

## ***Interactive comment on “Linking intrinsic and apparent relationships between phytoplankton and environmental forcings using machine learning – What are the challenges?” by Christopher Holder and Anand Gnanadesikan***

**Anonymous Referee #1**

Received and published: 13 August 2020

The authors try to find a ML method that can establish and help to explain the link between intrinsic and apparent relationships of phytoplankton and environmental forcing. Three different methods were tested: Multiple Linear Regression (MLR), Random Forests (RFs) and Neural Network Ensembles (NNEs). The tests were provided on three different Scenarios. The authors found that the NNEs reproduce well the observed biomass (biomass estimated from intrinsic relationships) based on the knowledge of only apparent relationships when both relationships operate on the same spatial and temporal timescales. All methods fail when the intrinsic and apparent relation-

C1

ships operated on different timescales. However, the main authors' conclusion is that ML methods still can give an information on shapes of intrinsic relationships. Using the Earth System Models (ESM) in their third Scenario they show that the ML methods can extract useful information from this model that can be used for an examination of interaction between input variables.

The article raises an interesting subject. However, it misses a clear explanation how the trained data and data for validation were constructed; the explanation of the role of input data, especially the physical meaning of the choice to fix them at 25th, 50th and 75th percentile for a sensitive analysis. I also think that the authors did not use all possible ML capacities for example, test more hidden layers in NNs, or add few more input variables, environmental ones, maybe a temperature.

Also, the main results that the NNEs can reproduce the general shape of intrinsic relationships does not have a systematic character: it is clearly seen that in some cases the NNEs shapes reproduce a different behavior compared to observation data. If the authors want to keep this conclusion, they will have to provide an additional analysis of conditions under which their conclusion is persistent.

The article in its current state needs a serious major correction.

Specific comments:

It is hard to understand reading the introduction what is the main aim of the article. I have found three:

- “A significant challenge that remains is determining how intrinsic relationships found in the laboratory scale up to the apparent relationships observed at the ecosystem scale (i.e., scaling the small to the large).”
- “What is less clear is: 1. Can robust relationships be found? 2. If so, what methods are most skillful in finding them? 3. How do you interpret the apparent relationships that emerge when they diverge from the intrinsic relationships we expect?”

C2

- “To investigate when and why the link between intrinsic and apparent relationships break...”

If the two first citations can be linked to the title “What are the challenges?”, the last one, what the authors actually did, is not according to the title “What are the challenges?”. I will advise to authors modify the structure of the introduction and emphasize what is the main idea of this article.

The conclusion also has to be rewritten in clear way: what are the challenges authors have found to link intrinsic and apparent relationships? The authors said: “Our main objective in this manuscript was to use ML to determine under what conditions intrinsic and apparent relationships between phytoplankton are no longer equal...” This objective was not clear in the introduction, and again, does not correspond well to the title of the article.

I advise you also to avoid the non-explained abbreviations in the abstract, like line 21: “ESM”, line 28: “MLR”.

It would be better not mentioning the results in the introduction (lines 100-103), and instead prepare your readers for the structure of the article.

In the section “2. Methods” I suggest you provide a scheme or formula what exactly you were using as input/predictors and output/target and how it links to your equations 1 and 2 etc. This will simplify the understanding. It is especially important for the description of the Scenario 2. It is unclear, are the values of biomass with which the authors compare their results calculated based on hourly values? Is the biomass for target in learning algorithm smoothed or calculated based on smoothed predictors? It will strongly affect the results.

The word “target” was not used in your article. It is common word in the ML domain and can also help to better understand the method especially for readers who only start to use ML techniques. Also, I would suggest to use the word “validation” instead

C3

of “testing”.

To perform a sensitive analysis the authors fixed two of three predictors at different percentiles. It misses the explanation why 25th, 50th and 75th percentiles were chosen, what is the physical meaning of this choice and how it influences the results? For example, does 75th percentiles represent the extremes and what does it mean? I think that it is important to explain it to better understand the results. Please, clarify that the sensitivity analysis was done already on trained ML model.

Did the authors try to increase the number of hidden layers in their NNs? It is known that the introduction of more hidden layers can improve the results. It would be interesting to see if there is any effect from the number of hidden layers in this particular problem.

The authors did not provide how they scaled their variables (lines 337-341). This procedure is known as normalization of variables. Normalization ensures that all predictors fall within a comparable range and avoids giving more weight to predictors with large variability ranges.

It is hard to agree that the NNEs and RFs represent well the behavior as for example on Figure 3 in left column the NNEs show a strong increase at the end of the macronutrient range that does not in agreement with the observation values; and there is a false decrease in the middle column at 25th percentile.

It is interesting to know if the authors have an idea about new parameter or variable that can bring back the information on hourly variability lost due to the time-averaging to improve the results? The article misses the total results for Scenarios 2 and 3 like on Figure 1.

In the Discussion of results for Scenario 3 the authors reasoned about BLING model behavior and did not mention their results. It would be interesting to know the authors' thoughts about NNEs behavior on Figure 7 middle column at 75th percentile.

C4

I have also found that the authors wrote too much on things that they did not do: lines 171-177.

Lines 260 – 273: I suggest to rewrite these two paragraphs, there is a mention of results that have not yet been presented; and I do not see the necessity to mention that the authors “previously had little experience with ML.”

Please use the figures captions like “a”, “b” etc.

Please, expand the figure captions, for example on Figure 2 it would be good to add that the black line is estimated from Eq. 1.

Figure 5 was mentioned as the last but it is placed before Figure 6 and 7.

Line 149, 156, 168, 356: please avoid to use the sign “ in scientific paper.

Line 149: sign “ should be before the point .

Line 206: it feels that “but” should be replaced by “and”.

Line 226, 235, 543: word “just” is unnecessary.

Line 250: “an ML” should be replaced by “a ML”.

Line 356: Miss a figure indication.

---

Interactive comment on Biogeosciences Discuss., <https://doi.org/10.5194/bg-2020-262>, 2020.