Computational Modeling of Phytoplankton Dynamics with Climatic and Ecological Ramifications

Abhinav K. Sharma

Massachusetts Academy of Math and Science

Advanced STEM with Scientific and Technical Writing

Instructor: Dr. Kevin Crowthers, Ph.D.

Worcester, MA. 01605

Author Note

The author declares that they had no conflicts of interest in carrying out research, performing empirical analysis, nor in the development of this report. For the purposes of concision, various supplementary files have been provided. These include access to more detailed data analysis, as well as the code used for this project. The dataset used can be accessed online for free using the Registry of Open Data provided by Amazon Web Services. The dataset used is provided in the appendices, as is the code to extract it (Supplementary File 1).

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 carbon emissions. With global warming modifying ocean conditions, understanding the drivers and impacts of changing phytoplankton dynamics is crucial. However, one-factor experiments have limited applicability due to heterogeneity in oceanic conditions and biological responses and preferences among phytoplankton. Conversely, multi-factor experiments produce confounding results. Therefore, a computational approach was taken wherein a series of models was developed. All data were derived from the NOAA's comprehensive World Ocean Database (WOD). Total oceanic chlorophyll concentration was used as an indicator for primary production. To assess accuracy in forecasting capabilities and potential correlations, a times series of each parameter was developed and each factor's relationship with chlorophyll was assessed, primarily using sinusoidal and linear regression. Model fitness was mild, as R-Squared values for the first and second set of models ranged from 0.077 to 0.847, and 0.07 to 0.54, respectively. Subsequently, driving parameters behind chlorophyll levels were identified using principal component analysis (PCA). Results indicated pH, followed salinity and pressure, as the most influential parameters. Overall results indicate that the proposed computational apparatus is viable for analyzing phytoplankton dynamics, but that iteration in the form of model modification and greater data implementation is necessary. This apparatus could also serve as a significant tool for policymaking related to aquatic ecosystem management.

Keywords: Phytoplankton, Computational Modeling, Global Warming, World Ocean Database

Acknowledgements

A great amount of gratitude is extended to Dr. Kevin Crowthers, who provided insight related to model development tools and parametric data, as well as overall project feedback. In addition, the family and friends of the author have provided continued emotional support throughout the process of carrying out this study and authoring this paper.

Computational Modeling of Phytoplankton Dynamics with Climatic and Ecological Ramifications

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, helping circulate 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 in fact 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 Aquatic Changes on Phytoplankton

With that in mind, the impact global warming has had on oceanic conditions themselves must first be considered. Climate change has led oceans to becoming 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, saltier benthic (deep-sea) water. This allows for the mixing of nutrients, the distribution of heat, and the regulation of climate. Analysis of past climate patterns reveals 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, but rather, vary extensively by region (Winder & Sommer, 2012). That means environmental conditions, which impact the nature of 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, that is, conditions where there are excessive micronutrients, leading to an unhealthy amount of growth in algal blooms that deplete ecosystem resources, whereas others under fresher or darker conditions (Winder & Sommer, 2012). There are a voluminous amount of environmental factors (e.g., light, heat, nutrients, pH, salinity, etc.) that impact phytoplankton dynamics (Winder & Sommer, 2012). Moreover, each species operates under different sets of ideal conditions. This raises a dilemma. To illustrate this, consider two phytoplankton species living in the same area. Suppose that one species can tolerate a pH range of 5.9 to 6.5, whereas another one 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 dynamics. Using the example given, would another factor, such as dissolved oxygen, have precedent over pH when it comes to these species seeking ideal conditions? Moreover, 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 aimed to address.

Examples of Parametric Variability

Parameters that influence phytoplankton conditions 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, with warming ocean temperatures, has been found to inhibit amino acid metabolism, as well as respiration and photosynthesis in phytoplankton, while enhancing fatty acid metabolism (Wan et al., 2023). This means that there is a slower rate of primary production and carbon sequestration, inhibiting phytoplankton's role both as the base of marine food chains and as climate regulators. However, seeing as ocean temperature shall change 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, as discussed above, varying levels of micronutrients, including these compounds, impact dynamics in different ways. It has been found that increased phosphorus levels has allowed for all metabolic processes to occur at faster rates, bolstering the ability of phytoplankton to sequester carbon and provide greater biomass for its predators. However, excessive phosphorus concentrations can be toxic and lead to eutrophication (Li et al., 2023). Moreover, 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 alongside 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.

In conjunction with the explanation offered in the previous section, these examples 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.

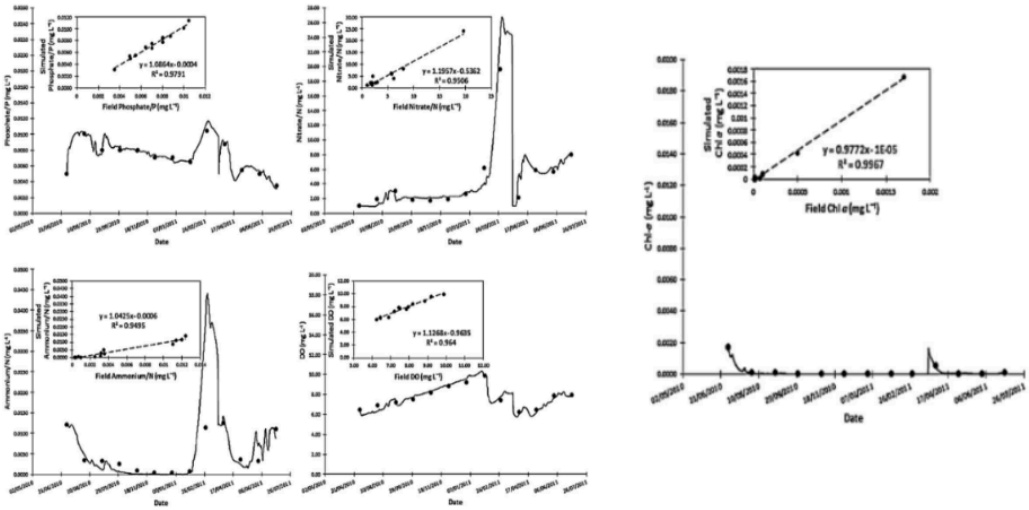
Presently, there are many limitations with computational models of phytoplankton dynamics. One major limitation is the misunderstanding of the role zooplankton play in the modeling process. Different models have made

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. The R2 values for phosphate, nitrate, ammonia, dissolved oxygen, and chlorophyll a were 0.9791, 0.9506, 0.9495, 0.964, and 0.9967, respectively.

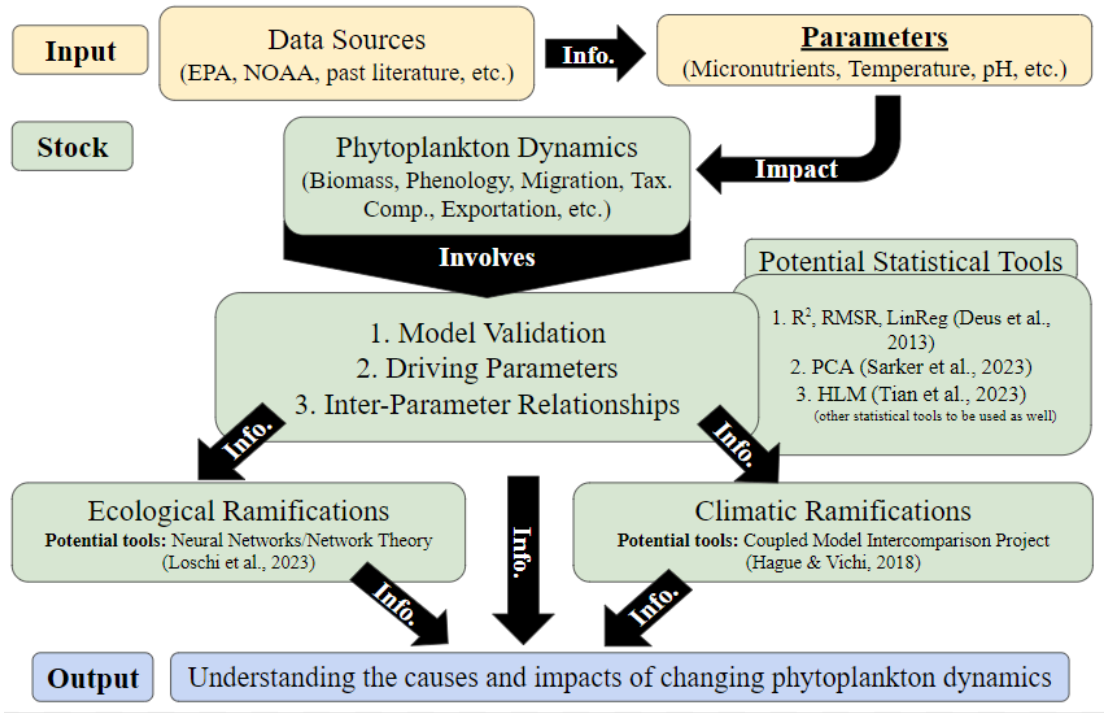
However, different models have been synthesized for different purposes. For example, some models have focused on the identification of driving parameters in phytoplankton dynamics. Using Principal Component Analysis (PCA), whereby the impact of parameters is measured using vectors, one study of coastal 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 the parameters into different levels and identifying statistically significant relationships, the one major inter-parameter relationship identified 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. Boit et al. 2012 suggests the gradual implementation of these factors through a series of successive neural networks. When applying this approach to Lake Constance, 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 been able to identify keystone species (Loschi et al., 2023). From a climatic lens, a focus has been placed on the accuracy of climate models in predicting bloom phenology, as well as other characteristics. The Coupled Model Intercomparison Project (CMIP), with its large scope, has been a particular area of focus. For example, 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). Overall, there exists ample literature describing 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 their populations, and in turn, the environment and climate could be produced. Figure 2 visualizes this overarching computational framework. This study applied this basic framework to data from the National Oceanic and Atmospheric Administration (NOAA)'s comprehensive 2018 World Ocean Database (WOD18). Specifically, the most spatiotemporally cosmopolitan dataset, the Ocean Station Dataset (OSD),

was analyzed. These data include millions of casts, spanning multiple centuries and covering virtually the entire ocean (Boyer et al., 2018). Given this impressive scope, this allows the study to take a holistic approach to analysis, partially helping to address the lack of data in computational models. Within the OSD, total oceanic chlorophyll was used as an indicator for primary production. Factors tested include oxygen, micronutrients, pH, salinity, temperature, pressure, and alkalinity. To assess potential forecasting capabilities and overall model strength, a time series of all parameters (including the stated indicator), was created mainly using sinusoidal regression. Subsequently, the relationship of each factor with the indicator was observed using linear regression. Lastly, driving parameters were identified using Principal Component Analysis (PCA).

Figure 2

Proposed Overarching Computational Framework for Modeling of Changing Phytoplankton Dynamics



Note. This model takes the form of a systems diagram wherein an input is provided for the system stock, operations are performed, and an output is provided. This study proposes that parametric data act as the input, that computational and statistical methods act as the operations within the stock, and that the insights provided on phytoplankton, that is, the study goal, to act as the output. All potential tools proposed above, while useful for achieving their respective ends, however, not all techniques were utilized within this paper.

This apparatus could serve as a viable streamlined process for experts studying 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 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: Methodology

Role of Student vs. Mentor

The author of this paper was the student, who was mentored by Dr. Kevin Crowthers. From July 2023 to February 2024, the bulk of the work, including idea generation and attainment, research, and model development, validation, and testing, was performed by the former. The latter was responsible for monitoring project progress, as well as offering advice, particularly on potential software usage for parametric testing.

Equipment and Materials

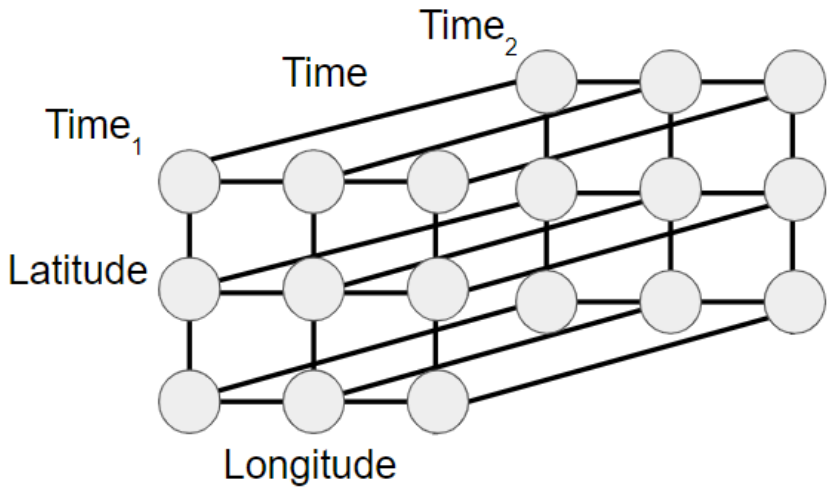
The primary dataset analyzed in this study was the 2018 World Ocean Database (WOD18) provided by the National Oceanic and Atmospheric Administration (NOAA). Both spatially and temporally, this dataset provides a highly cosmopolitan measurement of numerous environmental parameters, including water temperature, micronutrients, pH, salinity, among many others (Boyer et al., 2018). Access to the files of this dataset was attained through the Registry of Open Data provided by Amazon Web Services (AWS). The files used in this dataset were all updated within the AWS S3 explorer system on 17 October 2023 when obtained. Files were organized by year from 1900 to 2023, with pre-1900 data being referred to as 1800 (Amazon Web Services, 2024). The file of each year was systematically downloaded. Regression models were developed using Python code. The main programming interface used was Google Colaboratory, which, when used, was most recently updated on 8 January 2024, supporting Python 3.10.12 (Google Colaboratory, 2024). An online webpage was used to conduct PCA (Statistics Kingdom, 2017). Across all Python programs developed, the Pandas, Xarray, NumPy, SciPy, SciKitLearn, Seaborn, and MatPlotLib packages were utilized. Additionally, to observe the dimensions of the data files more closely, the Panoply software, provided by the National Aeronautic and Space Administration's (NASA's) Goddard Institute for Space Studies (GISS), was downloaded. The most current version, 5.3.1, was used, released 1 January 2024 (NASA GISS, 2024).

Decisions for Parametric Model Development Based off Data Structure

Two key observations were made during preliminary use and analysis of the dataset that led to two key decisions on the analytical procedure performed. The first observation relates to the data structure of the files used. These were NetCDF (.nc) files, which illustrate parametric measurements at specific points of latitude, longitude, and depth across a series of equal temporal increments (Figure 3). In the metadata attained for the .nc files, these spatiotemporal attributes were referred to as "coordinates." However, it was specifically noted within this directory that all dimensions, with the exception of the numerical measurements of parametric data, had their latitude, longitude, depth, and time alongside their measurement. Meaning, outside of the year as provided by the file, the spatiotemporal attributes of the primary data on environmental parameters could not be accessed. Therefore, a more holistic approach to analysis was taken, wherein the average global value of each factor was calculated for each year of data. This helped preserve temporal analysis and offered a viable overview of environmental relationships, though at the expense of omitting the nuances and consequences caused by dissimilar spatial attributes among parametric data.

Figure 3

Illustration of the NetCDF Data Structure



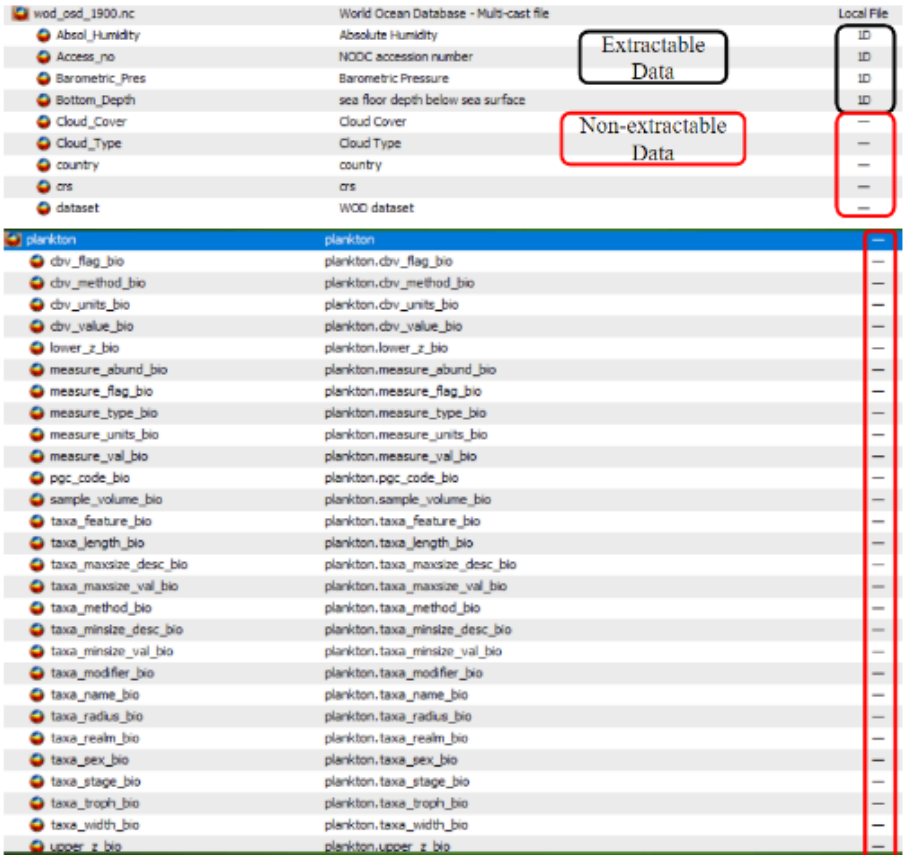
Note. This is a common illustration of the structure of a NetCDF file. The gray circles represent specific points in longitudinal and latitudinal space. Though not illustrated above, each of these coordinates also contains varying levels of depth. At all of these points in space, there exist parametric measurements. These measurements are projected out through a series of time increments, forming a three-dimensional figure.

The second observation made from the metadata was that planktonic data was inaccessible. Within the Panoply software, whereas extractable data was stored within one-dimensional arrays, the values of non-extractable data were unavailable. Figure 4 contains an illustration of this issue. Planktonic data fell under the latter category.

Consequently, it was decided to use average global oceanic concentration of total chlorophyll as an indicator of phytoplankton dynamics, namely primary production. This is because chlorophyll is a crucial pigment for carrying out the photosynthetic process, and in turn, all other metabolic processes. As such, a higher concentration of chlorophyll would indicate greater potential for primary production, whereas Though a valuable indicator, it is important to note that it is not a direct measurement of phytoplankton traits. Although data of all available parameters spanning 1900 to 2019 were downloaded, due to the limited temporal range of measurements for total oceanic chlorophyll, this study focused on data from 1954 to 2017. This also reduced the number of factors assessed.

Figure 4

Inaccessibility of Planktonic Data



Note. The extractable data (black) were stored as one-dimensional arrays, whereas the values of inextricable data could not be obtained. All planktonic data fell under the latter category.

Data Extraction and Cleaning

Using Google Colaboratory, a brief program was written to extract parametric data .nc files and save them as Comma Separated Value (.csv) files. Within each .csv file, the average, standard error, and sample size for each

year of each parameter was calculated and compiled into a separate spreadsheet file. Supplementary Files 1 and 2 in the Appendices section (all supplementary files may be found in the Appendices section), provide the exact code used. The main data cleaning involved the removal of non-numerical data within .nc files when converting them to .csv. This was achieved by the Python program written.

However, for total oceanic chlorophyll and alkalinity, the data processing was more complex. When originally creating a time series for the former, it was noted that model strength was inhibited by abnormally high measurements around the early 2000s. Supplementary File 3 (specifically Supplementary Figure 3), provides a visualization of this. Upon further investigation of the .csv files, it was noted that this was due to the abnormally high amount of outliers. In order to maximize model fitness, for chlorophyll data spanning 1998 to 2008, any and all measurements in excess of 20 μg/L were removed, and new averages, standard deviations, and sample sizes were determined. The next iteration of the time series had a stronger fit as a result. For alkalinity, many years had errors in how the data was recorded, in that the decimal place was improperly positioned. This led to values that were orders of magnitude too high for the dataset, and in turn, skewed summary statistics. As such, any such data was eliminated from the set, with summary statistics adjusted accordingly. Besides the processes described, all parametric data from 1954-2017 was preserved when performing data analysis.

Statistical Analysis

A variety of statistical tests and computational tools were used for the three major sets of models developed for this study. The time series for each parameter developed was created primarily using sinusoidal regression. The strength of each regression model was measured using a Pearson's Correlation, including both r and R^2. To assess correlation between each factor and total oceanic chlorophyll, linear regression, in conjunction with a Student's t-test for relationship significance and Pearson's correlation for relationship strength, was used. Finally, driving parameters were identified using PCA, along with supplementary techniques.

Sinusoidal Regression Including Pearson Correlation

Environmental features tend to be periodic in nature. The sine and cosine functions provide an effective way to model cyclical trends. Therefore, this specific type of a regression model was chosen, with R^2 and r measuring model strength and accuracy. This allowed for the evaluation of the validity of the overall computational system as well as projection abilities. Equation 1 represents the template function used for all time series models:

$$f(\kappa) = Asin((2\pi\gamma)\beta + \varepsilon) + \phi \quad (1)$$

Where $f(\kappa)$ is the function for the total chlorophyll concentration κ, A is the amplitude, $2\pi\gamma$ represents the length of the period (using radians, γ alone being in degrees), β is the given environmental factor (the next section enumerates the variable designation of each parameter), ε is the phase shift, and ϕ is the offset.

Additionally, using the offset as a midline for the sinusoid and the amplitude as a sort of ruler, an interval of all measurements projected by the sinusoid of each parameter was developed. Equation 3 represents the basic construction of the described sinusoidal interval:

$$\phi \pm A \tag{2}$$

Linear Regression Including Pearson Correlation and Student's t-test

For the relationship of every parameter with the indicator, total chlorophyll concentration, a Linear Regression model was developed. On a functional level, assessing each individual parameter's relationship with total chlorophyll acted as a precursor to identifying which of them had a significant influence on chlorophyll when all parameters were considered in tandem. In essence, performing linear regression acted as a prerequisite for then performing PCA. The significance of relationships were determined using a Student's t-test for Linear Regression at $\alpha = 0.05$. Both double- and single- tailed p-values were attained. This was done in conjunction with the use of R^2 and r to measure model accuracy and strength. Similar to the previous set of models, Equation 3 provides a template wherein each parameter's regression model was represented:

$$f(\kappa) = m\beta + \beta_0 \tag{3}$$

Where $f(\kappa)$ is the function for the total chlorophyll concentration κ, m is the predicted slope of the line, β is the given parameter, and β_0 is the y-intercept of the model.

PCA

In order to identify the driving parameters behind total chlorophyll concentrations, PCA was used. PCA is a dimension reduction technique that compresses multiple independent variables into fewer dimensions so as to summarize overarching data patterns and allow for ease of data visualization. Before performing PCA, the data must be standardized so that scale does not impede the accuracy of results. In this study, before PCA was performed, all data of the indicator (chlorophyll) and every parameter were standardized using minimum-maximum normalization. Within the setting of PCA, the total variance of the data is measured. This variance is captured by a finite set of portions of the data known as principal components. In two-dimensional representations, the first two principal components, that is, the two components that account for the highest amount of variance, denoted PC_1 and PC_2, are

placed on the horizontal and vertical axes respectively. Since information from the other principal components (PC_3, PC_4, … PC_n) is omitted, it is important that most variance is captured by PC_1 and PC_2. This is measured by each principal component's eigenvector values, which are derived from various matrix operations performed on the data. Then, to standardize the amount of variation each principal component captures, the eigenvalues are divided by the total variance of the dataset. A scree plot is used to depict the cumulative coverage of variance by all principal components. Along with a PCA plot, a scree plot was used to illustrate these notable properties of the principal components. In two-dimensional PCA, every parameter assessed captures some amount of either principal component, and holds either a positive or negative relationship with the directionality of component variances. This magnitude and directionality is represented by a pair of coordinates that form a vector. The greater the magnitude of variance represented by a parametric vector, the more influential that parameter is relative to the overall data, and in turn, driving the dependent variable. For the study, the magnitude of each parameter was calculated as depicted by Equation 4:

$$M(\beta) = \sqrt{(C_{PC1})^2 \cdot \nu_{PC1} + (C_{PC2})^2 \cdot \nu_{PC2}} \quad (4)$$

Where $M(\beta)$ is the function of the magnitude of parameter β, C_{PC1} is the contribution of the parameter to the variance of PC_1, while C_{PC2} is the contribution of the parameter to the variance of PC_2, and ν_{PC1} is the proportion of the total variance represented by PC_1, and ν_{PC2} is the proportion of the total variance represented by PC_2.

The magnitudes of each parameter were calculated using the above equation, and then subsequently ranked by descending magnitude values. The parameters with the highest calculated magnitude were identified as driving parameters of chlorophyll concentrations. Additionally, to assess the presence of inter-parameter relationships, a covariance matrix was used. A covariance matrix is an intermediate operation performed in the complex matrix calculations involved with PCA wherein all independent variables are arranged in a square array. The cells of this matrix contain the covariance between the row and column parameters. Values vary between 0 and 1, and can be either positive or negative based on the directionality of the relationship. A greater magnitude indicates a stronger relationship between the two parameters. The diagonal cells represent the variance of that individual parameter following dimension reduction procedures. The results from these three sets of computational models were then put into biogeochemical, ecological, and climatic context.

Section III: Results

As outlined in the methodology section, three sets of computational models were produced. The first set of models involved using sinusoids to create time series for the indicator, global oceanic chlorophyll, as well as all other factors. Table 1 provides the properties of each regression for each environmental measure. Parameters are ranked by descending R^2 values. In conjunction with a color gradient that spans from red to blue, where warmer colors correspond to higher R^2 values and cooler ones to lower values, this provides a gradient wherein the progressive decline in R^2 values and their corresponding parameters can be observed. In addition, offset and amplitude values, and the corresponding sinusoidal intervals, are depicted. It is important to note that this table is the first instance of variable denotation of environmental parameters. Henceforth, those designations are used in data presentation.

Table 1

Properties of Sinusoidal Time Series Models of Environmental Parameters

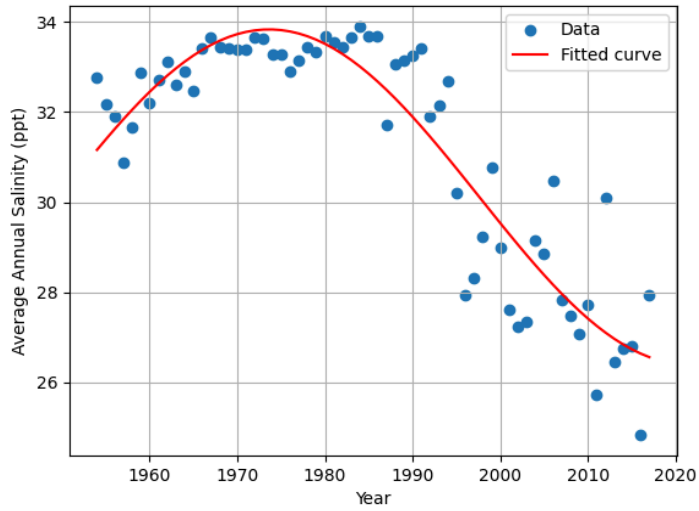
Parameter	Variable Designation	R^2 Value (r)	Offset (ϕ)	Amplitude (A)	Parametric Range Projected by Sinusoid ($\phi \pm A$)
Salinity (ppt)	s	0.847 (0.920)	30.112	3.724	(26.388, 33.836)
pH	φ	0.825 (0.908)	8.001	0.135	(7.866, 8.136)
Total Chlorophyll (µg/L)	κ	0.648 (0.805)	289.542	289.457	(0.085, 578.999)
Dissolved Oxygen (µmol/kg)	d	0.500 (0.707)	215.281	-12.543*	(202.738, 227.824)
Nitrate (µmol/kg)	η	0.381 (0.617)	-3650.602	3665.327	(-7315.929, 14.725)
Phosphate (µmol/kg)	q	0.336 (0.580)	1.191	0.118	(1.073, 1.309)
Temperature (℃)	t	0.327 (0.572)	9.640	1.045	(8.595, 10.685)
Silicate (µmol/kg)	h	0.270 (0.520)	33.095	6.120	(26.975, 39.215)
Alkalinity (milli-equivalent/liter CaCO3)	c	0.239 (0.489)	2.268	0.098	(2.170, 2.366)
Pressure (decibars)	ρ	0.077 (0.277)	489.463	109.099	(271.265, 598.562)

Note. *Although the amplitude for the sinusoid projecting dissolved oxygen is negative, this does not interfere with interval construction, as that is a matter of both adding and subtracting the magnitude of the amplitude from the offset, meaning the net output is the same. In addition to the ones listed above, variable "y" represents the year.

R^2 values of these regression models measure the proportion of the variation observed in the model predictions that are a result of the actual oceanographic data collected. This means a higher R^2 value indicates greater strength in model forecasting capabilities. In conjunction with this table, the sinusoidal regression for salinity, the parameter with the highest R^2 value, as well as pressure, the parameter with the lowest R^2 value, are provided. By representing the most and least fit models, the sinusoidal regression models for salinity and water pressure act as landmarks for the upper and lower bounds in forecasting capabilities observed among the time series models produced. Time series models of other parameters can be found in Supplementary File 3.

Figure 5

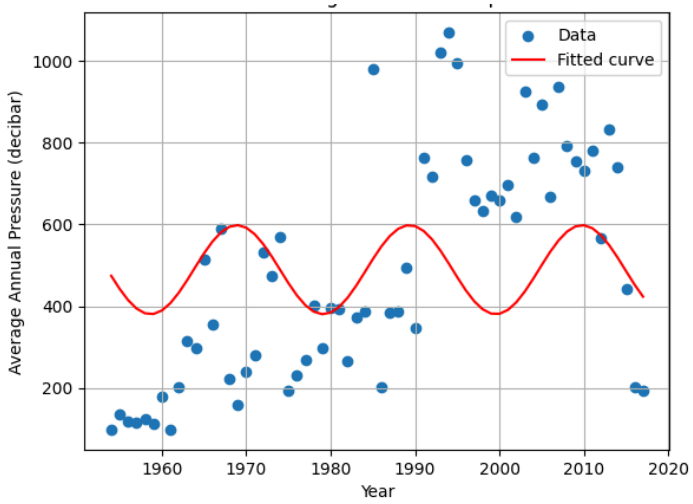
Time Series of Average Global Oceanic Salinity (ppt) from 1954-2017



Note. The time series for salinity (s) for year y is described by the sinusoidal regression function of f(s) = 3.724*sin(6.220y + 80.552) + 30.112; R^2 = 0.847. Blue points represent individual average salinity levels, while the red line depicts the sinusoid.

Figure 6

Time Series of Average Global Oceanic Pressure from 1954-2017 (Decibars)

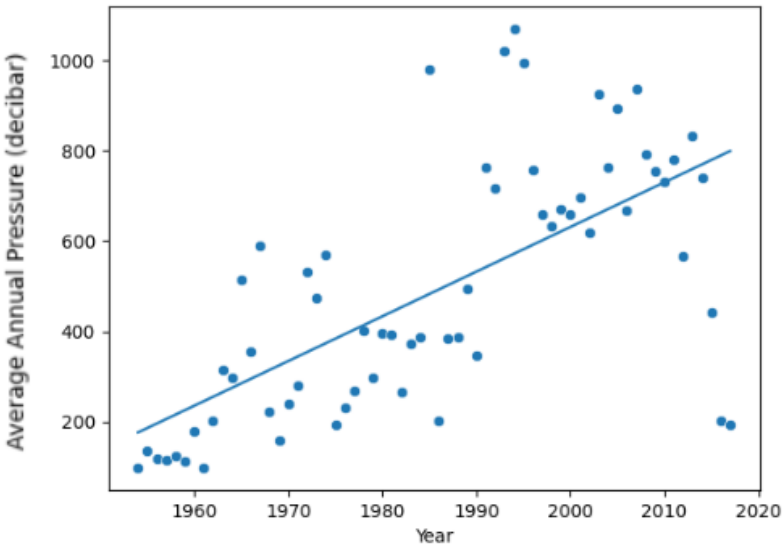


Note. The time series for pressure (ρ) for year y is described by the sinusoidal regression function of f(ρ) = 109.099*sin(6.951y + 43.661) + 489.463; $R^2 = 0.077$

Due to the extremely low R^2 value observed for water pressure, to increase forecasting capabilities, an alternative time series was constructed using linear regression. This model had a higher R^2 value and demonstrated a significant increase in water pressure from 1954 to 2017 (Figure 7, $\alpha = 0.05$, $p < 0.0001$***).

Figure 7

Alternative Time Series of Average Global Oceanic Pressure from 1954-2017 (Decibars)



Note. The alternative time series for pressure (ρ) for year y uses linear regression is described by the function of f(ρ) = 9.904*y - 19176.54; Since $R^2 = 0.44$, the forecasting fitness for water pressure has successfully been increased.

Following the analytical flow detailed in the methodology, the next operation performed was the creation of linear regression models representing each environmental factor's relationship with total oceanic chlorophyll concentrations. For every relationship combination, a scatterplot of chlorophyll concentrations and the corresponding parametric value for eaR^2 ch year from 1954 to 2017 was generated. From this, the regression line was formed. Table 2 supplies the strength (R^2) and significance (p) of each regression model, as well as the equation. In a similar fashion to the first set of models presented, Table 2 ranks parameters by highest to lowest R^2 values (which also correspond to increasing p-values, or decreasing statistical significance), with a color gradient visually enumerating the descent of this metric and corresponding factors along the way. Subsequently presented are the linear regression models with the highest and lowest R^2 values, which are s (salinity) and c (alkalinity), respectively. Meanwhile, the linear models of all other parameters are provided in Supplementary File 3.

Table 2

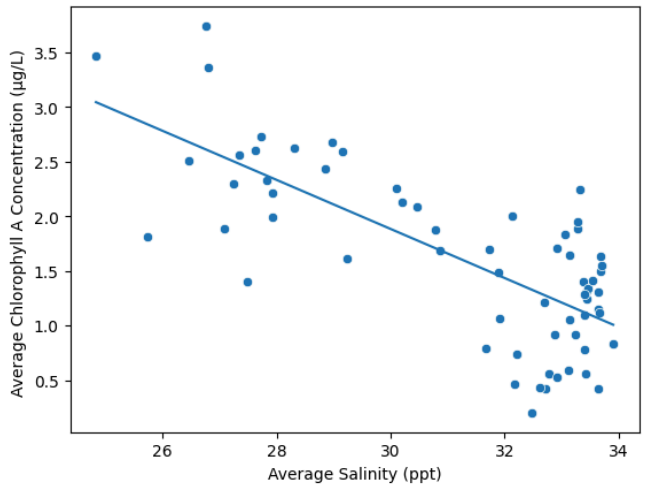
Linear Regression Information of Parametric Factors Related to Total Oceanic Chlorophyll

Variable	R^2 Value (r)	p-value (single-tailed p-value) $\alpha = 0.05$	Equation
s	0.54 (0.735)	0.0000 (0.0000)***	$f(\kappa) = -0.2248s + 8.6275$
φ	0.27 (0.520)	0.0000 (0.0000)***	$f(\kappa) = -3.9346\varphi + 33.0813$
d	0.25 (0.500)	0.0000 (0.0000)***	$f(\kappa) = 0.0333d - 5.6467$
ρ	0.22 (0.469)	0.0001 (0.00005)***	$f(\kappa) = 0.0014\rho + 0.9536$
q	0.10 (0.316)	0.0130 (0.0065)*	$f(\kappa) = 1.8526q - 0.5389$
η	0.09 (0.300)	0.0171 (0.0086)*	$f(\kappa) = 0.0634\eta + 0.8126$
h	0.08 (0.283)	0.0244 (0.0122)*	$f(\kappa) = 0.0324h + 0.6060$
t	0.07 (0.265)	0.0318 (0.0159)*	$f(\kappa) = -0.1641t + 3.2049$
c	0.07 (0.265)	0.0368 (0.0184)*	$f(\kappa) = 1.4551c - 1.6810$

Note. From highest to lowest R^2 value, the parameters are: salinity, pH, dissolved oxygen, pressure, phosphate, nitrate, silicate, temperature, and alkalinity. Significance levels of $p \leq 0.001$ are denoted with three astrices. When $p \leq 0.05$ is true, only one asterisk is used. No significance values between 0.01 and 0.001 were attained, so no significance values were denoted with two asterisk. Alkalinity is placed below temperature due to having a lower p-value. This addresses their equivalent R^2 values.

Figure 8

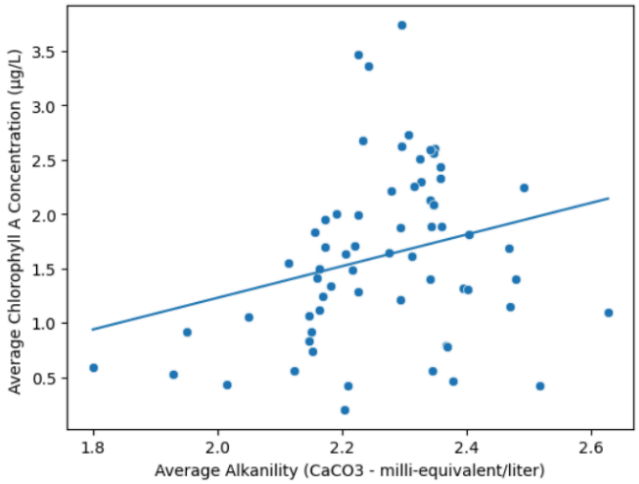
Linear Regression Model of Total Oceanic Chlorophyll (μg/L) Given Salinity (ppt)



Note. This regression model, with the highest R^2 value, possesses the strongest predictive capability between total chlorophyll and an environmental measure. Blue points represent corresponding salinity and chlorophyll measurements for each year between 1954 and 2017, while the equation, $f(\kappa) = -0.2248s + 8.6275$, is represented by the blue line.

Figure 9

Linear Regression Model of Total Oceanic Chlorophyll (µg/L) Given Alkalinity (meq/L of $CaCO_3$)

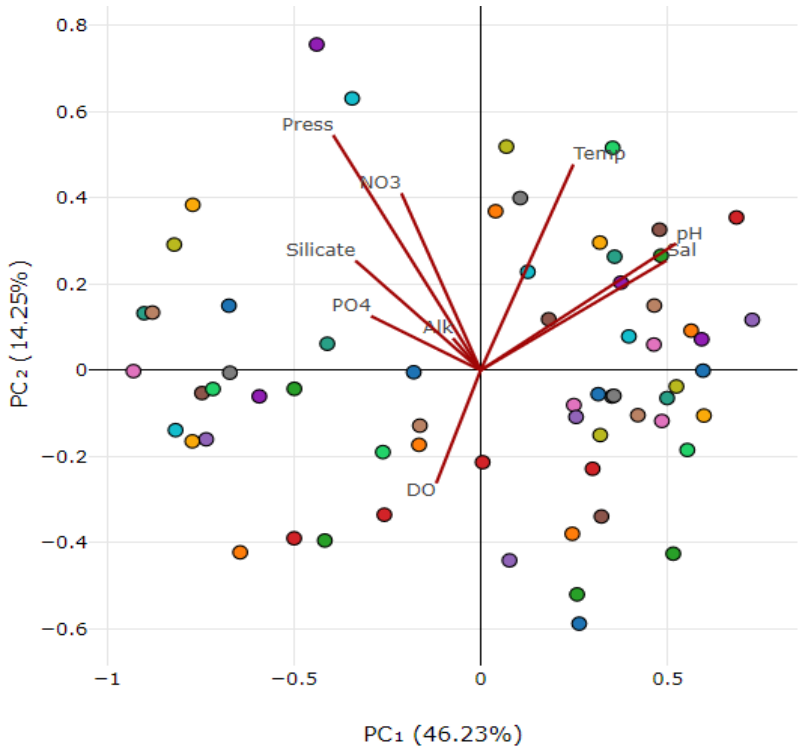


Note. This regression model, with the highest R^2 value, possesses the weakest predictive capability between total chlorophyll and an environmental measure. Blue points represent corresponding alkalinity and chlorophyll measurements for each year between 1954 and 2017, while the equation, $f(\kappa) = -0.2248s + 8.6275$, is represented by the blue line.

Finally, PCA was performed to identify driving parameters. Figure 10 provides the image of the raw PCA plot produced upon inputting data into the online webpage and calculating results.

Figure 10

PCA Plot of Driving Parameters Behind Total Oceanic Chlorophyll Chlorophyll Concentrations

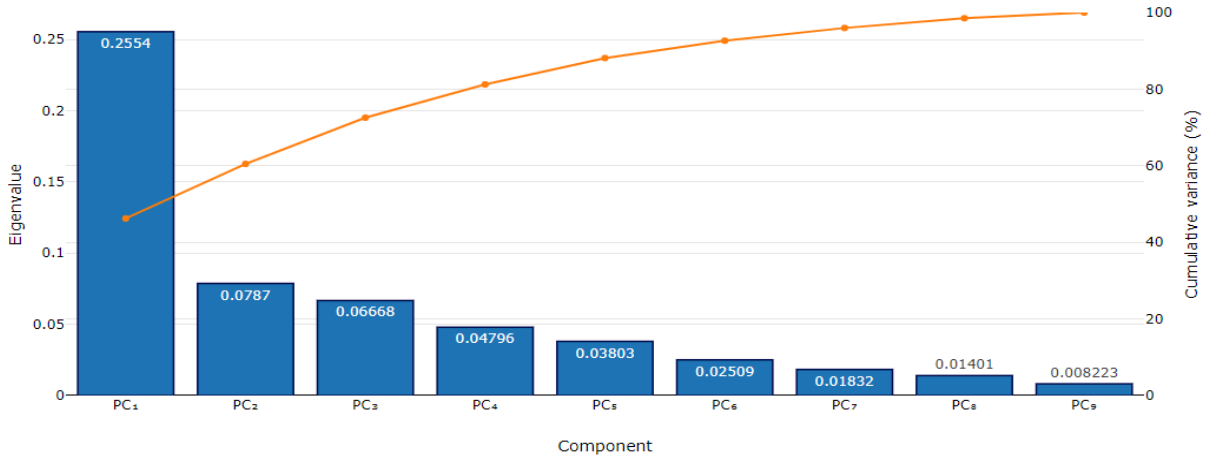


Note. The horizontal and vertical axes are represented by PC_1 and PC_2 respectively. The individual dots of varying color represent various instances of location of chlorophyll measurements relative to the first two principal components following dimension reduction. The red lines sprouting from the origin represent the vectors of each parameter's contribution to the variance of the first two principal components. Parametric abbreviations are designated as follows: pH (pH), salinity (sal), temperature (temp), nitrate (NO_3), pressure (Press), silicate (Silicate), phosphate (PO_4), alkalinity (alk), and dissolved oxygen (DO).

PC_1 accounts for 46.23% of overall variance of the dataset, whereas PC_2 accounts for 14.25%. Figure 11 provides a scree plot depicting the cumulative coverage of whole-dataset variance among the total nine principal components.

Figure 11

Scree Plot of Cumulative Dataset Variance Coverage Among Principal Components



Note. The blue bars with the white numbers represent the contribution of each principal component to total variance in the form of an eigenvector value. This is not to be confused with the standardized contribution of each component to total dataset variance, which is attained only after dividing these eigenvector values by the total dataset variance. The orange line represents the cumulative progression of coverage of this variance.

As with the previous two stages of computational models, a table utilizing a color gradient to represent the progression of statistical measurements is provided. Table 3 orders parameters from the highest to lowest magnitude, the measurements of which were based off of the operations enumerated in Equation 4.

Table 3

Magnitude of Parametric Contributions to the Primacy Principal Components

Variable	Contribution to PC_1	Contribution to PC_2	PC_1/PC_2 Magnitude*
φ	0.5229155	0.2942522	0.3724914
s	0.4967389	0.2542255	0.3511156
ρ	-0.3944697	0.5453883	0.3381172
h	-0.3350876	0.2544652	0.2472569
t	0.2481767	0.4774918	0.2469081
η	-0.2119327	0.4106519	0.2116481
q	-0.2929772	0.1256602	0.2047729
d	-0.1187105	-0.2619915	0.1276556
c	-0.07405966	0.07430118	0.0576397

Note. *As enumerated in equation for 4, these measurements were attained via the following formula (refer to Equation 4 for variable definitions): $M(\beta) = \sqrt{(C_{PC1})^2 \cdot \nu_{PC1} + (C_{PC2})^2 \cdot \nu_{PC2}}$. Given that all contributions are squared, regardless of the directionality of the contribution of parameters towards the first two principal components, the magnitude is expressed as a positive value.

Finally, the last data analysis technique presented is a covariance matrix depicting the inter-parameter relationships between the environmental factors within the WOD18 dataset that were tested for this study (Table 4).

Table 4

Covariance Among Oceanic Parameters in OSD Dataset of WOD18, 1954-2017

	Temperature (℃)	Salinity (ppt)	Dissolved O_2 (µmol/kg)	Pressure (decibars)	pH	Alkalinity (meq/L $CaCO_3$)	NO_3 (µmol/kg)	PO_4 (µmol/kg)	Silicate (µmol/kg)
Temperature (℃)	0.05679								
Salinity (ppt)	0.03166	0.08317							
Dissolved O_2 (µmol/kg)	-0.01647	-0.01959	0.04698						
Pressure (decibars)	-0.005877	-0.03823	0.01677	0.08013					
pH	0.03188	0.06853	-0.01324	-0.03891	0.09541				
Alkalinity (meq/L $CaCO_3$)	0.006023	-0.01187	0.001403	0.009997	-0.01088	0.02999			
NO_3 (µmol/kg)	-0.001303	-0.01835	0.003703	0.03955	-0.01663	0.0009591	0.03353		
PO_4 (µmol/kg)	-0.0178	-0.02756	-0.003317	0.02564	-0.03468	0.002638	0.017	0.05649	
Silicate (µmol/kg)	-0.02261	-0.03465	-0.007556	0.029	-0.03073	0.004527	0.02327	0.02543	0.06992

Note. Green cells represent the diagonal cells of the covariance matrix. These quantities, rather than inter-parameter covariance, represent the variance observable within the specified parameter. For example, the cell 0.08013 represents the variance seen within water pressure. By contrast, the cell below, -0.03891, represents the covariance between pH and water pressure.

Section IV: Discussion and Conclusions

Analysis and Ramifications of Time Series Models

Given the wide range of R^2 values, in conjunction with the varying practicality of the quantities projected by the sinusoidal intervals observed among all environmental variables, overall model fitness appears to be moderate, with extreme variability given the parameter of interest.

Salinity has the highest R^2 value, standing at 0.847 (Table 1). In context, this suggests that about 84.7% of the variability seen in model predictions of salinity is as a result of the relationship between predictions and the temporal progression of oceanic salinity. This strong connection is visually complemented by the relative proximity of data points to the sine wave (Figure 5). This provides evidence that some parameters can in fact be reliably forecasted using a computational time series. On the other hand, water pressure has the lowest R^2 value, sitting at 0.077 (Table 1). This means that only approximately 7.7% of model predictions are a result of the relationship held with the temporal distribution of water pressure. The weak connection between the variables is made apparent by the divergence of the majority of data from the projected sinusoid (Figure 6). However, it is clear that this lack of fitness can be rectified by using an alternative function. In the case of pressure, using linear regression as opposed to sinusoidal regression increases model fitness, with R^2 rising to 0.44 (Figure 7). Pressure acts as a parameter that provides conclusions contradictory to those of salinity. Whereas salinity's results indicate the possibility of a time series to accurately model parametric values, pressure's results provide evidence of the existence of cases where the exact opposite is true. Moreover, given that an improvement of model fit was attained via using an alternative form of regression, the necessity of employing multiple types of functions to maximize model fitness, as opposed to homogeneously using one function as done in this paper, is made clear. Salinity and pressure merely represent the upper and lower bounds of forecasting capabilities. Between $R^2 = 0.847$ and $R^2 = 0.077$ lie a range of R^2 values that represent varying abilities to provide accurate forecasting as well as varying degrees of divergence of data from the centralized sinusoid.

Another indication of model fitness is the sinusoidal intervals calculated by using the offset and amplitude values of the sinusoids. In other words, the values of the peaks and troughs of the sine wave were noted. In some instances, these values aligned very closely with the data. For example, salinity has a peak-trough interval ranging from 26.388 ppt to 33.836 ppt, which encompasses the range of the majority of data (Table 1; Figure 5). By contrast, other peak-trough intervals are impractical, including negative values in contexts that do not make sense, as well as going far beyond the range of the values of the empirics being attained and modeled. For example, the troughs of nitrate concentrations reach far into negative values, trough reaching -7315.929 μmol/kg (Table 1). Concentration

levels of a substance cannot be expressed as negative values, making these predictions impractical. This limits the temporal applicability of the sinusoidal model, as certain years, when plugged into the model, would result in these quantities. Once more, a gradient among model attributes can be observed, in this case of the practicality of each parameter's sinusoidal intervals. Similar model liabilities appear to be present with total chlorophyll concentrations, whereas other parameters, such as phosphate and temperature, have sine waves whose peaks and troughs properly encompass experimental values.

Analysis and Ramifications of Linear Regression Models

An analysis of R^2 and p values among the series of linear regression models depicting the relationship of different parametric variables with chlorophyll concentrations reveals that while relationships are weak, the directional value of each relationship is statistically significant.

The highest linear regression model had an R^2 value of about 0.54, that being for the relationship of chlorophyll levels given salinity, whereas all other models have ones below 0.3 (Table 2). This means that for most of the time, less than 30% of the observed variation in chlorophyll concentrations is due to its relationship with the given environmental variable. Looking at scatterplots for both the strongest and weakest relationships, most data diverge significantly from the trendline (Figure 8; Figure 9). Nonetheless, the directionality of these relationships are still statistically significant, with the p-values for all double-tailed t-tests for linear regression being less than $\alpha = 0.05$. The single-tailed p-values, being half the amount and focused specifically on relationship directionality, provide even stronger evidence for observed positive and negative relationships between chlorophyll and parameters (Table 2, $p < 0.05$*). By seeing how each individual parameter impacts chlorophyll concentrations, crucial ecological and climatic insights can be drawn, given the pigment's role in facilitating photosynthesis, which in turn influences the transfer of trophic energy, sequestration of carbon, cycling of biogeochemical nutrients, and other important functions for the global climate and environmental systems.

Chlorophyll concentrations hold a negative relationship with pH, temperature, and salinity (Table 2, $p < 0.05$*). Indeed, past literature has noted the overall increase in global oceanic temperatures and acidity (Berwyn, 2018). Moreover, salinity is known to inhibit chloroplast activity (Hnilickova et al., 2021). The implications of rising temperature, acidity, and salinity are not simple directional impacts on phytoplankton norms. Primary production capabilities (and more broadly, other traits), increase alongside temperature, pH, as well as any other parameter, until the optimum level is reached, after which there is a decline (Dedman et al., 2023). Using chlorophyll concentrations

as an intermediate indicator, this may imply that many phytoplankton species are under conditions suboptimal for optimally performing primary production. This may be observable in the form of slower metabolic rates and other biological indicators. However, lower chlorophyll concentrations would indicate lower metabolic capabilities for phytoplankton, due to the implied dearth of resources to photosynthesize. These data provide evidence that as ocean temperatures warm, primary production in phytoplankton shall decline. This means lower levels of energy being sent up the trophic pyramid, lower rates of nutrient cycling, and the inhibition of carbon sequestration and other climate regulation processes. In addition to this overall decline, primary production levels can be expected to become increasingly heterogeneous along the spatial gradient. Salinity and nutrient concentrations are projected to become less uniformly concentrated (Berwyn, 2018). With the former parameter holding a negative relationship with total chlorophyll (Table 2), this implies that the decrease in chlorophyll will be dissimilar among locations, provided different salinity levels. A similar logic may be applied to the latter variables, which have positive relationships with total chlorophyll (Table 2). This lack of spatial homogeneity in primary production further limits ecological stability. Therefore, as oceanic parameters continue to evolve, it appears that the stability and health of climatic and ecological systems shall continue to decline.

Meanwhile, all other parameters, mainly including oxygen and various micronutrients, hold a significant positive relationship with chlorophyll concentrations (Table 2, $p < 0.05$*). Indeed, when there is a greater presence of chlorophyll, that indicates that a greater amount of photosynthesis can occur, stimulating subsequent metabolic pathways that facilitate nutrient cycling, allowing for micronutrient concentrations to grow. This tie of chlorophyll to the stimulation of micronutrients could have led to the positive relationships observed. In essence, the data may serve as support for principles in biogeochemical cycling as well as similar areas.

These results indicate the need to address ways in which rises in ocean temperature can be perturbed, as well as the need to regulate the concentrations of nutrients and other variables. In conjunction with the time series models provided above, this provides a potent source for prediction and decision-making. Knowing the impact a given level of a parameter may have on an aquatic ecosystem can be crucial for policymakers and scientists. If it is known, from a strong time series model, that a certain level of, for example, phosphorus, shall lead to a harmful amount of eutrophication (or other phytoplankton trait), a conclusion reached from observing a model from the current set being presented, then decisive policy action may taken, provided these crucial details.

Analysis and Ramifications of PCA

Based off of PCA results, as well as data provided the covariance matrix and scree plot, it is apparent that pH, followed by salinity and pressure, on a global scale, are driving parameters behind chlorophyll concentrations, and in turn, primary production capabilities. Moreover, each parameter is independent of one another. However, these conclusions are limited by the low variance coverages of PC_1 and PC_2. Being a two-dimensional PCA, much variance information was lost by solely focusing on PC_1 and PC_2. Together, these contain an eigenvector value of only about 0.33 (Figure 11). When standardized, this means that only about 60% of variance of the overall dataset is covered (Figure 10). With nearly half of the data information lost from the process, the conclusions that can be drawn have a rather limited scope.

The contribution of each parameter to the variance of both PC_1 and PC_2, providing further evidence of inadequate variance coverage. Being measured on a scale with a magnitude of 1, the highest contribution to variance coverage of PC_1, captured by pH, was only 0.523, while the highest contribution to variance coverage of PC_2, captured by water pressure, was just 0.545 (Table 3). These contributions are at best, moderate in coverage. Seeing as most contributions are lower than this value, it is clear that most parameters do not particularly account for overall data variance. Nonetheless, in calculating the magnitudes (Equation 4) and ordering the results, it was found that pH, followed by salinity and pressure, are driving parameters of total oceanic chlorophyll concentrations, and in turn, primary production and other aspects of ecology (Table 3). While this may indicate a potential need to study the impact of these parameters on phytoplankton primary production and other dynamics, such a decision must be made cautiously. This because of both the lack of variance coverage from which these results are drawn, as well as the fact omitting factors would create a less representative understanding of empirical quantities and trends.

The covariance matrix values suggest data homogeneity within parameters and independence among different factors. Among the diagonal cells of the covariance matrix, the highest observable covariance stands at 0.083, with salinity. That means, among the values serving as measures of variance, that is, spread, the highest among these values was only 0.083, on a scale of 1 (Table 4). With all intra-parameter variance values being of this small of a magnitude, it appears that data values for parameters are homogenous. This may provide evidence into the consistency of the data. While this analysis was performed on a global scale using data spanning sixty years, even at scope this broad, some level of parametric homogeneity and consistency of ocean data is implicated. The diagonal cells, consequently, provide indirect evidence for the ocean as a system with properties that have a noticeable level of stability. Moreover, all other cells have even smaller covariance values for the inter-parameter relationships, the

vast majority failing to exceed a value of 0.1 (Table 4). With highly weak covariance values, this indicates that oceanic variables may be independent of one another. This is in terms of impacting the levels among one another, rather than with regards to phytoplankton dynamics. Given the focus of this paper and the results compiled among these three sets of computational models, it is clear that parameters exert a complex net impact on phytoplankton dynamics, even if they themselves may not impact their own values.

An additional observation is that salinity, and to an extent, pH and pressure, had some combination of notable R^2, p-values, covariance, and principal component contribution values across all three sets of computational models.

In terms of policy and scientific investigation, a tool such as PCA and covariance matrices would serve as preliminary analysis tools for aquatic ecosystems. This would help establish a framework of general understanding of a given ecosystem. From this vantage point, deeper empirical trends can be performed. Subsequently, this would allow for the development of both scientific and policy-related investigations, allowing for insights to be reached.

Additional Limitations and Ramifications, and Suggestions for Future Research

Beyond what has already been discussed in the methodology as well as in analyzing the data and its ramifications, a major limitation of this study is its scope. While the dataset is highly comprehensive geographically, helping in part to rectify past limitations induced by data unavailability, there are still severe spatiotemporal limitations. As mentioned, all parametric data lacked any provision of spatiotemporal attributes. Moreover, due to the lack of direct quantifications of phytoplankton dynamics, the singular dynamic of primary production had to be indirectly assessed using total oceanic chlorophyll concentrations. This led to the data only being assessed

Nonetheless, the ideas behind the computational framework applied in this study are still viable. Having a system wherein data is provided and the changes in phytoplankton dynamics, including with their causes and effects, can be assessed is highly powerful. Many aspects of this apparatus were achieved through this study. Data was attained from WOD18, known for being one of the most comprehensive sources for oceanographic metrics. In creating numerous time series models, a potential forecasting tool was provided for scientists and policymakers alike to predict future levels of oceanographic metrics, and in turn make potential research and policy decisions. However, further iterations making use of a variety of regression models is needed. In performing linear regression on total oceanic chlorophyll concentrations, phytoplankton primary productions could be analyzed, albeit indirectly. Within this process, ecological and climatic impacts were assessed. Future work may be bolstered by using exact models for

these metrics. For example, using neural networks to model food chain interactions given changing phytoplankton dynamics, or, climatically, using the Coupled Model Intercomparison Project or other models, as done in the past (Hague & Vichi, 2018). Driving parameters and inter-parameters relationships were characterized using PCA and covariance matrices, respectively. However, due inadequate variance capture of the principal components generated and analyzed, the resulting conclusions were limited in scope. Future data investigation should be performed, and it should be endeavored to increase variance coverage. Overall, a viable and applicable apparatus for studying phytoplankton dynamics has been provided and applied. However, future iterations will improve model fitness and result applicability.

Section V: Appendices

The appendices section include links to supplementary files containing the specific material referenced within the paper.

Original Dataset

Supplementary File 1

Supplementary File 2

Supplementary File 3

Supplementary File 4

References

Amazon Web Services. (2024). *Registry of Open Data on AWS*. Registry.opendata.aws; Amazon Web Services, Inc. https://registry.opendata.aws/

Anderson, S. I., Barton, A. D., Clayton, S., Dutkiewicz, S., & Rynearson, T. A. (2021). Marine phytoplankton functional types exhibit diverse responses to thermal change. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-26651-8

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

Boyer, T.P., O.K. Baranova, C. Coleman, H.E. Garcia, A. Grodsky, R.A. Locarnini, A.V. Mishonov, C.R. Paver, J.R. Reagan, D. Seidov, I.V. Smolyar, K. Weathers, M.M. Zweng,(2018): World Ocean Database 2018. A.V. Mishonov, Technical Ed., NOAA Atlas NESDIS 87. https://www.ncei.noaa.gov/sites/default/files/2020-04/wod_intro_0.pdf

Chang, C.-W., Miki, T., Ye, H., Souissi, S., Adrian, R., Anneville, O., Agasild, H., Ban, S., Be'eri-Shlevin, Y., Chiang, Y.-R., Feuchtmayr, H., Gal, G., Ichise, S., Kagami, M., Kumagai, M., Liu, X., Matsuzaki, S.-I. S., Manca, M. M., Nõges, P., & Piscia, R. (2022). Causal networks of phytoplankton diversity and biomass are modulated by environmental context. *Nature Communications*, *13*(1), 1140. https://doi.org/10.1038/s41467-022-28761-3

Dedman, C. J., Barton, S., Fournier, M., & Rickaby, R. E. M. (2023). The cellular response to ocean warming in Emiliania huxleyi. *Frontiers in Microbiology*, *14*. https://doi.org/10.3389/fmicb.2023.1177349

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

Google Colaboratory. (2024). *Colaboratory Release Notes*. Colaboratory. https://colab.research.google.com/notebooks/relnotes.ipynb

Hague, M., & Vichi, M. (2018). A Link Between CMIP5 Phytoplankton Phenology and Sea Ice in the Atlantic Southern Ocean. *Geophysical Research Letters*, *45*(13), 6566–6575. https://doi.org/10.1029/2018gl078061

Hern, S. C., Williams, L. R., Taylor, W. D., Lambou, V. W., Morris, M. K. (1979). Phytoplankton water quality relationships in U.S. lakes. (n.p.): Environmental Protection Agency, Office of Research and Development, [Office of Monitoring and Technical Support], Environmental Monitoring and Support Laboratory.

Hnilickova, H., Kraus, K., Vachova, P., & Hnilicka, F. (2021). Salinity Stress Affects Photosynthesis, Malondialdehyde Formation, and Proline Content in Portulaca oleracea L. *Plants*, *10*(5), 845. https://doi.org/10.3390/plants10050845

Käse, L., & Geuer, J. K. (2018). Phytoplankton Responses to Marine Climate Change – An Introduction. In S. Jungblut, V. Liebich, & M. Bode (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

Loschi, M., D'Alelio, D., Camatti, E., Aubry, F. B., Beran, A., & Libralato, S. (2023). Planktonic ecological networks support quantification of changes in ecosystem health and functioning. *Scientific Reports*, *13*(1). https://doi.org/10.1038/s41598-023-43738-y

Pezzullo, J. C. (2023). *Analysis of Variance from Summary Data*. StatPages. https://statpages.info/index.html

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

Statistics Kingdom. (2017). *Multiple Linear Regression Calculator*. (October 30, 2023)[web application].https://www.statskingdom.com/410multi_linear_regression.html

Tian, Y., Zhao, Y., Zhang, X., Li, S., & Wu, H. (2023). Incorporating carbon sequestration into lake management: A potential perspective on climate change. *Science of the Total Environment*, *895*, 164939–164939. https://doi.org/10.1016/j.scitotenv.2023.164939

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