ST412/512 FINAL PROJECT

Triton Desilva, Chandra La, Ugo Labbé

Introduction

Marine algae are essential to marine health; changes in algae abundance produce changes in the function, structure, and quality of marine relationships. Many factors affect algae, including nutrient levels (silicate, ammonia, nitrate/nitrite, phosphate), radiance, salinity, and temperature. A useful measure, chlorophyll a, has historically been used as an indication of the amount of algae present in an area. By investigating these factors from archived datasets[1-5] across the Pacific Arctic region, this study will attempt to answer the following question: What factors have a significant effect on chlorophyll a?

It is Important to note that assumptions will be made to analyze data. One assumption is that temporal and spatial factors have no effect. Though the data spans from 2015 - 2019, all data was taken within the month of July so that seasonal variation was minimized. Thus, this assumption is fairly reasonable. Spatial variations however may be present because different regions may vary in community composition, ultimately influencing chlorophyll levels. Therefore, this assumption is less reasonable and future data analysis should consider separating observations spatially.

Methods

The methods of analysis that we will be using is the Multiple Linear Regression, aiming to provide a model able to predict the most accurately possible the response variable, chlorophyll a, using the different explanatory variables: the different nutrient levels (silicate, ammonia, nitrate/nitrite, phosphate), salinity, radiance, salinity, and temperature.

The sample used had 1052 individuals.

We first need to assess the different assumptions of the data used : The linearity, constant variance, normality and independence of the observations.

Using RStudio, I couldn't assess the different assumptions with the base response variable, but transforming it with its log, the different conditions were met.

We checked the linearity and constant variance using a residual against fitted values plot, while we used a normal quantile-quantile plot for the normality condition. We can also say that each observation is taken independently.

To select the best model to predict the value of chlorophyll a, the stepwise method was used, using the lowest AIC (Akaike Information Criterion) value.

We also wanted to remove any influential point, if there was any. We used case influential statistics to see if any values had to be removed to improve the model. However, no point was found to be influential.

Results

Based on a significance level of 0.05, the results indicated that the following variables had a significant effect on the amount of chlorophyll a: silicate, ammonia, PAR, salinity, and temperature. The results also indicated the following significant interactions: silicate and nitrate/nitrite, silicate and PAR, silicate and salinity, silicate and temperature, nitrite/nitrate and ammonia, phosphate and PAR, phosphate and temperature, and salinity and temperature. The model created had an Adjusted R^2 of 0.3829, meaning that 38.29% of the variance in the target field is explained by the different inputs.

Interpreting the model, some interpretations that we can do are : an increase of temperature keeping all other variables the same increases the predicted transformed log of chlorophyll a, same goes with the salinity, while an increase of ammonia will reduce the predicted transformed log of chlorophyll a.

Summary

Our results show that silicate, ammonia, PAR, salinity, and temperature are all significant variables in explaining chlorophyll a levels. This is expected since algae, the chlorophyll a producer, is known to be highly dependent on these common biotic and abiotic factors. Interestingly, we also saw that not all of the common micronutrients that support life are significant in chlorophyll a levels. This may be due to the specific nutrient needs of the algae organisms that data was collected on. We also saw many of the interactions between variables being significant in predicting the response as well. This is likely due to some level of multicollinearity present in our explanatory variables, which is expected here since the presence of some of these growth factors is likely dependent on others. When fitting a model to the dataset, we faced difficulties with assessing whether or not the basic multiple linear regression assumptions were met. To solve this problem, we transformed our data using a log transformation which allowed us to confirm that our assumptions were met. Based on our data and conclusion, further research on algae would benefit our understanding of the primary producers in ocean ecosystems. Therefore, the question of how concentrations of chlorophyll a affect organisms present on different trophic levels would be appropriate for further research. As is common in many studies concerning biological life, we may have disincluded variables that do affect chlorophyll a levels without even knowing it. To try to minimize this problem, we included all data collected on commonly known nutrients used by algae. It may also be beneficial to see how our conclusions might change if we used a different strategy for selecting our model. For example, a model selected using BIC or Mallows Cp may lead us in a different direction, so further study can also still be done on this dataset.

Appendix/Sources

```
Coefficients:
                           Estimate Std. Error t value Pr(>|t|)    
(Intercept)              -3.184e+01  3.678e+00  -8.658  < 2e-16 ***
Silicate                  5.852e-01  2.377e-01   2.462 0.013981 *  
Nitrite_Nitrate          -7.225e-01  3.888e-01  -1.858 0.063396 .  
Phosphate                 2.401e-01  2.462e-01   0.975 0.329647    
Ammonia                  -1.233e-01  6.210e-02  -1.986 0.047343 *  
PAR                       1.993e-03  9.097e-04   2.191 0.028690 *  
Salinity_2                9.963e-01  1.179e-01   8.453  < 2e-16 ***
Temp_2                    1.518e+00  5.318e-01   2.854 0.004401 ** 
Silicate:Nitrite_Nitrate -3.000e-03  6.904e-04  -4.345 1.53e-05 ***
Silicate:Phosphate       -1.649e-02  8.987e-03  -1.835 0.066768 .  
Silicate:Ammonia         -1.030e-02  5.284e-03  -1.950 0.051484 .  
Silicate:PAR              4.219e-04  1.207e-04   3.495 0.000495 ***
Silicate:Salinity_2      -1.665e-02  7.505e-03  -2.218 0.026744 *  
Silicate:Temp_2          -6.279e-03  1.761e-03  -3.565 0.000380 ***
Nitrite_Nitrate:Ammonia   2.945e-02  1.166e-02   2.526 0.011691 *  
Nitrite_Nitrate:PAR       3.160e-04  1.905e-04   1.659 0.097438 .  
Nitrite_Nitrate:Salinity_2  2.397e-02  1.225e-02   1.957 0.050670 .  
Phosphate:PAR            -1.257e-02  2.476e-03  -5.076 4.58e-07 ***
Phosphate:Temp_2          7.941e-02  3.696e-02   2.148 0.031920 *  
Salinity_2:Temp_2        -4.808e-02  1.706e-02  -2.818 0.004922 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9123 on 1032 degrees of freedom
Multiple R-squared:  0.3941,	Adjusted R-squared:  0.3829 
F-statistic: 35.33 on 19 and 1032 DF,  p-value: < 2.2e-16
```

Figure 1: Model summary output.

[1]Lee W Cooper, Jacqueline M Grebmeier, Karen E Frey, & Svein Vagle. (2019). *Nutrient, oxygen 18/16 ratio, and chlorophyll data measured from water samples collected aboard the Sir Wilfred Laurier, Bering and Chukchi Seas, 2015*. Arctic Data Center. doi:10.5065/D6QN6544, version: doi:10.18739/A20000081.

[2]Lee W Cooper, Jacqueline M Grebmeier, Karen E Frey, & Svein Vagle. (2019). *Discrete water samples collected from the Conductivity-Temperature-Depth rosette at specific depths, Northern Bering Sea to Chukchi Sea, 2016*. Arctic Data Center. doi:10.18739/A23B5W875.

[3]Lee W Cooper, Jacqueline M Grebmeier, Karen E Frey, & Svein Vagle. (2020). *Discrete water samples collected from the Conductivity-Temperature-Depth rosette at specific depths, Northern Bering Sea to Chukchi Sea, 2017*. Arctic Data Center. doi:10.18739/A2P843X00.

[4]Lee W Cooper, Jacqueline M Grebmeier, Karen Frey, & Sarah Zimmerman. (2022). *Discrete water samples collected from the Conductivity-Temperature-Depth rosette at specific depths, CCGS Sir Wilfrid Laurier, Northern Bering Sea to Chukchi Sea, 2018*. Arctic Data Center. urn:uuid:29de9d96-514a-41ed-9151-e0303ccdcb95.

[5]Lee W Cooper, Jacqueline M Grebmeier, Karen Frey, & Sarah Zimmerman. (2022). *Discrete water samples collected from the Conductivity-Temperature-Depth rosette at specific depths, CCGS Sir Wilfrid Laurier, Northern Bering Sea to Chukchi Sea, 2019*. Arctic Data Center. urn:uuid:50f210c0-6529-4887-9db5-05a3bbe36069.