Geoscientific Model Development Discussions

Geosci. Model Dev. Discuss., referee comment RC1
https://doi.org/10.5194/gmd-2022-224-RC1, 2022

Comment on gmd-2022-224

Anonymous Referee #1

Referee comment on "Testing the reconstruction of modelled particulate organic carbon from surface ecosystem components using PlankTOM12 and Machine Learning" by Anna Denvil-Sommer et al., Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2022-224-RC1, 2022

Thank you to the authors for this well-written submission that I believe can become a nice step forward.

The authors nicely set up a real world experiment by sampling a model at real world observation locations. However, the authors then utilize drivers that are not always available with at these observation locations (and a very large set of drivers). Thus, the findings about whether ML can be utilized to better understand (and model) the transfer of POC to depth are not as applicable without a substantial change in observations that are made. My main concerns are:

(1) Was analysis done to calculate the correlation between the drivers? This was done between drivers and targets, but high correlations between drivers suggest the ML can do with utilizing fewer drivers. In the current state, there are too many drivers utilized for the findings to be of substantial use to the community without providing analysis that these drivers are available and colocated with the POC observations. Right now, the paper, with some minor revisions, would be useful to the BGC modeling community only.

(2) Was analysis done to determine which driver observations we do have (and at which time/location)? Is there a set of drivers that can be tested that would correspond to what is currently available (and could be utilized in the near future within ML)? Right now, I don't forsee the current findings to be directly applicable, as most of the driver sets are large. Do all the observations of POC also have observations of all the drivers? What set of drivers would be realistically available for use in ML? It would be most useful to start there and then add individual drivers to see which additions have the largest impact on the ML results.

(3) Whether with your best ML results, or with the most feasible set of observed drivers, I

am curious as to why no analysis was done about where additional observations would be of most value to the ML - where added observations would alter the global performance of the ML?

(4) Based on Figure 7, there are many features with extremely low importance. Did you go a step further and test your ML approaches and driver sets without these low-impact drivers?

In summary, as it currently stands, the article is most useful for BGC modelers. I think this will be a great contribution beyond this community, after care is taken to critically think (and analyze if any are redundant to the ML) which drivers are realistically available now at these observations, and if these prove to be inadequate for the ML results, determine the smallest set of additional environmental conditions that must be observed for the ML to give good results. If the current set of observed conditions are adequate, then many additional experiments can be done. Such as when and where do we need to sample to improve our ML model?