Analysis and Discussion

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 1). 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 2). 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 3). 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 1). 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 4; Figure 5). 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 7). When standardized, this means that only about 60% of variance of the overall dataset is covered (Figure 6). 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.