Insights and predictions of Phytoplankton Biomass in Equatorial Pacific using Random Forest Modeling

by

Cristian Swift
Email: Cjswift@uw.edu
University of Washington
School of Oceanography, Box 357940
Seattle WA 98195-7940
June 2nd, 2023

Abstract

Phytoplankton biomass drivers are integral to our understanding of how climate change will impact the carbon cycle as photosynthesis and carbon sequestration play a crucial role in mitigating climate change. The equatorial Pacific is a region with high primary productivity due to the convergence of easterly trade winds and equatorial currents, though few oceanographic cruises take place here. To address this lack of data, this study both assesses the predictive performance of random forest regression models for phytoplankton biomass and identify which environmental variables play the largest role in biomass. I hypothesized that nutrient concentration, salinity, and sea surface temperature are dominant determining factors of phytoplankton biomass because they represent key environmental conditions that influence the growth and distribution of phytoplankton in the world's oceans. Results suggest that *Prochlorococcus* demonstrates a large correlation between biomass and salinity, while iron plays a lesser role. Nitrate emerges as a vital predictor for *Prochlorococcus*, reflecting its role as a limiting nutrient. In contrast, *Synechococcus* relies heavily on phosphate, with nitrate contributing minimally due to specialized assimilation systems. Picoeukaryotes are primarily driven by nitrate, with salinity and temperature exerting less influence. By filling a knowledge gap in picophytoplankton biomass in the equatorial Pacific, this study contributes to our understanding of phytoplankton dynamics in the face of climate change and El Niño events. Further research is necessary to explore additional factors and enhance prediction models for a comprehensive understanding of equatorial Pacific phytoplankton dynamics.

Plain Language Summary

Phytoplankton are tiny marine organisms that play a crucial role in photosynthesis and carbon storage. Understanding what affects their growth is important for predicting how climate change will impact our planet. This study focused on the equatorial Pacific, where strong winds and currents create favorable conditions for phytoplankton because of the circulation of nutrients. However, there is limited research in this region. Using data obtained from a cruise in the equatorial Pacific in spring of 2023, computer models were built and used to predict the total mass of phytoplankton populations in the equatorial Pacific. Different sizes of phytoplankton were found to be influenced by different factors, making populations susceptible to climate change in alternate ways from each other.

Introduction

Phytoplankton biomass plays a critical role in the oceanic carbon cycle and the global climate system. Phytoplankton represent less than 1% of global photosynthetic biomass but are responsible for roughly half of the world's primary production and are a significant factor in the global carbon cycle, influencing climate patterns and atmospheric gas composition through the capture of CO2 and the production of O2 (Falkowski 2012). Phytoplankton biomass varies greatly in time and space, depending on a range of environmental factors, such as nutrient availability, light, temperature, and ocean currents (Behrenfeld et al. 2006). Changes in phytoplankton biomass can have significant impacts on the food web, carbon cycling, and the overall health of the ocean (Boyd et al. 2015). For example, low phytoplankton biomass can lead to reduced food availability for zooplankton and fish, which can in turn affect higher trophic levels and commercial fisheries (Hunt et al. 2002). Conversely, high phytoplankton biomass can result in human, ecological and economic consequences, depending on the species of phytoplankton in bloom impacts include amnesic shellfish poisoning, aquatic oxygen depletion, and closure of fisheries (Dai et al. 2023). Picophytoplankton are the smallest phytoplanktonic group ranging from 0.2-2 µm in size (Coello-Camba & Agustí, 2021). This includes *Prochlorococcus*, *Synechococcus*, and other picoeukaryotes, which are dominant in warm and oligotrophic regions of oceans (Coello-Camba & Agustí, 2021).

Picophytoplankton size is also an important factor that can impact their ecological role and influence the marine food web. Picophytoplankton are diverse physiologically with size ranges from 0.6 to 200 µm, with three size classes, picoplankton (diameter <2 µm), nanoplankton (diameter from 2 to 20 µm), and microplankton (diameter >20 µm) (Brotas et al. 2022). The rate at which phytoplankton sink is partially dependent on the organic matter's surface area to volume ratio as explained by Stokes Law, where a particle with a larger surface area to volume ratio sinks (smaller phytoplankton size) at a lower rate than a particle with a smaller surface area. Such pathways of export for picophytoplankton include aggregation and their incorporation into detritus as marine snow (Turner et al. 2015).

In the marine environment, there exists an inverse relationship between nutrient load and phytoplankton size. Smaller plankton are favored in oligotrophic oceans since they have a higher surface area to volume ratio making nutrient uptake relative to their size higher than larger phytoplankton (Acevedo-Trejos et al. 2018). Nitrate is a significant limiting nutrient for these organisms and its concentrations are correlated with other limiting nutrients such as phosphateas observed in the Redfield ratio (Tyrrell et al. 2019). These correlations make nitrate a good proxy for estimating nutrient concentrations in the water column.

Prochlorococcus and *Synechococcus* are two important examples of small phytoplankton that dominate in oligotrophic oceans and play a significant role in global primary production.

Prochlorococcus is the smallest known photosynthetic organism and is the most abundant photosynthetic organism in the world, responsible for up to 20% of global oxygen production (Partensky et al. 1999). This cyanobacterium is extremely well adapted to oligotrophic environments and has a high surface area to volume ratio, allowing for efficient nutrient uptake. *Synechococcus* is also abundant in oligotrophic oceans and can contribute up to 50% of primary production in these environments (Flombaum et al. 2013). Like *Prochlorococcus*, *Synechococcus* has a high surface area to volume ratio and is well adapted to oligotrophic environments. However, *Synechococcus* is more versatile in its nutrient requirements and can use a wider range of organic and inorganic compounds compared to *Prochlorococcus* (Wawrik et al. 2009). Both of these cyanobacteria are projected to increase in abundance in due to increased sea surface temperatures as a result of climate change (Flombaum et al. 2013). Generally, warmer temperatures can increase phytoplankton growth rates by increasing metabolic rates and nutrient uptake (Boyd et al. 2013). However, the relationship between temperature and phytoplankton biomass is complex and depends on various factors such as nutrient availability, light, and water column stability.

Nutrient availability is a critical factor that influences phytoplankton biomass in the marine environment. Among the nutrients that are essential for phytoplankton growth, nitrogen (N) and phosphorus (P) are the most commonly limiting in oligotrophic oceans (Bristow et al. 2017). The Redfield ratio, which refers to the ratio of dissolved inorganic carbon, nitrogen, iron, a N to P in phytoplankton and marine organic matter, provides a useful framework for understanding the relative importance of these nutrients for phytoplankton growth. Generally, the concentration of CO2 in seawater is not a limiting factor for phytoplankton growth, as it is present in high concentrations (Bristow et al. 2017). Nitrate (NO3-) is a significant source of N for phytoplankton, and its concentrations are correlated with other necessary nutrients such as phosphate. Accurate estimation of phytoplankton biomass is crucial for understanding the dynamics of marine ecosystems and their response to environmental changes. However, traditional methods for predicting phytoplankton biomass based on empirical models have limited accuracy and generalization, especially in remote regions such as the Equatorial Pacific (Friedrichs et al. 2009).

The Equatorial Pacific (5°S to 5°N latitude and 170°W to 120°W longitude) is a pelagic region of the Pacific Ocean and it's known for its unique features that support relatively high levels of primary productivity (Chavez et al. 2011). This productivity is largely driven by the upwelling of nutrient-rich waters from the deep ocean to the surface, which provides a source of nutrients for phytoplankton growth (Karl and Lukas, 1996). The upwelling is caused by the convergence of the easterly trade winds and the equatorial currents, which create a region of intense mixing and nutrient exchange between the surface and deeper waters (Chavez et al. 2011). The pelagic waters of the north and south of the Pacific are generally characterized by low concentrations of nutrients and primary production, often referred to as a marine desert (Morel et al. 2010). This area is influenced by the El Niño and the Southern Oscillation (ENSO). ENSO is a large-scale interaction between the ocean and the atmosphere, and it cycles between El Niño, "normal", and La Niña. El Niño is characterized by weakened equatorial easterlies winds, this results in negative sea level pressure and reduced upwelling. This results in increased atmospheric temperature along with higher SST due to a lack of cold, nutrient rich water upwelling (Wang & Fiedler, 2006). ENSO is most concentrated near the equator in the Pacific, with the largest changes in SST occurring there, and its variability has been increasing since the industrial revolution (Figure 1).

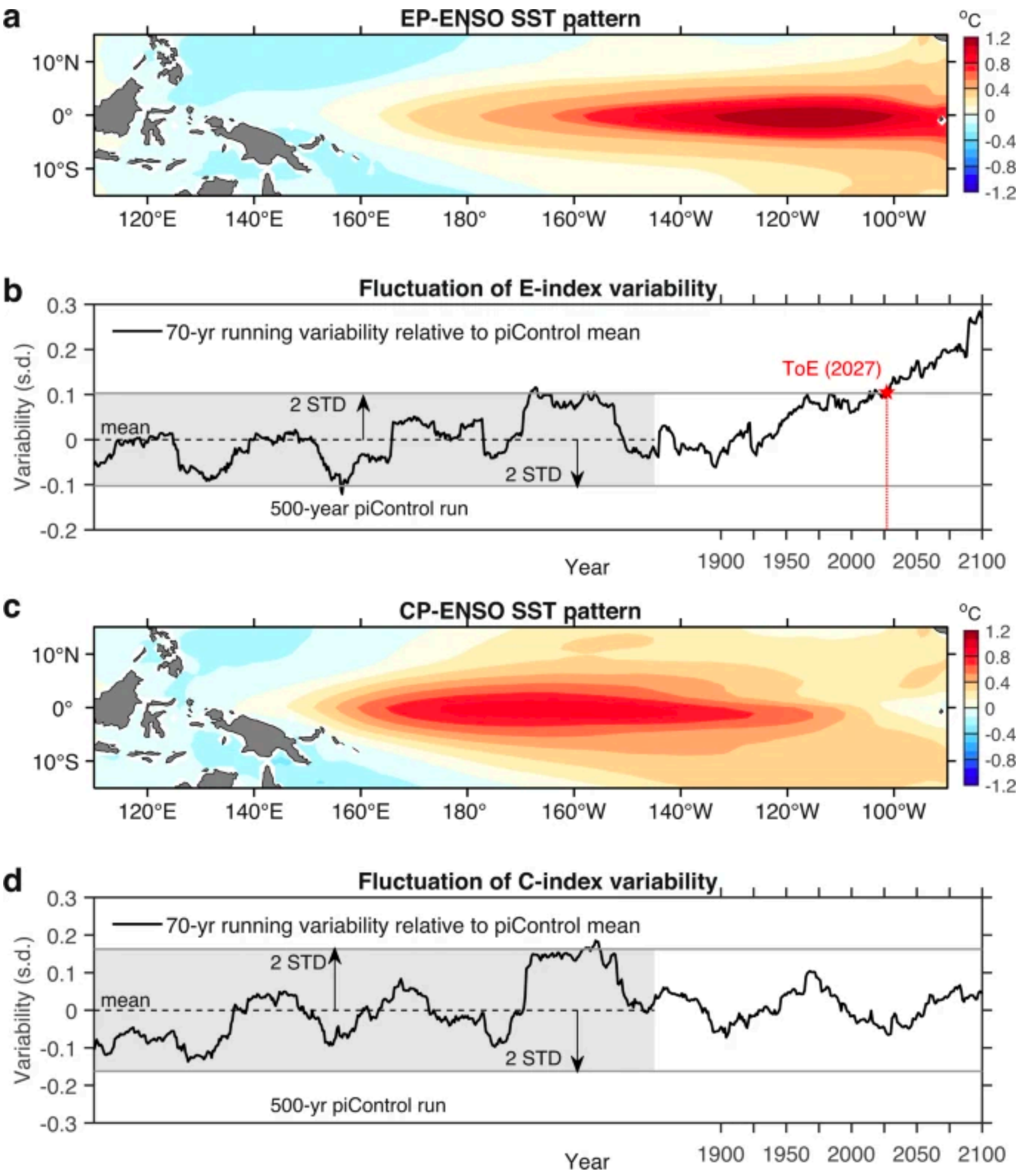

Figure 1: (Geng et al. 2022) El Niño SST variability from climate model GFDL-ESM4. Panel a shows the spatial pattern of ENSO in the eastern equatorial Pacific. Panel B describes El Niño variability relative to the mean of the E-index variability from pre-industrial levels. Panel C and D describe the same as panels A and B but relative to the central Pacific.

Few oceanographic research cruises take place in the equatorial Pacific, leaving a large gap of knowledge in the region. Especially considering global climate change leading to higher sea SST globally and increasingly frequent and intense El Niño events, it is important to understand and predict how biomass is changing and impacted by the changing climate. Satellite data has shown to be a promising remote sensing tool to predict phytoplankton biomass and random forest models have been show to be effective in predicting phytoplankton biomass (Zhang et al. 2021). Though there is a lack of existing models for the equatorial Pacific.

Through predicting biomass, we could expand biomass measurements on previous research cruises that did not have the instruments to collect this data. biomass is an important observation that can inform us of sequestration of surface carbon to the deep ocean in addition to environment nutrient load, photosynthetic efficiency, and even the marine food web structure. By expanding our knowledge of biomass through previous cruises taking place in the equatorial Pacific, we can gain a better understanding of inferred phytoplankton community changes over time.

To further address of data in the region for picophytoplankton biomass to be directly measured, a blend of remote sensing and in situ measures from a Pisces climatological model was used to create random forest models to predict the biomass of *Prochlorococcus*, Synechococcus, Eukaryotes. I hypothesized that Nutrient concentration, salinity, and sea surface temperature are dominant determining factors of phytoplankton biomass because they represent key environmental conditions that influence the growth and distribution of phytoplankton in the world's oceans.

Materials and methods:

Data Retrieval:

The data used for testing and training the model were obtained from R/V Thompson cruise TN413 (Figure 2) and Simon CMAP Pisces climatological data. TN413 departed from Honolulu, HI, on February 24th, 2023, and arrived in Suva, FJ, on March 12th, 2023, running a transect along the meridional section across the equator from 5°N to 5°S along 180. Temperature and salinity were collected in situ during the cruise from a real underway CTD. Phytoplankton biomass was calculated from SeaFlow observations of cell diameter and abundance. SeaFlow is an onboard flow cytometer in continuous operation, which enables shipboard monitoring of the quantity and optical characteristics of small phytoplankton (with equivalent spherical diameter less than 5 μm) (Francois Ribalet et al. 2019). To get phytoplankton biomass for each row of data, first calculate spherical volume from cell diameter ($\text{Biovolume}=4/3 *\pi * r^3$) then by calculating carbon quotas ($Qc = 0.261 * \text{Biovolume}^{0.86}$), and finally multiplying carbon quotas by cell abundance for each row. The other features that will be used in the random forest models are nitrate, phosphate, and iron concentrations. These were retrieved from a Simons CMAP query using the Mercator-Pisces Biogeochemistry Daily Forecast (cl1) model with colocalization done with TN413 Latitude and Longitude values. The Pisces had a spatial resolution of Spatial Resolution 1/4° X 1/4° and was captured at depths between 0.5 and 8 meters.

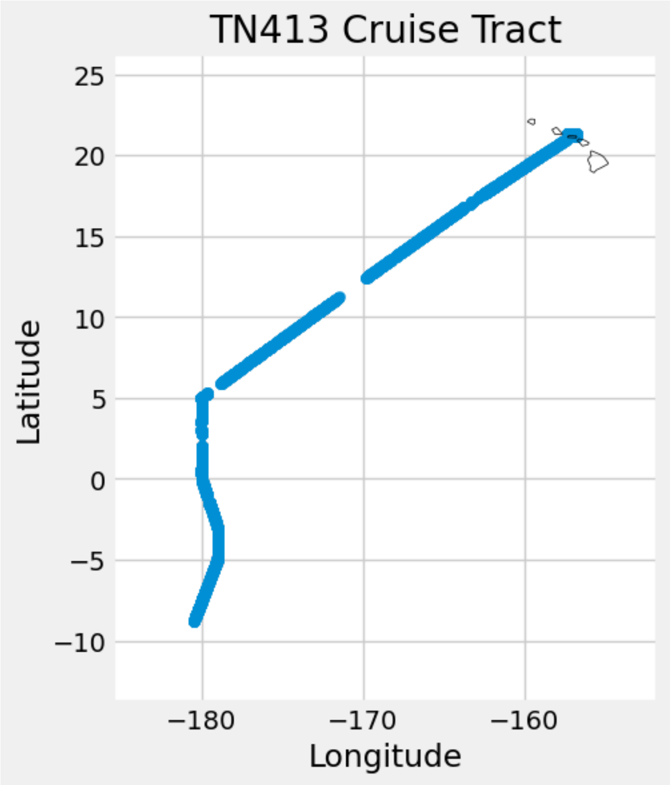

Figure 2: TN413 cruise tract from Honolulu, Hawaii to Suva, Fiji. Represents data used from the cruise with blank spots being where Nan values were detected.

Random Forest Regression Model:

Random forest regression is a machine learning algorithm commonly used for modeling and predicting various oceanographic variables such as sea surface temperature, chlorophyll-a concentration, and dissolved oxygen concentration. Random forest regression builds an ensemble of decision trees based on random subsamples of the training data, with each tree being grown to the maximum extent possible without pruning (Breiman, 2001). The algorithm is particularly useful for capturing nonlinear relationships and interactions between variables. (Huang et al. 2022). However, random forest models may not perform well with data values that are outside the range of values it has seen during the training process. This is because the decision tree splits are based on the features of the training data, and if there are values in the test data that are vastly different from those seen during training, the model may struggle to make accurate predictions.

I will use the RF regressor model since we are predicting biomass, which is a continuous measurement. Using anaconda python software, the RF will be a regression model from sklearn.ensemble.RandomForestRegressor. 70 percent of the data will randomly be used for training, and 30 percent of the data will be tested on by the resulting model which is a standard ratio used in random forest. The features are temperature, salinity, nitrate, iron and phosphate. SeaFlow data was divided into different phytoplankton populations (*Prochlorococcus*, Synechococcus, and Eukaryotes) and each population was trained on individual random forest regression models with the same features. In this study, the Random Forest model used is sklearn.ensemble.RandomForestRegressor, it lets us knows which features are most relevant for predicting biomass, this is referred to as "feature important", this study will use this as a way to talk about which environmental variables are most important to the different picophytoplankton. All code for this project is available in the appendix.

Results

Each phytoplankton population had different importance values assignment to each feature (Fig. 3). The *Prochlorococcus* model assigned about 27% of weight in predicting the biomass to salinity, meaning salinity was the most significant factor in predicting *Prochlorococcus* biomass. Nitrate was similar and followed at 26% and the other features at or below 18% (Fig. 3A). The *Synechococcus* model assigned about 60% of weight in the decision nodes to phosphate and 12% or less to other features (Figure 3B). The Eukaryotes model assigned about 45% of decision weights to nitrate, about 22% to phosphate, and 12% or lower to other features (Fig 3C). Salinity was only a significant predictor for *Prochlorococcus*, with the other models determining it to be only around 10% of weight in decisions for biomass. Temperature was the second largest contributor to predicting phytoplankton biomass in both *Prochlorococcus* and *Synechococcus* and ranked low for prediction Eukaryotes biomass. No models ranked iron very high as an important predictor for biomass.

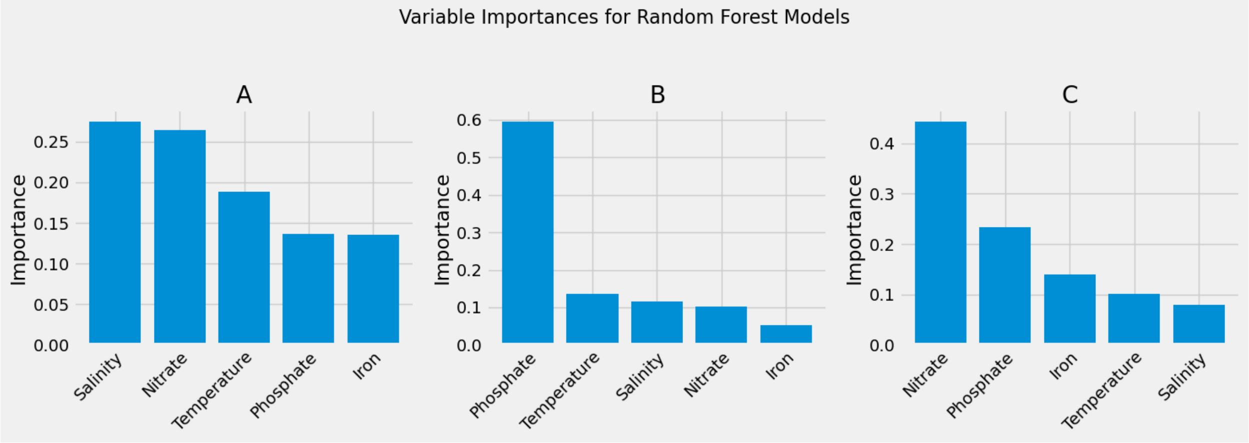

Figure 3: Feature importance values for determining what had the most influence on predicting phytoplankton biomass in the random forest model. Panel A represents Prochlorococcus, panel B represents Synechococcus, and panel C represents Eukaryotes (<2.5 micron µm. All feature importance's summed equate to 1.

Applying the phyto-specific models to the entire cruise dataset allows us to compare how biomass changes over latitude and at what latitude the model falls short on its predictions. In observed biomass there is an increase from -10º to 10º. Near Hawaii biomass was relatively low for all observed biomass of the phytoplankton from 10º to 20º. (figure 4).

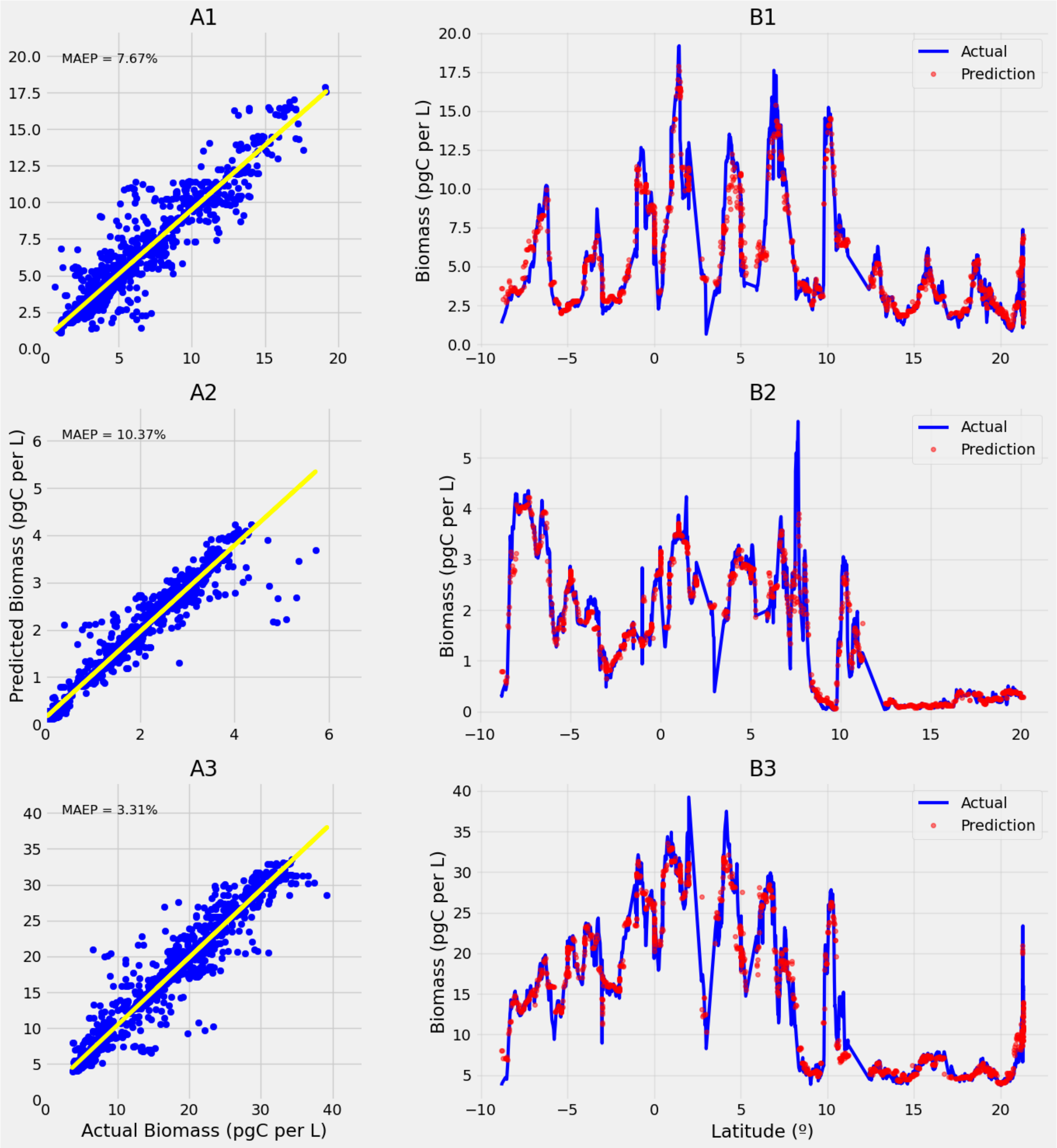

Figure 4: Panels in column A are predicted versus observed biomass with Mean Absolute Error Percentage (MAEP) displayed. Panels B represent latitude versus predicted and actual biomass over cruise tract latitudes random forest predictions versus observed biomass, blue line represents observed values from SeaFlow, and the red dots represent random forest predictions. Row one is representative of Prochlorococcus, row two is Synechococcus, and row three is Picoeukaryotes.

To assess the performance of the model predicted and observed biomass were plotted together to observe how linear their relationship (Figure 4 column A). In addition, actual and predicted biomass for each picophytoplankton population were mapped over latitude to determine where predictions along the cruise tract fall short and succeed (Figure 4 column B). The biomass of *Prochlorococcus* exhibited a wide range during the cruise, spanning from 2.5 pgC per liter to 19 pgC per liter, with the highest biomass observed between 0º to 10º latitude over the cruise tract (Figure 4A1). *Prochlorococcus'* random forest model has a Mean Absolute Error Percentage of 7.67%, meaning for a given value on average was off by 7.67% (Figure 4A1). In contrast, the biomass values of Synechococcus were overall lower than *Prochlorococcus*, ranging from near zero to 5.5 pgC per liter, with its highest biomass observed throughout -9º to -8º latitude (Figure 4B2). And Synechococcus' model had a MAEP of 10.37%, the highest among the three models (Figure 4A2).The Eukaryotes showed the greatest variability in biomass, ranging from 5 pgC per liter to 40 pgC per liter with their biomass being highest between latitudes -3º to 5º (Figure 4B3 (Figure 4B3). Picoeukaryotes also displayed the lowest MAEP, with it being 3.31% (Figure 4A3).

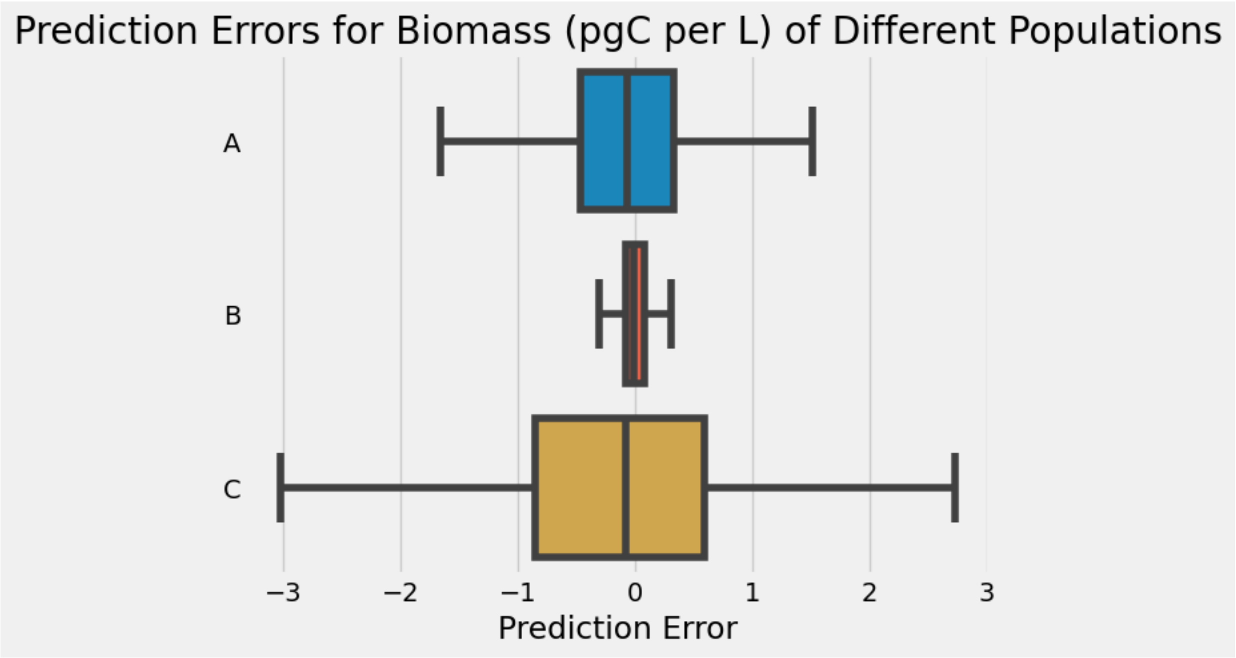

Figure 5: box and whisker plot with prediction errors for each phytoplankton population represented in SeaFlow. Errors calculated by subtracting observed phytoplankton biomass by random forest predicted phytoplankton biomass. Panel A represents Prochlorococcus, panel B represents Synechococcus, and panel C represents picoeukaryotes.

To further investigate model performance, predictions errors were calculated by subtracting predicted values from the observed values (Figure 5). *Prochlorococcus'* mean error was at -0.2 pgC per L and 95% the first quantile lies at -0.5 pgC per L and the third quantile lies at 0.3 pgC per L (Figure 5A). Synechococcus' mean error was at -0.1 pgC per L and 95% the first quantile lies at -0.2 pgC per L and their quantile lies at 0.1pgC per L (Figure 5B)**.** Picoeukaryotes mean error was at -0.2 pgC per L and 95% the first quantile lies at -0.8 pgC per L and their quantile lies at 0.6 pgC per L (Figure 5C).

Discussion:

This analysis demonstrates the effectiveness of random forest regression models for predicting phytoplankton biomass in the equatorial Pacific. All three phytoplankton populations analyzed had stark differences in their importance rankings (Figure 3). *Prochlorococcus*, Synechococcus, and Eukaryotes occupy different niches of the photosynthetic zone due to their varying sizes size (Coello-Camba & Agustí, 2021). This highlights the different varying drivers of biomass for *Prochlorococcus*, *Synechococcus*, and picoukaryotes (Falkowski, 2012).

Prochlorococcus' most influential environmental variable in the TN413 cruise tract was salinity (Figure 3A). *Prochlorococcus* chl a measurements and salinity have a good positive correlation as a previous study has found (Wang et al. 2022). Though this relationship is not shared with other picophytoplankton populations such as *Synechococcus* and picoeukaryotes (Figure 3). Iron was the least important predictor for *Prochlorococcus* biomass (Figure 3A), this is backed by *Prochlorococcus'* specialization in low iron waters in the equatorial Pacific (Rusch et al. 2010). Nitrate is also an important predictor for *Prochlorococcus* (Figure 3A), *Prochlorococcus*, Nitrate is a generally important for phytoplankton growth and often a limiting nutrient (Bristow et al. 2017).

Synechococcus does not share an important relationship with nitrate (Figure 3B), *Synechococcus* have Nitrate/Nitrite assimilation systems, allowing them to tolerate low nitrate levels in the water (Bird and Wyman 2003). *Synechococcus'* most important feature was by far phosphate at 60% (Figure 3B). In a study done in the Mediterranean Sea in the summer, also a warm generally oligotrophic body of water, they found *Synechococcus* were more P than N limited for their cell growth (Vaulot et al. 1996). The other features were not found to be nearly as important for *Synechococcus* at 12% or less.

Picoeukaryotes largest biomass driver is nitrate, in this study Eukaryotes were generalized due to limited data available, but they are a diverse group of phytoplankton when it comes to their size. And since these organisms are larger, they have a smaller surface area to volume ratio, making their nutritional requirements much higher than *Prochlorococcus* or *Synechococcus* (Wawrik et al. 2009). Salinity and temperature were found to be the least influential features for *Synechococcus* (Figure 3C), this is in contrast to what other studies which have found temperature to relational to biomass (Wang et al. 2022). Though this is only in relation to other features, which are all found to be generally important.

The model performed well with the *Prochlorococcus'* Mean Absolute Error Percentage (MAEP) being 7.67% (Figure 3A), *Synechococcus* at 10.37% (Figure 3B), and picoeukaryotes have the lowest MAEP score of 3.31% (Figure 3C). MAEP described, on average, the absolute value of how off the predictions were from the observed biomass. The predictions vs observations plot (Figure 3) also had fairly linear relationships for all the picophytoplankton populations, supporting the MAEP scores.

The application of the model to the entire dataset to predict on allowed us to analyze how well the model can predict along latitude for cruise TN413. The observed biomass showed substantial variability in the range of -10º to 10º latitude (Figure 4), likely due to the convergence of the easterly trade winds and the equatorial currents causing upwelling between the surface and deeper waters (Chavez et al. 2011). Biomass levels near Hawaii remained relatively low between 10º and 20º latitude likely because of its limited nutrients available (Morel et al. 2010).

The observed biomass showed substantial variability in the range of -10º to 10º latitude due to the convergence of the easterly trade winds and the equatorial currents causing upwelling between the surface and deeper waters (Chavez et al. 2011).). According to Behrenfeld et al. (2006), the equatorial region presents unique challenges for accurately predicting phytoplankton biomass due to the complex interactions between various environmental factors. The equator experiences significant changes in nutrient availability, and water column stability, which can affect the growth and distribution of phytoplankton. This is where we also saw the most off predicted values for all three plankton populations (Figure 4). This may suggest that there is another feature(s) that could be useful in predicting these populations biomass. My hypothesis was overall supported, with all random forest models for the three picophytoplankton populations being able to predict biomass well given nutrient, temperature, and salinity environmental data.

It is worth noting that our study focused on the equatorial Pacific, an area with limited research cruises and a significant knowledge gap in phytoplankton biomass. This gap in understanding is especially critical given the increasing sea surface temperatures and the frequency and intensity of El Niño events associated with global climate change (Flombaum et al. 2013). By expanding our knowledge of phytoplankton biomass in the equatorial Pacific, we can gain a better understanding of inferred phytoplankton community changes over time and their response to the changing climate and ENSO events (Wang & Fiedler, 2006).

Conclusion

In conclusion, this study highlights the effectiveness of random forest regression models in predicting phytoplankton biomass in the equatorial Pacific. The random forest models performed well in predicting phytoplankton biomass, as indicated by the low Mean Absolute Error Percentage (MAEP) scores for *Prochlorococcus* (7.67%), *Synechococcus* (10.37%) , and picoeukaryotes (3.31%). The analysis reveals the importance of salinity, phosphate, and nitrate as drivers of biomass for *Prochlorococcus*, *Synechococcus*, and picoeukaryotes, respectively. These findings contribute to our understanding of picophytoplankton biomass in this region, addressing a knowledge gap. The implications of this research are significant, considering the impact of climate change and El Niño events. Further research is needed to explore additional factors and improve prediction models for a more comprehensive understanding of phytoplankton dynamics in the equatorial Pacific.

Acknowledgments

I would like to acknowledge François Ribalet and Ginger Armbrus for being my mentor and giving me great guidance in my thesis work. I would also like to thank the teaching staff of the School of Oceanography for the 2023 school year for their help in reviewing this paper and being with us during the Senior Cruise aboard the R/V Thompson. I would also like to thank the crew of the R//V Thompson for allowing us to be their guest in their home for 3 weeks, feeding and supporting us all the way through. This project was funded by both the University of Washington School of Oceanography and the Leo Maddox Research Scholarship, thank you.

References

Behrenfeld, M. J., O’Malley, R. E., Siegel, D. S., McClain, C. R., Sarmiento, J. L., Feldman, G. C., et al. (2006). Climate-driven trends in contemporary ocean productivity, *444*(7120), 752–755. https://doi.org/10.1038/nature05317

Bird, C., & Wyman, M. R. (2003). Nitrate/NitriteAssimilation System of the Marine Picoplanktonic Cyanobacterium *Synechococcus* sp. Strain WH 8103: Effect of NitrogenSource and Availability on GeneExpression, *69*(12), 7009–7018. https://doi.org/10.1128/aem.69.12.7009-7018.2003

Bowling, C. (1994). Habitat and size of the Florida crown conch (Melongena corona Gmelin): Why big snails hang out at bars, *175*(2), 181–195. https://doi.org/10.1016/0022-0981(94)90025-6

Bristow, L. A., Mohr, W., Soeren Ahmerkamp, & Marcel. (2017). Nutrients that limit growth in the ocean, *27*(11), R474–R478. https://doi.org/10.1016/j.cub.2017.03.030

Buitenhuis, E. T., Corinne Le Quéré, Olivier Aumont, Beaugrand, G., Bunker, A., Hirst, A. G., et al. (2006). Biogeochemical fluxes through mesozooplankton, *20*(2), n/a-n/a. https://doi.org/10.1029/2005gb002511

Coello-Camba, A., & Agustí, S. (2021). Picophytoplankton Niche Partitioning in the Warmest Oligotrophic Sea, *8*. https://doi.org/10.3389/fmars.2021.651877

Dai, Y., Yang, S., Zhao, D., Hu, C., Xu, W., Anderson, D. M., et al. (2023). Coastal phytoplankton blooms expand and intensify in the 21st century, *615*(7951), 280–284. https://doi.org/10.1038/s41586-023-05760-y

Dunne, J. P., Armstrong, R. A., Gnanadesikan, A., & Sarmiento, J. L. (2005). Empirical and mechanistic models for the particle export ratio. *Global Biogeochemical Cycles*, *19*(4), n/a-n/a. https://doi.org/10.1029/2004gb002390

Falkowski, P. G. (2012). Ocean Science: The power of plankton, *483*(7387), S17–S20. https://doi.org/10.1038/483s17a

Flombaum, P., Gallegos, J. L., Gordillo, R. A., Rincon, J., Zabala, L. L., Jiao, N., et al. (2013). Present and future global distributions of the marine Cyanobacteria *Prochlorococcus* and *Synechococcus*, *110*(24), 9824–9829. https://doi.org/10.1073/pnas.1307701110

Francois Ribalet, Berthiaume, C. T., Hynes, A. M., Swalwell, J. E., Carlson, M., Clayton, S., et al. (2019). SeaFlow data v1, high-resolution abundance, size and biomass of small phytoplankton in the North Pacific, *6*(1). https://doi.org/10.1038/s41597-019-0292-2

Huang, H., Wang, W., Junping Lv, Liu, Q., Liu, X., Xie, S., et al. (2022). Relationship between Chlorophyll a and Environmental Factors in Lakes Based on the Random Forest Algorithm, *14*(19), 3128–3128. https://doi.org/10.3390/w14193128

L., B. (2016). Breiman, L. (2001) Random Forests. Machine Learning, 45, 5-32. - References - Scientific Research Publishing. Retrieved May 16, 2023, from https://www.scirp.org/(S(czeh2tfqw2orz553k1w0r45))/reference/referencespapers.aspx?referenceid=1734556

Rusch, D. B., Martiny, A. C., Dupont, C. L., Halpern, A. L., & J. Craig Venter. (2010). Characterization of *Prochlorococcus* clades from iron-depleted oceanic regions, *107*(37), 16184–16189. https://doi.org/10.1073/pnas.1009513107

Stawiarski, B., Buitenhuis, E. T., & Corinne Le Quéré. (2016). The Physiological Response of Picophytoplankton to Temperature and Its Model Representation, *3*. https://doi.org/10.3389/fmars.2016.00164

Turner, J. T. (2015). Zooplankton fecal pellets, marine snow, phytodetritus and the ocean's biological pump. *Progress in Oceanography*, *130*, 205–248. https://doi.org/10.1016/j.pocean.2014.08.005

Tyrrell, T. (2019). Redfield Ratio. *Encyclopedia of Ocean Sciences*, 461–472. https://doi.org/10.1016/b978-0-12-409548-9.11281-3

Vaulot, D., N Lebot, Marie, D., & Fukai, E. (1996). Effect of Phosphorus on the Synechococcus Cell Cycle in Surface Mediterranean Waters during Summer, *62*(7), 2527–2533. https://doi.org/10.1128/aem.62.7.2527-2533.1996

Wang, F., Wei, Y., Zhang, G., Zhang, L., & Sun, J. (2022). Picophytoplankton in the West Pacific Ocean: A Snapshot, *13*. https://doi.org/10.3389/fmicb.2022.811227

Zhang, J., Zhi, M., & Zhang, Y. (2021). Combined Generalized Additive model and Random Forest to evaluate the influence of environmental factors on phytoplankton biomass in a large eutrophic lake, *130*, 108082–108082. https://doi.org/10.1016/j.ecolind.2021.108082

Appendices

https://github.com/CristianSwift/SeniorThesis