Detailed Responses

Here, we provide detailed responses to the referee #2' comments. The comments from the referees are shown in black. Our responses to the critics of the referees are supplied in normal font and **blue**. The appropriate correction in the manuscript has been repeated in **red** font in the response letter.

Referee #2:

Thanks for the revised version. I have only a few corrections/comments and I hope the authors can incorporate them into the final version before acceptance.

Response:

We are very grateful for reviewing our manuscript and providing us with your recognition and valuable advices on our work. Your comments and suggestions will definitely help us improve the manuscript.

1. L200: It was noticed that the authors used SST CCI data before 2022 and OSTIA SST NRT products afterwards. Are the authors aware of the OSTIA reprocessed data https://doi.org/10.48670/moi-00168 covering 1988-2023, which I believe is a more consistent product with the OSTIA NRT data compared to the CCI SST. Have the authors checked on the difference of the OSTIA and CCI SST data? In some regions the difference can be significant.

Response:

Thank you for your detailed review.

Regarding the OSTIA reprocessed dataset you mentioned, we were indeed unaware of it in this study, and did not conduct an in-depth analysis of the differences between various SST products. Based on your suggestion, we performed a simple visual comparison and found that although there are differences between the OSTIA and CCI SST data, these differences are not significant at the spatiotemporal scales required for our research. Therefore, we believe that these differences have limited impact on the conclusions of our study.

Furthermore, we fully understand the importance of the variability in SST data for research accuracy. We deeply agree with your suggestion for further comparative analysis. However, considering the scope of this study, conducting a comprehensive comparison would exceed the anticipated workload of this research. We will consider a more comprehensive comparative analysis of the OSTIA and CCI SST datasets in future studies to ensure the robustness and reliability of the results.

2. Terminology inconsistency: It was found both in the response letter and in the revised manuscript, PFT stands for different terms (e.g. Plant functional types in the response letter, and Photosynthetic functional types in the ms??). I believe the authors know what PFT really denotes. Please check throughout the text to make sure the consistency of all terms.

Response:

We sincerely apologize for the errors and any confusion caused by the inconsistent use of terminology in our manuscript and response letter.

PFT stands for *Phytoplankton Functional Type*. We have conducted a thorough review of the entire manuscript, ensuring that all references to PFT are now consistent and correctly denote *Phytoplankton Functional Type*.

3. It was mentioned how the predictor variables were preprocessed (normalised etc) but not for the in situ PFT data as response variables. How were they normalised before put into the ensemble training?

Response:

Thank you for your comment.

All PFT data are processed using the log-10 transformation, which is common in the field of ocean color research. This transformation helps to narrow the range of data, bringing the distribution closer to a normal distribution, thereby enhancing the stability and accuracy of model predictions.

4. Uncertainty is a bit unclear - are they based on log-10 or natural logarithmic transformed data? Can you explain how to understand it or how to convert it to the common relative errors (%)

Response:

Thank you for your detailed review of the uncertainty issue.

The uncertainty in our study is based on log-10 transformed data, which is a common approach in the ocean color field for analyzing bio-optical parameters. We have clarified this in the revised manuscript (see page 13, line 261).

"It should be noted that all computations of the uncertainties in this study were conducted on log-10 transformed data, which follows conventional practice in the field of ocean color research (Xi et al., 2021)."

Xi, H. Y., Losa, S. N., Mangin, A., Garnesson, P., Bretagnon, M., Demaria, J., Soppa, M. A., D'Andon, O. H. F., and Bracher, A.: Global Chlorophyll a Concentrations of Phytoplankton Functional Types With Detailed Uncertainty Assessment Using Multisensor Ocean Color and Sea Surface Temperature Satellite Products, J Geophys Res-Oceans, 126, e2020JC017127, https://doi.org/10.1029/2020JC017127, 2021.

It is important to clarify that the uncertainty described in this paper is not the typical relative error (i.e., the percentage deviation between predicted and true values). Instead, it is estimated through the variance of predictions from multiple sub-models in the ensemble learning process, as shown in the following figure.

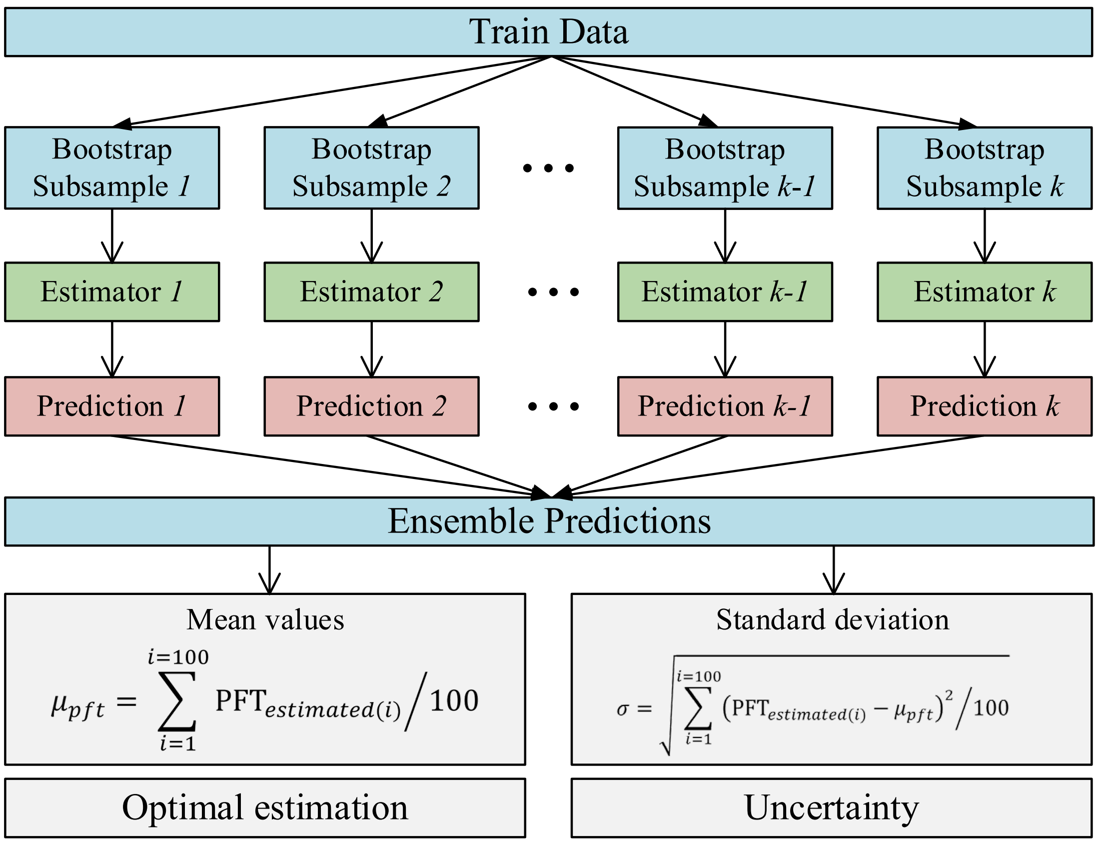

Fig. 1 Schematic view of Ensemble Learning.

The uncertainty in our study is conceptually different from the relative errors (%), and thus direct conversion between the two is difficult. This approach does not rely on simple point estimates but rather uses a distributional approach to characterize the dispersion of prediction results and model stability. In ensemble learning, uncertainty reflects the variability in model predictions when different training datasets or model architectures are used. This measure of uncertainty is more aligned with the requirements of evaluating ensemble model predictions in our study, as it better captures the model's robustness when faced with unseen data. Additionally, deep ensemble methods have been proven to be a Bayesian approach that can provide high-quality uncertainty estimates, outperforming methods like MC Dropout (Lakshminarayanan et al., 2017). Compared to traditional Bayesian methods such as Gaussian process regression, variational Bayesian, and Laplace approximation, ensemble learning offers significant advantages in flexibility, ease of implementation, and computational efficiency (Abdar et al., 2021). The concept of uncertainty in our study is different from the relative errors (%), and therefore cannot be directly converted.

Abdar, M., Pourpanah, F., Hussain, S., Rezazadegan, D., Liu, L., Ghavamzadeh, M., Fieguth, P., Cao, X. C., Khosravi, A., Acharya, U. R., Makarenkov, V., and Nahavandi, S.: A review of uncertainty quantification in deep learning: Techniques, applications and challenges, Inform Fusion, 76, 243-297, 10.1016/j.inffus.2021.05.008, 2021.

Lakshminarayanan, B., Pritzel, A., and Blundell, C.: Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles, Adv Neur In, 30, 2017.

5. Similar to my previous comment on Fig 12, Fig 13 doesn't show the uncertainty distribution in the north pole. Maybe put the maps of another month e.g. July or August in the Supplementary document.

Response:

Thank you for your comments. Follow your concerns, we have added the uncertainty distribution map for July 10, 2020, in the Supplementary to provide a more comprehensive view of the North Pole region.

“Additionally, Figure S3 in the supplementary materials illustrates the global distribution of uncertainties on July 10, 2020.

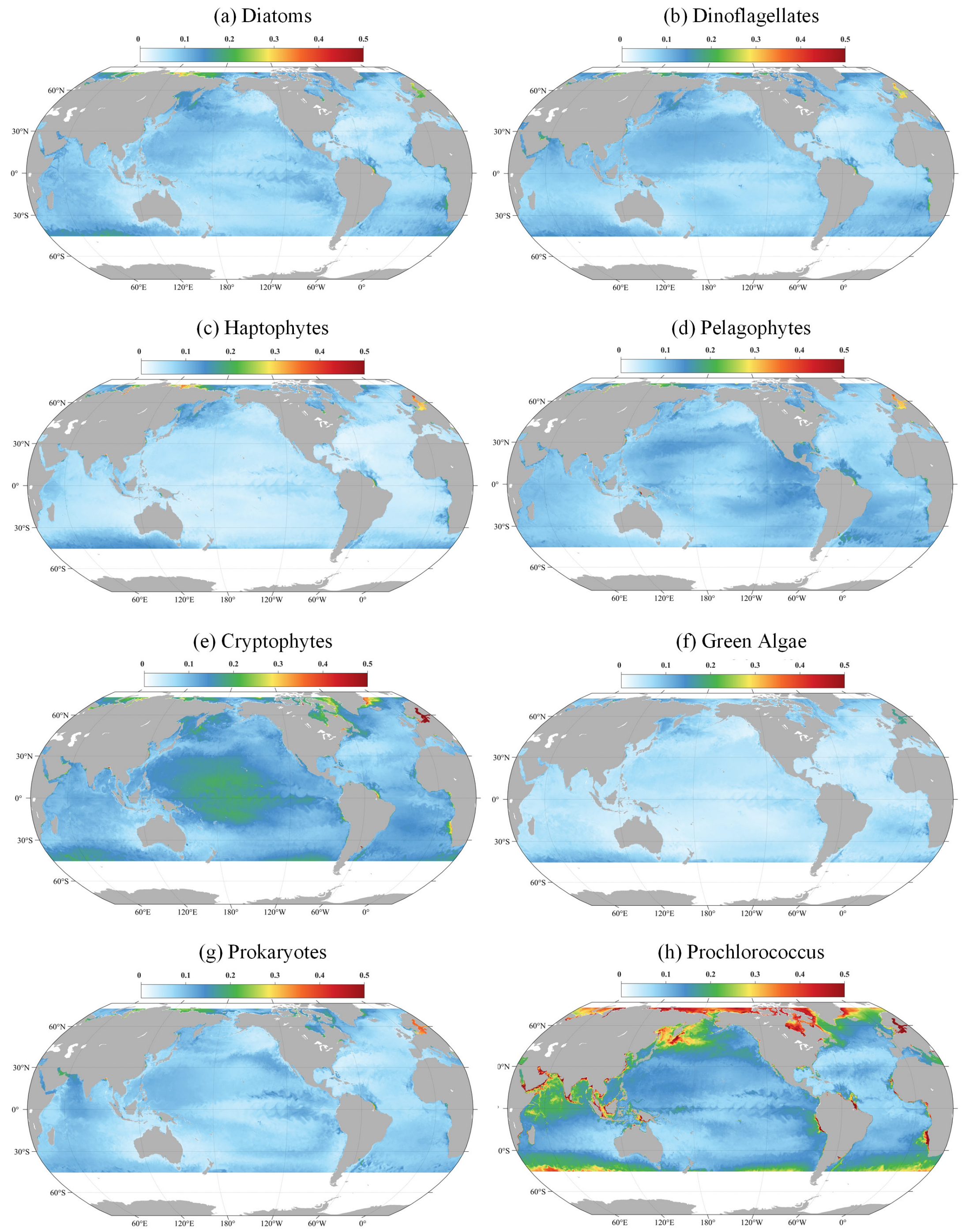

Figure S3 The global distribution (2020-07-10) of the uncertainties for (a) Diatoms, (b) Dinoflagellates, (c) Haptophytes, (d) Green Algae, (e) Prochlorococcus, (f) Prokaryotes, (g) Pelagophytes and (h) Cryptophytes."

6. L439 the uncertainty (is) relatively...

Response:

We apologize for the mistake. We have corrected it (see line 439 on page 25 of the revised manuscript).

"Overall, the uncertainty is relatively low in the open ocean, suggesting that the model performs with a high degree of confidence."

7. PFT maps - L430, Figs 12-13 Figure S2

Mind that: Prochlorococcus prediction in high latitudes is unrealistic as they are almost never found in oceans above 50 degrees if you have checked the global in situ HPLC data. Also see Flombaum et al 2013 and Xi et al. (2021). This should be minded and probably cut it off for the high latitudes and should also be discussed together with the uncertainty.

Response:

Thank you for pointing out this important concern.

We fully agree with you. As shown in Figure S3 of the supplementary materials, the uncertainty associated with Prochlorococcus predictions increases significantly at latitudes above 50°. This indicates that the predictions in these high-latitude regions may not be realistic. In future research and model improvements, we will consider introducing a threshold based on existing ecological studies and global in situ data analysis. This threshold will help filter out unrealistic predictions in high-latitude regions, particularly in areas with higher uncertainty, thereby achieving more accurate and ecologically consistent global PFT predictions. Additionally, incorporating more ecological prior knowledge as constraints during the neural network training process is crucial. This approach will not only help the model better understand and reflect ecological principles but also enhance its predictive capability and reliability in complex environments. We have added a brief discussion (see line 538 on page 31 of the revised manuscript).

"It is also necessary to consider introducing a threshold based on existing ecological studies and global in situ data analysis, which will help filter out predictions in areas with high uncertainty."

8. Regarding my last second comment about the difficulty of applying the STEE-DL model to future datasets - I suggested to have a brief discussion on this point in the revised version, however it is not found there.

Response:

Thank you for your comments. To facilitate the application of the STEE-DL model, we are developing a set of user-friendly data preprocessing tools that will help users effectively utilize our STEE-DL model with updated datasets. Following your concerns, we have added a brief discussion (see line 502 on page 30 of the revised manuscript).

"As environmental data continues to be updated, the STEE-DL model can be easily applied to future datasets, allowing for the continuous generation of PFTs, which will contribute to long-term global or local scale analyses."