Bruno Lopez

# Summer 2020 Reflection & Research

This summer I aimed to improve my data analysis skills because I felt like I wasn't progressing at a fast enough pace. Looking back at my paper 'Goals for Summer 2020' I felt like I had a good idea of what I wanted to do but wasn't exactly sure how to achieve it. Shifting career goals is something that I had thought about for a long time and, I finally decided to do it.

One of the goals I had set for myself this summer was to improve my Python and R skills while I worked on my project. The project that I worked on this summer was titled "Characterization of the Phytoplankton Phenology in the Subarctic Pacific Ocean". In this project, I compared two different methods quantifying at what time of the year does a phytoplankton blooms begin initiating off of the Gulf of Alaska. I will describe this project in greater detail later further into the paper.

Immediately when I started the project I was introduced to some programming concepts in R that I was unfamiliar with. I was very excited because right from the start I was learning new concepts. As time went on this trend continued and, I gained a lot of experience working with real-world data. My R skills improved rapidly and, I was happy with the progress that I was making. R wasn't the only language that I wanted to improve upon. I also wanted to strengthen my Python skill which was the weakest tool I had in my computer software toolbox. There were some tasks throughout my internship in which I found my struggling working in R which I knew I could do in Python but, I just did not know-how. I spent about a week reading textbooks and watching videos and, I finally had the skillset I needed to accomplish the task. I automated some file downloads from the internet using my new skills. It made the work that I was doing run about 10x faster.

Overall throughout the internship, I gained this programming knowledge which was the main goal I had set for myself this summer. As the program went on my goals began to shift in priority. I discovered that although programming was still near the very top of my goals my main goal had changed. I realized to become a true data scientist I would need to have a combination of programming mixed with analytical, writing, and presentation skills. Throughout this internship, these skills began to manifest themselves as I began to think of myself more as a data analyst than a statistical programmer. I have always been rather shy but, I worked on my presentation skills and was able to give some killer presentations.

Another goal I had set for myself this summer was to get the feel of how I would like to work 40 hours a week from 9-5. Honestly, at the beginning of the internship, this was something that I was dreading. I felt like I wouldn’t have enough time to do the things that I like to do. As the internship went on I found myself enjoying this schedule. There were a couple of reasons why I think I enjoyed working on this schedule as time went on. The first being that during the day, I spend many hours doing data analysis of some sort anyways. Being able to work with real-world data motivates me even further. I found myself completing my tasks well before my mentor thought I would. The second reason was that since there was no homework after 5 pm I could use this time to focus on my hobbies. I had felt drained because of school but this internship helped me rekindle my passion for data analysis. I often found myself excited to show my mentor my results.

The last goal I had for myself this summer was to network. Before this internship, I had limited contacts in the real world. I needed to expand upon this network especially, with other data analysts. I had gained confidence because I had to have conversations with multiple people every week during the summer. An unintended effect of this was that I was able to have smoother communications with others. Specifically when I was reaching out to recruiters or other researchers. I combined these skills with my data analysis skills checked off every box that I had set for myself this summer.

Something I wish I had learned at the beginning of Summer was how much work goes into setting up a project. My mentor and I often found ourselves changing the schema of our project when we encountered an issue. This happened multiple times, it made me realize the importance of having multiple back up options if something wasn't working correctly. I wished I had set up a time to do this when I began my project because I know I could have been even more productive than I already was.

## Research

This summer I worked on analyzing the phytoplankton phenology in the Subarctic Pacific Ocean. The main goal of this project was to see when Phytoplankton blooms began initiating. It was previously thought these blooms occurred at the beginning of Spring. I was testing out a hypothesis from the paper 'Phytoplankton Phenology in the North Atlantic: Insights From Profiling Float Measurements' (Yang et al 2019).

When quantifying phytoplankton biomass three variables are important for defining the beginning of the phytoplankton bloom. The first variable is the growth rate (μ) which, is influenced by characteristics such as light and nutrients in the ocean. The second variable is the loss rate (l) which, is influenced by characteristics such as grazing, and depletion of nutrients in the water. The third variable is the specific accumulation rate (r) which is the growth rate minus the loss rate (μ - l). The specific accumulation tells us a phytoplankton bloom has initialized when the value is higher than zero. The method I was using to determine when the phytoplankton bloom was beginning was that I was looking at the monthly climatology of the data which I will discuss a little later on

Our data was obtained from two different sources. The primary source of data came from Biogeochemical (BGC) Argo floats. These are these autonomous floats that are placed in the ocean. They pick up different metrics based on the sensors attached to them. Some of these variables are Ph, salinity. temperature, and light. All of these metrics play a role in quantifying the phytoplankton blooms but, they each compose their own piece of the overall project. Salinity and temperature make up density while the light is used to define the growth rate, mu. These metrics are sent to the Monterey Bay Aquarium Research Institute where quality control takes place. This means the data is cleaned to be more easily understood. These floats capture the data better arguably than other methods because they directly on the ocean as opposed to elsewhere. This brings me to the second form of data that is used in this analysis satellite imagery. While this data is taken from the sky is still produces accurate results when compared to the BGC Argo float data. Both of the different types of data were plotted on the same axis and produced similar results.

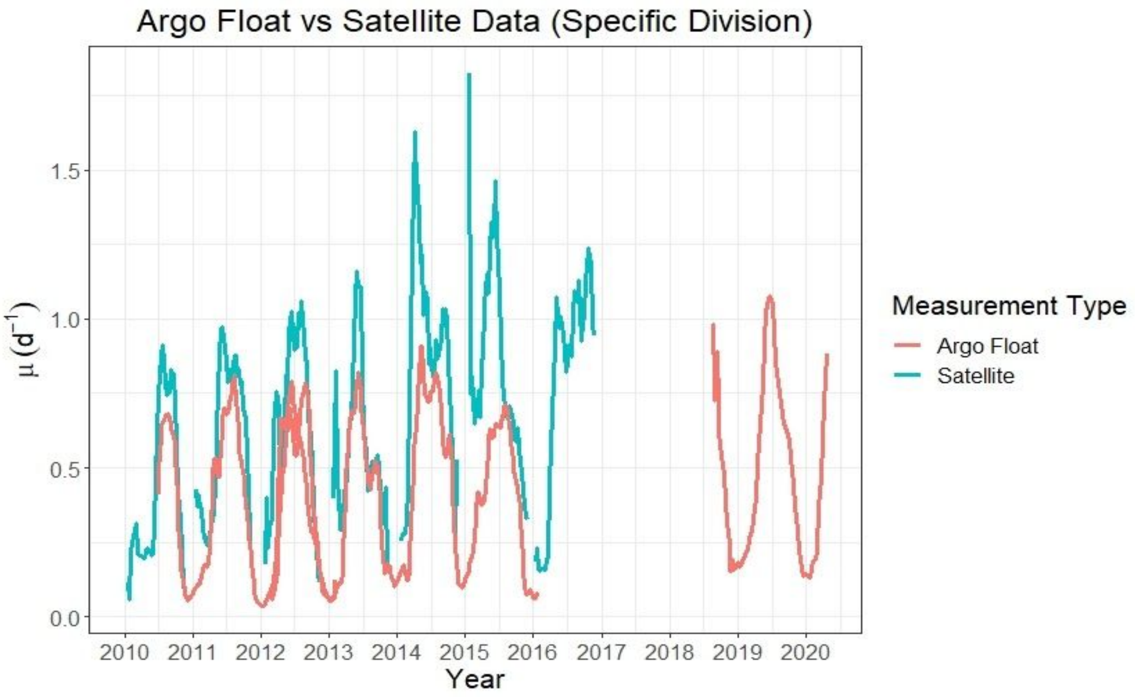

**Figure 1.** Growth Rate (μ) is taken from three different BGC Argo-floats in the Subarctic Pacific Ocean. The Argo floats took measurements from 2010-2020. One of these floats is still active in the Gulf of Alaska.

From these two types of data, we were able to quantify the phytoplankton blooms. As well as during what period they emerged. I grouped the data by date and took the median of all of the points for every date. I chose to use the median because there were some outliers in the data. This prevented those outliers from affecting the monthly climatology.

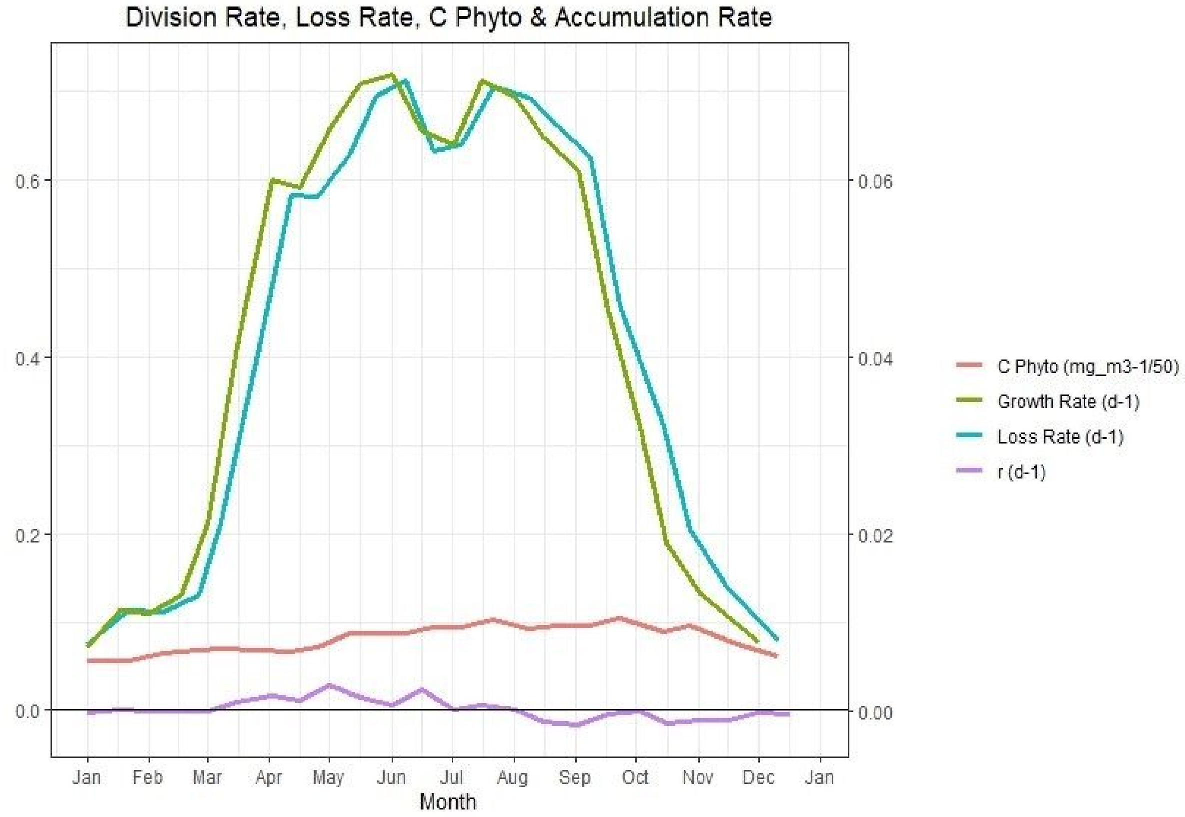

**Figure 2.** Monthly climatology of the BGC Argo-float data. Growth rate, loss rate, and r are all defined on the left axis, and C Phyto (Carbon Phytoplankton) is on the right axis. When r reaches a value greater than zero then a phytoplankton bloom has initialized.

The result that we got from this final figure shows that the bloom begins initiating in March. This did not align with the hypothesis that I initially set out to prove. One of the main reasons I believe this happened was because of the location where these measurements were taken. In the Yang et al. (2019) paper they focused on the North Atlantic. They also mentioned using their methods in different locations might produce different results.

Overall the project was a good experience because I got to work with real-world data as well as contribute to a meaningful project. Even though the result was different from what I expected a lot of the steps along the way produced great results I was satisfied with my internship this summer.