READ ME

Project: Relation between Sea Surface Salinity and Hurricane Prediction

General Info:

This is the final project for the graduate level course ME 5013 - Big Data Analysis in Spring 2021 at the University of Texas at San Antonio Mechanical Engineering Department.

Project statement:

Find a large data set and use python packages such as pandas, numpy, scipy, metpy, matplotlib, cartopy to draw analysis and test a hypothesis.

This specific project uses the NASA JPL SMAP Level 3 CAP Sea Surface Salinity Standard Mapped Image 8-Day Running Mean V5.0 Validated Dataset to compare and draw conclusions about the relation of sea surface salinity and hurricanes.

Hypothesis:

This data set contains information about sea surface salinity, temperature, and wind speed from all over the world since the satellite’s launch in 2015. Variables such as sea water salt concentration and variations has been found to be significant in weather forecasting. Experts in the field have found relationships between these variables during hurricanes. It would be interesting to see variations and correlations between salinity, temperature and windspeed to try and find patterns before hurricanes. If there is to be pattern in these variable changes some period of time before a natural disaster such as a hurricane, it could be further developed into adding onto current models or building new ones to predict and prepare for natural disasters. Using data provided by the U.S Geological Survey on the locations and time when hurricanes happened, the JPL SMAP SSS data could be retrieved using these parameters, make comparisons and analysis to look for patterns on salinity changes before hurricanes and see how the patterns for different hurricanes compare. The scope of the current project is reduced to Hurricanes Matthew and Irma, that occurred in the same Region in the Atlantic Ocean, in 2016 and 2017 respectively.

Python packages used: netCDF4, Xarray, Datetime, Pandas, Matplotlib, Numpy, Scipy.

Python code description: The code includes details about how to access, download, open the netCD4 data and manipulate it to extract the data of interest all using Xarray. It also includes plots of the salinity, wind speed and temperature mapped in the region of interest. The data set is 84 GB total including some nan values from earth surfaces that are not the sea or the times where the satellite could not retrieve data, hence the data was sliced to retrieve only the region of interest and then converted into a pandas data frame to perform further analysis.

Data set description:

The data set contains 12 variables: two with sea surface salinity data in practical salinity units (psu), sea surface temperature in Kelvin, two variables containing wind speed in meters per second (m/s), the Gaussian weighting factors, land fraction, ice\_fraction, uncertainty, latitude coordinate in degrees north, longitude coordinates in degrees east, and time in datetime64[ns] format.

The data is produced by the NASA JPL SMAP project. The data is averaged over a running mean of 8 days and validated.

Thorough details about the data can be found in:

https://cmr.earthdata.nasa.gov/search/concepts/C1972955240-PODAAC.html?token=EDL-O425c709adac4bf741b19e39363981a18e25ef8c65e09e065e51e09d5e71:OLpAZlE4HqIOMr0TYqg7UQ

Link to download:

https://search.earthdata.nasa.gov/downloads/7139064245

Student: Valeria Delgado Elizondo