Datathon - 1

[1]Arjun Verma : IMT2017008
CS732/DS732: Data Visualization
Course Instructors : Prof. Jaya Sreevalsan Nair
Technical Report - 1

Abstract. *This technical report contains a brief overview of the methodology involved in the visualization of the provided dataset of the 5-day moving average of values computed of Indian Ocean from the ocean model MOM, run by the Indian National Center for Ocean Information Services, INCOIS, Hyderabad. The report further attempts to make certain inferences from the visualizations generated. The additional submissions can be found here.*

1. Introduction

The provided dataset contains information about 5 different variables in the domain of oceanography. These variables include sea surface salinity (sss), sea surface temperature (sst), sea surface height anamoly (ssha), meriodional currents (mc) and zonal currents (zc). The techniques attempted for the visualization of the scalar variables : sss; sst; ssha involved contour mapping, color mapping and elevation mapping. The vector field visualizations were done by generating quiver plots. The methodology section would elaborate on the chosen technique for each field and why they were chosen. In the last section, we will attempt to make certain inferences from the visualizations generated.

2. Tools and Methodology

2.1. Tools

The exhaustive set of libraries used for generating the visualizations involve :

- os
- numpy
- pandas
- mpl_toolkits
- matplotlib
- scipy
- shutil
- pillow
- ogr

The usage of the aforementioned tools would be elaborated upon in the methodology section.

2.2. Methodology

In this section, we provide the distinct characteristics involved in each technique and the reasonings behind them. However, before we dive into the individual techniques, let's have a clarity on the basic workflow of the code.

2.2.1. General Methodology

For visualization of each of the variables given, we have three stages represented by three files of code to be run to get to the final output. The workflow proceeds in this order :

- For the first piece of code, the data is taken from the input files to produce images for each of the 147 time stamps.
- In the second piece of code, the images undergo pre-processing before they can be converted into a '.gif' using the pillow library.
- In the third piece of code, pillow library is used to generate the final output.

A basic framework is represented as :

Figure 1. Framework

Apart from this, the common charcteristics across the techniques involve :

- Reading from the txt files and extracting the relevant values.
- Generating a 2D grid to store the values at each (latitude, longitude) position.
- The stored values are normalized to keep the scaling consistent across time frames.
- Using Basemap from the mpl_toolkits library to get the underlying atlas and drawing the latitudes and longitudes on it.
- Choosing the appropriate technique to overlap with Basemap.
- Saving the 147 plt instances generated.

2.2.2. Sea Surface Salinity Visualization

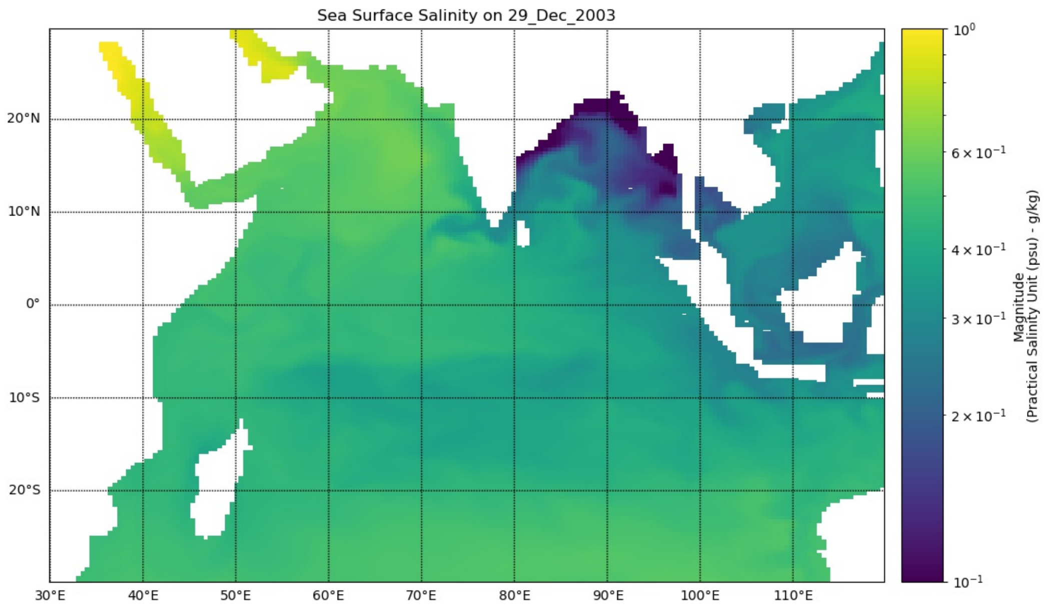

Figure 2. Sample SSS Frame

Distinct techniques (characteristics) for SSS visualization :

- Technique : Color Mapping
- Cmap : Viridis
- Normalization Technique : Max-Min Normalization
- Colorbar Scale : Logarithmic
- Values Utilized : Longitude, Latitude and SSS

The reasoning for each of the distinct techniques (characteristics) is as follows :

- Color Mapping was chosen as for analyzing salinity, it is important to analyze different regions with respect to each other and conclusion about a specific region and it's vicinities could be made.

- While selecting the colormap, it was important to pick up a perceptually uniform sequential colormap as the values of sss lie in a positive continuous interval. Viridis in particular was chosen to compensate for it's robustness in color vision deficiencies.

- Since the continuous range of values have no negative values, a normal max-min normalization proved sufficient.

- Outputs were initially generated using a linear scale, however on coming across an expert's thoughts in the domain of oceanography, the colorbar was stretched to a logarithmic scale. According to the expert, it is scientifically proven that the expression of the river plume at the surface has approximately log scaled area. To incorporate for these better visualization of river plume, the color scale was stretched logarithmically.

- The values utilized from the txt file were only the longitude, latitude and sss magnitude values. The time variable was simply extracted from the metadata in the filename.

2.2.3. Sea Surface Height Anamoly Visualization

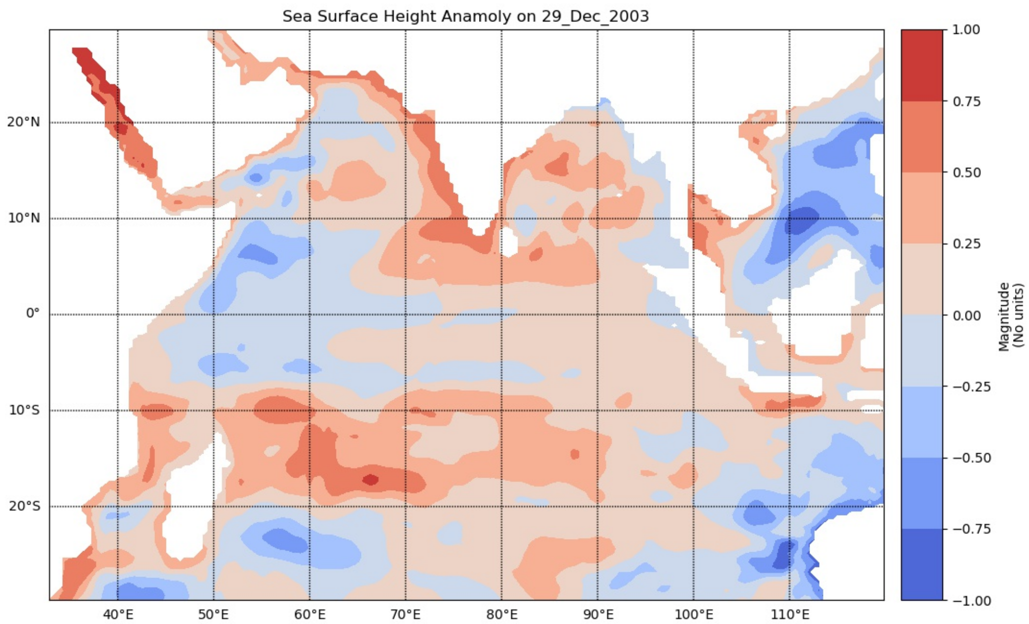

Figure 3. Sample SSHA Frame

<u>Distinct techniques (characteristics) for SSHA visualization :</u>

- Technique : Contour Mapping
- Cmap : Coolwarm
- Normalization Technique : 2 * {(val - min) / (max -min)} - 1
- Colorbar Scale : Linear
- Values Utilized : Longitude, Latitude and SSHA

The reasoning for each of the distinct techniques (characteristics) is as follows :

- Contour Mapping was chosen to group the regions of the same height together.

- If one notices the range of the magnitudes of ssha, one sees that the values are spread across the either axis; the positive and negative and hence I felt a diverging colormap would be of utmost relevance as we can see that values are diverging on either side of 0.

- To account for the negative values in the range, the aforementioned normalization technique was utilized. This technique maps the values to the range [-1, 1] and thus makes way for the application of the diverging colormap.

- Linear scaling of the colorbar seems sufficient for the purposes of ssha.

- The values utilized from the txt file were only the longitude, latitude and ssha magnitude values. The time variable was simply extracted from the metadata in the filename.

2.2.4. Sea Surface Temperature Visualization

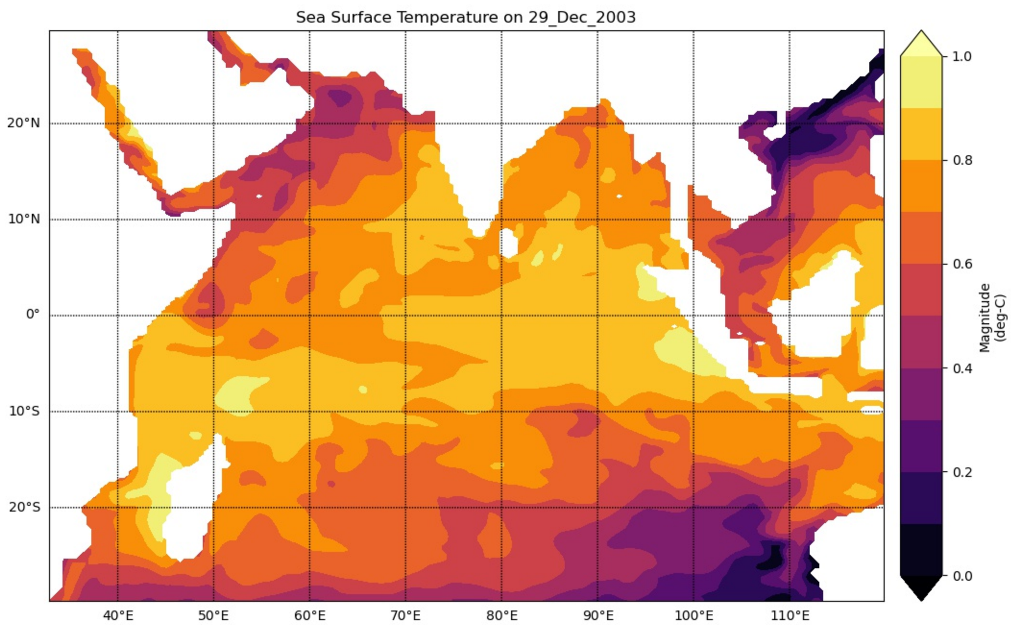

Figure 4. Sample SST Frame

Distinct techniques (characteristics) for SST visualization :

- Technique : Contour Mapping
- Cmap : Inferno
- Normalization Technique : Max-Min Normalization
- Colorbar Scale : Linear
- Values Utilized : Longitude, Latitude and SST

The reasoning for each of the distinct techniques (characteristics) is as follows :

- Contour Mapping was chosen to group the regions of the same temperature together.

- While selecting the colormap, it was important to pick up a perceptually uniform sequential colormap as the values of sst lie in a positive continuous interval. Inferno is a widely used cmap for oceanographic temperature and hence it was chosen.

- Since the continuous range of values have no negative values, a normal max-min normalization proved sufficient.

- Linear scale seems sufficient for the purposes of sst.

- The values utilized from the txt file were only the longitude, latitude and sst magnitude values. The time variable was simply extracted from the metadata in the filename.

2.2.5. Ocean Currents Visualization

Distinct techniques (characteristics) for ocean currents visualization :

- Technique : Quiverplot
- Cmap : Coolwarm
- Normalization Technique : Max-Min Normalization
- Colorbar Scale : Linear
- Values Utilized : Longitude, Latitude, Meridional currents and Zonal currents

The reasoning for each of the distinct techniques (characteristics) is as follows :

- Quiverplot was chosen to visualize the current directions along with their magnitudes across the map.

- Coolwarm was chosen to draw a major distinction between the relatively highest and lowest magnitudes as the values for the currents seem to lie very close to each other. So, it was important to capture these values at the extreme ends and the values in the middle lose their relevance as they are extremely similar.

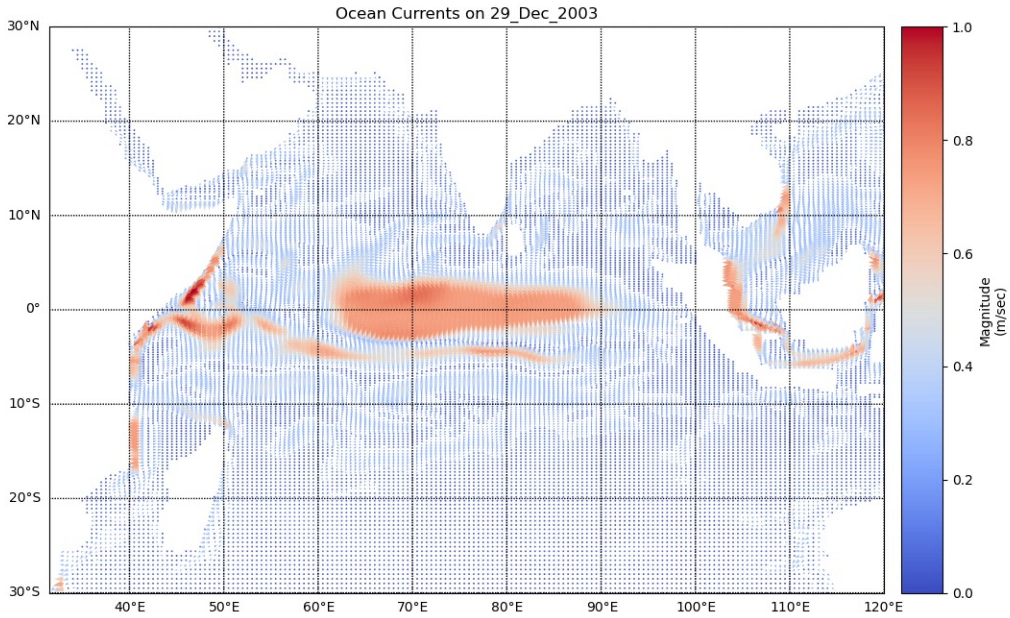

Figure 5. Sample Ocean Current Frame

- Since the continuous range of values (magnitude of current) have no negative values, a normal max-min normalization proved sufficient.
- Linear scale seems sufficient for the purposes of sst.
- The values utilized from the txt file were only the longitude, latitude , zonal currents and meridional currents where [zonal, meridional] formed the [u,v] components of the vector respectively. The time variable was simply extracted from the metadata in the filename.

2.2.6. Tsunami Case Study

This section covers the general methodology applied when inferences related to the Tsunami were being made.

Key characteristics of the plots :

- The epicentre of the Tsunami (Northern coast of Sumatra) has been marked as a black point.
- The radius of impact of the Tsunami has been marked in black circles. The radius has been limited upto the Indian subcontinent for the purposes of this report and the indicative tasks.
- The countries impacted have their outlines highlighted in orange.

Key methodology characteristics :

- The function for plotting circles depicting radius of impact is based on the Haversine Formula.
- The outlines of the countries affected were generated by reading the shapefiles obtained from the official GADM website.
- While making inferences, only plots for the dates : 24th and 29th December, 2004 were focused on, for each of the variables.

The sample images for the case study of the Ocean Currents variable are provided below. Note, this methodology was repeated across all the variables to study the effects of the Tsunami. This was just an intermediate step for the purposes of personal inferences.

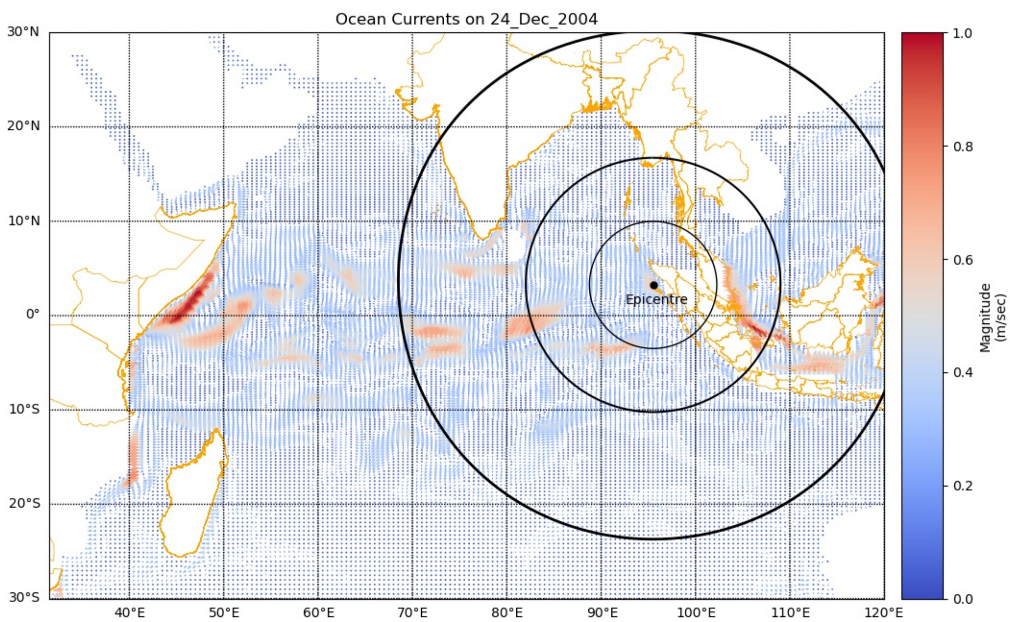

Figure 6. 24th December, 2004

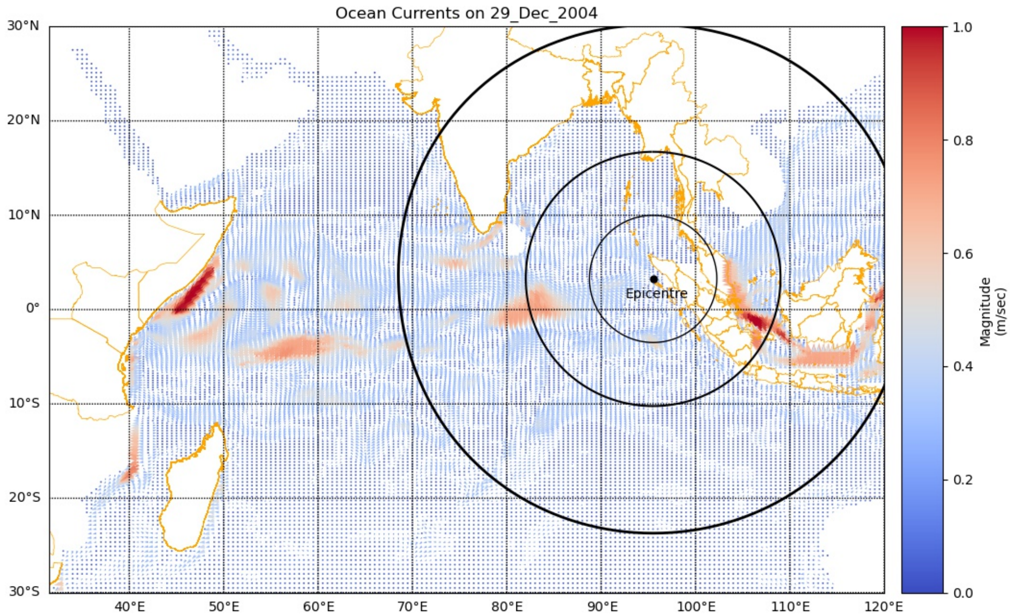

Figure 7. 29th December, 2004

3. Inferences

This section has been divided into two parts. One accounts for general inferences which have no relations to the Tsunami of Dec, 2004 while the other solely focuses on that.

3.1. General Inferences

1. If one has a look at the SST visualization, and focuses on a particular region, say the Arabian Sea, one sees that the temperature of the ocean rises as expected during months of April, May before cooling down during monsoon in July, August. It rises for a short while again, during October , which I account to the annual October Heat experienced in the region before slightly cooling down in winter.

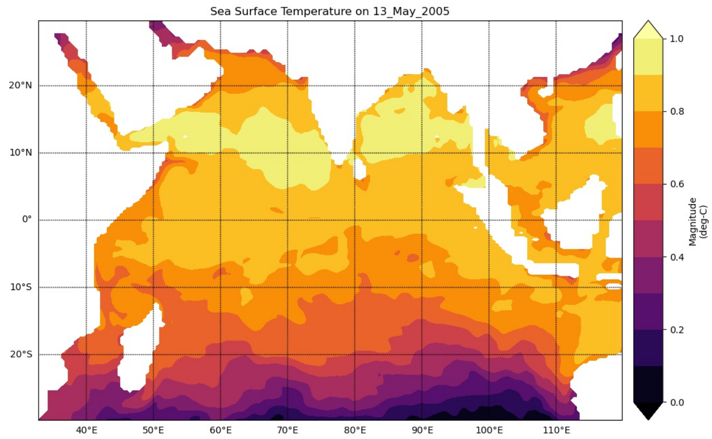

Figure 8. SST during May, 2005

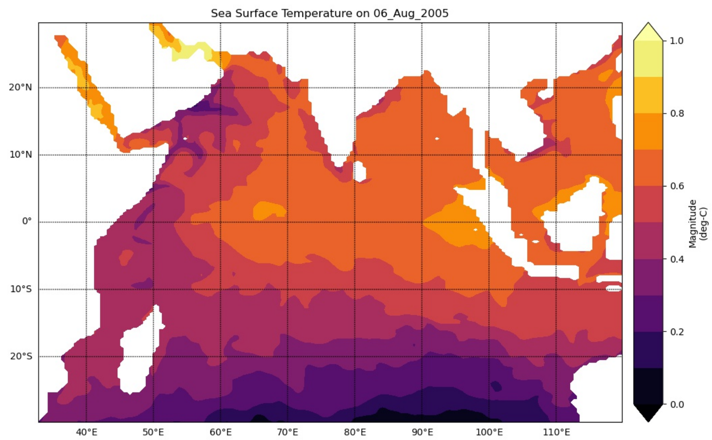

Figure 9. SST during August, 2005

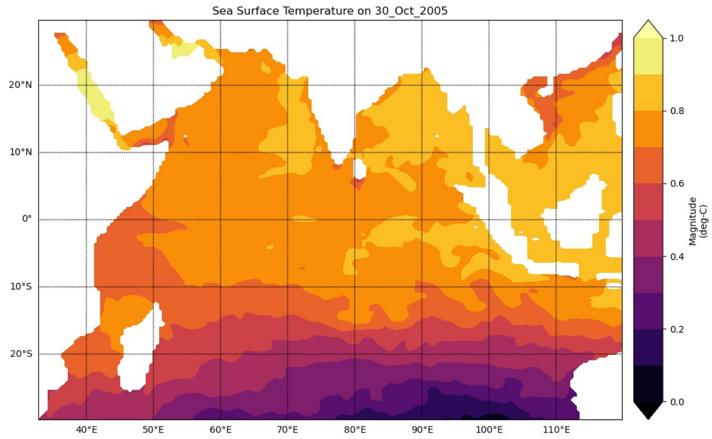

Figure 10. SST during October Heat, 2005

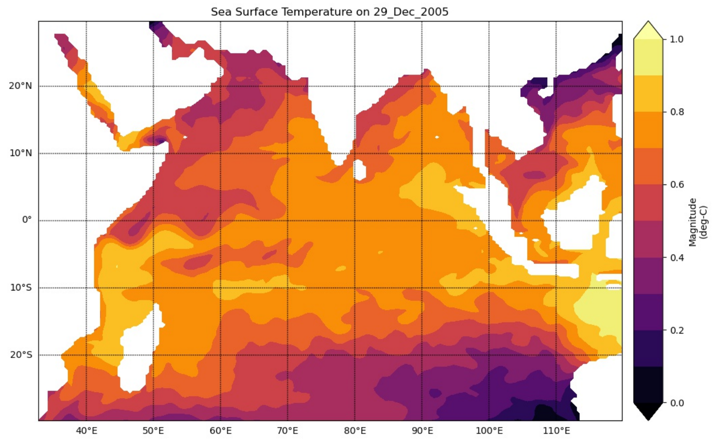

Figure 11. SST during December, 2005

2. On analyzing the quiverplot, one can easily account for the winter (December) falls received by the state of Tamil Nadu by noticing the high current magnitudes along Tamil Nadu's coast.

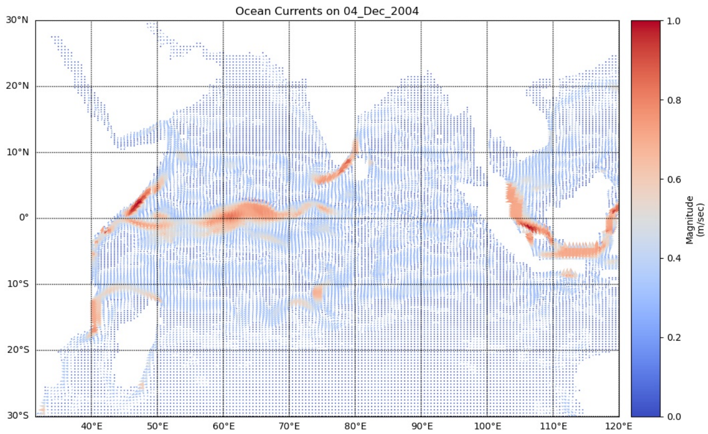

Figure 12. December 04, 2004

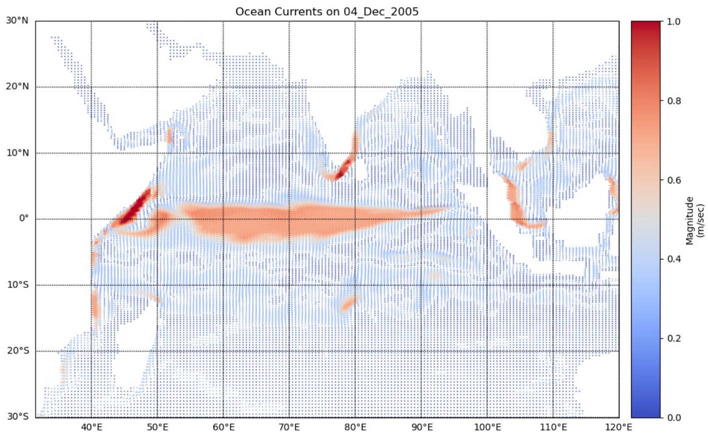

Figure 13. December 04, 2005

3. Having a thorough glance at the Sea Surface Salinity visualization, one can see that the salinity increases in the Arabian Sea during the monsoon months (Jun - Aug). This is in complete accordance with the fact that an increase in rate of evaporation results in an increase in salinity.

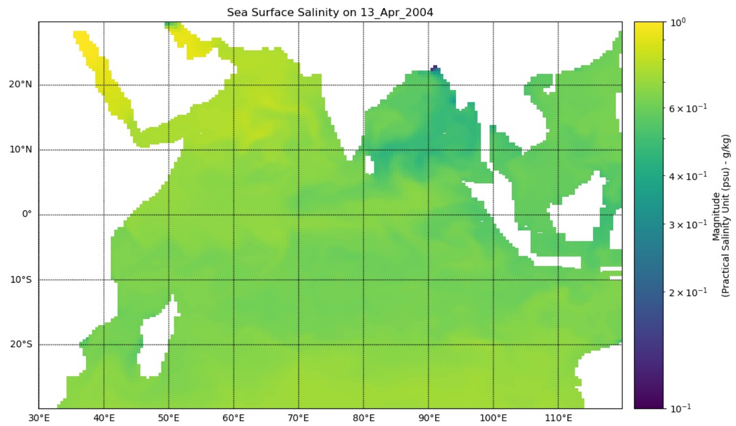

Figure 14. April (2004)

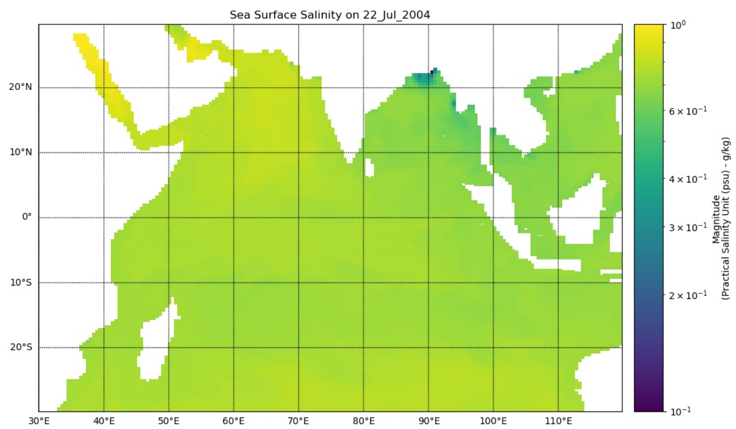

Figure 15. July (2004)

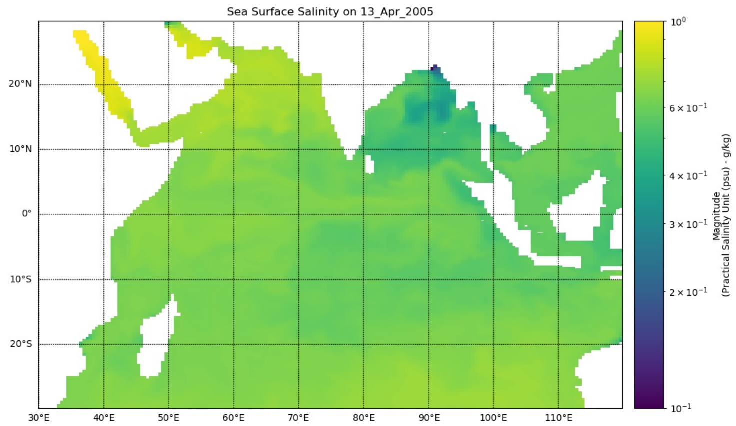

Figure 16. April (2005)

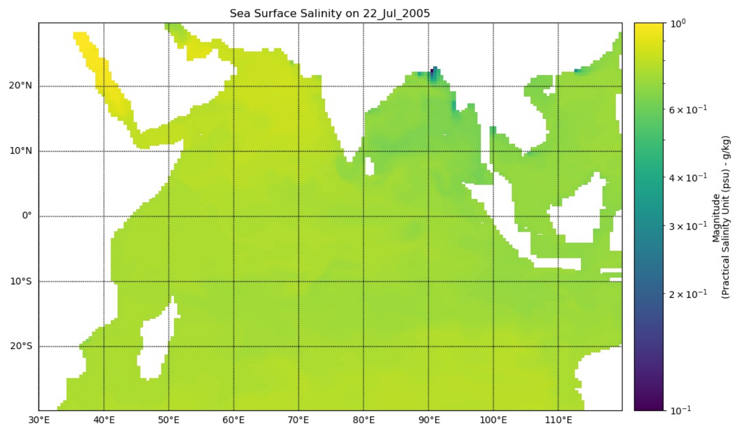

Figure 17. July (2005)

3.2. Tsunami Related Inferences

1. If one were to have a look at the SSHA (height anamoly) visualization one would see that there is a massive increase in the the values of ssha during the December of 2004 when compared to the December of 2005 which is much calmer. Moreover, one can see that the ssha increase is propagating westward originating from **Sumatra, Indonesia** which was considered to be the epicentre of the Tsunami. I believe SSHA is the only variable which shows a massive correlation with the raging Tsunami.

<u>December, 2004 (Tsunami Struck) :</u>

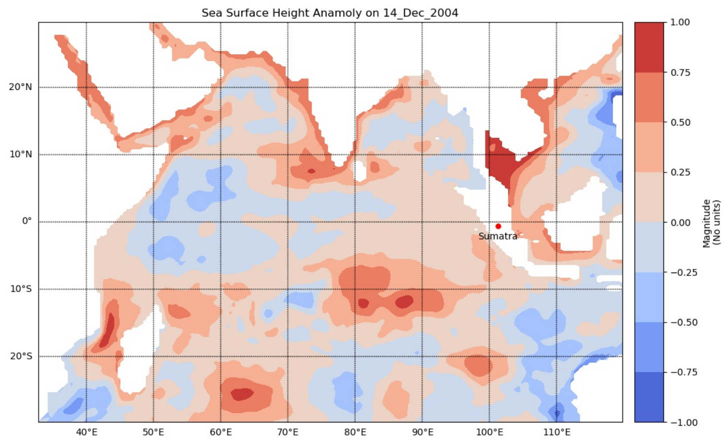

Figure 18. December 14, 2004

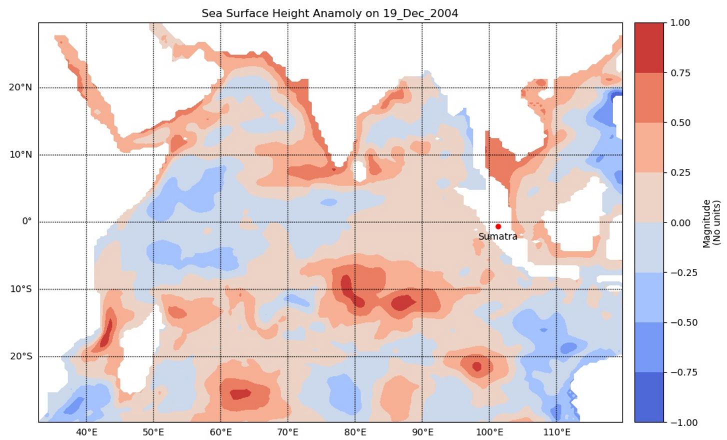

Figure 19. December 19, 2004

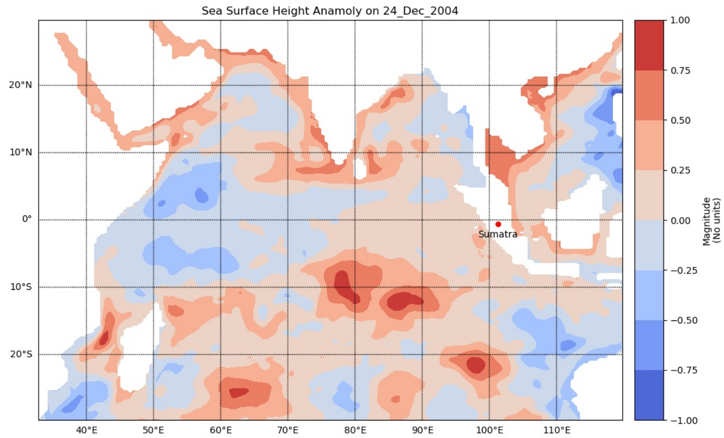

Figure 20. December 24, 2004

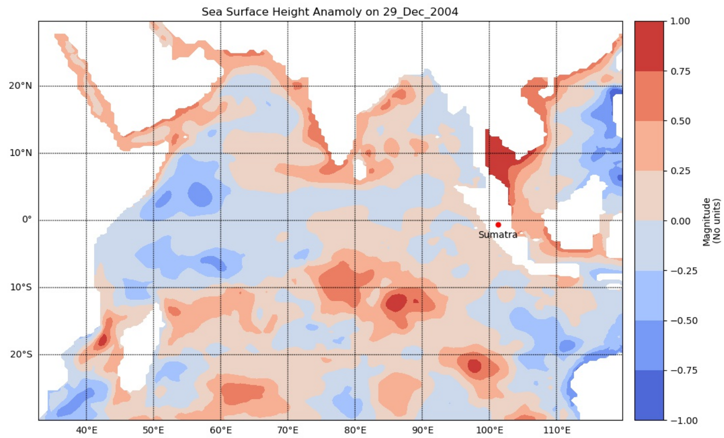

Figure 21. December 29, 2004

December, 2005 (No Tsunami) :

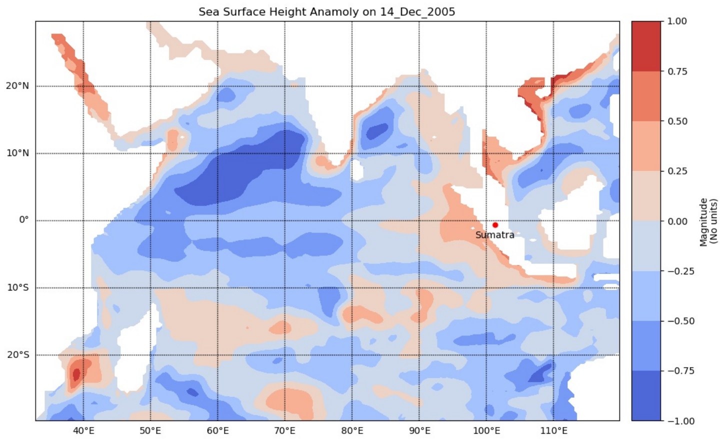

Figure 22. December 14, 2005

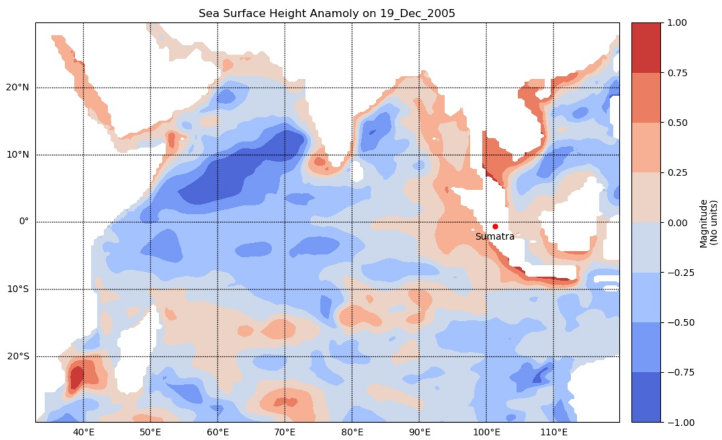

Figure 23. December 19, 2005

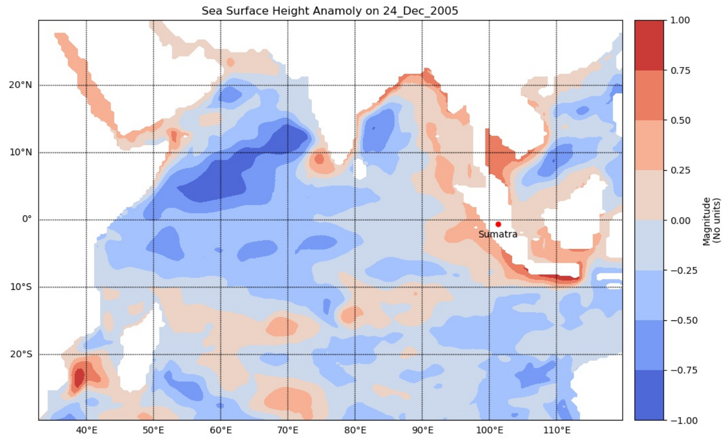

Figure 24. December 24, 2005

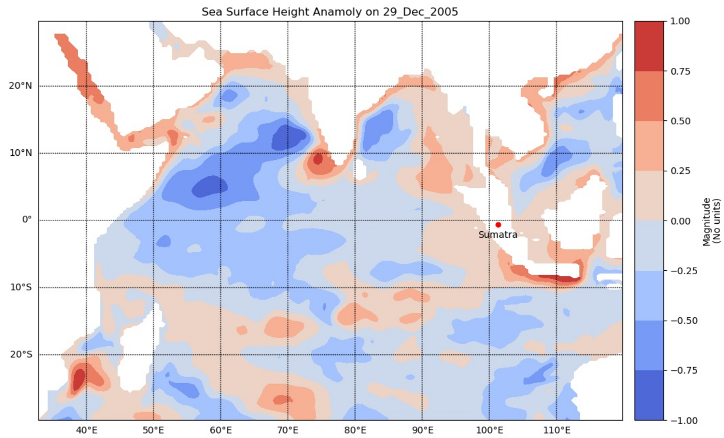

Figure 25. December 29, 2005

2. If one analyzes the ocean current patterns, one can clearly see a stark difference between the 29th December of 2003,2005 and 29th December of 2004, where the latter shows a much higher magnitude in contrast to the two.

29 December (No Tsunami) :

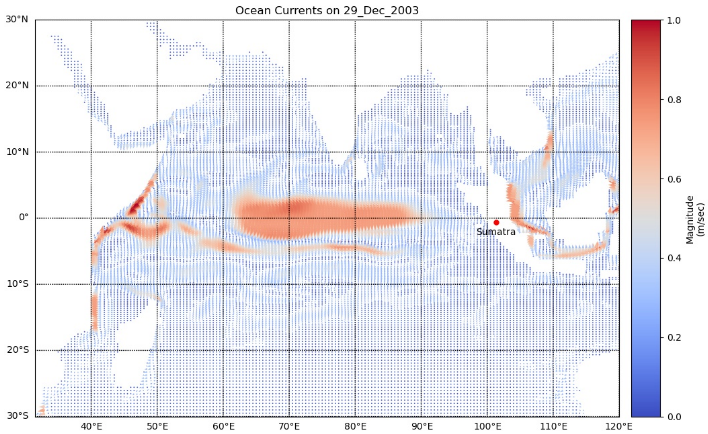

Figure 26. 29 December , 2003

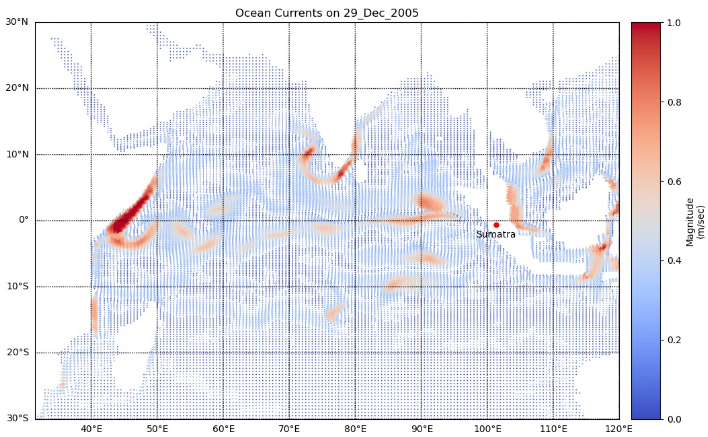

Figure 27. 29 December , 2005

29 December (Tsunami Struck) :

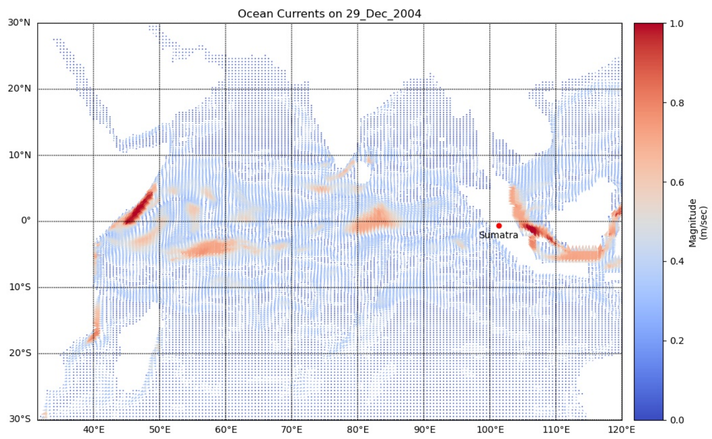

Figure 28. 29 December , 2004

4. Conclusion

In conclusion, it is worth to mention that the report has successfully elaborated on all the deliverables. The tools and methodology section entails the questions :

1. Which parts of the dataset were you able to use, and how have you been able to use?

2. Which visualizations did you choose, why, what technologies (Python libraries, others) did you use for the visualizations ?

All the indicative tasks have been successfully completed :

1. Generated Scalar Field Visualizations
2. Generated Vector Field Visualizations
3. Answered the questions :
   - Did the ocean variables reflects changes after the 2004 Indian Ocean tsunami on December 29, 2004? (Answered in the subsection Tsunami Related Inferences)

   - Did any ocean variables reflect the tsunami effects itself even though we don't have data for the exact day? (Answered in the subsection Tsunami Related Inferences)

   - As an informed citizen of the country, what did you learn about the Indian Ocean based on this dataset? (Answered in the subsection General Inferences)