# Data Visualization – Datathon 1

Sushranth Hebbar
IMT2017042
sushranth.hebbar@iiitb.org

## I. Introduction

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 was provided.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, and zonal current, during the period December 2003 – December 2005. The datasets are at 5-day interval, thus giving 147 timesteps. The objective of this report is to process the data, generate visualisations and draw inferences from these visuals.

## II. Methods

The process of tackling this problem can be divided into four stages:- Installing software packages, Data Preprocessing, Generating visuals for each dataset and drawing inferences.

### A. Software

First, a new enivronment was created using miniconda. Then python3.7, Basemap, numpy, pandas and matplotlib were installed. Other editing softwares like opencv and ffmpeg were later installed for generating videos from the datasets. Jupyter notebook was used as an intermediate platform for providing quick results.

### B. Preprocessing

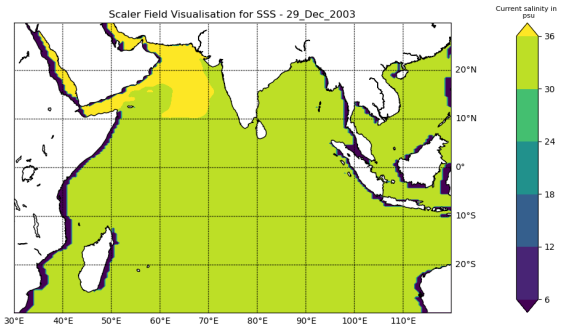
Now, each timestamp of a variable had extra information like the source of the data, the units of measurements and the columns used in the file. This occupied the first 10 lines. After, that there were data points. So, the first step was to remove these lines from every timestamp of every variable. Now, each timestamp is ready to be loaded into a dataframe by pandas.

The first observation was the presence of a value called bad flag which had to be dealt with as it was in the way of future computations. It was analogous to values like Nan or None which if kept in the dataset will distort future visualisations. So, the frequency of these values was obtained for every timestamp of very variable. This turned out to be the same value and it was over 11000. The next step was to determine if all these points were on land or water. Using basemap's utility function was enough to answer this question. It turns out that among 11000 points, only 1500 of them belong on the water surface and the rest of the points were on land.

This means that there were about 1500 points on the water surface which had unknown values, possibly due to a malfunction in the buoys. This was corrected by replacing all the bad flag values with 0 for all the timestamps of every variable except SST for which the average temperature was used. Since a majority of the bad flag valued points are on land, the above replacement makes sense. The other 1500 points which constitute less than 4% of the dataset will not cause any major fluctuations in the visuals.The next step was to obtain unique values of latitude and longitude from their respective columns and sort them so that they can be passed on to the respective basemap function. The value of the variable column was also converted from a 1D array into a 2D array in accordance with the syntax of the basemap functions. Finally, an image is generated corresponding to a particular time stamp of a variable. This whole process is repeated for all the timestamps to generate a total of 147 images for each variable. Finally, these images are stiched together into a video. Refer to link.txt to obtain a google drive link to those videos. The above steps are common to all variables.

### C. SSS

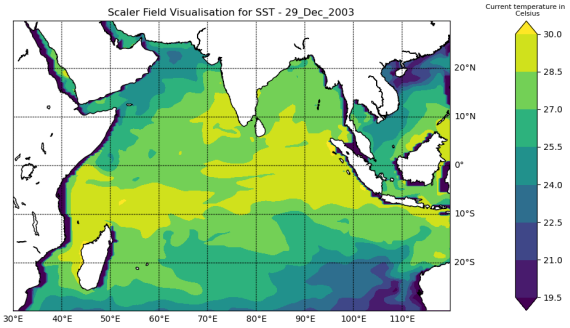
The contour fill function was used to obtain the below visual. The range of values are provided on the right in the form of a colourbar and the unit of measurement used was psu. The values on land were approximated as zero.




### D. SST

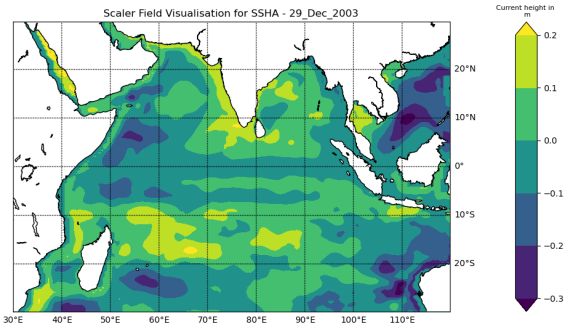
The contour fill function was used to obtain the below visual. The range of values are provided on the right in

the form of a colourbar and the unit of measurement used was Celsius. The values on land were approximated with the average temperature recorded during that timestamp.




### E. SSHA

The contour fill function was used to obtain the below visual. The range of values are provided on the right in the form of a colourbar and the unit of measurement used was metre. The values on land were approximated to zero.

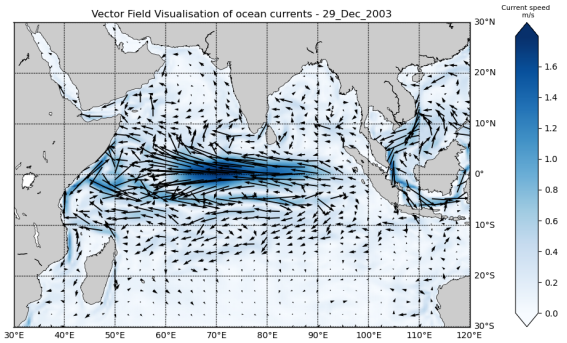


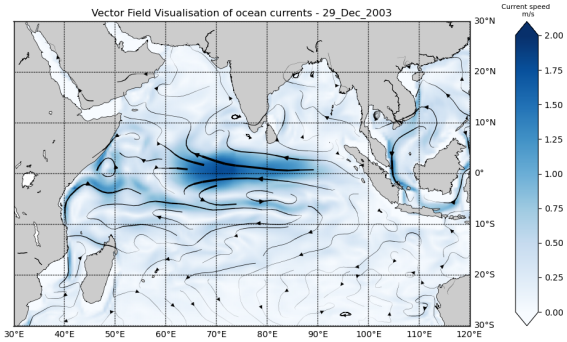

### F. Quiver Plots

The quiver function was used to generate a vector field for the ocean currents. For the sake of clarity the points were sampled with a step value set to five. The range of values are provided on the right in the form of a colourbar and the unit of measurement used was metre/sec. The values on land were approximated to zero. A colormesh was also applied to obtain the bluish background of the image. The absolute value of the current was calculated and used as a scalar field value in the colormesh function.

### G. Stream Plots

The streamplot function was used to generate a vector field for the ocean currents. The range of values are provided on the right in the form of a colourbar and the unit of measurement used was metre/sec. The values on land were approximated to zero. A colormesh was also applied to obtain the bluish background of the image. The absolute value of the current was calculated and used as a scalar field value in the colormesh function. A slight constraint was that there should be equal spacing between each of the latitudes and between each of the longitudes. But that was not the case with the latitudes. So, the latitudes had to be slightly adjusted and this resulted in a few of the stream lines being plotted on top of the continents.



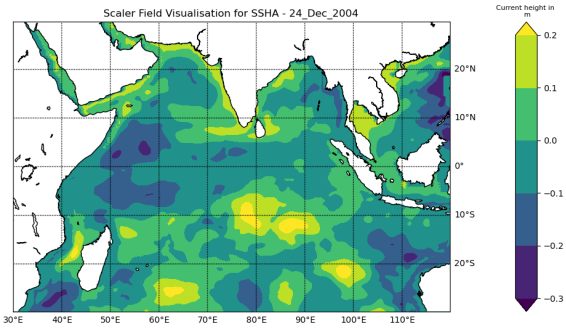



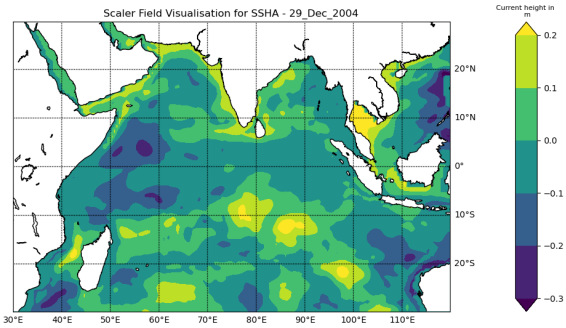

## INFERENCE

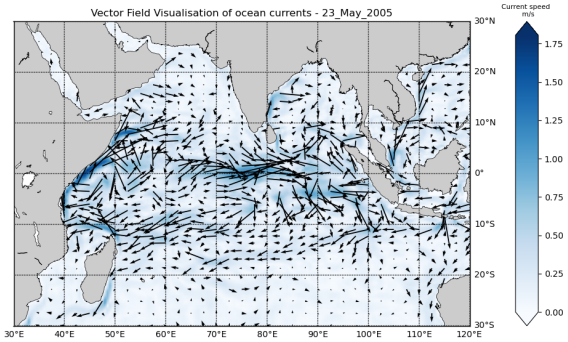
### H. Tsunami Effects

It's hard to say with certainty that any ocean variable reflected the tsunami effects because of the time gap between the each time stamp which is around five days. Theoretically, if we assume that there does exist such a variable, then it would have to be the SSHA variable. Now consider the below images.

It does not look like there is any drastic change in the variable before the tsunami and after the tsunami hit. The timegap of five days is simply too high. So, data closer to the time of occurrence of the tsunami might reflect the effects of the tsunami.




## *I. Monsoon*

Consider the below images that were captured during different seasons. In the month of February, i.e, before monsoon, the direction of the ocean currents is towards the west and it look like it is retreating. But during the month of May, i.e, during the start of the monsoon, the direction of the ocean currents is towards the east.