clarification of the relationship between these two descriptions would improve the overall clarity and facilitate a better understanding of the project purpose.

- Section 2.2 describes the datasets employed, including the training, validation, and test sets, and provides details on input data preprocessing. This information is also partially introduced in Section 2.1, in paragraph 2.1.2, which presents the training, validation, and test sets in a particularly clear manner. It may therefore be worth considering its relocation to Section 2.2, if feasible, to improve the overall organization of the manuscript. In addition, Table B2 appears to offer valuable support for clarity and readability, but it is not currently included in the manuscript; its inclusion is recommended. Furthermore, the order in which the datasets are presented in Section 2.1.2 differs from that used in Section 2.2. Adopting a consistent order

throughout the manuscript could enhance readability and help the reader in better understanding the role and usage of each dataset.

- SHAP analysis (Section 2.3) quantifies the contribution of each spatial location's input value to the predicted LME-mean chlorophyll anomaly. This is achieved by comparing the model's predictions with and without the grid point of interest, while accounting for all possible subsets of the remaining grid points. However, the procedure used to modify the input data in order to represent the absence of a given grid point is not clearly described. For example, it is unclear whether the value at that location is set to zero or replaced using an alternative strategy. Providing further clarification on this aspect would enhance the transparency and robustness of the analysis.

RESULTS:

- Although the implemented revisions have improved the clarity of the explanation, it remains unclear whether the reference model is defined prior to, or derived from, the sensitivity analysis. On one hand, at line 212, the authors state that *"in each sensitivity experiment (blue bars), a single component of the reference model was modified,"* which suggests that the reference model serves as the baseline configuration for the sensitivity experiments. On the other hand, at line 223, it is stated that *"the reference model, optimized through these sensitivity experiments, represents the configuration that achieved the best balance of spatial robustness and computational efficiency,"* implying that it is the outcome of the sensitivity analysis itself. Clarifying the role and development of the reference model—specifically whether it represents the initial baseline or the final optimized configuration—would improve the consistency of the description and strengthen the interpretability of the sensitivity analysis experiments.

- At line 223, the authors define the reference model as "*the model which best balances spatial robustness and computational efficiency*", compatible with the results reported in table 1. Although that, figure 2 reports higher predictive skill for the model with 5 layers. This discrepancy suggests that additional criteria may have been considered in selecting the reference model. Providing further clarification on this aspect would improve the transparency, reproducibility, and overall understanding of the experimental design.

- In the review, the authors provided the rationale for adopting a 2D-to-0D approach. Including the motivation expressed in the answer also in the Discussion section would

better justify the reason behind the choice of this approach, clarifying its advantages and enhancing readability.