Computational Modeling of Phytoplankton Dynamics with Climatic and Ecological Implications

Grant Proposal

Abhinav K. Sharma

Massachusetts Academy of Math and Science at Worcester Polytechnic Institute

STEM with Scientific and Technical Writing

Dr. Kevin Crowthers

29 January 2024

85 Prescott St. Worcester, Massachusetts 01605

Abstract

Phytoplankton lie at the base of marine food webs and are major regulators of climate and biogeochemical cycling, accounting for over half of primary production and the absorption of 30% of anthropogenic carbon emissions. However, changes in ocean conditions and the biological responses and preferences of phytoplankton are highly heterogeneous. Therefore, to better understand the impact of global-warming induced oceanic changes on phytoplankton populations, this study aims to develop a series of computer models that simulate the causes and effects of changing phytoplankton dynamics (e.g., biomass, phenology, etc.). The first model incorporates multiple environmental factors (e.g., temperature, pH, salinity, etc.), that impact phytoplankton dynamics. From this, driving parameters, as well as inter-parameter relationships and the impacts phytoplankton populations experience are determined. Using these results, the next series of models delineates climatic and ecological ramifications, using neural networks and existing climate models respectively. To begin this process, preliminary parameter analysis was performed using an EPA dataset. Here, the homogeneity of biological preferences was tested among fifty-seven genera from the Eastern United States. One-way ANOVA and Post-Hoc Tukey test results indicate that mean biological preferences for temperature, turbidity, dissolved oxygen, and pH do not vary significantly, while the rest of the thirteen variables tested do ($\alpha = 0.05$, $p < 0.0001$). These results serve as a sample for informing parameter selection for initial model development and validation. More broadly, the results of this entire apparatus could serve as a valuable decision-making tool for policymakers with regards to water body management.

Keywords: Phytoplankton Dynamics, Computational Modeling, Global Warming, Climate Model, Neural Networks

Computational Modeling of Phytoplankton Dynamics with Climatic and Ecological Implications

Grant Proposal

How can the complex, multifarious impacts of global-warming induced changes in oceanic conditions on phytoplankton be understood? How shall these changes to phytoplankton impact global climate and marine ecosystems? By developing a series of computational models that incorporates parametric data on environmental factors, utilizes statistical and computational techniques for validation and result determination, and harnesses the resulting impacts on phytoplankton for forecasting purposes, driving factors of phytoplankton dynamics as well as climatic and ecological implications can be identified.

An Introduction to Phytoplankton

Phytoplankton encompass a broad range of aquatic, microscopic, photosynthetic species of viruses, bacteria, fungi, protists, animals, and archaea. They are responsible for about half of all global primary production, the production of nutritional organic matter from inorganic compounds via photosynthesis and other metabolic processes (Käse & Geuer, 2018). Phytoplankton are key to biogeochemical cycling, circulating nitrogen, phosphorus, silica, and other micronutrients (Sarker et al., 2023). They also absorb 30% of anthropogenic carbon emissions (Rohr et al., 2023). Beyond photosynthesis, carbon sequestration is also performed through exportation, a process where, after death, cellular matter sinks to the ocean floor, forming carbon sinks. Phytoplankton regulate climate not only through controlling carbon circulation, but also through light reflection. Certain functional groups produce dimethylsulfoniopropiothetin, a complex, sulfur-containing molecule. This compound decomposes into dimethylsulfide, which in turn decomposes into compounds that reflect solar radiation (Deppeler & Davidson, 2017). It is believed that biochemical processes such as this one helped cause the first major ice ages on Earth (Käse & Geuer, 2018). Additionally, phytoplankton lie at the base of marine food chains, serving as prey for various species of zooplankton and fish (Käse & Geuer, 2018; Loschi et al., 2023). Therefore, phytoplankton are an integral part of the global climate and environmental systems, making the ability to understand how their operations and functionalities are to change because of global warming incredibly crucial.

Understanding The Impact of Global-Warming Induced Oceanic Changes on Phytoplankton

With that in mind, the impact global warming has had on oceanic conditions themselves must first be considered. Climate change has left oceans warmer, more acidic, anoxic, and stratified. Sea levels are rising, while salinity and micronutrient concentrations are losing uniformity. Moreover, ocean currents have begun to slow down (Berwyn, 2018). The thermohaline cycle involves the cycling of warmer, fresher, and less dense pelagic (surface) water with colder, denser and saltier benthic (deep-sea) water. This process is crucial for mixing nutrients, distributing heat, and regulating climate. Analysis of past climate patterns indicates that a slower thermohaline cycle has been associated with more extreme climate patterns (Berwyn, 2018). However, it is important to note that changes in ocean conditions are not uniform. Rather, changes vary extensively by region (Winder & Sommer, 2012). That means environmental parameters that influence phytoplankton populations are not homogenous, adding a layer of complexity when determining the impacts they are to face.

Similarly, phytoplankton are undergoing some overarching changes. Common trends include shifting phenology, a change in preferences towards smaller, more buoyant cells, and poleward migration (Ratnarajah et al., 2023). However, under the surface, population modifications are far more complex. For example, certain groups are favored under eutrophic conditions, where excessive micronutrients foster unhealthy algal blooms that deplete ecosystem resources, whereas others are favored under fresher or darker conditions (Winder & Sommer, 2012). Moreover, there are a voluminous amount of environmental factors (e.g., light, heat, nutrients, pH, salinity, etc.) that impact phytoplankton dynamics, with each species operating under a different set of ideal conditions (Winder & Sommer, 2012).

This heterogeneity in traits and environmental conditions raises a dilemma. To illustrate this, consider two phytoplankton species living in the same area. Suppose that one species tolerates a pH range of 5.9 to 6.5, whereas another tolerates a range of 6.7 to 7.3. With ocean acidity changing heterogeneously, if one area of the ocean has a pH of 6, and another area a pH of 7, then each species would migrate to the area matching their respective preferences, heavily modifying taxonomic composition, biomass, exportation, and other dynamics. However, there are other influential environmental factors, making it important to consider how multiple factors simultaneously impact phytoplankton populations. For example, would another factor, such as dissolved oxygen, have precedent over pH when it comes to these species seeking ideal conditions? Additionally, these migrations would leave predators bereft of a major source of food. How would that impact the entire ecosystem? What climatic shifts may result? The circumstances and questions raised by a scenario like this capture the essence of what this study aims to address.

Phytoplankton Genomics and Examples of Parametric Variability

Parameters that influence phytoplankton are present at the molecular, genomic, cytological, and ecological level. Changes in their values can impact various important biological characteristics, including primary production and metabolic rates. For instance, biochemical processes like DNA methylation, whereby a methyl functional group is applied to the fifth carbon in the carbon ring of the nitrogenous base of cytosine, has been found to inhibit amino acid metabolism, as well as respiration and photosynthesis given warming oceans (Wan et al., 2023). This means that there is a slower rate of primary production and carbon sequestration, inhibiting upward movement of energy along the trophic pyramid and climate regulation. However, with ocean temperature changing heterogeneously, the extent to which this trend occurs shall vary.

Meanwhile, micronutrients also play a major role in influencing metabolic rates. For example, phosphorus is an integral component of all forms of metabolism, making phosphorus-containing compounds crucial for phytoplankton. However, varying levels of micronutrients, including these compounds, impact dynamics in different ways. Increased phosphorus levels can allow for all metabolic processes to occur at faster rates, enhancing phytoplankton’s ecological services. However, excessive phosphorus concentrations can be toxic and lead to eutrophication (Li et al., 2023). Moreover, tolerance towards toxicity and metabolic rates vary across different species.

Another example of significant environmental variability is water temperature. Different genera of phytoplankton exhibit different responses to warming ocean temperatures. For example, using a modified Eppley Curve, an exponential function that models the relationship between growth rates and water temperature, one analysis found that, while growth rates are expected to increase with temperature, the rate at which the growth rate increases for diatoms was greater than that of dinoflagellates, cyanobacteria, and coccolithophores (Anderson et al., 2023). Additionally, dissimilar thermal attributes are predicted to result in differential migration patterns among different functional groups.

These examples clearly illustrate that for any environmental parameter, there is a great amount of nuance when it comes to the impact that phytoplankton face. This nuance only expands when multiple variables are considered in tandem. It is extremely difficult to perform an experiment that involves multiple independent variables, as confounding factors would easily arise. The alternative would be to perform an experiment using only one variable, which would fail to account for the multifactor interactions that occur. The results of such a procedure across different instances would also vary, failing to paint a solid picture of the impact of that one parameter (Chang et al., 2022).

Computational Modeling of Phytoplankton Dynamics: Progress and Current Limitations

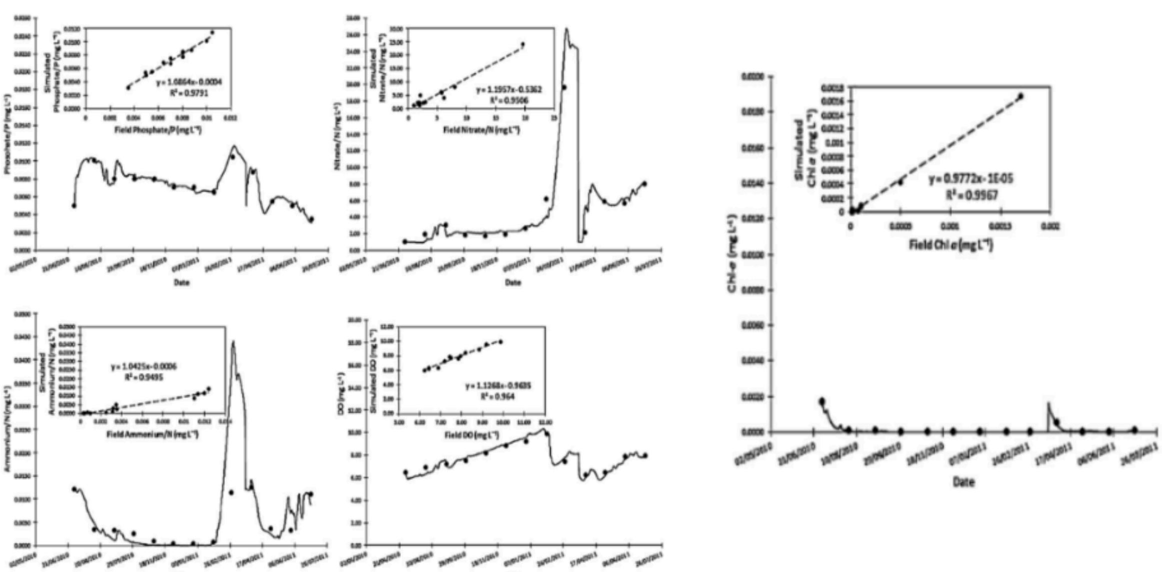
As a result, a computational modeling approach is imperative, as it can be used to capture the nuances of this situation, and provide greater insight into what the observed results signify. In essence, this is what the goal of this project is: to take the complex relationships in phytoplankton populations, and organize, synthesize, and contextualize them, delineating ramifications.

Presently, there are many limitations with computational models of phytoplankton dynamics. One major limitation is the misunderstanding of the role zooplankton play in the modeling process. Different models have made different assumptions about how zooplankton interact in ecological systems, leading to divergent predictions in climate and food web scenarios (Rohr et al., 2023). Indeed, it has been found that more robust data collection methods and raw data on zooplankton is necessary (Ratnarajah et al., 2023). It is a dearth in overall data that limits the predictive power of these computer models. There is a particular lack of data from the Southern hemisphere (Deppeler & Davidson, 2017).

That is not to say that accurate models have not been developed. In fact, there have been models developed for small bodies of water, such as the Tucuruí reservoir in Pará, Brazil (Deus et al., 2013). This computer model was based off of field data on chlorophyll a, dissolved oxygen, ammonia. Through linear regression analysis including R^2, root mean square error, and the slope of regression lines comparing computer predictions to actual results, it was determined that the model was in fact accurate. Figure 1 (Deus et al., 2013) depicts the linear regression between the predicted and field values of these parameters. With extremely high R^2 values, the model was deemed fit to perform other functions within study. This provides a strong example for how the accuracy in computer model predictions can be assessed, allowing for model results and ramifications to be validated. Indeed, validation relies on some form of statistical analysis, which varies from model to model.

Figure 1

An Example of Computational Model Validation Techniques: Tucuruí Reservoir as a Case Study



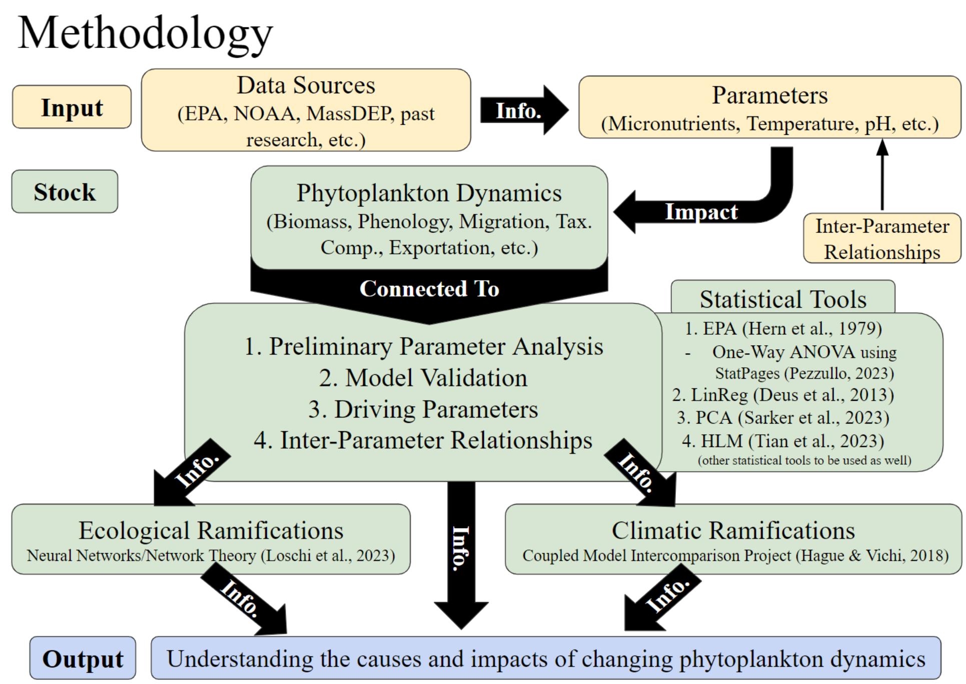
Note. Each parameter contains a larger graph depicting the raw comparison between field data and computer predictions. From lop left to bottom right, the parameters shown are phosphate, nitrate, ammonia, dissolved oxygen and chlorophyll a. Embedded within are the linear regressions that compare the computer model predictions against the actual field data. Therein lie the R^2 values which serve to evaluate model accuracy.

However, different models have been synthesized for different purposes. For example, some models have focused on the identification of driving parameters in phytoplankton dynamics. For instance, using Principal Component Analysis (PCA), whereby the impact of parameters is measured using vectors, one study of coastal Bangladesh found that salinity, followed by micronutrient concentrations, turbidity, and water temperature played the most significant roles in regulating abundance and spatial variability in phytoplankton (Sarker et al., 2023). Other models have focused on inter-parameter relationships. One study of several lakes in Wuhan, China used a hierarchical linear model. After sorting parameters into different levels and identifying statistically significant relationships, the one major inter-parameter relationship found was a negative one between grasslands and water temperature (Tian et al., 2023). From an ecological lens, neural networks have been developed to model the changing flow caused by changing phytoplankton conditions. At a broad level, these networks take in various rates related to energy and matter transfer as parameters, the values of which can be modified to simulate different scenarios. These factors can be gradually implemented through a series of successive neural networks. When applying this approach to Lake Constance in Central Europe, the fit of the model to predict observed dynamics was maximized, providing a format through which food webs of other systems can be created (Boit et al., 2012). Other studies, such as one of the Venice Lagoon, have identified keystone species (Loschi et al., 2023). From a climatic lens, a focus has been placed on the accuracy of climate models in predicting many attributes, including bloom phenology. The Coupled Model Intercomparison Project (CMIP), with its large scope, has been a particular area of focus. One study found that bloom phenology in the Southern ocean is not accurately predicted, as the sea ice concentration levels used in the model were not reflective of on-site levels (Hague & Vichi, 2018). As a whole, there is an ample amount of literature that describes a myriad of empirical relationships and computational models of the various aspects of the changing characteristics of phytoplankton as well as those ramifications. What is lacking, however, is a unified apparatus to unite these models.

Given the background information and limitations presented, this paper seeked to create a series of computational models bound together as one entire system whereby parametric information on phytoplankton populations could be introduced and results for dynamics, ecology and climate could be produced. The first step of this model involved the implementation of data from multiple sources, including the National Oceanic and Atmospheric Administration (NOAA), Environmental Protection Agency (EPA), past research, and other databases. In combining data from multiple sources, a repository of data on a global scale was forged, helping to rectify the issue of data availability previously described. Following the identification of driving parameters and inter-parameter relationships, the impact of parameters on phytoplankton dynamics (e.g., biomass, phenology, migratory patterns, etc.) was determined. These impacts, in turn, in addition to other metrics, served as parameters for the neural network of the food web and the climate model, providing implications in the respective areas. Figure 2 describes the overall methodology employed through the lens of a system diagram.

Figure 2

Systems Diagram of Study Methodology



Note. This system diagram provides an overview of the methodology. The input includes the data sources that provide relevant parametric information. This information feeds into the stock, which includes the impact on phytoplankton characteristics, and the various computational and statistical tools used for analysis. The output is this study's goal of modeling the causes and impacts of changing phytoplankton conditions.

This apparatus serves as a viable streamlined process that could be used by experts in the field to help study phytoplankton populations and their role in the environment and climate. Moreover, it has the potential to serve as a tool for policy makers with regards to marine water body management. For example, Tian et al. 2023 used results from a multi-agent based model to recommend a controlled increase in micronutrient concentrations and fish that feed exclusively on zooplankton (Tian et al., 2023). As a whole, this study has provided a potentially potent framework whereby the causes and impacts of phytoplankton conditions can be effectively observed.

Section II: Specific Aims

Specific Aim 1

Model validation through comparison of the constructed computer model with field data, using a widely known parameter.

Specific Aim 2

Utilize statistical tools to identify driving parameters in the observed phytoplankton dynamics. In addition, assess the factors that influence the driving parameters in different areas, and how all parameters impact one another to exert a net impact on phytoplankton throughout different global regions.

Specific Aim 3

Determine climatic and ecological ramifications. Construct a neural network for a food web containing phytoplankton genera and their predators. Climatic ramifications are to be modeled through testing climate scenarios that result from the results from Specific Aim 2 within the framework of current Climate models (i.e., CMIP5).

Expected Findings

The expected outcome of this work is a synthesis of findings of the complex factors influencing phytoplankton dynamics and their climatic and ecological ramifications. The model or models produced are expected to serve as important tools for decision-makers regarding water body management, ideally at a federal or even international level. Moreover, it is hoped that this model will provide unprecedented predictive power for hypothetical scenarios and the impact of specific combinations of parameters.

Section III: Project Goals and Methodology

Relevance/Significance

Since phytoplankton play a major role in biogeochemical cycling, climate regulation, and ecological stability, they are important facets of the global climate and environmental systems. Thus, capturing the complexity of the multifarious stressors they are facing is key to understanding how they will be impacted overall, and what that means for the environment and climate in turn. If developed to be accurate and efficient, this model would have significant ramifications for policymakers. Namely, since phytoplankton play a major role in the formation and changes of an algal bloom, in having a model to predict population characteristics given parametric data, policymakers would be able to make informed decisions about lake management. For instance, one study focusing on lake ecosystems in Wuhan, China devised an accurate computational model that was able to run hypothetical scenarios. From these scenarios, concrete policy recommendations were made, including control of the nutrient stoichiometry between nitrogen and phosphorus, as well as increasing the presence of exclusively zooplankton-feeding fish (Tian et al., 2023). If a model on a larger scale is successfully devised, then policymaking applications would be greater in magnitude, and more versatile to a wider range of different types of bodies of water. Carrying out this project would have significantly beneficial impacts both in science and policy.

Innovation

Presently, there exists a wide variety of computational techniques to model phytoplankton dynamics. However, different models have different focuses. Some computer models may focus on a specific body of water. For example, computational modeling has been used for phytoplankton dynamics both in the coastal waters of Bangladesh and Wuhan (Sarker et al., 2023; Tian et al., 2023). However, being two very different regions with different conditions, those models had very different structures and made use of very different techniques. Function is another major avenue through which models vary. For instance, while both examples aimed to make sense of the multiple stressors phytoplankton face by sifting through many environmental parameters, a model dedicated to depicting a food web between phytoplankton genera and predators would have significantly different architecture.

By seeking to specifically create a series of computational models, these heterogeneously designed computer algorithms may be brought under one central system. Since this project aims to model phytoplankton dynamics on a global scale, taking into account many parameters, the model shall be highly versatile and adaptable to the unique circumstances of a given aquatic ecosystem. Multiple different purposes, including identifying driving parameters, differentiating impact among various species, and projecting climatic and environmental implications, can also be achieved with this model. Hence, by consolidating many important empirical functions through a large scope, this project shall produce a unique way to computationally model phytoplankton dynamics.

Methodology

Specific Aim #1

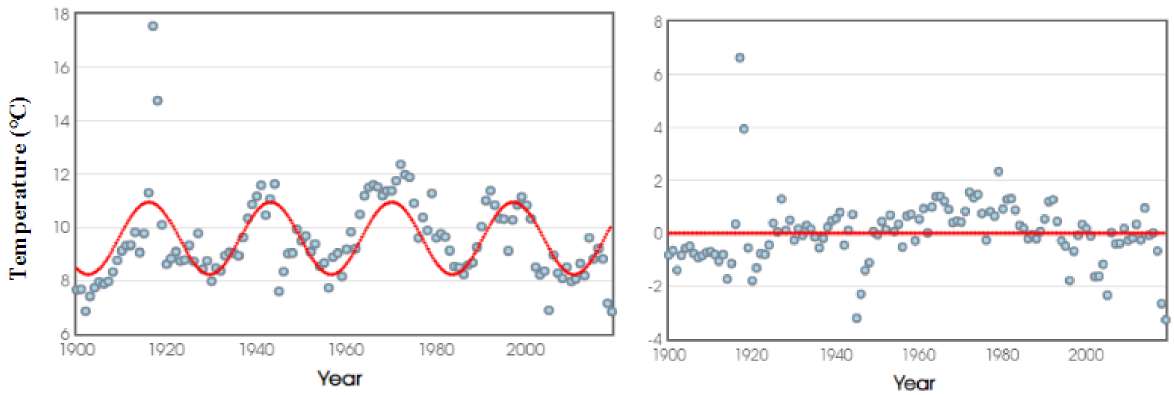
The first task in constructing the computational model shall be validation of the model. Validation shall be achieved through comparing computer model predictions with field data using a widely known parameter.

Justification and Feasibility. The findings of any computational model offer no value unless it is ensured that the model is accurate. Therefore, devising some plan to implement statistical techniques that verify model validity shall be imperative. This approach is highly feasible as there exists a wide variety of software tools on which a myriad of statistical algorithms may be run. For example, the above mentioned study regarding the Tucuruí reservoir used three metrics, including root means square error, R^2, and the slope of the least squares regression lines, to evaluate model accuracy (Deus et al., 2013). It is likely that other tools, such as a Student's t-test, among other hypothesis tests, may be used to assess the significance of the differences in observed and computer data.

Summary of Preliminary Data. The primary dataset to be analyzed in this study is the 2018 World Ocean Database (WOD18) provided by the National Oceanic and Atmospheric Administration (NOAA). Both spatially and temporally, this dataset provides a highly cosmopolitan measurement of numerous environmental parameters, including water temperature, micronutrients, pH, salinity, among many others (Boyer et al., 2018). A time series of each parameter within the dataset was created. Figure 3 illustrates the times series developed for average global oceanic temperature, given data spanning from 1900 to 2019. This specific time series is modeled using sinusoidal regression.

Figure 3

Time Series of Average Global Oceanic Temperature (℃) with Residual Plot



Note. R^2 = 0.3868 (r = 0.6219). Temperature = 1.3548 * sin(0.2327(Year) + 1.7855) + 9.5894

Present model fitness appears to be rather mild (R^2 = 0.3868, Figure 3). A major factor that could have reduced model fitness were two abnormally high ocean temperatures observed during 1917 and 1918. Indeed, these years had a smaller amount of samples taken, possibly causing this skew. In future iterations, it may be necessary to remove outliers to improve model fitness. With a midline and amplitude of 9.5894℃ and 1.3548℃ respectively, the model predicts that average global oceanic temperature fluctuates between 8.23℃ and 10.94℃. This occurs over a period of about twenty-five years (Figure 3).

Expected Outcomes. It is expected that once the statistical tool or tools used to validate the model is applied, insight into model validity can be attained. For example, as explained with Figure 1 (Deus et al., 2013) in the introduction, conclusions about model accuracy could be reached using R^2. Thus, it is expected that the metric or metrics within the tool chosen for model validation be indicators of model validity.

Potential Pitfalls and Alternative Strategies. There exist a wide variety of techniques to ensure model validity. Each test offers a unique set of advantages and disadvantages with regards to the factors taken into account. For example, while the linear model can effectively use metrics such as R^2 to assess model validity, it is often subject to many confounding factors, which may inhibit its quality as a marker for computational accuracy. However, alternative statistical and computational tools may be used in conjunction with linear regression to account for factors that it fails to. Conversely, linear regression would be able to account for actors these other statistical tools might miss. With this understanding, an approach using multiple validation techniques shall be devised.

Specific Aim #2

Once the model is validated, implement more species and more parameters. Utilize statistical tools to identify driving parameters in the observed phytoplankton dynamics. In addition, assess the factors that influence the driving parameters in different areas, and how all parameters impact one another to exert a net impact on phytoplankton throughout different global regions.

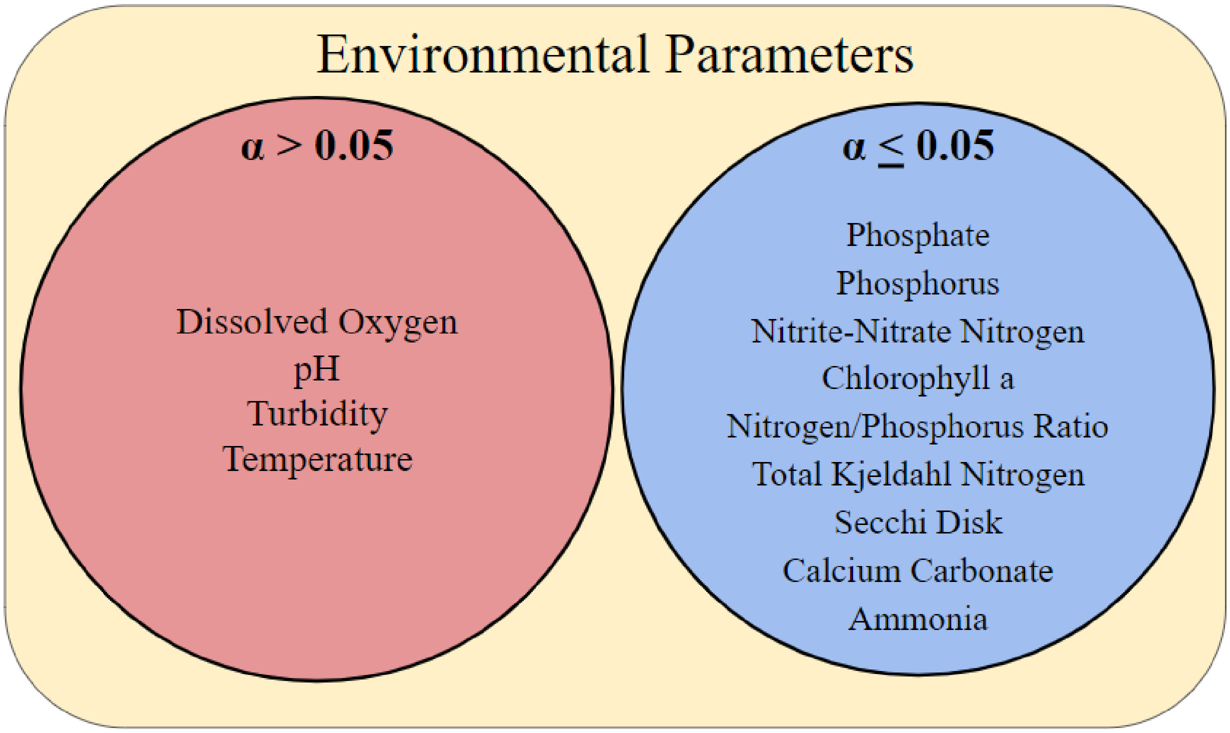
Justification and Feasibility. As previously discussed, there are numerous factors that influence phytoplankton dynamics. Experimenting with these parameters on phytoplankton simultaneously is not viable, as that would lead to too many confounding variables to be able to discern any meaningful relationship or trend. Hence, incorporating these factors into a computational model would help eliminate this ambiguity. To better understand the observed changes in phytoplankton dynamics, the driving factors behind these changes need to be understood. That way, any possible mitigation strategies to aid ecosystems can be created. Since computational modeling offers the ability to disentangle the driving parameters, their relationships, and the factors behind them, it is a viable approach for this project.

Summary of Preliminary Data. The first empirical procedure performed for this project involved performing a series of One-Way ANOVA tests and Post-HOC Tukey tests on a dataset attained from the EPA. The aim was to determine, among the fifty-seven genera listed in the sample, whether there were any differences in mean parameter values. For example, one of the thirteen parameters tested was mean water temperature for each genera. The above-mentioned statistical tests were then used to test if mean temperature values found in each genera significantly varied, and where (that is, which specific genera) that variation could be found. Findings indicate that mean factor values vary significantly among genera, while others do not. These findings may be useful for making decisions in initial model development.

A dataset from the EPA was attained (Hern et al., 1979). The data used was taken from Table 3 of the original paper. The original dataset was reorganized such that each genus was listed in alphabetical order, followed by their frequency of observation and mean parameter values. Sample standard error was attained by dividing each mean parameter value for each genus by the square root of the sample size. Then, the data was further reorganized into multiple sheets for each parameter. Using an online software, one-way ANOVA, along with post-hoc Tukey tests were performed using summary statistics (sample size, mean, standard error). This software not only produced individual pair results for the post-hoc tests, but it also provided a series of 95% confidence intervals for the spread of the environmental factor for each genus of phytoplankton, as well as ANOVA tables (Pezzullo, 2023).

The results indicate that inter-genera preferences do not significantly vary for dissolved oxygen, temperature, pH and turbidity, whereas all other parameters do (Figure 4). Among the parameters tested, there exists a gradient of heterogeneity in the form of differential significance values as well as differential proportions of significantly different Post-Hoc Tukey pairs (Table 1). The homogenous parameter identified may be ideal for initial model development, as homogeneity in preferences could correlate homogeneity in empirical data. More consistent numbers shall make initial model validation a smoother process.

Figure 4
Assortment of Insignificant and Significant Parameters



Note. The yellow box is used to denote the set of all environmental parameters. Within this sampling space, the red and blue circles depict the insignificant and significant parameters, respectively ($\alpha = 0.05$).

Table 1

One-Way ANOVA and Post-Hoc Tukey Test Results for Parameters

Parameter	**Significance Level For One-Way ANOVA**	**Proportion of Significantly Different Pairs ($\alpha = 0.05$) From Post-HOC Tukey Tests**
Phosphate (µg/L)	0	0.5432
Phosphorus (µg/L)	2.11E-220	0.4467
Nitrite-Nitrate Nitrogen (µg/L)	2.04E-192	0.3690
Chlorophyll a (µg/L)	6.53E-163	0.3766
Nitrogen/ Phosphorus Ratio	3.13E-128	0.3045
Total Kjeldahl Nitrogen (µg/L)	1.05E-64	0.1823
Secchi Disk (in.)	9.25E-42	0.1397
Calcium Carbonate (µg/L)	7.72E-31	0.0877
Ammonia (µg/L)	2.59E-06	0.0025
Temperature (℃)	8.84E-01	0.0000
Turbidity (% Trans.)	9.16E-01	0.0000
DO (µg/L)	1.00E+00	0.0000
pH	1.00E+00	0.0000

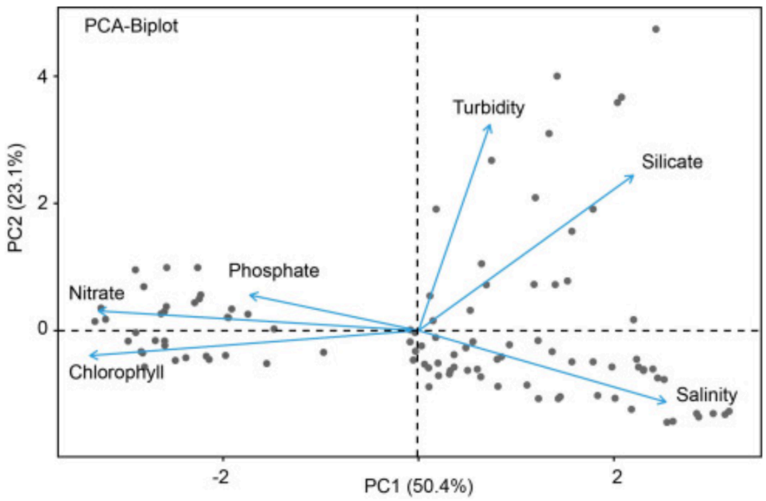
Note. Parameters are listed top to bottom in order of decreasing significance level (greater p-value). A color gradient from blue to red accompanies this progression. Since fifty-seven genera were tested, there were 1,596 unique pairs to compare in the Post-HOC Tukey tests. The proportions in the rightmost column are thus out of 1,596. To aid in clarity, a visual was provided at the left of each parameter.

This data and analysis have a few limitations given that the timeframe of when the data were collected (1970s), and that the scope was small, only focusing on fifty-seven genera in the Eastern United States. However, this data serves as a solid preliminary investigation of what parameters to use for initial model development and validation. To these ends, parameters found to be homogenous are likely to be better candidates than those that are heterogenous. This is because having less varied data initially will make model construction and validation easier. Moreover, homogenous factors, when changed, will likely impact a wider range of phytoplankton in a more consistent way than factors that vary. Hence, preliminary predictive modeling will also be easier using such parameters.

Expected Outcomes. Through the use of principal component analysis (PCA), as well as a hierarchical linear model of environmental parameters, driving parameters, as well as their relationship with one another, and the factors that influence them can be detected. Using PCA, multiple different parameters can be condensed into vectors in order to simplify the complex, multi-variable dynamics of a dataset. In doing so, the magnitudes of the vectors can be analyzed, and from there, the driving parameters can be identified. For example, Figure 5 (Sarker et al., 2023) depicts a PCA performed on environmental parameters measured for the impact on phytoplankton for coastal Bangladesh. Turbidity, silicate, and salinity, with the largest magnitude, were identified as the most important driving parameters in the study. PCA is a robust example of how driving parameters may be identified.

Figure 5

An Example of Using Principal Component Analysis to Identify Driving Factors



Note. This graph depicts the PCA performed on the seven parameters measured in coastal Bangladesh. Each vector describes the extent to which each parameter impacts the observed phytoplankton dynamics.

Using a hierarchical linear model (HLM) can help identify relationships between parameters. An HLM is a type of multivariable regression that splits up parameters into different levels. For example, for the model created for phytoplankton in Wuhan, China, the broader, ecological parameters were placed at level 2 within the hierarchy, whereas, smaller-scale physicochemical parameters were placed at level 1. This neat assortment of factors allowed the authors to determine the driving parameters by using t-tests for linear regression between the level 1 parameters and the phytoplankton dynamics (which, in this study, was primary production), level 2 and primary production, as well as between levels 2 and 1. It is this comparison that takes place between level 1 and 2 (and higher levels, if used) that allows parameter relationships, and from there, the factors behind the driving parameters, to be determined (Tian et al., 2023). Although, it is important to understand that the factors behind the driving parameters are not deduced by statistics, but rather, by the science of the situation. For example, since the Tucuruí reservoir was man-made, the hydrodynamical forces causing the inflow and outflow of water into the reservoir was what drove phytoplankton population changes. The man-made artificiality of the dam made the rapid inflow and outflow the driving parameter (Deus et al., 2013). This simple deduction was done using analysis of the environment rather than statistics. The point, then, of using an HLM, is to help sort parameters and identify their significance in an organized manner, contextualizing the factors behind driving parameters. In the context of this study, it was found that nitrogen, phosphorus, water temperature, and the *Rotifera* genus of zooplankton were driving parameters (Tian et al., 2023).

Potential Pitfalls and Alternative Strategies. HLMs and PCAs are only two techniques that can disentangle the driving parameters of phytoplankton dynamics. HLMs pose an advantage in that they can compartmentalize, and consequently, draw conclusions not only about what factors are driving parameters, but relationships that lie therein. However, it is harder to integrate a very large number of parameters into an HLM. Conversevely, although PCA does not compartmentalize factors, its vectors provide a more concise and visual way of understanding which parameters are significant. Similar to Specific Aim #1, great care in maximizing the use of statistical tools to achieve the specified end must be taken. Consequently, overall error is minimized as each test collectively contributes information about all relevant metrics.

Specific Aim #3

Determine climatic and ecological ramifications. Construct a neural network for a food web containing phytoplankton genera and their predators. Climatic ramifications are to be modeled through testing climate scenarios that result from the results from Specific Aim 2 within the framework of current Climate models (i.e., CMIP5).

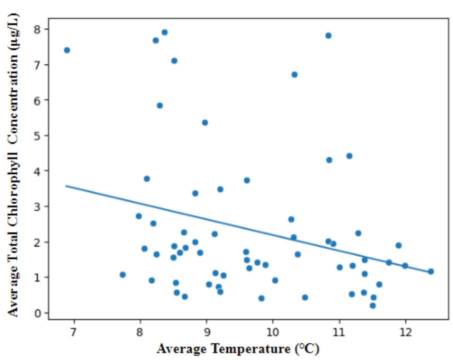
Justification and Feasibility. At this point, the next step would be to determine what impacts the changes phytoplankton face would have on marine ecosystems as well as the climate. Knowing this information is crucial, given the major role phytoplankton play in biogeochemical cycling, marine food webs, and climate regulation. Understanding what those ramifications are is important. There have been past instances where neural networks and network theory principles have been utilized in constructing and analyzing food webs with phytoplankton. For example, one study done off the coast of Venice was able to successfully model a food web of the area. Each node was assigned a set of parameters, most of which related to the transfer of energy and organic matter among predators, prey, decomposer, and other ecological players. The authors were able to model how energy transfer was changed due to environmental parameters, and how those changes impacted the ecosystem (Loschi et al., 2023).

Meanwhile, CMIP5, known as the Coupled Model Intercomparison Project, is a project that climatologists have used to aggregate data about climate change, as well as climate models. By using this interface to predict scenarios of phytoplankton-induced changes to climate, climatic ramifications can be discerned.

Summary of Preliminary Data. In performing data analysis on the data from WOD18, total chlorophyll concentration was identified as an indicator for primary production in phytoplankton. This is because chlorophyll is a crucial pigment for photosynthesis, which in turn allows for the occurrence of all other metabolic processes. As a preliminary step towards identifying each parameter's relationship with total chlorophyll concentrations, linear regression models were created on Google Colaboratory using various packages from Python (Google Colaboratory, 2024). Figure 6 depicts a linear model of total chlorophyll concentrations given oceanic temperature.

Figure 6

Linear Regression of Average Global Oceanic Chlorophyll Concentrations (μg/L) Given Average Global Oceanic Temperature (℃)



Note. Chl = -0.4438(Temperature) + 6.6180; $R^2 = 0.08$ ($r = 0.2828$); $p = 0.0235$*

Similar to the other preliminary analysis of WOD18, model fitness is limited ($R^2 = 0.08$, Figure 6). However, this could once again be due to outliers. Data on total chlorophyll concentrations from the early 2000s was abnormally high. When correlated with temperature, this caused several points to occur far above the least-squares regression line, reducing model fitness. Even so, there exists a significant linear relationship between average temperature and total chlorophyll concentrations ($\alpha = 0.05$, p =

0.0235*, Figure 6). The slope of the model indicates a predicted decrease of 0.4438 μg/L in chlorophyll concentrations for every increase in temperature by 1℃ (Figure 6). With average global oceanic temperatures projected to increase, this means that chlorophyll concentrations shall continue to decline. With less abundance of this pigment, phytoplankton would perform photosynthesis at lower rates. In turn, the production of organic matter would decrease, depleting the availability of trophic energy. Moreover, carbon sequestration would be inhibited, as lower rates of photosynthesis would reduce fixation of atmospheric carbon dioxide.

Expected Outcomes. It is expected that, given some input from relevant parameters, which would include driving parameters identified from the previous subtask as well as those relevant to energy transfer, the neural network would present how energy transfer from phytoplankton to the higher parts of the trophic pyramid would be modified. It is expected that the change in abundance of the species is noted. Other ecological ramifications of the observed changes in the food web should also be inferrable from the neural network model. It is expected that, when phytoplankton-induced climate scenarios are modeled within the existing CMIP5 framework (and potentially other climate databases), that climatic implications of changes in phytoplankton dynamics can be attained.

Potential Pitfalls and Alternative Strategies. Network theory could potentially be a tenuous way to computationally model a food web. The use of a neural network would necessitate the use of numerous other correcting mechanisms and other functions to be operational and accurate. If the neural network is too complex, which may be the case for a scenario as nuanced as this one, making a neural network may be immensely difficult. It may be fruitful to search for an alternative approach to computationally modeling a food web, such as multi-agent based modeling. Finally, CMIP5 may not be the ideal software by which to model the ramifications of phytoplankton-induced climate scenarios. It may be necessary to research other climate modeling software to model these ramifications.

Section IV: Resources/Equipment

Major Parts of the Project (rough outline) will continue to evolve over time and should be updated frequently.

Variables of Study

The independent variable shall be the configuration of the computer model with respect to the environmental factors included as parameters, the sources of data used, and the computational techniques employed.

The dependent variables shall be the model validity, which shall act as a prerequisite to other dependent variables, which shall include the changing phytoplankton dynamics in their own right, the driving factors behind those changes, the variables that influence these driving factors, as well as the climatic and environmental implications.

There shall be multiple rounds of iterations for the computer model.

Materials List

This study plans to make use of various computational tools to model phytoplankton dynamics such as:

- NetLogo
- TensorFlow and other Artificial Intelligence modeling tools

This study plans to make use of various sources of data to feed into the model as input for predictions and scenarios such as:

- Environmental Protection Agency
- Massachusetts Department of Environmental Protection
- United States Department of Agriculture
- Datasets from previous experiments and other sources

Procedure

The following is a rough outline for how this study shall be carried out.

- Using software tools, such as the ones stated above, create and validate a simple computer model, using only one environmental factor from one of the sources from above or an alternative source.
- Identify and employ statistical tests and tools to assess model validity (explained in data analysis section)

- Once a model is validated, incorporate more parameters, and more data about more species. Ensure that a global scale is covered in doing so.
- Identify results for phytoplankton dynamics
    - Biomass, export and primary production, migratory patterns, etc.
- Apply these findings to a network modeling food web of phytoplankton with other organisms in order to discern broader ecological ramifications
    - Possible establishment of new parameters, which could be the results from phytoplankton dynamics
- Use ANOVA, Post-HOC, PCA, and other statistical tools to identify driving factors in the phytoplankton dynamics observed. Additionally, the factors influencing these driving parameters are to be determined.
- Use results from these procedures and possibly other data to determine climatic implications

The data collected from computer simulations is planned to be subject to a series of statistical tools in order to assess model validity. Previous instances of computational models of phytoplankton have used regression analysis, comparing actual data to model predictions. Moreover, in order to identify driving factors of phytoplankton dynamics, it is planned to use ANOVA tests to compare the significance of the impact of different parameters on phytoplankton dynamics. Subsequent Post-Hoc tests would be performed to investigate this further, as has been done in previous instances (Sarker et al., 2023). Similar statistical tools may have to be applied to the networks created for depicting ecological ramifications. Additionally, with so many parameters at play in this study, performing procedures such as principal component analysis, as well as non-metric multidimensional scaling, are to be imperative in order to simplify data so as to make it understandable, and thereby reveal all of its insights. One way climatic ramifications could be assessed empirically is through looking at rates in metabolism, primary and export production, light reflection, among other metrics. Climatic implications can also be determined more indirectly through factors including biomass, diversity, and chlorophyll concentrations.

Section V: Ethical Considerations

Given that the scope of this project is computational, it should not pose any safety concerns. No individuals, animals, nor any other facet of the natural or anthropological environment should be harmed by this project. Changes in ocean conditions are all simulated. Consequently, there are no artificial modifications to any actual water body will undergo. The planned methodologies for this study strongly deter any possible risk with regards to ethics and safety.

Section VI: Timeline

September

- Brainstorming Project Ideas
- Establishment of Phytoplankton as an Area of Focus
- Establishment of Dual-Focus
    - Modeling Phytoplankton Biochemistry and Genomics Given Micronutrient Concentrations and other Environmental Factors
        - Impact on primary production, metabolic processes, among other areas
    - Computationally Modeling Phytoplankton Dynamics (biomass, phenology, migratory patterns, etc.)
        - Impact on climate and food web
        - Using environmental conditions as parameters.

October

- Narrowing Down Project Focus
    - Identifying Computationally Modelling as Primary Approach
- Researching on Computational Techniques, Software, Possible Sources of Data For Input
    - NetLogo
    - TensorFlow and other Artificial Intelligence Modelling and Machine and Deep Learning Tools
    - Network Theory, Neural Networks, Louvain Method,
    - EPA, MassDEP, and other potential databases

November

- Choosing Software Tools and Sources of Data
- Initial Model Validation
    - Done using one wide-known variable, such as Dissolved Oxygen of Water Temperature
    - Establishing statistical tools used to validate model for this project, such as R^2, RMSE, hypothesis tests, among other tools
- Develop Testing Strategy. This should begin to be performed. It should follow this vague outline:
    - Once model is validated, incorporate more parameters
        - Salinity, pH, micronutrients, turbidity, zooplankton, among other factors
    - Incorporate data for multiple species of phytoplankton
    - Utilize techniques such as network theory, multi-agent programming, among other tools, to model the impact of changing phytoplankton dynamics
        - Apply parameters to these models to assess trophic ramifications.
    - Employ chosen statistical tools to evaluate results with the following metrics:
        - Model accuracy
        - Driving Factors of the Observed Changes in Phytoplankton Dynamics
        - Impact on Higher Trophic Levels

December

- Continue executing the developed test strategy. Ensure that:
    - Phytoplankton dynamics, including
    - Changing phytoplankton dynamics can applied and/or parameterized into network models including higher level organisms in order to assess ecological impact
- Begin to statistically analyze results
    - Use tests and techniques to identify driving factors
- Provide Data/Evidence on Climatic Ramifications as well
- Models should be being iterated upon, and data should be being collected across different models.

January

- Finalize models and results
- Complete data analysis on ecological and climatic ramifications, identifying driving factors
- Applying these procedures through the lens of tracking eutrophication and local, ideally even higher level, policymaking.
- Identify future areas of focus

February

- Develop necessary products (clean up Project Notes and Logbook, STEM Fair poster, etc.)
- Present results

Section VII: Concluding Remarks

The goal of this project is to understand the causes and effects of changing phytoplankton conditions. Having this understanding is necessary, given these species' crucial ecological, biogeochemical, and climatic functions for the global environmental and climatic systems. The apparatus proposed could serve as a unique, innovative, streamlined process that allows scientists to easily study phytoplankton conditions. Moreover, it can serve as a tool for policymakers with regards to water body management. Therefore, this necessitates the allocation of necessary resources to this project.

References

Berwyn, B. (2018, May 7). *Scientists Say Ocean Circulation Is Slowing. Here's Why You Should Care.* Inside Climate News. https://insideclimatenews.org/news/07052018/atlantic-ocean-circulation-slowing-climate-change-heat-temperature-rainfall-fish-why-you-should-care/?gclid=CjwKCAjw2K6lBhBXEiwA5RjtCZMQVm0nxKHzSuutQ_Cgz9mZ1peI8xnzAhVnN0VCr8vudXHq2Sa3IhoC5O0QAvD_BwE

Deppeler, S. L., & Davidson, A. T. (2017). Southern Ocean Phytoplankton in a Changing Climate. *Frontiers in Marine Science*, *4*. Frontiers. https://doi.org/10.3389/fmars.2017.00040

Deus, R., Brito, D., Kenov, I. A., Lima, M., Costa, V., Medeiros, A., Neves, R., & Alves, C. N. (2013). Three-dimensional model for analysis of spatial and temporal patterns of phytoplankton in Tucuruí reservoir, Pará, Brazil. *Ecological Modelling*, *253*, 28–43. https://doi.org/10.1016/j.ecolmodel.2012.10.013

Käse, L., Geuer, J.K. (2018). Phytoplankton Responses to Marine Climate Change – An Introduction. In: Jungblut, S., Liebich, V., Bode, M. (Eds.), *YOUMARES 8 – Oceans Across Boundaries: Learning from each other* (pp. 55-71). Springer, Cham. https://doi.org/10.1007/978-3-319-93284-2_5

Li, J., Zhang, K., Li, L., Wang, Y., Wang, C., & Lin, S. (2023). Two-sided effects of the organic phosphorus phytate on a globally important marine coccolithophorid phytoplankton. *Microbiology Spectrum*, *11*(5). https://doi.org/10.1128/spectrum.01255-23

Ratnarajah, L., Abu-Alhaija, R., Atkinson, A., Batten, S., Bax, N. J., Bernard, K. S., Canonico, G., Cornils, A., Everett, J. D., Grigoratou, M., Ishak, N. H. A., Johns, D., Lombard, F., Muxagata, E., Ostle, C., Pitois, S., Richardson, A. J., Schmidt, K., Stemmann, L., &

Swadling, K. M. (2023). Monitoring and modelling marine zooplankton in a changing climate. *Nature Communications*, *14*(1). https://doi.org/10.1038/s41467-023-36241-5

Rohr, T., Richardson, A., & Shadwick, E. (2023, June 15). *Oceans absorb 30% of emissions, driven by a huge carbon pump: Tiny marine animals are key to cycle, says study*. Phys.org. https://phys.org/news/2023-06-oceans-absorb-emissions-driven-huge.html

Sarker, S., Haque, A. B., Chowdhury, G. W., & Huda, A. N. M. S. (2023). Environmental Controls of phytoplankton in the river dominated sub-tropical coastal ecosystem of Bangladesh. *Regional Studies in Marine Science*, *66*, 103114. ScienceDirect. https://doi.org/10.1016/j.rsma.2023.103114

Wan, J., Zhou, Y., Beardall, J., Raven, J. A., Lin, J.-H., Huang, J., Lu, Y., Liang, S., Ye, M., Xiao, M., Jing, Z., Dai Xiao-ying, Xia, J., & Jin, P. (2023). DNA methylation and gene transcription act cooperatively in driving the adaptation of a marine diatom to global change. *Journal of Experimental Botany*, *74*(14), 4259–4276. https://doi.org/10.1093/jxb/erad150

Winder, M., & Sommer, U. (2012). Phytoplankton response to a changing climate. *Springer EBooks*, *221*, 5–16. Springer Link. https://doi.org/10.1007/978-94-007-5790-5_2