Linking intrinsic and apparent relationships between phytoplankton and environmental forcings using machine learning – What are the challenges?

Christopher Holder[1], Anand Gnanadesikan[1]

[1] Morton K. Blaustein Department of Earth and Planetary Sciences, Johns Hopkins University, Baltimore, MD 21218, United States of America

Correspondence to: Christopher Holder (cholder2@jh.edu)

Abstract. Controls on phytoplankton growth are typically determined in two ways: by varying one driver of growth at a time such as nutrient or light in a controlled laboratory setting (intrinsic relationships) or by observing the emergence of relationships in the environment (apparent relationships). However, challenges remain when trying to take the intrinsic relationships found in a lab and scaling them up to the size of ecosystems (i.e., linking intrinsic relationships in the lab to apparent relationships in large ecosystems). We investigated whether machine learning (ML) techniques could help bridge this gap. ML methods have many benefits, including the ability to accurately predict outcomes in complex systems without prior knowledge. Although previous studies have found that ML can find apparent relationships, there has yet to be a systematic study that has examined when and why these apparent relationships will diverge from the underlying intrinsic relationships. To investigate this question, we created three scenarios: one where the intrinsic and apparent relationships operate on the same time and spatial scale, another model where the intrinsic and apparent relationships have different timescales but the same spatial scale, and finally one in which we apply ML to actual ESM output. Our results demonstrated that when intrinsic and apparent relationships are closely related and operate on the same spatial and temporal timescale, ML is able to extract the intrinsic relationships when only provided information about the apparent relationships. However, when the intrinsic and apparent relationships operated on different timescales (as little separation as hourly to daily), the ML methods underestimated the biomass in the intrinsic relationships. This was largely attributable to the decline in the variation of the measurements; the hourly time series had higher variability than the daily, weekly, and monthly-averaged time series. Although the limitations found by ML were overestimated, they were able to produce more realistic shapes of the actual relationships compared to MLR. Future research may use this type of information to investigate which nutrients affect the biomass most when values of the other nutrients change. From our study, it appears that ML can extract useful information from ESM output and could likely do so for observational datasets as well.

1 Introduction

Phytoplankton growth can be limited by multiple environmental factors (Moore et al., 2013) such as macronutrients, micronutrients, and light. Limiting macronutrients include nitrogen (Eppley et al., 1973; Ryther and Dunstan, 1971; Vince and Valiela, 1973), phosphorus (Downing et al., 1999), and silicate (Brzezinski and Nelson, 1995; Dugdale et al., 1995; Egge and Aksnes, 1992; Ku et al., 1995; Wong and Matear, 1999). Limiting micronutrients can include iron (Boyd et al., 2007; Martin, 1990; Martin and Fitzwater, 1988), zinc, and cobalt (Hassler et al., 2012). Additionally, limitations can interact with one another to produce colimitations (Saito et al., 2008). Examples of this include the possible interactions between the micronutrients iron, zinc, and cobalt (Hassler et al., 2012) and the interaction between nitrogen and iron (Schoffman et al., 2016) such that local sources of nitrogen can have a strong influence on the amount of iron needed by phytoplankton (Maldonado and Price, 1996; Price et al., 1991; Wang and Dei, 2001). Spatial and temporal variations, such as mixed layer depth and temperature, affect such limitations, and have been related to phytoplankton biomass using different functional relationships (Longhurst et al., 1995).

Limitations on phytoplankton growth are usually characterized in two ways – which we term intrinsic and apparent. Intrinsic relationships are those where the effect of one driver (nutrient/light) at a time is observed, while all others are held constant (often at levels where they are not limiting). An example of such intrinsic relationships is the Michaels-Menten growth rate curves that emerge from laboratory experiments (Eppley and Thomas, 1969). Apparent relationships are those which emerge in the observed environment. An example of apparent relationships is those that emerge from satellite observations, which provide spatial distributions of phytoplankton on timescales (say a month) much longer than the phytoplankton doubling time, which can be compared against monthly distributions of nutrients. 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). Differences may arise between the two because apparent relationships reflect both intrinsic growth and loss rates, which are near balance over the long monthly timescales usually considered in climatological analyses. Biomass concentrations may thus not reflect growth rates. Differences may also arise because different limitation factors may not vary independently.

Earth System Models (ESMs) have proved valuable in linking intrinsic and apparent relationships. The intrinsic relationships are programmed into ESMs as equations that are run forward in time and the output is typically provided as monthly-averaged fields. The output of these ESMs is then compared against observed fields such as chlorophyll and nutrients and can be analyzed to find apparent relationships between the two. If the ESM output is close to the observations we find in nature, we say that the ESM is performing well. However, as recently pointed out by Löptien and Dietze (2019), ESMs can trade-off biases in physical parameters with biases in biogeochemical parameters (i.e., they can arrive at the same answer for different reasons). Using two versions of the UVic 2.9 ESM, they showed that they could increase mixing (thus bringing more nutrients to the surface) while simultaneously allowing for ~~this nutrient~~ to be more efficiently cycled – producing similar distributions of surface properties. However, the carbon uptake and oxygen concentrations predicted by the two models diverged under climate change. Similarly, Sarmiento et al. (2004) showed that physical climate models would be expected to produce different spatial distributions of physical biomes due to differences in patterns of upwelling and downwelling, as well as the annual cycle of sea ice. These differences would then be expected to be reflected in differences in biogeochemical cycling, independent of differences in the biological models. These studies highlight the importance of constraining not just individual biogeochemical fields, but also their relationships with each other. 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?

Recently, researchers have turned to machine learning (ML) to help in uncovering the dynamics of ESMs. ML is capable of fitting a model to a dataset without any prior knowledge of the system and without any of the biases that may come from researchers about what processes are most important. As applied to ESMs, ML has mostly been used to constrain physics parameterizations, such as longwave radiation (Belochitski et al., 2011; Chevallier et al., 1998) and atmospheric convection (Brenowitz and Bretherton, 2018; Gentine et al., 2018; Krasnopolsky et al., 2010, 2013; O'Gorman and Dwyer, 2018; Rasp et al., 2018).

With regards to phytoplankton, ML has not been explicitly applied within ESMs but has been used on phytoplankton observations (Bourel et al., 2017; Flombaum et al., 2020; Kruk and Segura, 2012; Mattei et al., 2018; Olden, 2000; Rivero-Calle et al., 2015; Scardi, 1996, 2001; Scardi and Harding, 1999) and has used ESM output as input for an ML model trained on phytoplankton observations (Flombaum et al., 2020). Rivero-Calle et al. (2015) used random forest (RF) to identify the drivers of coccolithophore abundance in the North Atlantic through feature importance measures and partial dependence plots. The authors were able to find an apparent relationship between coccolithophore abundance and environmental levels of $CO_{2,}$ which was consistent with intrinsic relationships between coccolithophore growth rates and ambient CO_2 reported from 41 laboratory studies. They also found consistency between the apparent and intrinsic relationships between coccolithophores and temperature. While they were able to find links between particular apparent relationships found with the RFs and intrinsic relationships between laboratory studies, it remains unclear when and why this link breaks.

ML has been used to examine apparent relationships of phytoplankton in the environment (Flombaum et al., 2020; Rivero-Calle et al., 2015; Scardi, 1996, 2001) and it is reasonable to assume that ML could find intrinsic relationships when provided a new independent dataset from laboratory growth experiments. However, it has yet to be determined under what circumstances the apparent relationships captured by ML are no longer equal to the intrinsic relationships that actually control phytoplankton growth. In this paper, we identify two drivers of such divergence. The first is colimitation that limits the biological responses actually found in the ocean, which causes non-parametric ML methods to produce apparently non-physical results. The second is climatological averaging of the input and output variables, which can distort these relationships in the presence of non-linearity.

To investigate when and why the link between intrinsic and apparent relationships break, we applied ML methods to three scenarios. For the first, we constructed a simple model in which the intrinsic and apparent relationships operated on the same time and spatial scale and were only separated by a scaling factor, but in which the environmental drivers had realistic inter-relationships. In the second, we modified the first scenario to allow the intrinsic and apparent relationships to operate on different timescales – allowing us to evaluate the impact of time-averaging on the retrieval of intrinsic relationships. In the third, we took the output from an established biogeochemical model in which the biomass is a non-linear function of growth rate to demonstrate the potential information that can be extracted from ESM output using ML.

2 Methods

2.1 Scenario 1: Intrinsic and apparent relationships on the same timescale

In the first scenario, we wanted to determine how well different ML methods could extract intrinsic relationships when only provided information on the apparent relationships and when the intrinsic and apparent relationships were operating on the same timescale. In this scenario, the apparent relationships were simply the result of multiplying the intrinsic relationships between predictors and biomass by a scaling constant.

We designed a simple phytoplankton system in which biomass was a function of micronutrient, macronutrient, and light limitations based on realistic inter-relationships between limitations (Eq. 1):

$$\mathrm{B} = \mathrm{S}_{*} \times \min(L_{micro}, L_{macro}) \times L_{Irr} \qquad (1)$$

where B is the value for biomass (mol kg^{-1}), S_{*} is a scaling factor, and $\mathrm{L}_{\mathrm{micro,macro,irr}}$ are the limitation terms for micronutrient (micro), dissolved macronutrient (macro), and light (irradiance; irr), respectively. The scaling factor (1.9x10^{-6} mol kg^{-1}) was used, so the resulting biomass calculation was in units of mol kg^{-1}. While simplistic, this is actually the steady-state solution of a simple phytoplankton-zooplankton system when grazing scales as the product of phytoplankton and zooplankton concentrations and zooplankton mortality is quadratic in the zooplankton concentration.

Each of the limitation terms (L in Eq. 1) were functions of Michaelis-Menten growth curves (Eq. 2):

$$L_N = \frac{\mathrm{N}}{\mathrm{K_N+N}} \qquad (2)$$

where $\mathrm{L_N}$ is the limitation term for the respective factor, N is the concentration of the nutrient/intensity of the light, and $\mathrm{K_N}$ is the half-saturation constant specific to each factor. In terms of our nomenclature, Eq. 1 defines the apparent relationship between nutrients, light, and biomass such as might be found in the environment, while Eq. 2

is the intrinsic relationship between nutrient and growth rate such as might be found in the laboratory or coded in an ESM.

For the concentrations of each factor (*N* in Eq. 2), we took the monthly-averaged value for every lat/lon pair (i.e., 12 monthly values for each lat/lon pair) from the Earth System Model ESM2Mc (Galbraith et al., 2011). ESM2Mc is a fully coupled atmosphere, ocean, sea ice model into which is embedded in an ocean biogeochemical cycling module. Known as BLING (Biogeochemistry with Light, Iron, Nutrients, and Gases; Galbraith et al., 2010), this module carries a macronutrient, a micronutrient, and light as predictive variables and uses them to predict biomass using a highly parameterized ecosystem (described in more detail below). The half-saturation coefficients (K_N in Eq. 2) for the macronutrient and micronutrient were also borrowed from BLING with values of $1\text{x}10^{-7}$ mol kg^{-1} and $2\text{x}10^{-10}$ mol kg^{-1}, respectively. The half-saturation coefficient for light was set at 34.3 W m^{-2}, which was the global mean for the light limitation factor in the ESM2Mc simulation used later in this paper.

The final dataset consisted of three input/predictor variables and one response term with a total of 77,328 "observations." The input variables given to each of three ML methods (Multiple Linear Regression, Random Forests, and Neural Network Ensembles, described in more detail below) were the concentrations (not the limitation terms) for the micronutrient, macronutrient, and light. The response variable was the biomass we calculated from Eq. 1 and 2.

The dataset was then randomly split into training and testing subsets, with 60% of the observations going to the training subset and the remainder going to the testing subset. This provided a convenient way to test the generalizability of each ML method by presenting them with "new" observations from the test subset and ensuring the models did not overfit the data. The input and output values for the training subset were then used to train a model for each ML method. Once each method was trained, we provided the trained models with the input values of the testing subset to acquire their respective predictions. These predictions were then compared to the actual output values of the test subset. To assess model performance, we calculated the coefficient of determination (R^2), the mean squared error (MSE), and the root mean squared error (RMSE) between the ML predictions and the actual output values for the training and testing subsets.

Following this, a sensitivity analysis was performed. We allowed one predictor to vary across its min-max range while holding the other two input variables at their 25th, 50th (median), and 75th percentile values. This was repeated for each predictor. This allowed us to isolate the impact of each predictor on the biomass – creating "cross-sections" of the dataset where only one variable changes. For comparison, ~~these~~ values were also run through Eq. 1 and 2 to calculate the “true” response of how the simple phytoplankton model would behave. This allowed us to view which of the models most closely reproduced the underlying intrinsic relationships of the simple phytoplankton model.

This method of sensitivity analysis is in contrast to partial dependence plots (PDPs), which are commonly used in ML visualization. PDPs show the marginal effect that predictors have on the outcome. They consider every combination of the values for a predictor of interest and all values of the other predictors, essentially covering the entire data space. The predictions of a model are then averaged and show the marginal effect of a predictor on the outcome – creating responses moderately comparable to “averaged cross-sections.” Because of this averaged response, PDPs may hide significant effects from subgroups within a dataset. A sensitivity analysis avoids this disadvantage by allowing separate visualization of subgroup relationships.

Using the predictions produced from the sensitivity analyses, we also computed the half-saturation constants for each curve. Using the Matlab function “fitnlm,” the half-saturation constants were determined by fitting a non-linear regression model to each sensitivity analysis curve matching the form of a Michaelis-Menten curve (Eq. 3):

$$B = \frac{\alpha_1 N}{\alpha_2 + N} \tag{3}$$

where B corresponds to the biomass predictions from the sensitivity analyses, N represents the nutrient concentrations from the sensitivity analyses, and α_1 and α_2 are the constants that are being estimated by the non-linear regression model. α_2 was taken as the estimation of the half-saturation coefficient for each sensitivity analysis curve.

2.2 Scenario 2: Intrinsic and apparent relationships on different timescales

In Scenario 1, the intrinsic and apparent relationships differed only by a scale factor and operated at the same time and spatial scale. However, in reality, input variables (such as light) vary on hourly time scales, while satellite observations and ESM model output are often only available on monthly-averaged timescales. So the reality is that even if a system is controlled by intrinsic relationships, the apparent relationships gained from climatological variables on long timescales will not reproduce these intrinsic relationships since the average light (irradiance) limitation is not equal to the limitation given the averaged light value (Eq. 4).

$$\overline{L_{Irr}} = \overline{\frac{Irr}{K_{Irr}+Irr}} \neq \frac{\overline{Irr}}{K_{Irr}+\overline{Irr}} \tag{4}$$

where the overbar denotes a time-average, and Irr stands for irradiance (light). We wanted to investigate how such time averaging biased our estimation of the intrinsic relationships from the apparent ones; i.e., how does the link between the intrinsic and apparent relationships change with different amounts of averaging over time?

For the short timescale intrinsic relationships, we took daily inputs for the three predictor variables for one year from the BLING model. We further reduced the timescale from days to hours to introduce daily variability for the irradiance variable relative to the latitude, longitude, and time of year (Eq. 5).

$$\mathrm{Irr_{Int}}(\mathrm{t}) = \frac{12\pi \mathrm{Irr_{daily}}}{T_{Day}} \sin\left(\frac{\pi(t - t_{Sunrise})}{T_{Day}}\right) \text{ when } 0 < t < T_{Day} \qquad (5)$$

where $\mathrm{Irr_{Int}}$ is the hourly interpolated value of irradiance, $\mathrm{Irr_{daily}}$ is the **daily-mean** value of irradiance, t is the hour of the day being interpolated, $\mathrm{t_{Sunrise}}$ is the hour of sunrise, and $\mathrm{T_{Day}}$ is the total length of the day. The resulting curve preserves the day to day variation in the daily mean irradiance due to clouds but allows a realistic variation over the course of the day. The hourly values for the micronutrient and macronutrient were assigned using a standard interpolation between each of the daily values. These hourly interpolated values were then used to calculate the hourly biomass from Eq. 1 and 2. Note that we are not claiming the biomass itself would be zero at night but assume that on a long enough timescale, it should approach the average of the hourly biomass.

To simulate apparent relationships, we smoothed the hourly values for both biomass and the input variables into daily, weekly, and monthly averages for each lat/lon point. To reiterate, the intrinsic and apparent relationships in Scenario 2 differed in timescales, but not in spatial scales. Each dataset was then analyzed following steps similar to those outlined in Scenario 1; constructing training and testing subsets, using the same variables for input to predict the output (biomass), and using the same ML methods. To assess each method's performance, we calculated the R^2 value, MSE, and RMSE between the predictions and observations for the training and testing subsets. We also performed a sensitivity analysis and calculated half-saturation constants similar to those described above.

2.3 Scenario 3: BLING biogeochemical model

As a demonstration of their capabilities, the ML methods were also applied directly to monthly averaged output from the BLING model itself using the same predictors in Scenarios 1 and 2, but using the biomass calculated from the actual BLING model. As described in Galbraith et al. (2010), BLING is a biogeochemical model where biomass is diagnosed as a non-linear function of the growth rate smoothed in time. The growth rates, in turn, have the form

$$\mu = \min\left(\frac{N_{micro}}{K_{micro}+N_{micro}}, \frac{N_{macro}}{K_{macro}+N_{macro}}\right) \times \left(1 - \exp\left(-\frac{Irr}{Irr_K}\right)\right) \qquad (6)$$

where $N_{macro,micro}$ are just the same concentrations of nutrients as in Scenarios 1 and 2, Irr is the irradiance and Irr_k is a scaling for light limitation – very similar to what was done in Eq. 1 and 2 with a slight difference in the handling of light (note that the Michaelis-Menten form of light limitation in the previous scenarios can be obtained by expanding $\frac{1}{\exp\left(\frac{Irr}{Irr_k}\right)}$ as a two-term Taylor series and that in this case $K_{Irr} = Irr_k$). A more substantive difference is that the light limitation term is calculated using a variable Chl:C ratio following the theory of Geider et al. (1997). The variation of the Chl:C ratio would correspond to a K_{Irr} in Scenarios 1 and 2 which adjusts in response to both changes in irradiance (if nutrient is low) or changes in nutrient (if irradiance is high) as well as changes in temperature. Given the resulting growth rate μ, the total biomass then asymptotes towards

$$B = \left(\frac{\tilde{\mu}}{\lambda} + \frac{\tilde{\mu}^3}{\lambda^3}\right) S_* \qquad (7)$$

where λ is a grazing rate, the tilde denotes an average over a few days and S_* is just the biomass constant that we saw in the previous two scenarios. Growth rates and biomass are then combined to drive the uptake and water-column cycling of micronutrient and macronutrient within a coarse-resolution version of the GFDL ESM2M fully coupled model (Galbraith et al., 2011), denoted as ESM2Mc.

As described in Galbraith et al. (2011) and Bahl et al. (2019), ESM2Mc produces relatively realistic spatial distributions of nutrients, oxygen, and radiocarbon. Although simpler in its configuration relative to models such as TOPAZ (Tracers of Ocean Productivity with Allometric Zooplankton; Dunne et al., 2013), it has been demonstrated that in a higher-resolution physical model BLING produces simulations of mean nutrients, anthropogenic carbon uptake, and oceanic deoxygenation under global warming that are almost identical to such complicated models (Galbraith et al., 2015).

We chose to use BLING for three main reasons. The first is that we know it produces robust apparent relationships between nutrients, light, and biomass by construction – although these relationships can be relatively complicated – particularly insofar as iron and light colimitation is involved (Galbraith et al., 2010). As such, it represents a reasonable challenge for an ML method to recover such non-linear relationships. The second is that we know how these relationships are determined by the underlying intrinsic relationships between limiting factors and growth. Models with more complicated ecosystems (including explicit zooplankton and grazing interactions between functional groups) may exhibit more complicated time-dependence that would confuse such a straightforward linkage between phytoplankton growth limitation and biomass. The third is that despite its simplicity, the model has relatively realistic annual mean distributions of surface nutrients, iron, and chlorophyll, and under global warming, it simulates changes in oxygen and anthropogenic carbon uptake that are similar to much more complicated ESMs (Galbraith et al., 2015).

2.4 ML Algorithms

We chose to use Random Forests (RFs) and Neural Network Ensembles (NNEs) in this manuscript because they are two of the more popular ML algorithms. Although other ML methods exist, the list of possible choices is rather long. With the main purpose of this paper being to examine the link between intrinsic and apparent relationships on different time and spatial scales, it was decided that the number of ML algorithms being compared would be limited to RFs and NNEs given their popularity in studying ecological systems. The results of the ML methods were compared against Multiple Linear Regression (MLR) to demonstrate the better performance of ML as compared to more conventional empirical methods. Although the stronger performance of ML may seem clear to experienced ML experts, it was not immediately evident to us since we previously had little experience with ML. Therefore, MLR is included here for demonstrative purposes for less experienced ML users.

It should be noted that we are not trying to suggest that MLR is always ineffective for studying ecological systems. MLR is a very useful and informative approach for studying linear relationships within marine ecological systems (Chase et al., 2007; Harding et al., 2015; Kruk et al., 2011). However, we highly encourage our readers to try ML as it can provide insight into the non-linear portions of a dataset.

2.4.1 Random Forests

RFs are an ensemble ML method utilizing a large number of decision trees to turn “weak learners” into a single “strong learner” by averaging multiple outputs (Breiman, 2001). In general, RFs work by sampling (with replacement) about two-thirds of a dataset and constructing a decision tree. At each split, the random forest takes a random subset of the predictors and examines which variable can be used to split a given set of points into two maximally distinct groups. This use of random predictor subsets helps to ensure the model is not overfitting the data. The process of splitting the data is repeated until an optimal tree is constructed or until the stopping criteria are met, such as a set number of observations in every branch (then called a leaf / final node). The process of constructing a tree is then repeated a specified number of times, which results in a group (i.e., “forest”) of decision trees. Random forests can also be used to construct regression trees in which a new set of observations traverse each decision tree with its associated predictor values and the result from each tree is aggregated into an averaged value.

Here, we used the same parameters for RF in the three scenarios to allow for a direct comparison between the scenarios and to minimize the possible avenues for errors. Each RF scenario was implemented using the TreeBagger function in MATLAB 2019b, where 500 decision trees were constructed with each terminal node resulting in a minimum of five observations per node. An optimization was performed to decide the number of decision trees that minimized the error while still having a relatively short runtime of only several minutes. For reproducible results, the random number generator was set to “twister” with an integer of “123”. Any remaining options were left to their default values in the TreeBagger function.

2.4.2 Neural Networks

NNs are another type of ML that has become increasingly popular in ecological applications (Flombaum et al., 2020; Franceschini et al., 2019; Guégan et al., 1998; Lek et al., 1996a, 1996b; Mattei et al., 2018; Olden, 2000; Özesmi and Özesmi, 1999; Scardi, 1996, 2001; Scardi and Harding, 1999). Scardi (1996) used NNs to model phytoplankton primary production in the Chesapeake and Delaware Bays. Lek et al. (1996a) demonstrated the ability of NNs to explain trout abundance using several environmental variables through the use of the "profiling" method, a type of variable importance metric that averages the results of multiple sensitivity analyses to acquire the importance of each variable across its range of values.

Feed-forward NNs consist of nodes connected by synapses (or weights) and biases with one input layer, (usually) at least one hidden layer, and one output layer. The nodes of the input layer correspond to the input values of the predictor variables, and the hidden and output layer nodes each contain an "activation function." Each node from one layer is connected to all other nodes before and after it. The values from the input layer are transformed by the weights and biases connecting the input layer to the hidden layer, put through the activation function of the hidden layer, modified by the weights and biases connecting the hidden layer to the output layer, and finally entered into the final activation function of the output node.

The output (predictions) from this forward pass through the network is compared to the actual values, and the error is calculated. This error is then used to update the weights with a backward pass through the network using backpropagation. The process is repeated a specified number of times or until some optimal stopping criteria are met, such as error minimization or validation checks where the error has increased a specified number of times. For a more in-depth discussion of NNs, see Schmidhuber (2015).

For this particular study, we use neural network ensembles (NNEs), which are a collection of NNs whose predictions are averaged into a single prediction. It has been demonstrated that NNEs can outperform single NNs and increase the performance of a model by reducing the generalization error (Hansen and Salamon, 1990).

To minimize the differences between scenarios, we used the same framework for the NNs in each scenario. Each NN consisted of three input nodes (one for each of the predictor variables), 25 nodes in the hidden layer, and one output node. The activation function within the hidden nodes was a hyperbolic tangent sigmoid function and the activation function within the output node used a linear function. The stopping criteria for each NN was set as a validation check such that the training stopped when the error between the predictions and observations increased for six consecutive epochs. An optimization was performed to decide the number of nodes in the hidden layer that minimized the error while maintaining a short training time. Additionally, sensitivity analyses were performed using different activation functions to ensure the choice of activation function had minimal effect on the outcome and apparent relationships found by the NNEs.

Each NNE scenario used the feedforwardnet function in MATLAB 2019b. Any options not previously specified remained at their default values in the feedforwardnet function. The NNEs contained ten individual NNs for each scenario. For reproducibility, the random number generator was set to “twister,” and the random number seed was set to the respective number of its NN (i.e., 1, 2, 3, up to 10).

Each variable was scaled between -1 and 1 based on its respective maximum and minimum. This step ensures that no values are too close to the limits of the hyperbolic tangent sigmoid activation function, which would significantly increase the training time of each NN. These scalings were also applied to the RF and MLR methods for consistency between methods and the scaling did not affect the results of either method (results not shown). The results presented in this paper were then transformed back to their original scales to avoid confusion from scaling.

3 Results

3.1 Scenario 1: Intrinsic and apparent relationships on the same timescale

In Scenario 1, the RF and NNE both outperformed the MLR as demonstrated by higher R^2 values, lower MSE, and lower RMSE (Table 1). The decreased performance of the MLR is not ~~inherently~~ surprising~~,~~ given the non-linearity of the underlying model, but it does demonstrate that the range of nutrients and light produced as inputs by ESM2Mc is capable of producing a non-linear response. Additionally, each method showed similar performances between the training and testing subsets suggesting adequate capture of the model dynamics in both subsets.

From the spatial distributions of the true response and the predictions from each method, it can be observed that the RF and NNE showed the closest agreement with the true response (Fig. 1). Although MLR was able to reproduce the general trend of the highest biomass in the low latitudes and low biomass in the high latitudes, it was not able to predict higher biomass values.

In addition to examining whether the different ML methods got the “right” answer, we also interrogated these methods to look at how different predictors contributed to the answer, and whether these contributions matched the intrinsic relationships between the predictors and growth rate as we had put into the model. The MLR (red dashed lines) shows very little response to changes in macronutrient (left column), an unrealistic negative response to increases in micronutrient (central column), and a reasonable (albeit linear) match to the light response (right column). By contrast, the response to any predictor for the NNE (green dashed lines) showed agreement with the true response of the model (black lines) in all circumstances insofar as the true response was always within the standard deviation of the NNE predictions. The RF prediction of the response to a given predictor (blue dashed lines) showed agreement with the true response when the other predictors are fixed at the lower percentiles (top two rows) but began deviating in the higher percentiles.

When we computed an "effective" half-saturation for the nutrient curves in the top row of Fig. 2, we got values for K_N that were far lower than the actual ones specified in the model (Table 4). The "effective" half-saturation of when other predictors are held at their 25th percentile for the micro- and macronutrient were underestimated by one and two orders of magnitude, respectively. It was only at the higher percentiles that the micronutrient "effective" half-saturation was adequately captured when the macronutrient was not limiting. Furthermore, the "effective" half-saturation of the macronutrient was not captured even when the other variables were held at their 75th percentiles because the 75th percentile of the micronutrient still limited growth.

3.2 Scenario 2: Intrinsic and apparent relationships on different timescales

As in Scenario 1, the RF and NNE outperformed the MLR based on the performance metrics for the daily, weekly, and monthly time-averaged scenarios (Table 2). The comparable performances between the training and testing subsets suggest a sufficient sampling of the data for each method to capture the dynamics of the underlying model.

Examining the monthly apparent relationships found for each method and comparing them to the true intrinsic relationships shows that none of the methods were able to reproduce the true intrinsic relationships, with one exception being the 25th percentile plot of the micronutrient (Fig. 3). This result was consistent across the different timescales, and the sensitivity analysis showed little difference in the predicted relationships between the daily, weekly, and monthly averaged timescales for the NNEs (Fig. 4). Interestingly, the NNE and RF appeared to asymptote near the proper concentration for the micro- and macronutrients (Fig. 3). For example, the true response of the macronutrient has a sharp asymptote at low concentrations, and the NNE and RF appear to mimic this asymptote, even though the predicted biomass concentration is lower than the true biomass (Fig. 3). Furthermore, the ML methods were able to mimic the non-linearity of the system, which is an important result regardless.

When the "effective" half-saturation constants were computed for the daily, weekly, and monthly NNEs, many of the light and micronutrient half-saturations were of the same magnitude as the true value (Table 4). This is an interesting result given that the predicted biomass concentrations were much lower than the true response.

3.3 Scenario 3: BLING biogeochemical model

When run in the full ESM, the BLING biogeochemistry does end up producing surface biomass, which is a strong function of the growth rate (Fig. 6a) with a non-linear relationship as in Eq. 7. As the growth rate, in turn, is given by Eq. 6, we can also examine how the monthly mean limitation terms for nutrient and light compare with the means given by computing the limitations with monthly mean values of nutrients, Irr, and Irr_k. As shown in Fig. 6b, the nutrient limitation is relatively well captured using the monthly mean values, although there is a tendency for the monthly means to underestimate moderate values of nutrient limitation. Further analysis shows that this is due to the interaction between micro- and macro- nutrient limitation – with the average of the minimum limitation being somewhat higher than the minimum of the average limitation. However, using the actual monthly mean values of Irr, and Irr_k (Fig. 6c) causes the light limitation to be systematically biased high.

When MLR and ML were applied to the output of one of the BLING simulations, the RF and NNE again outperformed the MLR in all of the performance metrics for the training and testing subsets (Table 3). The RF performed slightly better than the NNE (R^2 of 0.973 vs. 0.942) on the training subset, but this difference was lessened in the testing subset (R^2 of 0.945 vs. 0.939). Although there were slight differences in the RF performance between the training and testing subsets, the values of the performance metrics were of the same magnitude. The similar performance for each method across the training and testing subset expresses the adequate capture of the dataset's variability.

The sensitivity analysis shows the biomass continues to increase with an eventual asymptote even in the 75th percentile plots (Fig. 7). However, the NNE curve for biomass is strongly hindered in the light and macronutrient plots even at higher percentiles, while large increases are observed in the micronutrient plots when light and macronutrient are at higher concentrations.

4 Discussion

4.1 Scenario 1: Intrinsic and apparent relationships on the same timescale

In the first scenario, our main objective was to determine if ML methods could extract intrinsic relationships when given information on the apparent relationships and reasonable spatiotemporal distributions of colimitation when the intrinsic and apparent relationships were operating on the same timescale.

Despite the fact that it agreed well with the observations, the RF prediction deviated from the true response to a given variable when other variables are held at higher percentiles (Fig. 2). This can likely be explained by the range of the training subset and how RFs acquire their predictions. When presented with predictor information, RFs rely on the information contained within their training data. If they are presented with predictor information that goes outside the range of the dataspace of the training set, RFs will provide a prediction based on the range of the training set. When performing the sensitivity analysis, the values of the predictors in the higher percentiles were probably outside the range of the training subset. For example, the bottom left plot of Fig. 2 shows how RF deviates from the true response as the concentration of the macronutrient increases – actually decreasing as nutrient increases despite the fact that such a result is not programmed into the underlying model. Although there may be observations in the training subset where the light and micronutrient are at their 75th percentile values when the macronutrient is low, there likely are not any observations where high levels of the macronutrient, micronutrient, and light are co-occurring. Without any observations meeting that criteria, the RF provided the highest prediction it could based on the training information. We discuss this point in more detail below.

In contrast to the RF's inability to extrapolate outside the training range, the NNE showed its capability to make predictions on observations on which it was not trained (Fig. 2). Note, however, that while we have programmed Michaelis-Menten intrinsic dependencies for individual limitations into our model, we do not get Michaelis-Menten type curves back for macro- and micronutrients when the other variables were set at low percentiles. The reason is that Liebig's law of the minimum applies to the two nutrient limitations so that when the micronutrient is low, it prevents the entire Michaelis-Menten curve for the macronutrient from being seen.

When the "effective" half-saturation was computed for the macro- and micronutrient curves in Fig. 2, they were far lower than the true values in the lower percentiles because of colimitations between the macro- and micronutrients (Table 4). While mathematically obvious, this result has implications for attempts to extract (and interpret) K_N from observational datasets, such that one would expect colimitation to produce a systematic underestimation of K_N.

With respect to our main objective for Scenario 1, it was evident that only the NNE was able to extract the intrinsic relationships from information on the apparent relationships. This was due in large part to its capability of extrapolating outside the range of the training dataset, whereas RFs were constrained by training data, and MLR was limited by its inherent linearity and simplicity.

4.2 Scenario 2: Intrinsic and apparent relationships on different timescales

In Scenario 1, the intrinsic and apparent relationships were simply related by a scaling factor. In practice, the relationships are more difficult to connect to each other. For the second scenario, both the output biomass and predictors (light, macronutrient, and micronutrient) were averaged over daily, weekly, and monthly timescales. Our main objective was to investigate how the link between intrinsic and apparent relationships changed when using climatologically averaged data – as is generally the case for observational studies.

When comparing the apparent relationships of the time-averaged datasets with those of the hourly intrinsic relationships, the methods almost always underestimated the true response to light and nutrient (Fig. 3 and 4). This result is not entirely unexpected. The averaging of the hourly values into daily, weekly, and monthly timescales quickly leads to a loss of variability, especially for light (Fig. 5). In fact, the variability was lost in the daily time averaging with the longer timescales showing only small differences in the possible range of values (Fig. 5). The loss of variability means that the light limitation computed from the averaged light is systematically higher than the averaged light limitation. To match the observed biomass, the asymptotic biomass at high light has to be systematically lower (see Appendix A for the mathematical proof). Differences were much smaller for nutrients as they varied much less over the course of a month in our dataset. Our results emphasize that when comparing apparent relationships in the environment to intrinsic relationships from the laboratory, it is essential to take into account which timescales of variability averaging has removed. Insofar as most variability is at hourly time scales, daily-, weekly-, and monthly-averaged data will produce very similar apparent relationships (Fig. 4). But if there was a strong week-to-week variability in some predictor, this may not be the case.

Although the ML methods were unable to reproduce the intrinsic relationships, they were able to model the general trend of the relationships (i.e., higher concentrations of each predictor lead to higher biomass; eventual asymptotes in the macro- and micronutrient). Additionally, the NNE and RF appeared to asymptote at the same nutrient concentrations as that of the true response (Fig. 3). This type of result can help to answer questions such as: which nutrients have the greatest impact on biomass when other nutrients change? This effectively allows one to examine the interactions between variables.

The computed “effective” half-saturation constants were interestingly of the same magnitude as the true value (Table 4). This is a clear demonstration of the potential hazards one may face when inferring K_N from observational datasets, as mentioned previously in Scenario 1. A further implication from Scenario 1 is reinforced in the computation of the “effective” half-saturation of the macronutrient, such that it is underestimated by an order of magnitude relative to the true value because of micronutrient limitation (Table 4).

4.3 Scenario 3: BLING biogeochemical model

To demonstrate their capabilities, each method was also applied directly to the monthly averaged output of one of the BLING simulations. The main purpose of the final scenario was to demonstrate the capabilities of the ML methods when applied to actual ESM output with the reasoning that if the ML methods were unable to provide useful information on BLING, they would also fail on more complex models.

The large increases in biomass in the micronutrient plots and hindrance of biomass in the light and macronutrient plots suggest that the system is limited by the concentration of micronutrient (Fig. 7). The biomass remained low even when macronutrient and light were at favorable levels because even when at the 75th percentile value, the micronutrient was still limiting (Fig. 8). Conceptually this makes sense since the micronutrient limitation in the BLING model hinders growth, but also limits the efficiency of light-harvesting (Galbraith et al., 2010). Additionally, the computation of the “effective” half-saturation constants demonstrates that the half-saturation constant for light drops sharply as nutrients drop (Table 4).

5 Conclusions

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, to identify whether such divergence depends on the ML method or how the input data is handled, and to understand how such divergence is related to underlying biological dynamics.

In Scenario 1, we demonstrated that NNEs were capable of extracting the intrinsic non-linear relationships from the apparent relationships when apparent and intrinsic relationships were operating on the same timescale and when they were linearly related by a scaling factor. However, this relationship broke down in Scenario 2, when time-averaging caused a systematic overestimate of light limitation. We note that while Scenario 2 illustrates that the ability to recover the intrinsic relationship with light may be compromised by temporal averaging, spatial averaging could have a similar impact. If, for example, we imagine coastal regions in which nutrient delivery is very patchy, a spatially averaged relationship between biomass and nutrient may also show similar biases. So it appears that the extent to which ML methods can extract the intrinsic relationships depends on the extent to which the variability of the system is captured; i.e., more coverage of the parameter space at higher temporal resolution would yield more accurate estimates of the intrinsic relationships.

Although RFs and NNEs were unable to extract the exact intrinsic relationships due to time-averaging, they were able to model the general trend of the relationships in Scenario 2. This mimicking of the non-linear relationships can still be a valuable tool for examining a dataset, in that one can assess which combinations of nutrients most affect the biomass and can get a relative estimate of the uncertainty in the prediction; effectively, allowing one to examine the interactions between variables and their effect on the outcome. This was further demonstrated in Scenario 3 when it was observed that even at high concentrations of light and macronutrient, biomass was limited by the concentration of micronutrient. This observation was not immediately expected or evident to us when we applied these methods to the BLING model. Similar insights might be found in other ESM output or observational datasets.

In addition to climatological averaging, it was also observed that colimitation could affect the apparent relationships found by ML. In each Scenario, we observed instances where biomass was low even when the concentrations of one of the drivers were high. This was due to one of the other drivers being limiting. Had we not known what the true intrinsic relationships were, it may have appeared that the ML methods were producing unrealistic results. For example, if the real world behaved like the right-hand column of Fig. 7, we might conclude that phytoplankton were strongly photo-inhibited, even though our results with BLING (which does not have explicit photoinhibition) demonstrate that this is not a necessary conclusion. This demonstrates the caution one must take in interpreting these kinds of systems.

Both RFs and NNEs performed well when the predictions they were asked to make were within the range of the training data. However, the sensitivity analyses illustrated the impact of RFs inability to extrapolate outside that range and that RF's suggested systematic decreases in biomass at high values of a limiting variable. Nonetheless, RFs were able to capture the same relationships as the NNEs when the sensitivity analysis was querying environments within the range of the training data. It seems that as long as RFs are presented with information across the range of the dataset, RFs will perform just as well as NNEs in a sensitivity analysis. This strengthens the conclusions of Rivero-Calle et al. (2015) in that physiologically reasonable relationships between forcing variables and biomass found using RF are reliable so long as the forcing variables (in this case pCO_2 and temperature) vary over their entire range independently of other variables (nutrients and light). However, when variation in pCO_2 is related to variation in nutrients and light (i.e., in the seasonal climatology where pCO_2 is high in the winter, light is low, and nutrients are high) RFs are unable to extract a clear signal of pCO_2 limitation.

This paper examined two of the more popular ML algorithms, but many other methods exist as well. Future research should attempt to use some of the other methods to see how they perform. However, one of the main takeaways would likely be the same regardless of the ML method; the training data should contain sufficient coverage of the range of forcing and the spatiotemporal variability within a system in order to capture the intrinsic relationships.

This paper also limited the number of predictor variables for each scenario so that the sensitivity analyses could be easily visualized. In the real world, phytoplankton may be limited by more physical and biological processes, making the visualization of the sensitivity analyses impractical due to the sheer number of possible interactions that would have to be considered. In cases such as those, it would be beneficial to perform some form of importance analysis or dimensionality reduction to remove insignificant predictor variables, after which sensitivity analyses could be done on the remaining predictors.

ML techniques have several benefits that could make them useful for biological oceanographers and ecosystem modelers. Many ML methods (including the two presented here) do not require any prior knowledge of a system to construct a model. Additionally, new methods are continually being developed for viewing the dynamics of the ML models. Given these advantages, ML could provide a compact form for representing relationships between ecosystem parameters such as biomass and primary productivity and their environmental drivers (nutrients and light) in observational data and complex models. Preliminary work indicates that we can use NNEs in particular to: 1. Compare model relationships with those derived from observational datasets, rather than simply using spatial patterns of errors. 2. Evaluate whether differences between models reflect important differences in biological parameters or whether they are due to differences in the physical circulation. We would expect that two different physical models run with the same biological scheme would produce the same relationships. 3. Evaluating whether global warming really would be expected to drive ecosystems outside their historical parameter range. We will report on these results in a future manuscript.

Appendix A

Illustration of why time variation causes underestimation of the dependence of biomass on a limiter

$$B = S_* * \frac{Irr}{K_{irr}+Irr} = S_* * \frac{\overline{Irr}+Irr\prime}{K_{irr}+\overline{Irr}+Irr\prime} \tag{A1}$$

where the overbar refers to a time-average and the prime to a variation from this time average. Insofar as the variations are small.

$$B = S_* * \frac{\overline{Irr}+Irr'}{(K_{irr}+\overline{Irr})*(1+Irr'/(K_{irr}+\overline{Irr}))} \approx S_* \frac{\overline{Irr}+Irr'}{(K_{irr}+\overline{Irr})} * (1 - Irr'/(K_{irr} + \overline{Irr})) \tag{A2}$$

Averaging

$$\bar{B} \approx S_* \left\{ \frac{\overline{Irr}}{(K_{irr}+\overline{Irr})} - \frac{\overline{Irr'^2}}{(K_{irr}+\overline{Irr})^2} \right\} < S_* \frac{\overline{Irr}}{(K_{irr}+\overline{Irr})} \qquad \text{(A3)}$$

so that if we are trying to fit a curve of the form

$$\bar{B} \approx S_*^{ave} \left\{ \frac{\overline{Irr}}{(K_{irr}+\overline{Irr})} \right\} \qquad \text{(A4)}$$

We would expect that $S_*^{ave} < S_*$.

Code and Data Availability

The Matlab scripts for the construction of the figures and tables, the scripts for training and testing the MLR, RF, and NNE algorithms, and the source files for each scenario are available in the Zenodo data repository (https://doi.org/10.5281/zenodo.3932388, Holder and Gnanadesikan, 2020).

Author Contribution

CH implemented the ML algorithms, analyzed the results for each scenario, and wrote the majority of the manuscript. AG helped in developing the simple phytoplankton models for Scenarios 1 and 2, provided the biogeochemical model output used in Scenario 3, and helped in the analysis of the results.

Competing Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors would like to thank Eric Baumann and Dr. Nicole DeLuca for their comments and recommendations for this manuscript.

Financial Support

This research was supported in part by the National Science Foundation (NSF) Integrative Graduate Education and Research Traineeship (IGERT) (Grant No. 1069213) and by the NSF Division of Ocean Sciences (OCE) (Grant No. 1756568).

References

Bahl, A., Gnanadesikan, A. and Pradal, M.-A.: Variations in Ocean Deoxygenation Across Earth System Models: Isolating the Role of Parameterized Lateral Mixing, Glob. Biogeochem. Cycles, 33(6), 703–724, doi:10.1029/2018GB006121, 2019.

Belochitski, A., Binev, P., DeVore, R., Fox-Rabinovitz, M., Krasnopolsky, V. and Lamby, P.: Tree approximation of the long wave radiation parameterization in the NCAR CAM global climate model, J. Comput. Appl. Math., 236(4), 447–460, doi:10.1016/j.cam.2011.07.013, 2011.

Bourel, M., Crisci, C. and Martínez, A.: Consensus methods based on machine learning techniques for marine phytoplankton presence–absence prediction, Ecol. Inform., 42, 46–54, doi:10.1016/j.ecoinf.2017.09.004, 2017.

Boyd, P. W., Jickells, T., Law, C. S., Blain, S., Boyle, E. A., Buesseler, K. O., Coale, K. H., Cullen, J. J., de Baar, H. J. W., Follows, M., Harvey, M., Lancelot, C., Levasseur, M., Owens, N. P. J., Pollard, R., Rivkin, R. B., Sarmiento, J., Schoemann, V., Smetacek, V., Takeda, S., Tsuda, A., Turner, S. and Watson, A. J.: Mesoscale Iron Enrichment Experiments 1993-2005: Synthesis and Future Directions, Science, 315(5812), 612–617, 2007.

Breiman, L.: Random forests, Mach. Learn., 45(1), 5–32, 2001.

Brenowitz, N. D. and Bretherton, C. S.: Prognostic Validation of a Neural Network Unified Physics Parameterization, Geophys. Res. Lett., 45(12), 6289–6298, doi:10.1029/2018GL078510, 2018.

Brzezinski, M. A. and Nelson, D. M.: The annual silica cycle in the Sargasso Sea near Bermuda, Deep Sea Res. Part Oceanogr. Res. Pap., 42(7), 1215–1237, doi:10.1016/0967-0637(95)93592-3, 1995.

Chase, Z., Strutton, P. G. and Hales, B.: Iron links river runoff and shelf width to phytoplankton biomass along the U.S. West Coast, Geophys. Res. Lett., 34(4), L04607, doi:10.1029/2006GL028069, 2007.

Chevallier, F., Chéruy, F., Scott, N. A. and Chédin, A.: A Neural Network Approach for a Fast and Accurate Computation of a Longwave Radiative Budget, J. Appl. Meteorol., 37(11), 1385–1397, doi:10.1175/1520-0450(1998)037<1385:ANNAFA>2.0.CO;2, 1998.

Downing, J. A., Osenberg, C. W. and Sarnelle, O.: Meta-Analysis of Marine Nutrient-Enrichment Experiments: Variation in the Magnitude of Nutrient Limitation, Ecology, 80(4), 1157–1167, doi:10.2307/177063, 1999.

Dugdale, R. C., Wilkerson, F. P. and Minas, H. J.: The role of a silicate pump in driving new production, Deep Sea Res. Part Oceanogr. Res. Pap., 42(5), 697–719, doi:10.1016/0967-0637(95)00015-X, 1995.

Dunne, J. P., John, J. G., Shevliakova, E., Stouffer, R. J., Krasting, J. P., Malyshev, S. L., Milly, P. C. D., Sentman, L. T., Adcroft, A. J., Cooke, W., Dunne, K. A., Griffies, S. M., Hallberg, R. W., Harrison, M. J., Levy, H., Wittenberg, A. T., Phillips, P. J. and Zadeh, N.: GFDL's ESM2 Global Coupled Climate-Carbon Earth System Models. Part II: Carbon System Formulation and Baseline Simulation Characteristics*, J. Clim., 26(7), 2247–2267, doi:10.1175/JCLI-D-12-00150.1, 2013.

Egge, J. and Aksnes, D.: Silicate as regulating nutrient in phytoplankton competition, Mar. Ecol. Prog. Ser., 83, 281–289, doi:10.3354/meps083281, 1992.

Eppley, R. W. and Thomas, W. H.: Comparison of Half-Saturation Constants for Growth and Nitrate Uptake of Marine Phytoplankton 2, J. Phycol., 5(4), 375–379, doi:10.1111/j.1529-8817.1969.tb02628.x, 1969.

Eppley, R. W., Renger, E. H., Venrick, E. L. and Mullin, M. M.: A Study of Plankton Dynamics and Nutrient Cycling in the Central Gyre of the North Pacific Ocean, Limnol. Oceanogr., 18(4), 534–551, 1973.

Flombaum, P., Wang, W.-L., Primeau, F. W. and Martiny, A. C.: Global picophytoplankton niche partitioning predicts overall positive response to ocean warming, Nat. Geosci., 13(2), 116–120, doi:10.1038/s41561-019-0524-2, 2020.

Franceschini, S., Tancioni, × Lorenzo, Lorenzoni, M., Mattei, × Francesco and Scardi, M.: An ecologically constrained procedure for sensitivity analysis of Artificial Neural Networks and other empirical models, PLoS One San Franc., 14(1), e0211445, doi:http://dx.doi.org/10.1371/journal.pone.0211445, 2019.

Galbraith, E. D., Gnanadesikan, A., Dunne, J. P. and Hiscock, M. R.: Regional impacts of iron-light colimitation in a global biogeochemical model, Biogeosciences, 7(3), 1043–1064, doi:https://doi.org/10.5194/bg-7-1043-2010, 2010.

Galbraith, E. D., Kwon, E. Y., Gnanadesikan, A., Rodgers, K. B., Griffies, S. M., Bianchi, D., Sarmiento, J. L., Dunne, J. P., Simeon, J., Slater, R. D., Wittenberg, A. T. and Held, I. M.: Climate Variability and Radiocarbon in the CM2Mc Earth System Model, J. Clim., 24(16), 4230–4254, doi:10.1175/2011JCLI3919.1, 2011.

Galbraith, E. D., Dunne, J. P., Gnanadesikan, A., Slater, R. D., Sarmiento, J. L., Dufour, C. O., Souza, G. F. de, Bianchi, D., Claret, M., Rodgers, K. B. and Marvasti, S. S.: Complex functionality with minimal computation: Promise and pitfalls of reduced-tracer ocean biogeochemistry models, J. Adv. Model. Earth Syst., 7(4), 2012–2028, doi:10.1002/2015MS000463, 2015.

Geider, R. J., MacIntyre, H. L. and Kana, T. M.: Dynamic model of phytoplankton growth and acclimation: responses of the balanced growth rate and the chlorophyll a: carbon ratio to light, nutrient-limitation and temperature, Mar. Ecol. Prog. Ser., 148, 187–200, 1997.

Gentine, P., Pritchard, M., Rasp, S., Reinaudi, G. and Yacalis, G.: Could Machine Learning Break the Convection Parameterization Deadlock?, Geophys. Res. Lett., 45(11), 5742–5751, doi:10.1029/2018GL078202, 2018.

Guégan, J.-F., Lek, S. and Oberdorff, T.: Energy availability and habitat heterogeneity predict global riverine fish diversity, Nature, 391(6665), 382–384, doi:10.1038/34899, 1998.

Hansen, L. K. and Salamon, P.: Neural network ensembles, IEEE Trans. Pattern Anal. Mach. Intell., 12(10), 993–1001, doi:10.1109/34.58871, 1990.

Harding, L. W., Adolf, J. E., Mallonee, M. E., Miller, W. D., Gallegos, C. L., Perry, E. S., Johnson, J. M., Sellner, K. G. and Paerl, H. W.: Climate effects on phytoplankton floral composition in Chesapeake Bay, Estuar. Coast. Shelf Sci., 162, 53–68, doi:10.1016/j.ecss.2014.12.030, 2015.

Hassler, C. S., Sinoir, M., Clementson, L. A. and Butler, E. C. V.: Exploring the Link between Micronutrients and Phytoplankton in the Southern Ocean during the 2007 Austral Summer, Front. Microbiol., 3, doi:10.3389/fmicb.2012.00202, 2012.

Holder, C. D. and Gnanadesikan, A.: Linking intrinsic and apparent relationships between phytoplankton and environmental forcings using machine learning - What are the challenges?, doi:10.5281/zenodo.3932388, 2020.

Krasnopolsky, V. M., Fox-Rabinovitz, M. S. and Belochitski, A. A.: Development of neural network convection parameterizations for numerical climate and weather prediction models using cloud resolving model simulations, in The 2010 International Joint Conference on Neural Networks (IJCNN), pp. 1–8., 2010.

Krasnopolsky, V. M., Fox-Rabinovitz, M. S. and Belochitski, A. A.: Using Ensemble of Neural Networks to Learn Stochastic Convection Parameterizations for Climate and Numerical Weather Prediction Models from Data Simulated by a Cloud Resolving Model, Adv. Artif. Neural Syst., doi:10.1155/2013/485913, 2013.

Kruk, C. and Segura, A. M.: The habitat template of phytoplankton morphology-based functional groups, Hydrobiologia, 698(1), 191–202, doi:10.1007/s10750-012-1072-6, 2012.

Kruk, C., Peeters, E. T. H. M., Nes, E. H. V., Huszar, V. L. M., Costa, L. S. and Scheffer, M.: Phytoplankton community composition can be predicted best in terms of morphological groups, Limnol. Oceanogr., 56(1), 110–118, doi:10.4319/lo.2011.56.1.0110, 2011.

Ku, T.-L., Luo, S., Kusakabe, M. and Bishop, J. K. B.: 228Ra-derived nutrient budgets in the upper equatorial Pacific and the role of “new” silicate in limiting productivity, Deep Sea Res. Part II Top. Stud. Oceanogr., 42(2), 479–497, doi:10.1016/0967-0645(95)00020-Q, 1995.

Lek, S., Delacoste, M., Baran, P., Dimopoulos, I., Lauga, J. and Aulagnier, S.: Application of neural networks to modelling nonlinear relationships in ecology, Ecol. Model., 90(1), 39–52, doi:10.1016/0304-3800(95)00142-5, 1996a.

Lek, S., Belaud, A., Baran, P., Dimopoulos, I. and Delacoste, M.: Role of some environmental variables in trout abundance models using neural networks, Aquat. Living Resour., 9(1), 23–29, doi:10.1051/alr:1996004, 1996b.

Longhurst, A., Sathyendranath, S., Platt, T. and Caverhill, C.: An estimate of global primary production in the ocean from satellite radiometer data, J. Plankton Res., 17(6), 1245–1271, doi:10.1093/plankt/17.6.1245, 1995.

Löptien, U. and Dietze, H.: Reciprocal bias compensation and ensuing uncertainties in model-based climate projections: pelagic biogeochemistry versus ocean mixing, Biogeosciences, 16(9), 1865–1881, doi:https://doi.org/10.5194/bg-16-1865-2019, 2019.

Maldonado, M. T. and Price, N. M.: Influence of N substrate on Fe requirements of marine centric diatoms, Mar. Ecol. Prog. Ser., 141, 161–172, doi:10.3354/meps141161, 1996.

Martin, J. H.: Glacial-interglacial CO2 change: The Iron Hypothesis, Paleoceanography, 5(1), 1–13, doi:10.1029/PA005i001p00001, 1990.

Martin, J. H. and Fitzwater, S. E.: Iron deficiency limits phytoplankton growth in the north-east Pacific subarctic, Nature, 331(6154), 341–343, doi:10.1038/331341a0, 1988.

Mattei, F., Franceschini, S. and Scardi, M.: A depth-resolved artificial neural network model of marine phytoplankton primary production, Ecol. Model., 382, 51–62, doi:10.1016/j.ecolmodel.2018.05.003, 2018.

Moore, C. M., Mills, M. M., Arrigo, K. R., Berman-Frank, I., Bopp, L., Boyd, P. W., Galbraith, E. D., Geider, R. J., Guieu, C., Jaccard, S. L., Jickells, T. D., La Roche, J., Lenton, T. M., Mahowald, N. M., Marañón, E., Marinov, I., Moore, J. K., Nakatsuka, T., Oschlies, A., Saito, M. A., Thingstad, T. F., Tsuda, A. and Ulloa, O.: Processes and patterns of oceanic nutrient limitation, Nat. Geosci., 6(9), 701–710, doi:10.1038/ngeo1765, 2013.

O’Gorman, P. A. and Dwyer, J. G.: Using Machine Learning to Parameterize Moist Convection: Potential for Modeling of Climate, Climate Change, and Extreme Events, J. Adv. Model. Earth Syst., 10(10), 2548–2563, doi:10.1029/2018MS001351, 2018.

Olden, J. D.: An artificial neural network approach for studying phytoplankton succession, Hydrobiologia, 436(1), 131–143, doi:10.1023/A:1026575418649, 2000.

Özesmi, S. L. and Özesmi, U.: An artificial neural network approach to spatial habitat modelling with interspecific interaction, Ecol. Model., 116(1), 15–31, doi:10.1016/S0304-3800(98)00149-5, 1999.

Price, N. M., Andersen, L. F. and Morel, F. M. M.: Iron and nitrogen nutrition of equatorial Pacific plankton, Deep Sea Res. Part Oceanogr. Res. Pap., 38(11), 1361–1378, doi:10.1016/0198-0149(91)90011-4, 1991.

Rasp, S., Pritchard, M. S. and Gentine, P.: Deep learning to represent subgrid processes in climate models, Proc. Natl. Acad. Sci., 115(39), 9684–9689, doi:10.1073/pnas.1810286115, 2018.

Rivero-Calle, S., Gnanadesikan, A., Castillo, C. E. D., Balch, W. M. and Guikema, S. D.: Multidecadal increase in North Atlantic coccolithophores and the potential role of rising CO2, Science, 350(6267), 1533–1537, doi:10.1126/science.aaa8026, 2015.

Ryther, J. H. and Dunstan, W. M.: Nitrogen, Phosphorus, and Eutrophication in the Coastal Marine Environment, Science, 171(3975), 1008–1013, 1971.

Saito, M. A., Goepfert, T. J. and Ritt, J. T.: Some Thoughts on the Concept of Colimitation: Three Definitions and the Importance of Bioavailability, Limnol. Oceanogr., 53(1), 276–290, 2008.

Sarmiento, J. L., Slater, R., Barber, R., Bopp, L., Doney, S. C., Hirst, A. C., Kleypas, J., Matear, R., Mikolajewicz, U., Monfray, P., Soldatov, V., Spall, S. A. and Stouffer, R.: Response of ocean ecosystems to climate warming, Glob. Biogeochem. Cycles, 18(3), doi:10.1029/2003GB002134, 2004.

Scardi, M.: Artificial neural networks as empirical models for estimating phytoplankton production, Mar. Ecol. Prog. Ser., 139(1/3), 289–299, 1996.

Scardi, M.: Advances in neural network modeling of phytoplankton primary production, Ecol. Model., 146(1), 33–45, doi:10.1016/S0304-3800(01)00294-0, 2001.

Scardi, M. and Harding, L. W.: Developing an empirical model of phytoplankton primary production: a neural network case study, Ecol. Model., 120(2), 213–223, doi:10.1016/S0304-3800(99)00103-9, 1999.

Schmidhuber, J.: Deep learning in neural networks: An overview, Neural Netw., 61, 85–117, doi:10.1016/j.neunet.2014.09.003, 2015.

Schoffman, H., Lis, H., Shaked, Y. and Keren, N.: Iron–Nutrient Interactions within Phytoplankton, Front. Plant Sci., 7, doi:10.3389/fpls.2016.01223, 2016.

Vince, S. and Valiela, I.: The effects of ammonium and phosphate enrichments on clorophyll a, pigment ratio and species composition of phytoplankton of Vineyard Sound, Mar. Biol., 19(1), 69–73, doi:10.1007/BF00355422, 1973.

Wang, W.-X. and Dei, R. C. H.: Biological uptake and assimilation of iron by marine plankton: influences of macronutrients, Mar. Chem., 74(2), 213–226, doi:10.1016/S0304-4203(01)00014-7, 2001.

Wong, C. S. and Matear, R. J.: Sporadic silicate limitation of phytoplankton productivity in the subarctic NE Pacific, Deep Sea Res. Part II Top. Stud. Oceanogr., 46(11), 2539–2555, doi:10.1016/S0967-0645(99)00075-2, 1999.

Tables

Table 1: Scenario 1 comparison of MLR, RF, and NNE method performance for the training and testing sets.

	Training Data			Testing Data		
	R-squared	MSE	RMSE	R-squared	MSE	RMSE
MLR	0.4141	1.09×10^{-14}	1.05×10^{-7}	0.4092	1.10×10^{-14}	1.05×10^{-7}
RF	0.9988	2.53×10^{-17}	5.03×10^{-9}	0.9977	5.00×10^{-17}	7.07×10^{-9}
NNE	0.9998	3.18×10^{-18}	1.78×10^{-9}	0.9998	3.19×10^{-18}	1.79×10^{-9}

Table 2: Scenario 2 comparison of MLR, RF, and NNE method performance for the training and testing sets. Each method was trained and tested on the daily, weekly, and monthly time-averaged apparent relationship data.

		Training Data			Testing Data		
		R-squared	MSE	RMSE	R-squared	MSE	RMSE
Daily	MLR	0.3312	4.77×10^{-15}	6.90×10^{-8}	0.3254	4.84×10^{-15}	6.96×10^{-8}
	RF	0.9847	1.12×10^{-16}	1.06×10^{-8}	0.9695	2.22×10^{-16}	1.49×10^{-8}
	NNE	0.9707	2.09×10^{-16}	1.45×10^{-8}	0.9700	2.15×10^{-16}	1.47×10^{-8}
Weekly	MLR	0.3170	4.39×10^{-15}	6.63×10^{-8}	0.3172	4.35×10^{-15}	6.60×10^{-8}
	RF	0.9842	1.04×10^{-16}	1.02×10^{-8}	0.9699	1.94×10^{-16}	1.39×10^{-8}
	NNE	0.9695	1.96×10^{-16}	1.40×10^{-8}	0.9702	1.90×10^{-16}	1.38×10^{-8}
Monthly	MLR	0.3122	4.13×10^{-15}	6.42×10^{-8}	0.3230	4.06×10^{-15}	6.37×10^{-8}
	RF	0.9863	8.45×10^{-17}	9.19×10^{-9}	0.9737	1.60×10^{-16}	1.26×10^{-8}
	NNE	0.9732	1.61×10^{-16}	1.27×10^{-8}	0.9732	1.61×10^{-16}	1.27×10^{-8}

Table 3: Scenario 3 comparison of MLR, RF, and NNE method performance for the training and testing sets.

	Training Data			Testing Data		
	R-squared	MSE	RMSE	R-squared	MSE	RMSE
MLR	0.0672	6.51×10^{-16}	2.55×10^{-8}	0.0691	6.39×10^{-16}	2.53×10^{-8}
RF	0.9727	2.02×10^{-17}	4.49×10^{-9}	0.9445	3.92×10^{-17}	6.26×10^{-9}
NNE	0.9417	4.07×10^{-17}	6.38×10^{-9}	0.9386	4.22×10^{-17}	6.50×10^{-9}

Table 4: Estimated half-saturation coefficients using NNEs for each scenario and nutrient/light.

			NNE		
			Macronutrient	Micronutrient	Light
True Value			1.00 x 10^{-7}	2.00 x 10^{-10}	34.30
Scenario 1		25th Percentile	6.80 x 10^{-9}	-5.55 x 10^{-11}	34.05
		50th Percentile	1.06 x 10^{-8}	1.31 x 10^{-10}	34.89
		75th Percentile	1.91 x 10^{-8}	2.54 x 10^{-10}	34.23
Scenario 2	Daily	25th Percentile	1.03 x 10^{-8}	-1.13 x 10^{-10}	26.73
		50th Percentile	3.22 x 10^{-8}	1.78 x 10^{-10}	27.97
		75th Percentile	3.35 x 10^{-8}	9.55 x 10^{-10}	20.98
	Weekly	25th Percentile	6.99 x 10^{-9}	-1.15 x 10^{-10}	30.17
		50th Percentile	3.21 x 10^{-8}	1.87 x 10^{-10}	26.26
		75th Percentile	5.05 x 10^{-8}	8.33 x 10^{-10}	24.63
	Monthly	25th Percentile	7.70 x 10^{-9}	-1.35 x 10^{-10}	27.32
		50th Percentile	3.16 x 10^{-8}	2.01 x 10^{-10}	20.97
		75th Percentile	7.39 x 10^{-8}	1.09 x 10^{-9}	22.19
Scenario 3		25th Percentile	3.50 x 10^{-8}	-2.11 x 10^{4}	1.85
		50th Percentile	8.89 x 10^{-8}	6.94 x 10^{-10}	5.80
		75th Percentile	1.64 x 10^{-7}	2.41 x 10^{-9}	7.78

Figures

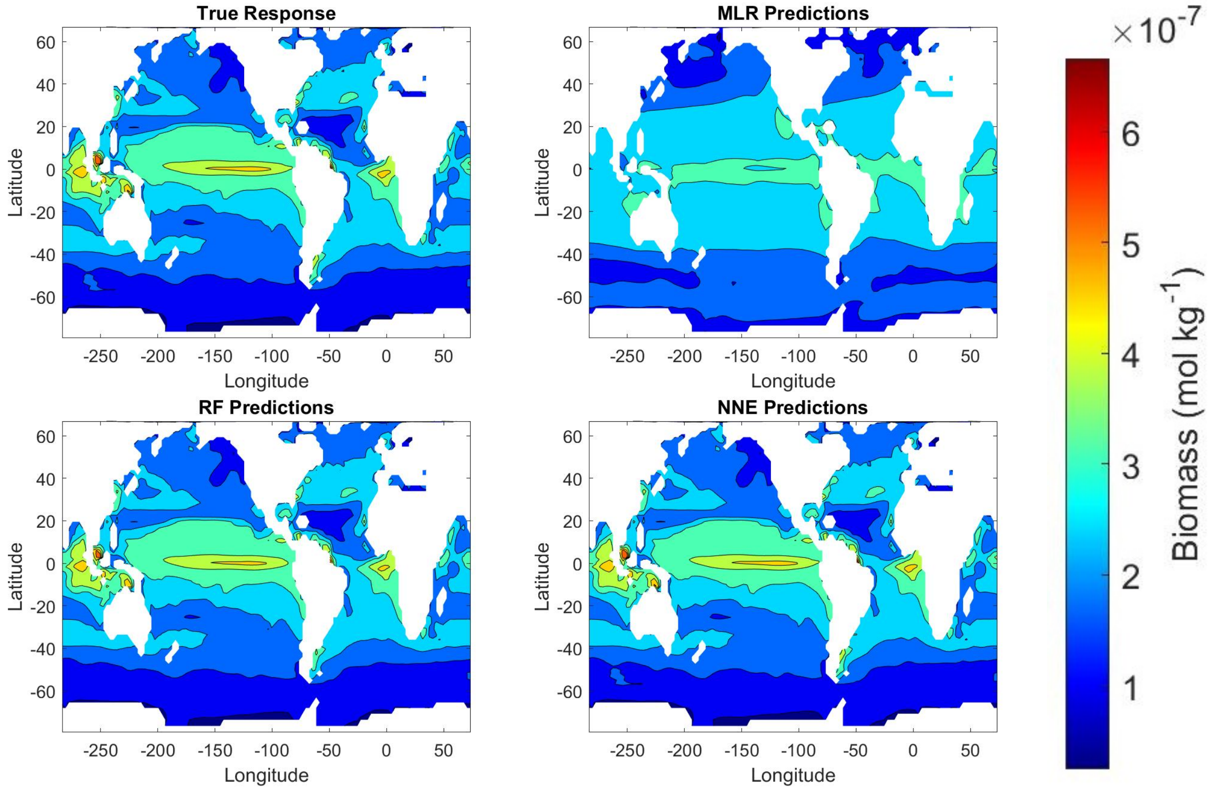


Figure 1: Contour plots comparing the true response for the yearly-averaged biomass (top left) and the associated predictions for MLR (top right), RF (bottom left), and NNE (bottom right).

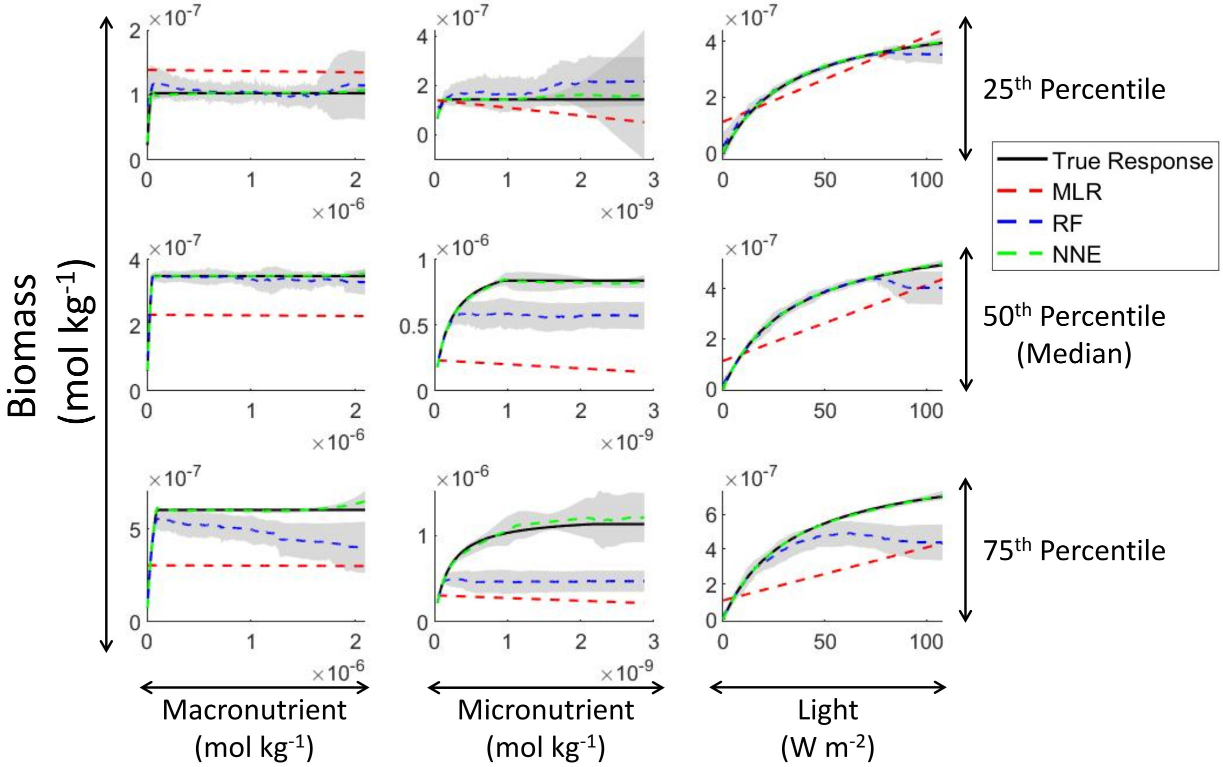


Figure 2: Sensitivity analysis for Scenario 1 with the columns corresponding to the predictors and the rows corresponding with the percentile value at which the other predictors were set. The black line shows the true intrinsic relationship and the dashed lines show the predicted apparent relationships for each method.

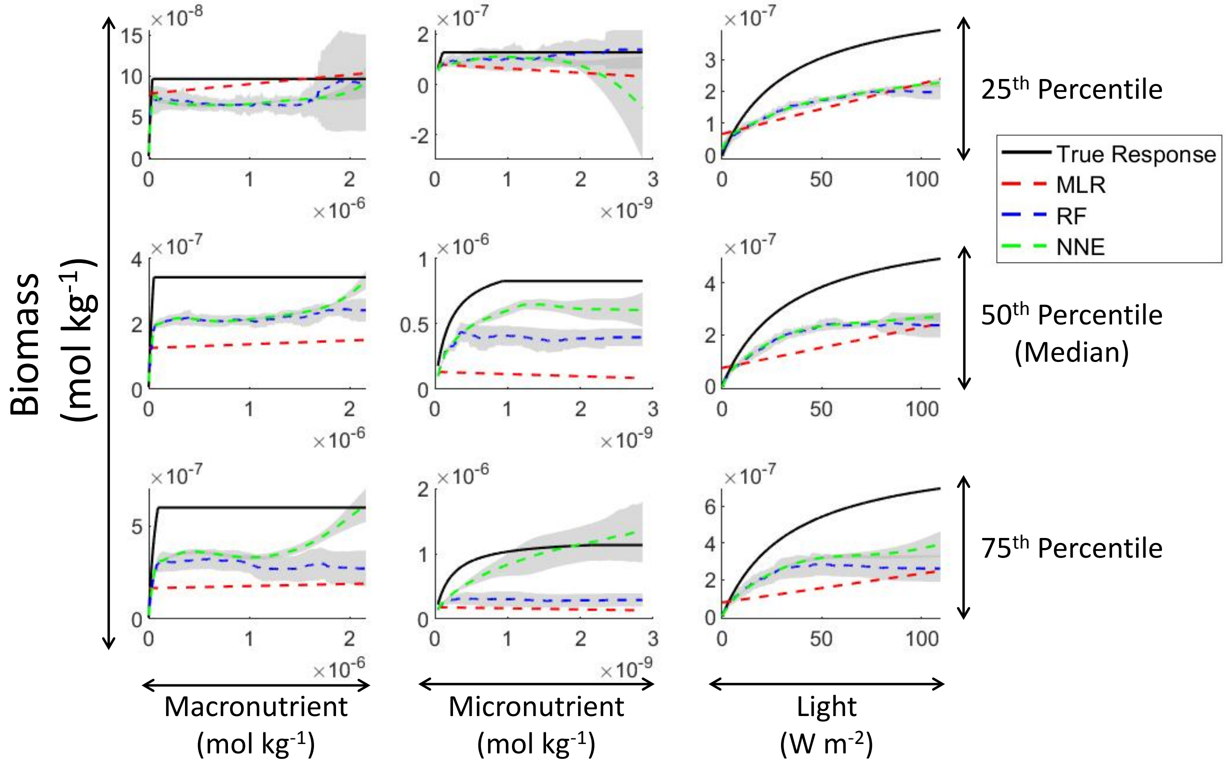


Figure 3: Sensitivity analysis for Scenario 2 with the columns corresponding to the predictors and the rows corresponding with the percentile value at which the other predictors were set. The black line shows the true intrinsic relationship and the dashed lines show the predicted **monthly** apparent relationships for each method.

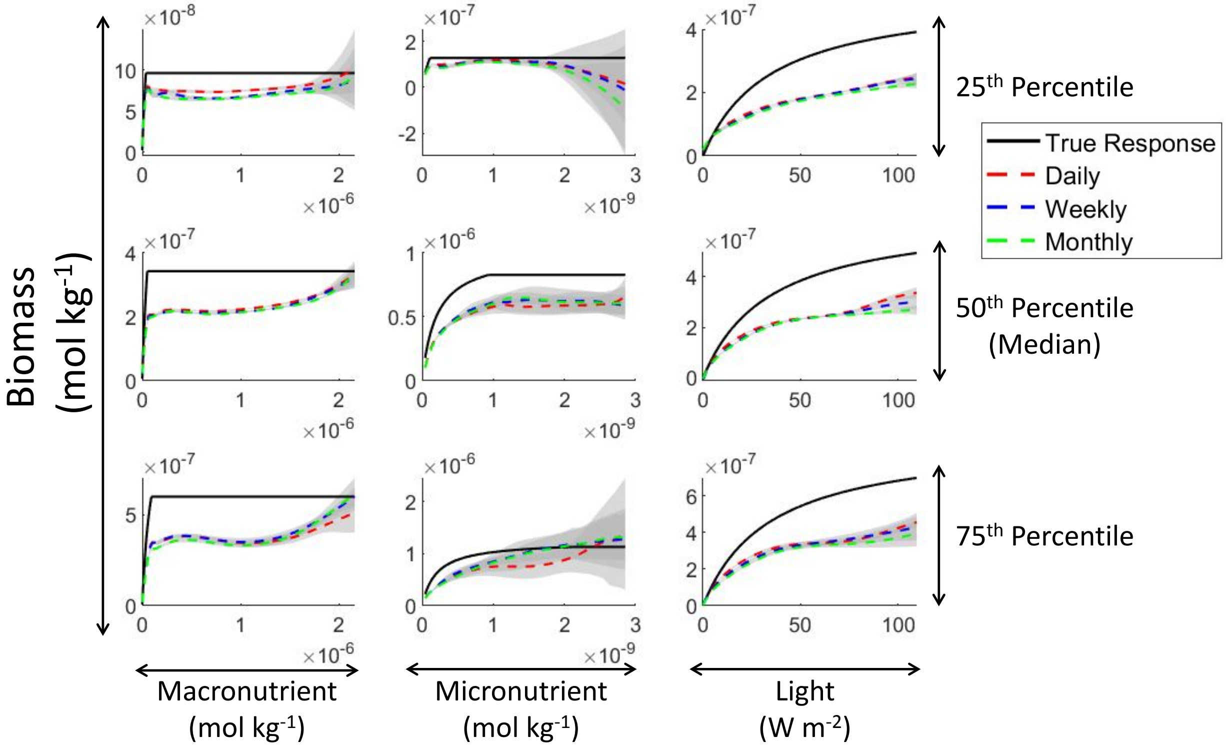


Figure 4: Sensitivity analysis for Scenario 2 with the columns corresponding to the predictors and the rows corresponding with the percentile value at which the other predictors were set. The black line shows the true intrinsic relationship and the dashed lines show the predicted apparent relationships for the NNEs corresponding to the daily, weekly, and monthly timescales.

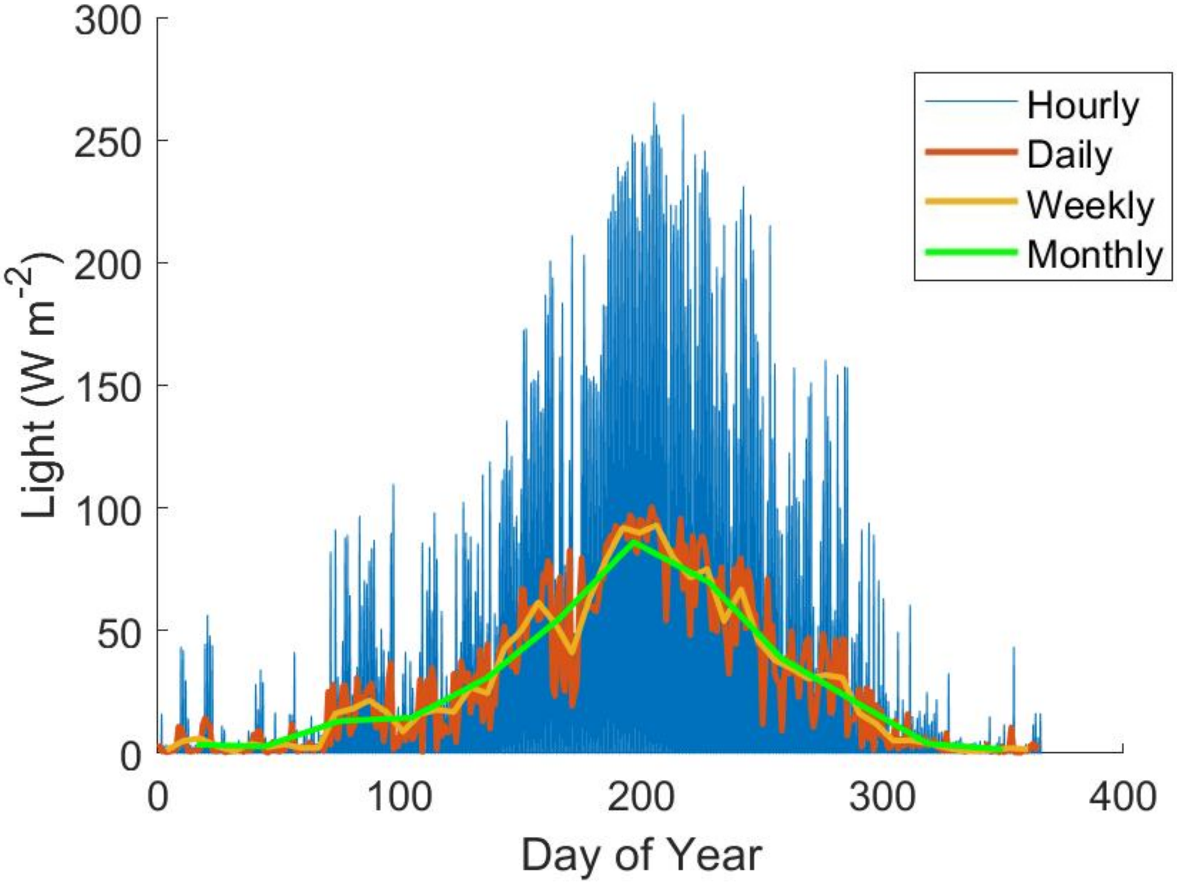
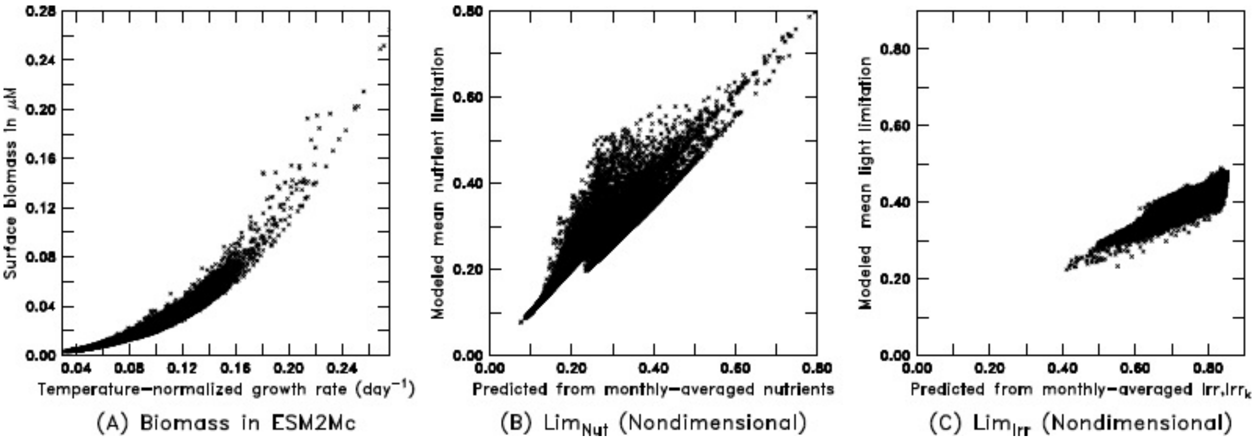


Figure 5: Line plot showing the differences in light levels for a point in the North Atlantic (39.08°N 40.5°W) for the various timescales in Scenario 2.



(A) Biomass in ESM2Mc

(B) $\mathrm{Lim}_{\mathrm{Nut}}$ (Nondimensional)

(C) $\mathrm{Lim}_{\mathrm{Irr}}$ (Nondimensional)

Figure 6: Scatter plots from the BLING model (a: surface biomass vs. temperature-normalized growth rate; b: mean nutrient limitation vs. monthly-averaged nutrients; c: mean light limitation vs. monthly-averaged Irr, Irr_k).

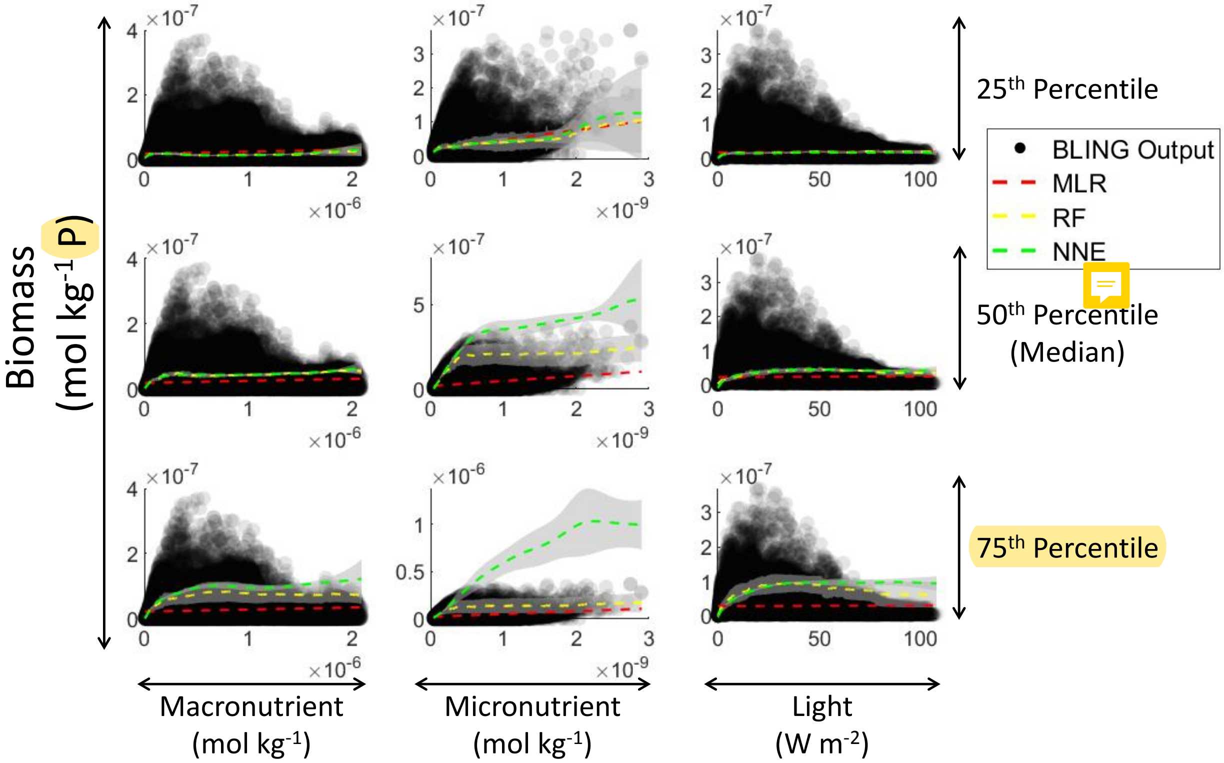


Figure 7: Sensitivity analysis for Scenario 3 with the columns corresponding to the predictors and the rows corresponding with the percentile value at which the other predictors were set. The gray circles show the observations from the BLING model and the dashed lines show the predicted apparent relationships for each method.

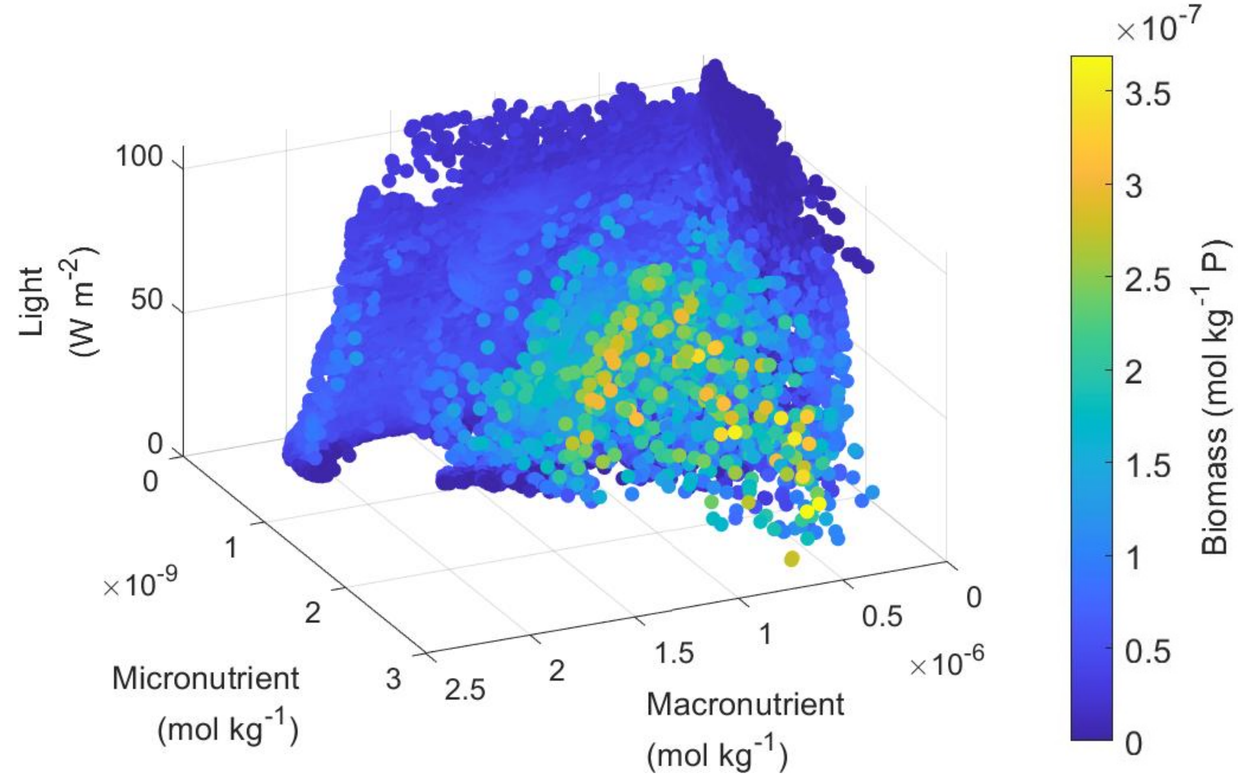


Figure 8: A 3-D scatter plot showing the concentrations from Scenario 3 for the macronutrient, micronutrient, and light with the color of the data points corresponding to the biomass concentrations.