Data Visualization - Assignment 1

Vignesh Bondugula
IMT2019092
vignesh.bondugula@iiitb.ac.in

Abstract

This report contains the 2D scalar and vector field visualizations of INCOIS dataset of the Indian Ocean. It explains the use of color maps, contour maps for scalar fields and quiver plots for vector fields. It also explains how combining these visualisations might help to derive necessary conclusions.

1 Introduction and Data Overview

The visualizations in this report are done on the INCOIS dataset of the Indian Ocean. The dataset consists of 5 folders, each for different variables, namely,

- Sea Surface Salinity (SSS)
- Sea Surface Temperature (SST)
- Sea Surface Height Anomaly (SSHA)
- Meridional Current
- Zonal Current

The data is during the period December 2003 – December 2005. The datasets are at 5-day interval, hence overall there are 147 timesteps. The datasets for SSS, SST, and SSHA have data for 187 longitudes and 188 latitudes. Thus, each data file for a time-stamp has 187*188=35156 scalar values.

The zonal and meriodional current values can be treated as a single 2-dimensional vector field, where, zonal current speed (east-to-west speed, along latitude) can be assumed to be u- (or x-) component, and meridional current speed (north-to-south speed, along longitude), as v- (or y-) component. The datasets for current values have data for 181 longitudes, and 189 latitudes. Thus, each data file for a time-stamp has 181*189=37209 scalar values in each file. The packages used to visualize the plots are,

- Numpy
- Pandas
- Plotly, Dash
- Gif

Various visualization techniques were used such as Contour Mapping, Color Mapping, Quiver plots, Combinations of two visualizations, Animations

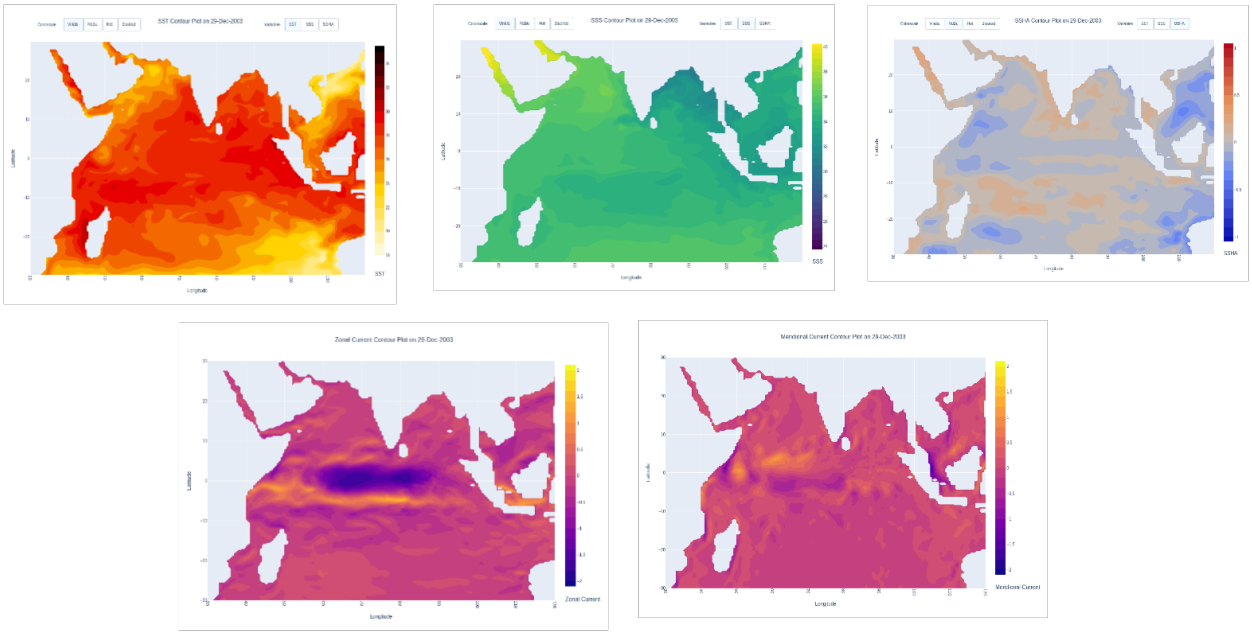
2 Visualizations

The data is extracted from the files using the numpy and pandas packages. The class DataGenerator generates the data and stores it in a pickle file, which can be read easily whenever necessary. The Scalar field data i.e SSS, SST and SSHA along with their latitudes and longitudes for each timestamp are extracted and put in a dataframe. All the dataframes for each timestep are stored in the pickle. Similarly, the Meridonal and Zonal currents are extracted and stores in dataframe for each timestep. All the vector fields' dataframes are stored in another pickle.

2.1 Scalar Field Visualization

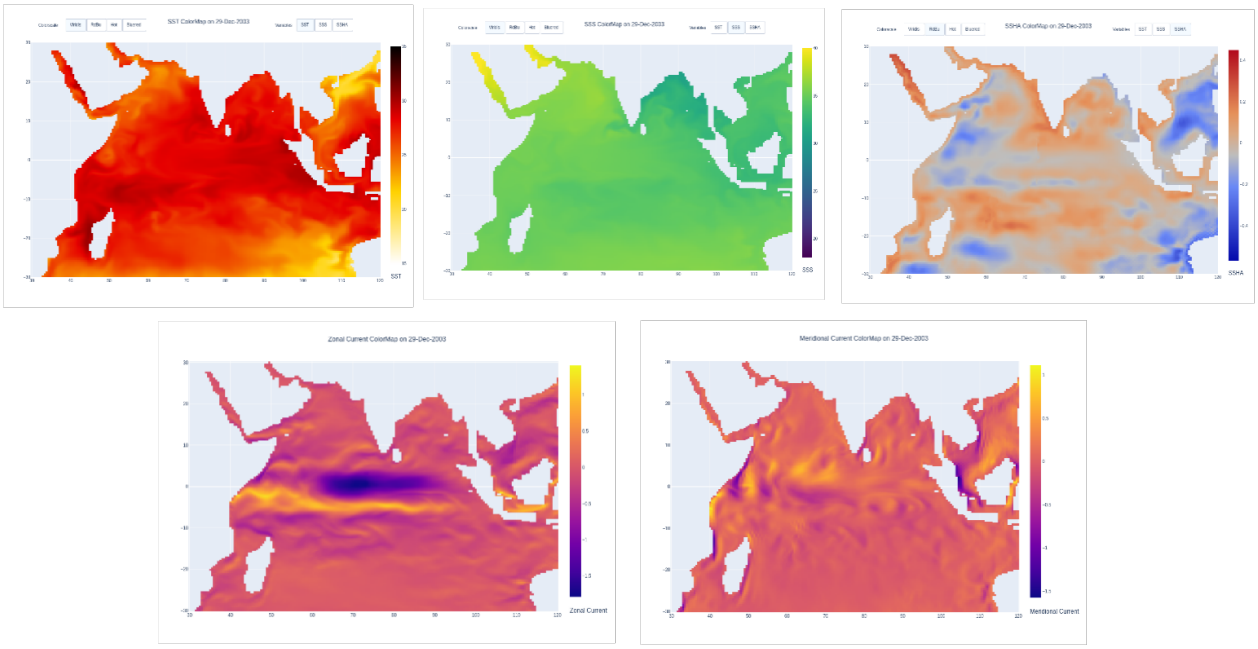
2.1.1 Contour Mapping

The Contour map from graph objects in plotly package was used to generate the plots. The maximum and the minimum values for the colour maps of each variable have been fixed after studying the upper limits and lower limits of the given data. An interactive framework is made using the plotly and dash packages to see SSS, SST and SSHA plots upon choice. The color maps also can be changed in an interactive fashion so we can select whichever color map suits the most. 5 contour plots are generated for each timestep. Contour plots for one of the timesteps are shown below,



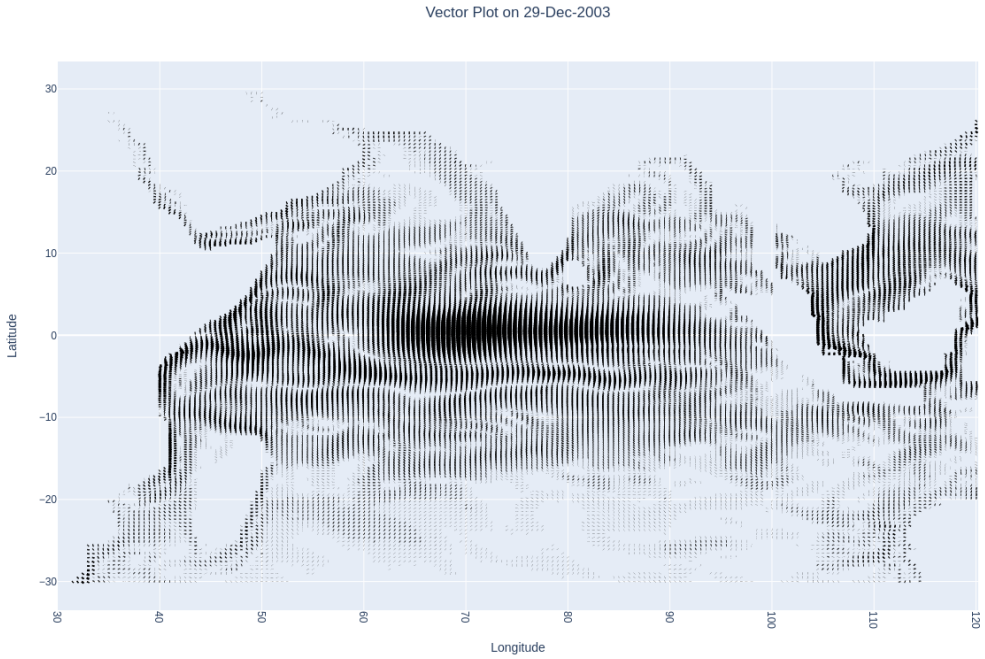
2.1.2 Color Mapping

The HeatMap from graph objects in the plotly package was used to generate the plots for color mapping. Various color maps has be tried out and experimented with, which includes sequential, diverging and cyclic color maps. The images and insights of which are presented in the later sections. The same interactive framework explained in the previous subsection is used for color maps to try different variables as well as different color maps. 5 Color maps are generated for each timestep. Color maps for one of the timesteps are shown below,



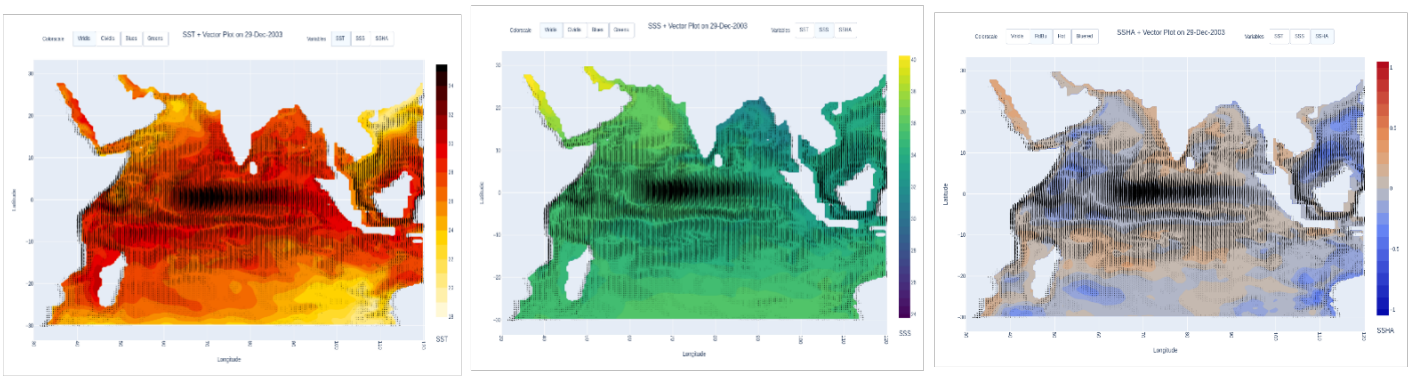
2.2 Vector Field Visualization

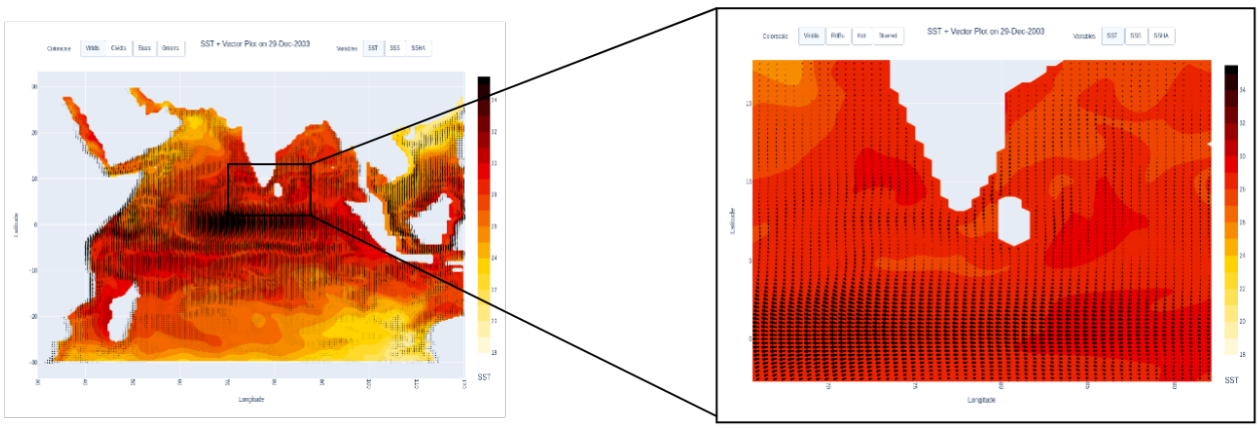
The create_quiver from figure factory in the plotly package is used to generate the quiver plot of the vector field. Zonal current speed (east-to-west speed, along latitude) is assumed to be u- (or x-) component, and meridional current speed (north-to-south speed, along longitude), as v- (or y-) component while plotting the quivers. One vector field plot is generated for each timestep. Quiver plot for one of the timesteps is shown below,



2.3 Combining Scalar and Vector Field Visualizations

Now to understand the vector plots and scalar plots in detail and to derive useful insights we combined both the contour plots of all 3 scalar variables and the quiver plot of the vector field. The interactive framework helps in easy transformation of different contour plots on the vector field with different color maps. The quivers are seen clearly when zoomed in the interactive framework. The insights gained by doing so are explained in the later sections. Three combined plots are generated for each timestep. Combined plots for one of the timesteps are shown below,





2.4 Animation

To understand the time varying effects on both the scalar and vector fields I have generated a animation using frames functionality in plotly. But due to data being big the animation was not smooth. Hence, I have used gif package to create gifs for different timestamps on both the fields. I have created two animations for each type.

- In first animation I have taken two timesteps each month to observe long term oceanic effects.
- In second animation I have considered only the timestamps during 19th Nov, 2004 to 28th Jan, 2005 to particularly observe and analyse changes during the Tsunami on December 2004.

3 Tasks

How did you decide which time instances to use?

- I have taken two time stamps for each month which makes a total time stamps of 49 so that we can observe long term effects of oceanic currents. Taking every time stamps(i.e 5 day timestep) is not necessary and computationally heavy, therefore I chose a timestep of 15 days which gives 2 time stamps per month.
- I have also taken every time stamp from 19th Nov, 2004 to 28th Jan, 2005 to observe the Tsunami on December 2004 as mentioned earlier.

For contours, will you use the same contour values for all the selected timesteps?

We would not use the same contour values for different timesteps. Contours essentially represent a collection of points that have the same value. This value can be anywhere between the start and end. For each variable, I have calculated start and end by calculating the max value and min value in the entire data set, then specified it while making the contour map of respective variable. Certain particular values that are present in one timestep need not be present in the other. For multiple timesteps, we cannot guarantee that the same contour values will be taken. Therefore, we cannot specify the same contour values across different timesteps.

For color mapping, will you use the same min-max values to generate the color palette?

Yes, we will use the same min-max values to generate the color palette since we do not want the min max values in the scale during the animation. The scale has to be fixed for each timestep. Hence, for each variable, I have calculated the max value and min value in the entire data set, then specified it while making the color palette of color map of respective variable.

Did any color palette outperform the others? How would you rationalize the performance?

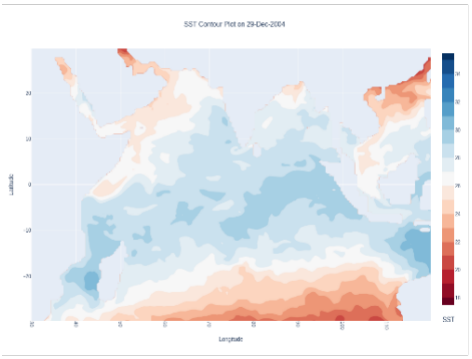
There are three kinds of color maps to chose from, namely,

- **Sequential :** This is a palette of just one or two hues, with its ends ranging from a dark shade of the hue to a lighter shade. This is typically used with data that has an intrinsic ordering, with one end representing lower magnitude and the other indicating higher magnitude of the value.
- **Divergent :** This palette has two contrasting colours on either of its ends, with a neutral shade in between. Divergent colour palettes are best used when the data is compared against a particular value. Example: to represent the pH of a liquid, red indicates a strong acid, blue indicates a strong base and a greenish shade in between indicates a neutral liquid.
- **Qualitative :** This comprises of various, almost unrelated colours and can be used to represent values that cannot be ordered in any way. For example, to denote statistics of different cities, we can assign colours at random to the cities - pink for Bangalore, blue for Mumbai and orange for Delhi.

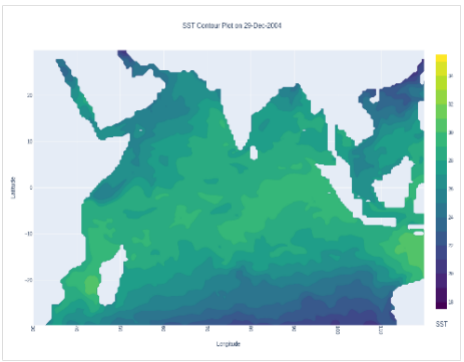
I have chosen sequential color maps for SSS and SST are positive and can be ordered. Since there is no notion of "neutral" or "middle element" when it comes to density and mass, I chose sequential instead of divergent. I have chosen **Hot_r** for SST because it is the perfect color scale for indicating temperature. High temperature is shown with red color and the lower

temperatures are indicated by white. The redness increase as the temperature increases. For SSS, I have chosen **Viridis** as it seemed to capture the information of the sea salinity the best to us. Any other sequential map would have also worked fine.

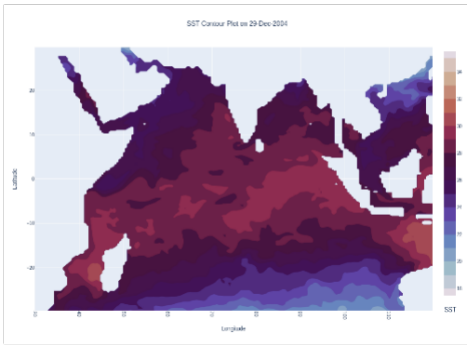
I have chosen Diverging color map for SSHA, since the average value is regarded as the zero value and anomaly is measured as either a positive or negative difference from this average value. I chose **RdBu** diverging color map, which has red and blue at either of its ends.



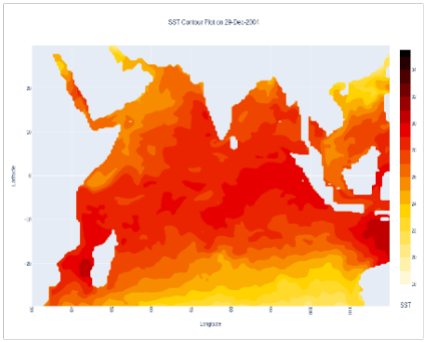
RdBu(Diverging)



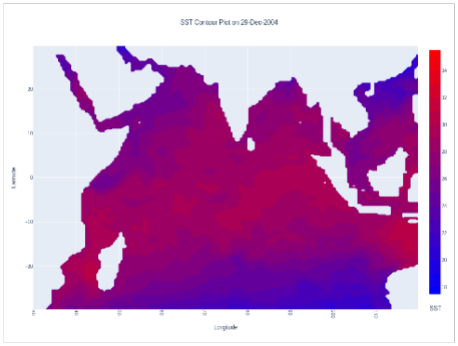
Viridis(Sequential)



Twilight(Cyclic)



Hot_r(Sequential)

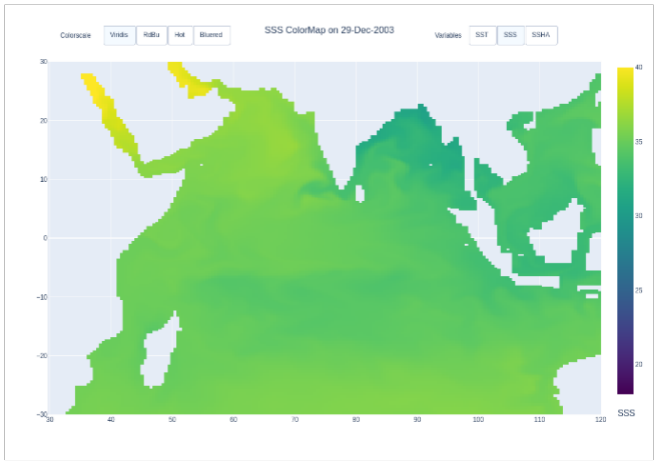


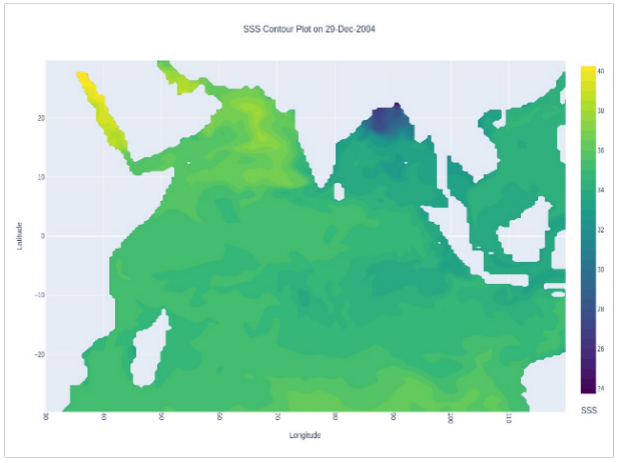
Bluered(Sequential)

Cyclic and Qualitative maps were found not suitable for any given data. The issue with cyclic mapping was that the min and max colors were similar, making it harder to interpret the colors.

Did such a visualization enable you to make joint inferences of different fields?
Yes, the visualizations of combining various fields helped me to make joint inferences of both the fields. Mainly, it helped me to see the correlation between the variables. I found the temperature are closely related to the current patterns visible over Indian Ocean. This is because the temperature difference creates a pressure difference which effects the oceanic currents flow. We can observe this in the animations.

Do your visualizations help you infer any aspects of the tsunami in December 2004?
Yes, visualizations helped me infer the aspects of tsunami in December 2004. As we can see in the below figure the Sea Surface Salinity appears to have increased a bit after the occurrence of tsunami. Even though the change is not drastic, it seems to remain high until January-February which we can see in the animations generated. Also the temperatures were lower than in usual Decembers.





4 Conclusion

The visualizations gave us lots of insights. Since the data was really big it would not be possible to understand the information and gain knowledge without using different types of visualisations. Many more insights could have been made with much more field knowledge related to oceanic currents.