PHYTOPLANKTON AGGREGATE EVENTS

AND

HOW THEY RELATE TO SEA SURFACE SLOPE

Emily E. Anderson

CE 394K.3

Fall 2015

4 December 2015

Objective

Using skills gained in this ArcGIS class I wanted to create a map showing how measured phytoplankton aggregate events compared to the sea surface slope surrounding each event for the month of July. I wanted to see if aggregate events were more likely to occur in areas with greater slope. My area of interest was the North Pacific Subtropical Gyre located northeast of the Hawai'i. This area is known for its sporadic late summer or early fall phytoplankton blooms (population explosions of single celled oceanic plants). I used data gained by the autonomous aquatic vehicle: *Honey Badger* to determine when and where phytoplankton aggregate events occurred. I then plotted those locations on daily sea surface slope maps.

I wanted to do this project to gain a better understanding of when and where bloom events occur in the ocean. I would like to compare various sea surface data to phytoplankton blooms to pin-point what factors promote bloom development. If I could learn to predict when and where blooms were more likely to develop I could help other scientists plan research cruises. Right now the areas in which these blooms could develop is very large and it is impossible to determine exactly when a bloom might develop. Planning a research cruise is impossible when you do not know when to book the cruise or where to go. I hope this the first step in being able to predict when and where these blooms of interest develop.

I also wanted to create a visual aid that would allow me to view the changing sea slope over time and how the aggregate events appeared with these changes. I wanted to use an ArcGIS animation to accomplish this.

Data Sources

Data was obtained from several sources to complete this project. I primarily needed data about sea surface slope and phytoplankton aggregate events. I used the data collected from the satellite controlled autonomous aquatic vehicle: *Honey Badger* to calculate where phytoplankton aggregate events were likely to have occurred. I used the base fluorescence (F_o) to measure the amount of chlorophyll in the water. The amount of chlorophyll would give me a good idea of the volume of primary producers present. This data was stored in the free data server: ERDDAP. This data was easy to download and use.

I was unable to locate direct sea surface slope data, but I was able to find sea surface height data as measured by a satellite. This data was also stored in the ERDDAP servers. The data name was: *NRL HYCOM+NCODA, GLBu0.08/expt 91.1, Global, 1/12 deg, Apr 2014 to Now + Forecasts, at Surface*. I narrowed the range to my area of interest in the North Pacific Subtropical Gyre: 19-30° N and 203-216° E. The rasters are *very* coarse grained with each cell about 7x8 km but the surface ocean does not change as dramatically as land features so there has been a low need for fine resolution height images. In class we used Data Elevation Models (DEMs) to generate slope rasters and I planned to use this method to create my sea surface slope data from the sea surface height data.

Methodology

My first step was to establish an index to record the locations where aggregate events were likely to have happened based on the F_o data from *Honey Badger*. To do this I downloaded all the data for July into Microsoft Excel. I then calculated the standard deviation of my samples and multiplied this value by 3 to find the outliers, instances where the measured value was 3 times over the standard deviation are strong signals and locations where phytoplankton aggregates were likely to have occurred. These locations of suspected aggregate events were referred to as "spikes". After all the July data was processed I removed the instances where there were no spikes. Leaving only features where aggregate events were suspected. These spikes, with time stamp and geographical information, were saved as a .csv file. I imported the .csv file into ArcMAP and plotted the points as a feature class. This allowed me to view these spike locations in relation to each other and in relation to the basemap.

The sea surface height data was downloaded from ERDAPP as tabular data sets. Each set contained one day's sea surface height data. In July only 29 out of 31 days had data (the 15^{th} and 24^{th} did not have data). The tabular data was imported into ArcMAP as a .csv file. I used the "Display X,Y" command to plot the data as a feature class with Geographic Coordinate System: GCS_WGS_1984. I then used the ArcToolbox → Conversion Tools → To Raster → Feature to Raster. For "Input Feature" I chose the sea surface height feature class, the "Field" drop down was set to "SSH (m)", the output raster was renamed SSH_dd_July (dd was replaced by the day represented by the SSH data), and .08 was used for the "Output cell size". The default output had the slopes binned but I converted it into a stretched symbology raster. I had then created one SSH raster. I repeated the steps above for all the days in July that had SSH data on ERDDAP. I only had one month of data to process so it did not take too long once I figured it all out.

With the SSH rasters created I was then able to convert them into slope rasters using ArcToolbox → Spatial Analyst Tools → Surface → Slope. I input each of the SSH rasters and changed the output names to Slope_dd_July. Thus all the slope rasters I needed were at this point completed.

But all these rasters were individualized and not connected to each other. I needed to connect them together to create my animation.

After looking for advice on how to make ArcGIS animations I decided to create a slope catalogue. Creating the slope catalogue was easy and filling it with the appropriate rasters was also very intuitive and easy to do in large numbers. But the animation was still not working right, the time slider was having difficulty identifying the rasters as separate days because the file names were not in a time-stamp format. So I created a new table in Excel with my raster catalogue object ids and a new timestamp column in a format that ArcMAP could read. I then imported this new table into ArcMAP and joined my raster catalogue and this new table. With the tables joined I was then able to get the time slider to work, showing the slope rasters in chronological order. I opened Properties on the Slope Catalogue and navigated to the Time tab where I clicked “Enable time on this layer”. I selected “Each feature has a single time field” for the Layer Time with the joined “Date” for the Time Field. I selected the format that the Date was in, set the “Time Step Interval” to 1.00 Days and hit “Calculate”. It successfully read the data from 7/1/2015 to 7/31/2015. I then set the Time Zone to UTC because I knew the time zone that the data was in. I left the other fields to their default.

The raster catalogue was done but I wanted to show each of the spiking events per day on their appropriate daily slope raster. Luckily each of the spike events maintained their original time-stamp which was recognizable to ArcMAP. I enabled the time on the event features with similar steps to the slope catalogue, but this time I selected “Display data cumulatively” to keep showing the past data. I wanted to show the path *Honey Badger* took through the area of interest.

My animation was complete at this point. Everything was working as I wanted it so I exported the video using the Time Slider's "Export to Video" option. It converted the video to an .avi file, which was easily imported into PowerPoint. I learned that exporting the video takes a long time. The export process took just over 2 hours and the resulting video was a little jumpy in places. I think the jumpiness of the video might be the result of a glitch during export. I am unsure if re-exporting the video would improve the quality of if there is another way to improve the quality.

After completing the animation I wanted to take a close look at what the slope measurements were around each spiking event. I wanted to see how frequently each of the slopes had a spike event. I opened a new file in ArcMAP and imported all of the individual slopes and a separate feature class for each day of spiking. I started with the July 1st data by showing the slope and the spikes for that day. I recorded in an Excel file what the slope was under each of the spike events and continued this process for the rest of July data. I then binned the slope data and created a histogram to show the number of times each slope was associated with an event.

I attempted to use the Zonal Statistics Tools but decided to stop using this tool because it failed to return data for some days and only returned partial data on other days. For each day I attempted to run the Zonal Statistic as Table Tool using each days' features and the slope raster for that day. I set the zones by the Timestamp (unique for each event) and told it to pull the values from the slope raster. For each day I would run the Zonal Statistic and each day I would get a "success" message, but only half of the days had any data at all in the resulting table, and some of the days only a portion of the events would show up. When all of the results for a day appeared I liked the data it gave, but I had to abandon it because it did not provide consistent or reliable results.

Results

Converting the SSH data into slope data was successful and showed some very nice maps (Fig. 1 & Fig. 2). These maps demonstrate that the oceans' surfaces are not flat, but covered by gradual hills and valleys. These hills and valleys are created by several forces including the rotation of the earth, storm systems, regions of upwelling and downwelling, and wind patterns. These changes in surface height are slight (rarely over 1 m) but create regions of the ocean that are interesting to study.

I was also able to map *Honey Badger*'s path in July (Fig. 3) and calculate how far it traveled (1075 km). The markers indicate the strength of the F_o signal calculated as the F_o value – (moving average + 3 standard deviations) or how far above the outlier threshold the measurement was. Each of the markers, regardless of magnitude, was used when I pulled the slope data.

The slope data frequencies did not support my hypothesis that higher slopes would have more spiking events (Fig. 4). I would be interested in testing my hypothesis on a larger data set, but for July data, my hypothesis is not supported.

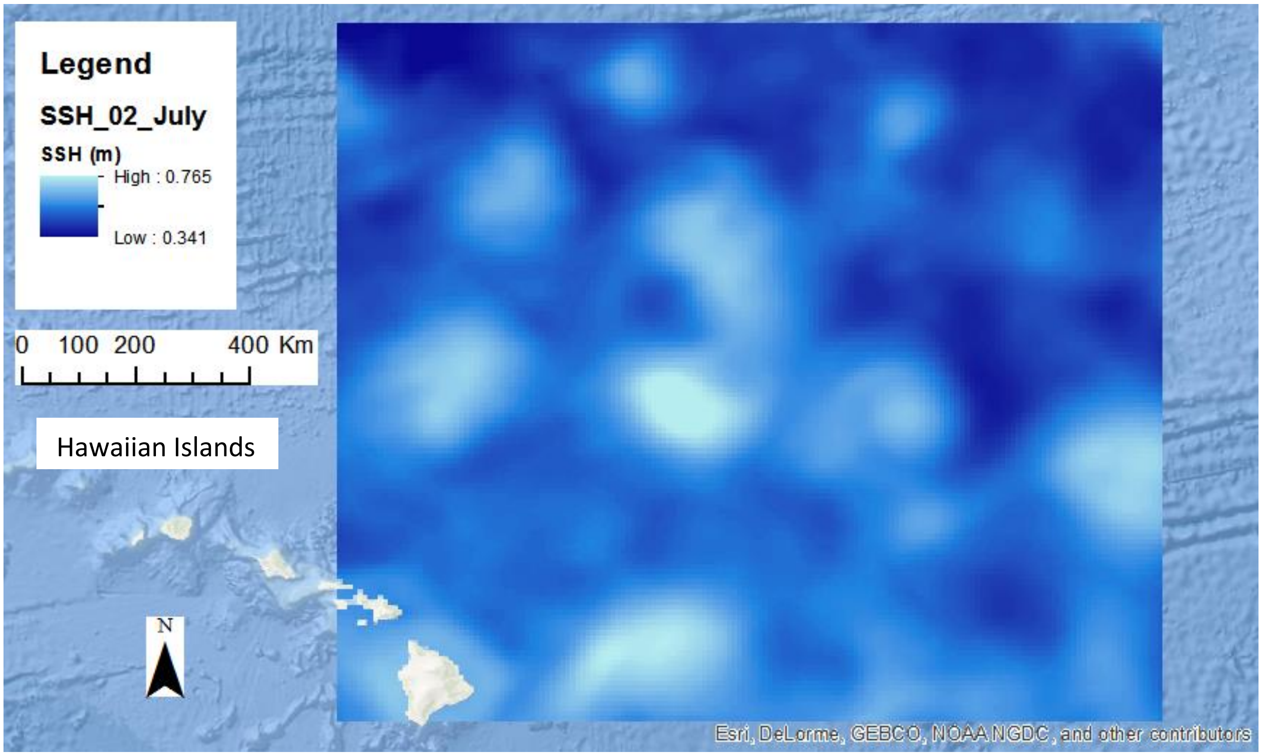

*Figure 1: **Sea Surface Height (SSH) Raster for July 2, 2015.** This figure nicely demonstrates that the oceans' surfaces are not flat. They have gradual and broad hills and valleys.*

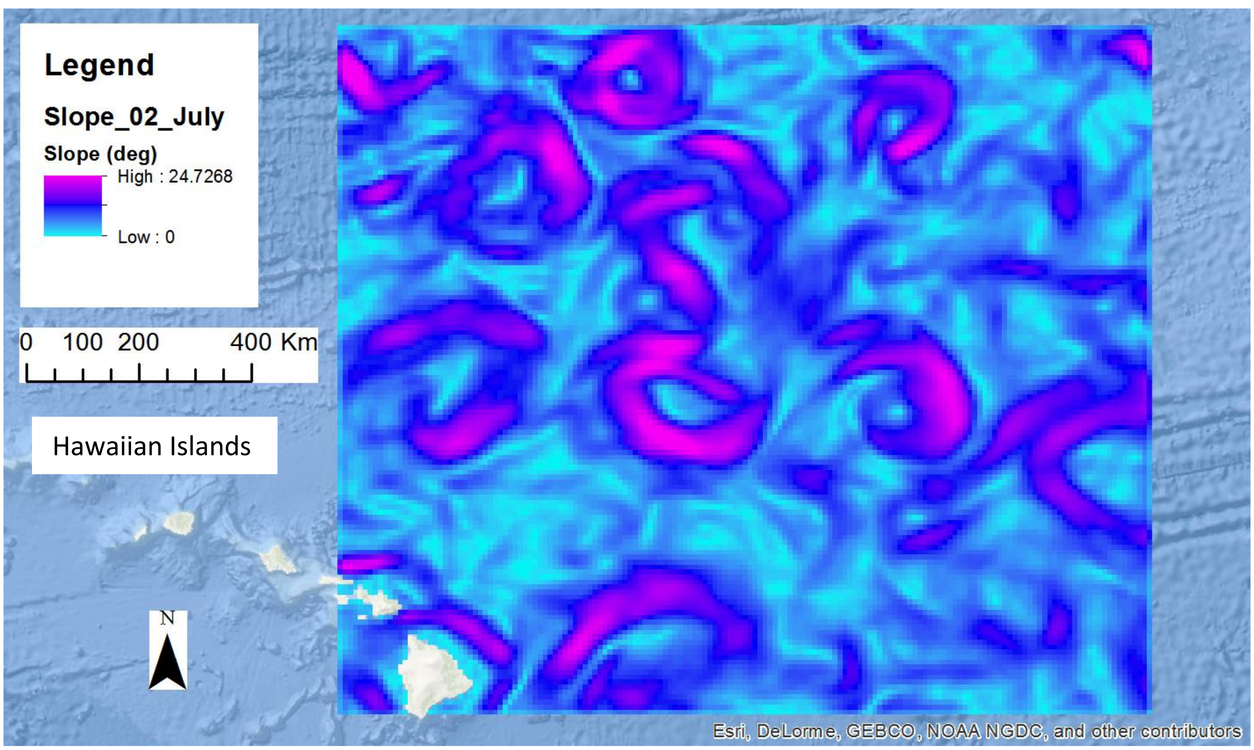

*Figure 2: **Sea Surface Slope Raster for July 2, 2015.** This figure clearly shows the slope associated with Sea Surface Height. These slopes are not as extreme as slopes seen on land, but are easily seen as features of the oceans' surfaces.*

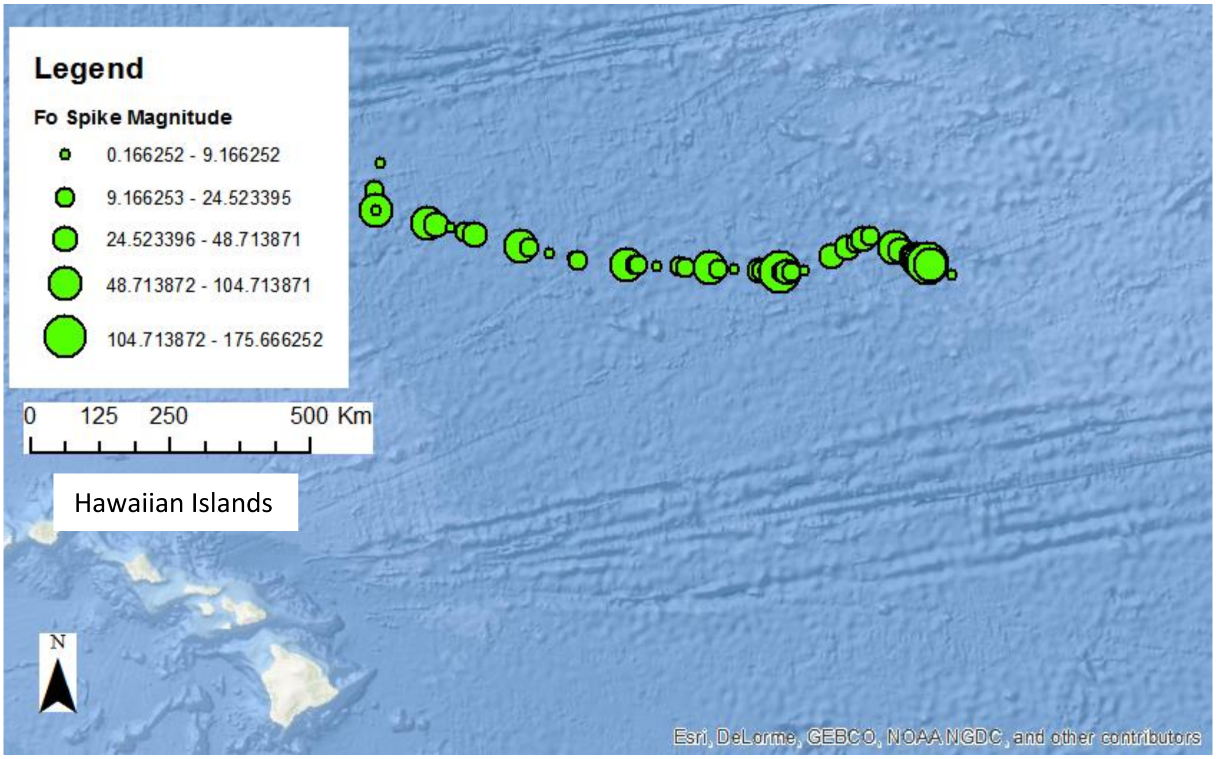

Figure 3: ***Honey Badger's path in July 2015.*** *The size of the markers show how much the greater the F_o Measurement was compared to three times the SD (the magnitude of the spike strength).*

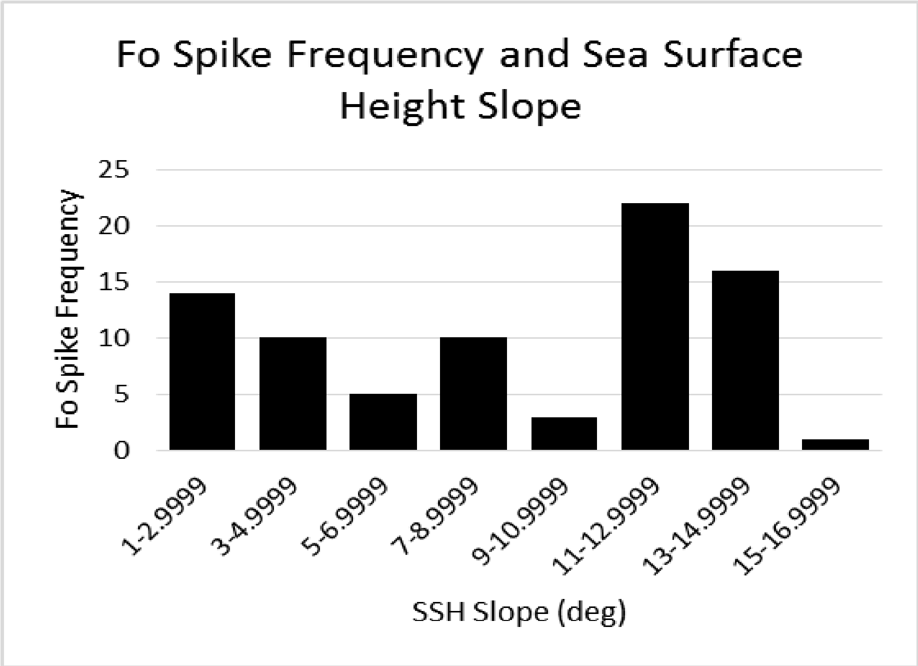

Figure 4: F_o Spike Frequency and Sea Surface Height Slope. The graph does not support my hypothesis that higher slopes have higher frequencies of spiking events.

Discussion

ArcMAP was a very helpful tool in visualizing my data. The mapping and animation capabilities make this a worthwhile tool to know how to use. I was disappointed that my data did not confirm my hypothesis that higher slopes had a higher frequency of the spiking events, but I will note that the data set was small. There are other data sets on ERDDAP where I could expand the time scale of this study. There are satellites that measure the color of the water for the chlorophyll content, this is not data from an *in situ* device, but it covers most of the earth and data is gathered every day. For these larger data sets a functioning Zonal Statistic Tool would help expedite the analysis process.

I originally wanted to bring elements of coding into this project to help the data transformations along, but I learned I know very little about coding and that I am very picky about what I name my outputs. I didn't know how to use coding to systematically go through a whole folder of separate .csv files and transform them into slope rasters with a specific name each. I think coding would be a very useful skill to know, but I do not know it, yet. I will be taking a course next semester that should help me through some of the basics of coding. For the purposes of this project I felt it was more time efficient to just do the data transformations one at a time rather than try to figure out how to code it.

I also want to look into how slope aspect might affect the spiking events. I would expect that south facing slopes would show a greater spiking frequency, but future investigations are needed to test this new hypothesis.

References

Marine Geospatial Ecology Tools. < https://code.env.duke.edu/projects/mget/wiki/HYCOM > Last accessed: 04 December 2015

ERDDAP. < http://coastwatch.pfeg.noaa.gov/erddap/index.html > Last Accessed: 04 December 2015.