Datathon-1

Seethamraju Purvaj (IMT2017039)

Data Visualization - **CS 732**

August 31, 2020

Abstract

This report contains the inferences and observations done as aprt of the Datathon-1 . It involves exploring the oceanography data in the indian ocean i.e. Sea surface salinity, sea surface temperature, sea surface height anamoly , meridional currents and zonal currents.

Introduction

Most of the work for this taks included exploring various libraries and tools for visualizing geographic data in python. These included matplotlib, Basemap, cmocean, etc. The final codes are in matplotlib and Basemap plots. Some brief overview of the information of various variables that was done as a part of the task for better understanding is also covered. The variable data is present over a span of almost 2 years. Hence one could also try to plot the temporal plots to see the changes in the variables over time. Most of the plots are temporal as they make more sense for variables like below which keep changing throughout the year. The spatial plots that make sense are the plots related to variable vs time. However we have around 35K points and it will be tough to consider specific ones for visualization. Basemap provided a wide range of options. For this task it was chosen because it provided minimal, yet significant details and plots for better visualizations making them self explanatory. It was ensured that the visualizations are as self explanatory as possible so that the whole information is conveyd at a glance.

1 Data Overview

The overall data is related to 5 variables mainly.

- Sea Surface Salinity(SSS)
- Sea Surface Height Anamoly(SSHA)
- Sea Surface Temperature (SST)
- Meridonial Currents
- Zonal Currents

2 Methods

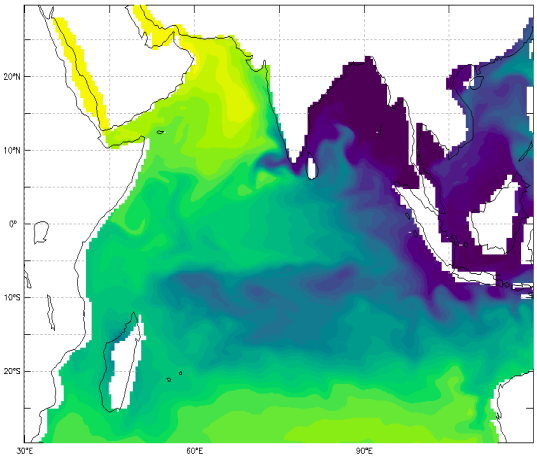
There are mainly 5 kinds kinds of plots done.

- Color Mapping(pmeshcolor)
- Contour Mapping(contour)
- Elevation (plot_surface)
- Streamline Plots (streamline)
- Quiver Plots (quiver)

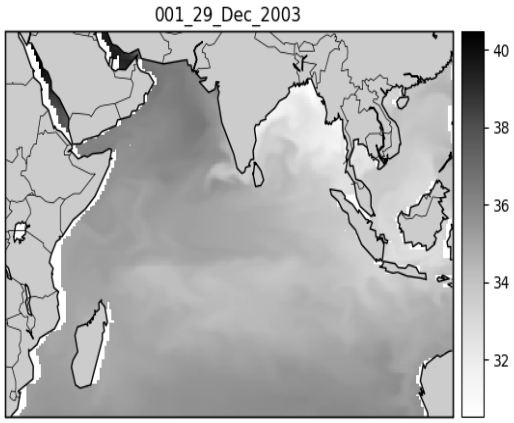
Each type of plot is used for variables that the methods seems to fit the most. For example elevation mapping seems to be apt for SSHA and not others. For every plot the data was taken and a meshgrid was created and then. Intuitively each poit on the Earth's surface is mapped to an xy point on the plane. The variable value at that point will be given to that point in the grid and the grid will be plotted. The contour and color map plots for same variable reflect the same data. The contour mappings try to visualize a boundary on the values of the variable. For example we can visualize the overall temperature at a particular location where as color maps are more of a continuous visualization.

3 Sea Surface Salinity

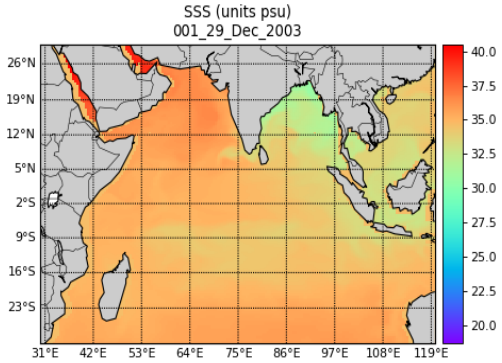
The data that was represented by -1e+34 were converted to nan and then the color maps and contour maps were plotted. This represents the amount of salt content dissolved in the water. Measured in psu. The original plot(Figure 1) was taken as a reference to see which color map suits the best. we tried to select the color range in such a way that any small change in the value can be reflected in the plot. There was a lot of variation in values that had to be reflected in the visualization because of which two color sequential colormaps were not used (like grey scale color map).



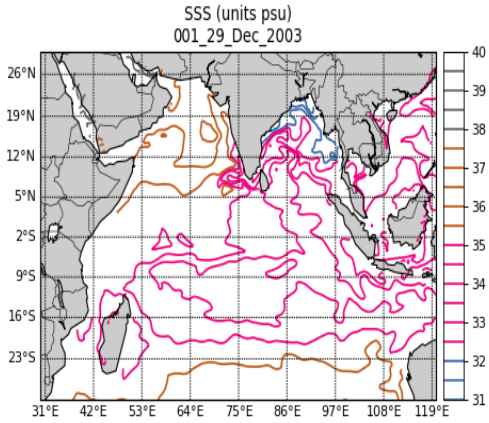
(a) Original Plot



(b) Gray Scale



(c) My Plot

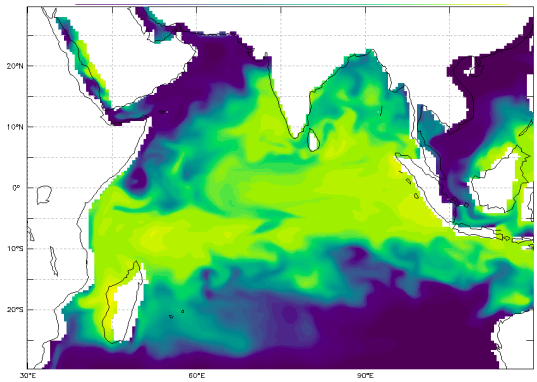


(d) My Plot

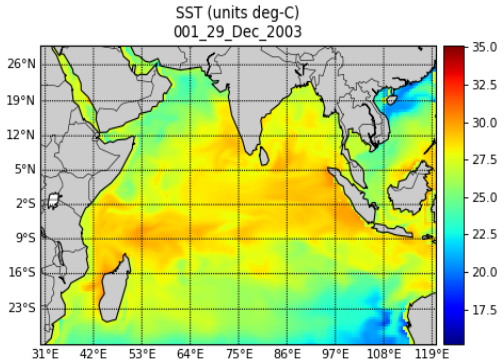
Figure 1: Sea Surface Salinity

4 Sea Surface Temperature

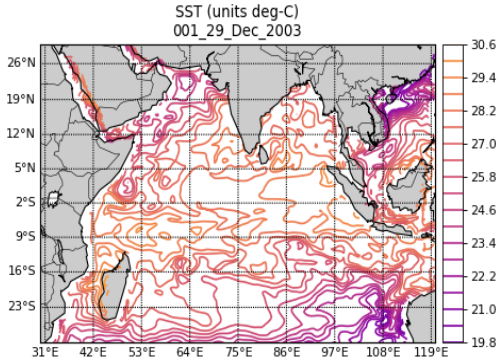
The data that was represented by -1e+34 were converted to nan and then the color maps and contour maps were plotted. This represents the temperature of the ocean water . Usually thermal satellites help in getting such sorts of data. The original plot(Figure 2) was taken as a reference to see which color map suits the best. The color range was chosen in such a way that any small change in the value can be reflected in the plot. Also the color map was chosen in such a way that visualization reflects the physical aspect of the variable. (heat corresponds to something dark red in color). Also the temporal plots seemed to agree with the season cycle that India has (higher temperature throughout the summer months near the peninsular region.) As we go south it is remains cold due to the Arctic effect.



(a) Original Plot



(b) My Plot

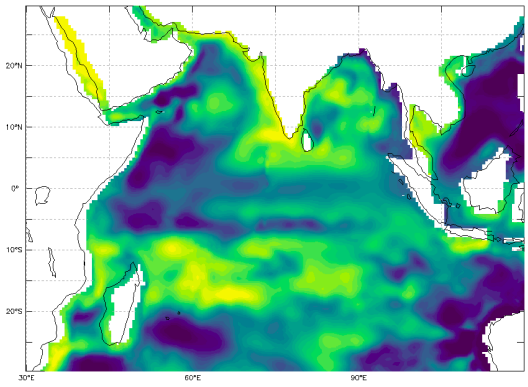


(c) My Plot

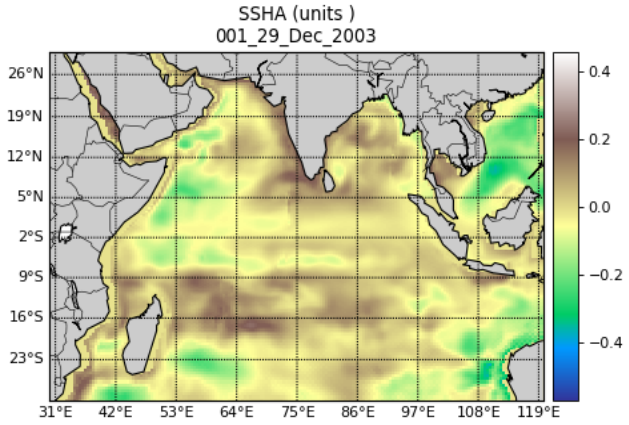
Figure 2: Sea Surface Temperature

5 Sea Surface Height Anamoly

The data that was represented by -1e+34 were converted to nan and then the color maps and contour maps were plotted. Though there wasn't much detailed information of how this is useful for oceanography, it had something to do with the height anamoly of sea. Hence the elevation mapping in 3d format had to be used for clear visualization. However we were able to capture all views of the map in temporal plots. Also the color mapping was done to view the visualization in 2D.

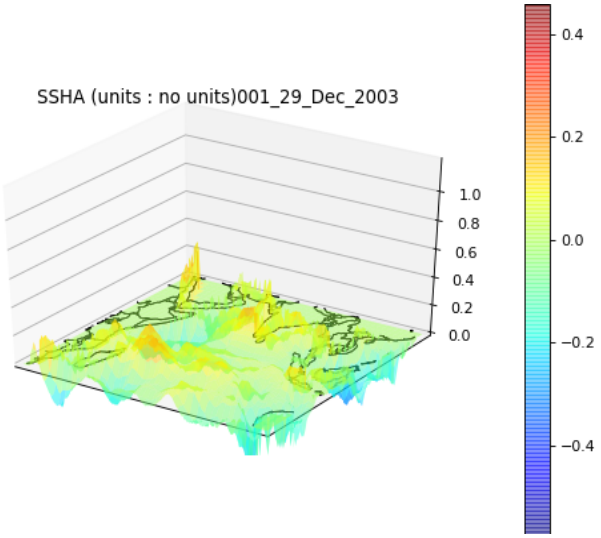


(a) Original Plot

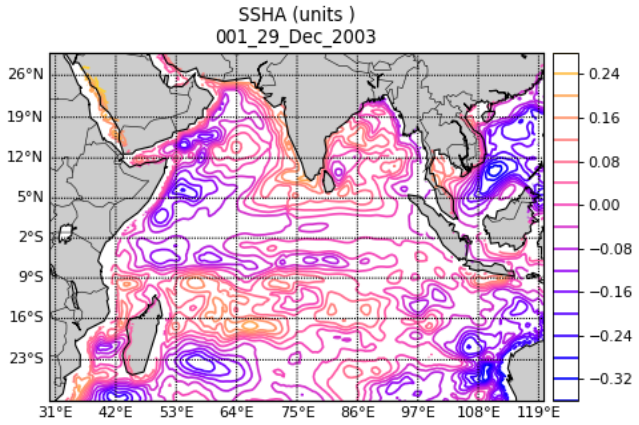


(b) 2D Color Map

Figure 3: Sea Surface Height Anamoly



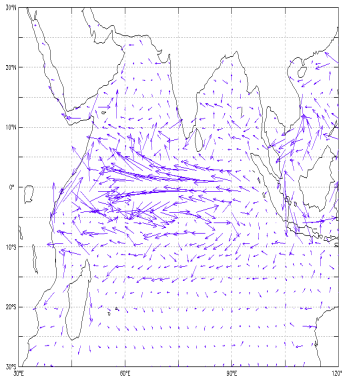
(a) 3D Elevation Map



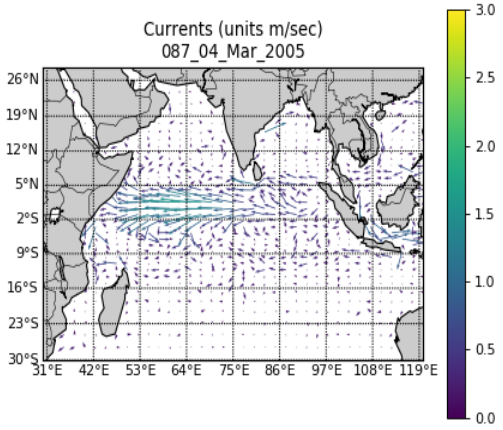
(b) 3D Elevation Map

6 Zonal and meridional Currents

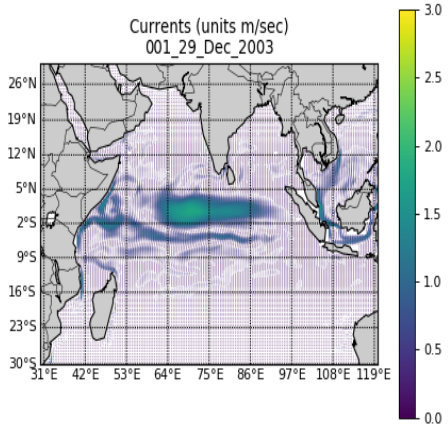
The data that was represented by -1e+34 were converted to nan and then the quiver plots were used(Figure 5). However there was excess of data (which generally is a rare thing in data science) due to which the data had to be sub sampled to get a clear visualization. The color map is used to map the norm of the vector. Streamlined plots were also plotted to understand the currents, their directions better.



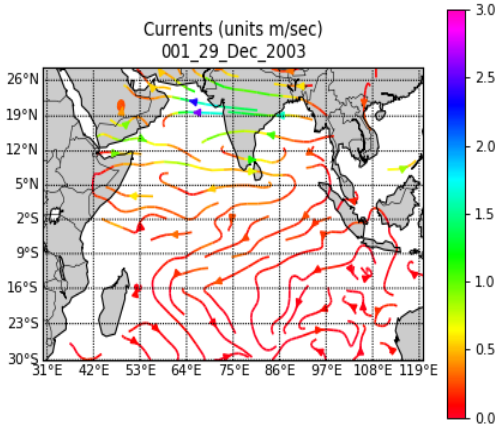
(a) Original Plots



(b) Currents plot (sub sampling)



(c) Currents plot based on Data

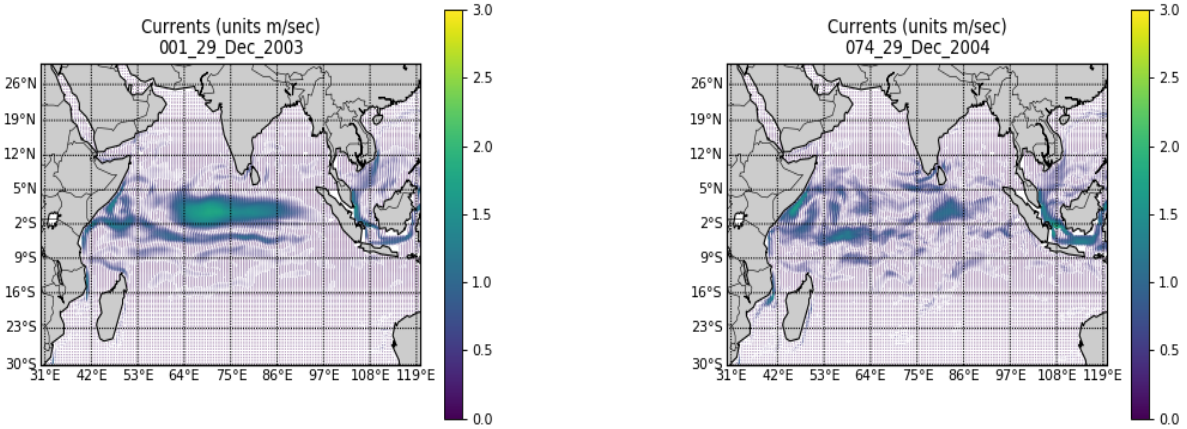


(d) Streamline Plots

Figure 5: Current Plots

7 Inferences and Questions

1) **Effect of variables before and after the 2004 Tsunami (Questions 1 and 2)**.
Most of the variables seem to have no direct effect because of the Tsunami. The tsunami was slightly reflected in the currents. Though not significant. There was a slight increase in salinity when the plots for the same dates of two different years were seen side by side. The 2004 tsunami had affected Indonesia as well. The sea surface temperature seemed to have been a bit less compared to the temperature the year before on the same day. The zonal and meridional currents seemed to have missed their previous year's pattern after the tsunami.



(a) Dec 29th 2003

(b) Dec 29th 2004

Figure 6: Same day 2003 Vs 2004. (Before and After Tsunami.)

2) **Inferences regarding patterns. Question 3** For Salinity the values near the East Indian coast were less which was generally due to the merging of the rivers into the Bay of Bengal. Saline content near the Gujarat gulf has a lot of saline content owing to the huge industrial exposure on that coastline. The currents' flow to the bottom of India(i.e. middle of the Indian Ocean) affect the climatic conditions of south India largely the wind directions.

8 Conclusion

Most of the plots seem to have reflected the phenomena in real world. Given the spatial and temporal visualizations these pictures would mean a million words to the people who have the domain knowledge and help them evaluate the patterns more accurately.

9 Files

- **contour.py** : the file that gives the contours for each of the variables for all the days.
- color.py : the file that gives the color mappings for each of the variables for all the days.
- **elevate.py** : the file that gives the elevation model for SSHA.
- **currents.py** : the file that gives the vector filed visualization for zonal and meridional currents based on sub sampled data.
- **streamlined.py** : the file that gives the streamline plots for zonal and meridional currents.

  And the temporal visualizations for each of the plots.

References

(1) Matplotlib
(2) Basemap
(3) Data and reference images