Review of "Using Neural Network Ensembles to Separate Biogeochemical and Physical Components in Earth System Models" by Holder et al. (gmd-2021-167)

October 8, 2021

In their article, Holder et al. use the approach of neural network ensembles (NNE) to extract relationships between predictor (nutrients, irradiance, temperature) and target (small and large phytoplankton biomass) variables within ocean biogeochemical models. Specifically, they investigate whether the NNE approach is capable of determining why different models produce different results. They study three test cases, where they either alter the physical formulation controlling the circulation or biological equations. Thereby, they focus on the two different types of relationship, i.e., intrinsic vs. apparent relationships. They conclude that the NNE approach is capable of characterizing these relationships and can thus be considered as a parsimonious representation of the system, including extrapolative power.

Overall, this study provides a valuable contribution of how one can leverage "Machine Learning" approaches to better understand the dynamics of a complex model, such as ocean biogeochemical or Earth system models. Also, not being an expert in ocean biogeochemical modeling, I consider the presented methods and analyses to be robust. The manuscript is well written, however some passages need restructuring and the manuscript needs to be tuned for its target audience — please see my comments.

I recommend minor revisions of the manuscript before publication.

1 General Comments:

1.1 *Overall, the manuscript based on the title and the bigger part of the abstract aims at a larger readership working with Earth system models (ESM), the main part of the manuscript is, however, very focused on ocean biogeochemistry modeling. For example, the abstract is very ESM-general until line 12, but then jumps into a very specific problem on phytoplankton. Earth system modelers, who are not so familiar with ocean biogeochemistry, might be a bit lost here and in general throughout the article. I suggest to either sharpen the focus of the manuscript to only aim for the ocean biogeochemistry community, or to be more inclusive for Earth system modeler in general. The latter solution would require that you clearly state that the ocean biogeochemistry problem investigated in this study is used as a case study to demonstrate your approach, introduce the reader more to the problem of small vs. large phytoplankton prediction, and how one could adapt your approach/case study to other aspects of the Earth system.*

2 Specific Comments:

2.1 *Please stick to the tenses, i.e. do not switch between present and past tense when describing your results. I recommend that you always use present tense when describing your study at hand, i.e. when describing your methods, your results etc., and only use past tense when referring to already published studies.*

2.2 *L1: The title is too general. There are also biogeochemical and physical components in the land-surface models. Better to add "Ocean" in the title.*

2.3 *L22: The abstract misses a concluding sentence. Please add a sentence that gives a general outlook of your study and highlights its significance for the discipline of Earth system modeling.*

2.4 *L27: It is limited not only by imperfect knowledge, but also by the fact that we cannot resolve the processes in current models and current HPC facilities.*

2.5 *L46: Maybe better "are indeed being modelled".*

2.6 *L50: Include a sentence here that shortly explains the concept behind the NNE.*

2.7 *L51: Again, the use of tenses in this manuscript is a bit misleading. It is better to write: "... (NNEs)* ***are*** *able to extract ..." instead of "... were able ...". It's not that they lost the capability to do so in the meantime.*

2.8 *L64: Better "high irradiance" instead of "high light".*

2.9 *L71: The paper is, on the one hand, specific about ocean biogeochemical modeling and, on the other hand, it tries to be more general about Earth system modeling. One could add a statement here that you look into phytoplankton physiology as a case study, but the approach is also applicable to other problems in the Earth system.*

2.10 *L73–83: This section reads a bit like you already discuss your results. It would work better if you used present tense and explain the different approaches which are applied in this research, and why.*

2.11 *L110: "***ocean** *biogeochemical components of ESMs"*

2.12 *L112; Equation 1: Please replace "Light" with* $\mathbf{I}$ *for irradiance.*

2.13 *L131: "computationally cheap".*

2.14 *L132: either "in" or "within" the model.*

2.15 *Case Descriptions: Could you include for each case an equation describing how the NN is set up? E.g. something like Biomass = NN(Irradiance, Nutrients, Temperature) with proper variable names?*

2.16 *L190: Can you more clearly explain what the "LgSm" acronym is referring to?*

2.17 *L231: "NNEs possess some capability of extrapolating outside the range of the data on which they are trained." Very important point - you should provide a citation here!*

2.18 *L232: With RF you mean Random Forests, I assume. Can you make it clear?*

2.19 *L248–250: Why did you not set up your NN system with training, test and validation datasets? So, validation dataset to prevent overfitting, and test dataset to test generalizability?*

2.20 *L260: Maybe you can write that hyperparameters tuning showed that the setup is not very sensitive to the selection of different hyperparameters.*

2.21 *L262: Did you also use a different scheme for normalization, e.g. normalization to zero mean and unit standard deviation.*

2.22 *L340: For me, the extrapolative power of your NNE approach is a very encouraging result. You show that a NN can learn the dynamics of the system from the PI run and is able to extrapolate to extrem forcing like 4xCO2 - maybe one should make a bigger deal out of this and highlight in the abstract.*

2.23 *L437: You have not introduced the abbreviations Chl:C. I know, it is clear for the reader with ocean biogeochemistry background, but your title addresses a larger readership. So, please introduce all abbreviations.*

2.24 *L498: Better rename to "Summary & Conclusions".*

2.25 *L518: Rephrase "we can be relatively confident" to something like "their predictions can be considered reliable."*

2.26 *Conclusions: Overall, I find them too long and not to-the-point. Can you boil it down to a few concise statements?*

2.27 *Figure 1 & 2: I cannot comment on the specifics of the ocean biogeochemical models. Ideally, another referee with the needed expertise should comment on these aspects.*

2.28 *Sensitivity Analysis Figures: The colored lines are the actual model run output, right? Or is it the mean NNE? The grey shading is the NNE, right? Could you put this in the legend? If the actual model output is not included in the figure, where do you show the performance for NNE versus actual model output except R^2 and RMSE values in the tables.*

2.29 *Sensitivity Analysis Figures: Why do you show in e.g. Figure 3 small and large phytoplankton biomass together and Figure 4 only small phytoplankton biomass. Can you not remove small phytoplankton biomass from Figure 3 and corresponding subsequent figures?*

2.30 *Sensitivity Analysis Figures: I find the black arrows at the axis to be a bit misleading - do you need them?*

2.31 *Sensitivity Analysis Figures: What does "ex." in the captions mean? Example? Better use e.g. then.*

2.32 *Figure 5: I'd prefer if you added the unit next to the colorbar.*