Development of a tool to project and compare mean sea level variability between two recording stations

Hannah Walcek
Avnika Manaktala

Abstract: This project creates a tool that is able to compare mean sea level changes between two recording stations by conducting graphical data analysis, data quality checking, summarizing descriptive statistics and performing an ARIMA forecasting for each station.

Keywords: sealevelchange, NOAA, comparingsealevel

Project Introduction

Global mean sea level has risen about 8-9 inches since 1880 as a result of global warming according to the NOAA (2019). With these changing conditions comes a new challenge for most countries around the world which is staying afloat. These rising sea levels threaten a vast number of lives as 8 out of 10 major cities around the world are found near a coast. These rising sea levels also cause destruction in the form of storm surges such as Hurricane Katrina and "Superstorm" Sandy.

Being able to predict sea level changes seems to still be a challenge for the scientific world today. While long-term global changes can be mapped out it's impacts on each individual region should be closely monitored to help save lives of people living in coastal areas. Predicting these changes would help countries gain a preview of what is to be expected and help them better prepare resources for land management.

The National Oceanic and Atmospheric Adminstration (NOAA) detects mean sea level changes in recording stations around the world. These mean sea level datasets should be analysed as this can generate a predictive analysis which we aim to do which will give us a better idea on what changes are set to come. With the knowledge of mean sea level changes, we can be better prepared for the looming effects of climate change by allocating resources more efficiently to prepare for this change.

This project aims to create a tool that performs graphical data analysis, data quality checking, summarizes descriptive statistics and performs ARIMA forecasting for two recording stations from the NOAA website as chosen by the user.

Description of datasets

Preliminary exploration of the data available includes the importation of two datasets: one csv from Apra Harbor, Guam and the other from Juneau, Alaska. These two datasets were selected due to their vastly different locations, as Guam represents the tropical regions and Alaska represents the polar regions. We hope that the variability in the location of these stations will help us create a tool that can model the datasets regardless of the environment that they were collected in.

Description of the data

The table below shows the variables and their units that are provided in each dataset:

Variable	Unit
Year	-
Month	-
Monthly_MSL	Metres
Linear_Trend	Metres
High_Conf.	Metres
Low_Conf.	Metres

The highlighted row shows the monthly mean sea level (MSL) values for the station and this was the row used for our analysis. The Linear_Trend shows the values associated with the linear trend of this dataset and the High_Conf. and Low_Conf. Values show the values associated with the 95% confidence interval. The Juneau, Alaska csv file also includes the variable "Unverified" which does not have a description provided by the source.

The availability of this dataset changes between the stations as it depends on when the station was established. However, this station data is available for most coastal regions of the U.S. and the Pacific Islands. The Apra Harbor, Guam station provides data from January 1948 to December 2018, and the Juneau, Alaska station provides data from January 1936 to January 2020 at the time of download (March 2020).

Description of Graphical Analysis Methods

For this submission we have chosen two distinct datasets, and plotted them as a line plot to show the variability in the monthly MSL with time. The first dataset is from Apra Harbor, Guam and the second is from Juneau, Alaska.

Since these datasets are based on the variability of time we first created a time step to display the datasets as a time series plot. We took a monthly time step from the start dates for each dataset (January 1948 for Guam and January 1936 for Alaska) and plotted these two stations.

Figures

	Year	Month	Monthly_MSL	Linear_Trend	High_Conf.	Low_Conf.
0	1948	1	-0.001	0.007	0.056	-0.043
1	1948	2	-0.019	0.006	0.056	-0.043
2	1948	3	-0.014	0.006	0.056	-0.043
3	1948	4	0.056	0.006	0.056	-0.043
4	1948	5	0.068	0.006	0.055	-0.043
...	...	...	...	...	...	...
847	2018	8	0.057	0.126	0.198	0.053
848	2018	9	0.054	0.126	0.199	0.053
849	2018	10	-0.026	0.126	0.200	0.053
850	2018	11	-0.055	0.126	0.200	0.052
851	2018	12	-0.005	0.127	0.201	0.052

Figure 1: Head and tail of Apra Harbor, Guam data frame imported from csv using the Pandas package. Dates are shown ranging from January 1948 to December 2018. In total there are 852 measurements.

	Year	Month	Monthly_MSL	Unverified	Linear_Trend	High_Conf.	Low_Conf.
0	1936	1	1.051	U	1.050	1.067	1.033
1	1936	2	0.916	U	1.049	1.066	1.032
2	1936	3	0.929	U	1.048	1.065	1.031
3	1936	4	0.963	U	1.047	1.063	1.030
4	1936	5	1.052	U	1.046	1.062	1.029
...	...	...	...	...	...	...	...
1004	2019	9	-0.111	NaN	-0.065	-0.049	-0.081
1005	2019	10	-0.213	NaN	-0.066	-0.050	-0.082
1006	2019	11	-0.205	NaN	-0.067	-0.051	-0.083
1007	2019	12	-0.004	NaN	-0.068	-0.052	-0.084
1008	2020	1	-0.059	NaN	-0.069	-0.053	-0.085

Figure 2: Head and tail of Juneau, Alaska data frame imported from csv using the Pandas package. Dates are shown ranging from January 1936 to January 2020. In total there are 1009 measurements. This dataset differs from the previous one in that it includes an "Unverified" column indicating some measurements have not been subjected to the National Ocean Service's quality control procedures. These values may need to be omitted.

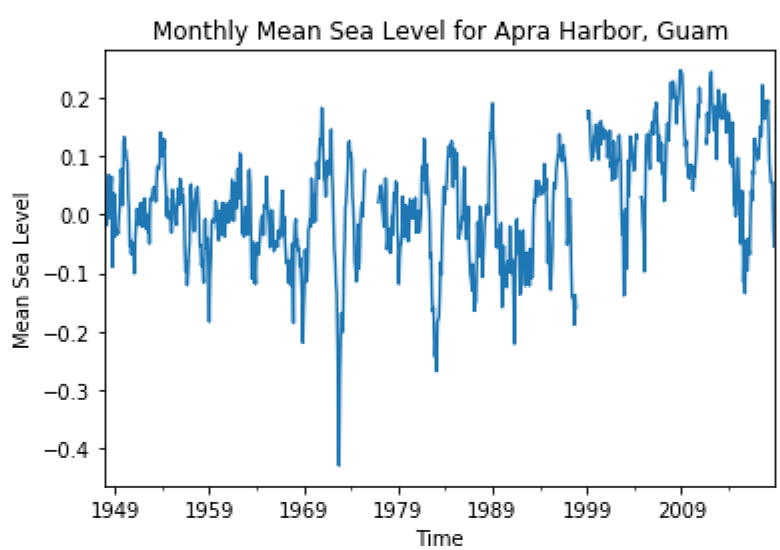

Figure 3: Monthly Mean Sea Level over time plotted as a time series for Apra Harbor, Guam. This was selected to show the current variability in mean sea level in the area which can be compared to other stations. This dataset was specifically chosen to represent variability of mean sea level in the tropics.

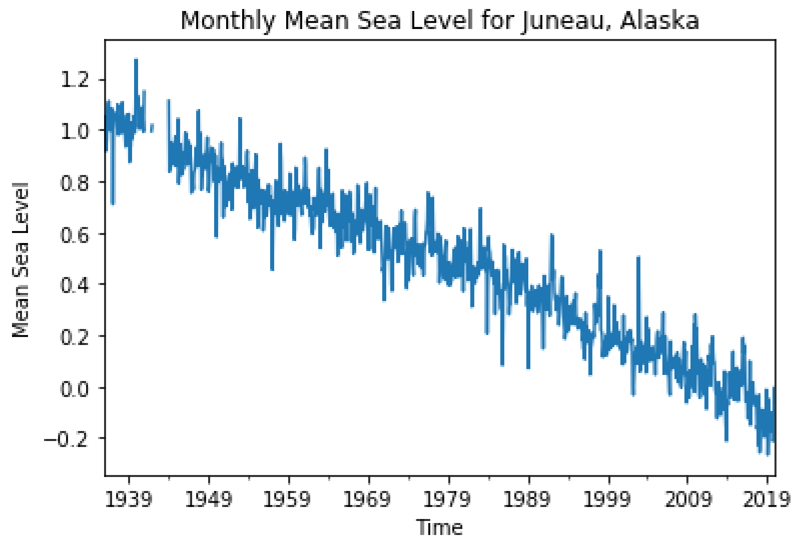

Figure 4: Monthly Mean Sea Level over time plotted as a time series for Juneau, Alaska. This was selected to show the current variability in mean sea level in the area which can be compared to other stations. This dataset was specifically chosen to represent variability of mean sea level in the polar regions.

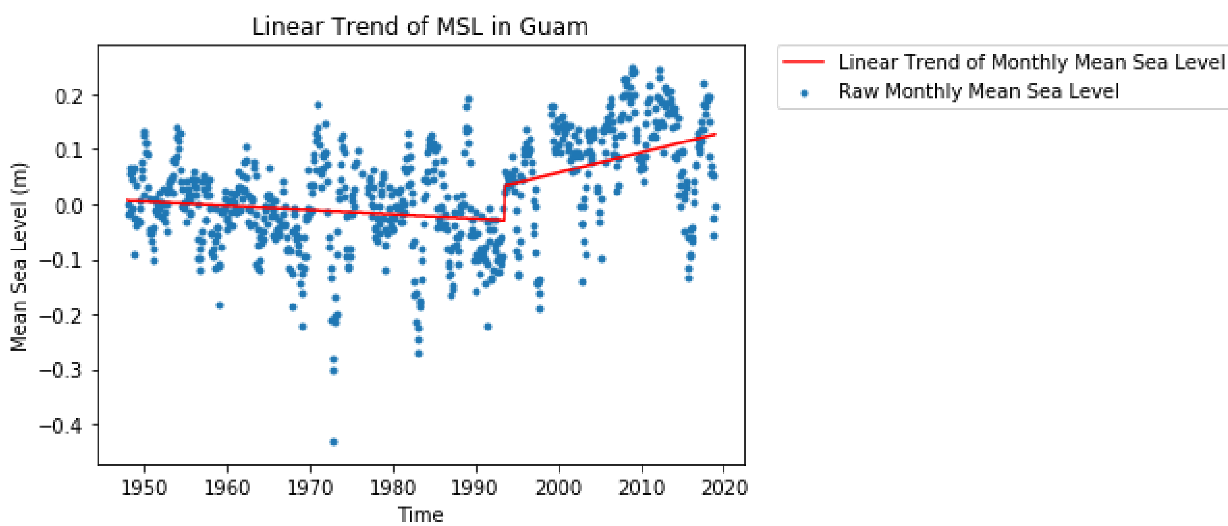

Figure 5: Linear trend (red) with respect to Monthly Mean Sea Level (blue) for Guam. The sudden increase in MSL around 1995 is attributed to the occurence of an earthquake in the region. Overall, this graph shows an increase in MSL.

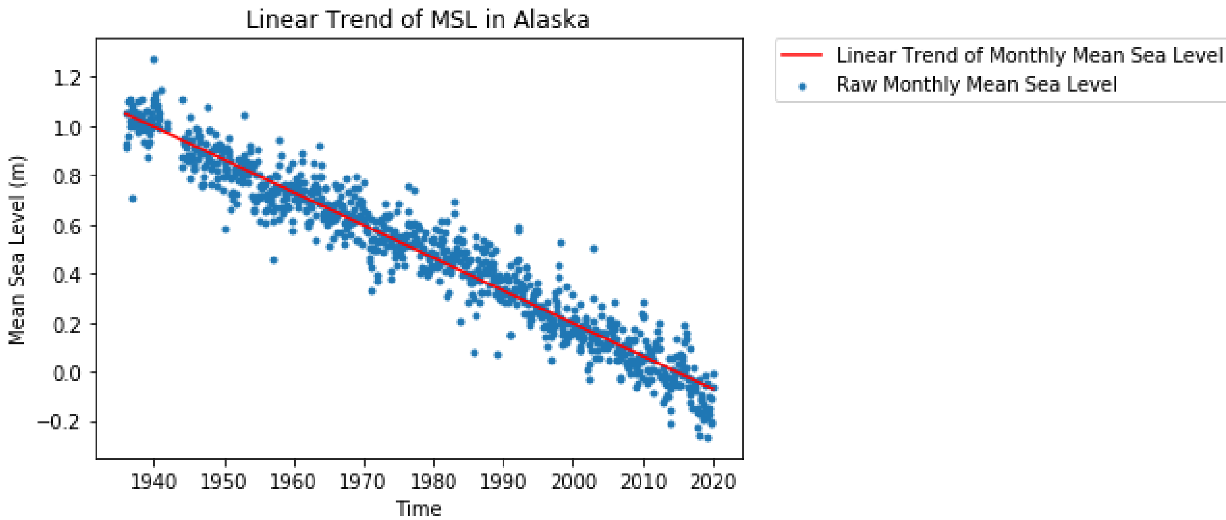

Figure 6: Linear trend (red) with respect to Monthly Mean Sea Level (blue) for Alaska. The graph shows a steady decline in MSL with no sudden changes in the variability.

<u>Methodology for Data Quality Checking:</u>

The data quality checking for this project mainly focuses on the Monthly_MSL variable within the datasets. Two checks were made on this variable as listed below:

1. Removing no data values: No data values are denoted as whitespace characters in the raw file. These were removed from the datasets to ensure that these do not affect the results obtained.
2. Range check for Monthly_MSL: The Monthly_MSL values were checked to see if they were lesser than 1.7 mm as according to NOAA, the maximum possible change in mean sea level is 1.7 mm/year.
3. Removing outliers by examining the high and low confidence columns. Creating the column “Conf_Width” which describes the difference between high and low confidence. By creating an IQR score, we were able to set parameters for what constitutes an unacceptable confidence width.

<u>Results of Data Quality Checking:</u>

Final changed values counts.....

	Month	Monthly_MSL	Linear Trend	High_Conf.	Low_Conf.
1. No Data	37.0	0.0	37.0	37.0	37.0
2. Range Fail	0.0	0.0	0.0	0.0	0.0
3. Confidence Fail	0.0	0.0	0.0	0.0	0.0

Figure 7: Replaced Value info for Guam data. This table shows that the main operation performed for this dataset is the removal of no data values. This was expected as these datasets have already been cleaned by the NOAA.

```
Final changed values counts.....
                       Month  Monthly_MSL  Linear Trend  High_Conf.  Low_Conf.
1. No Data              0.0         0.0           0.0         0.0        0.0
2. Range Fail           0.0         0.0           0.0         0.0        0.0
3. Confidence Fail      0.0         0.0           0.0         0.0        0.0
```

Figure 8: Replaced Value Info for Alaska Data shows that no changes have been made to the dataset. This was expected as the values have been processed before being uploaded onto the NOAA website.

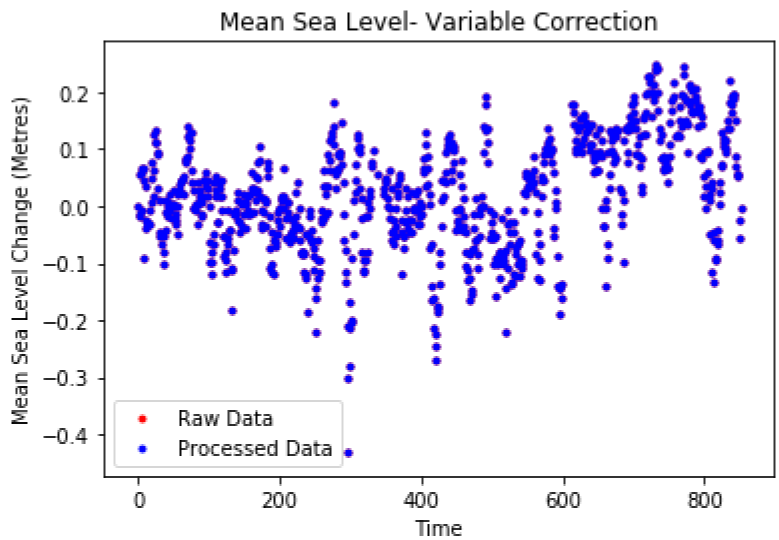

Figure 9: Mean sea level correction for Guam shows no differences in the datasets due to the reasons stated above.

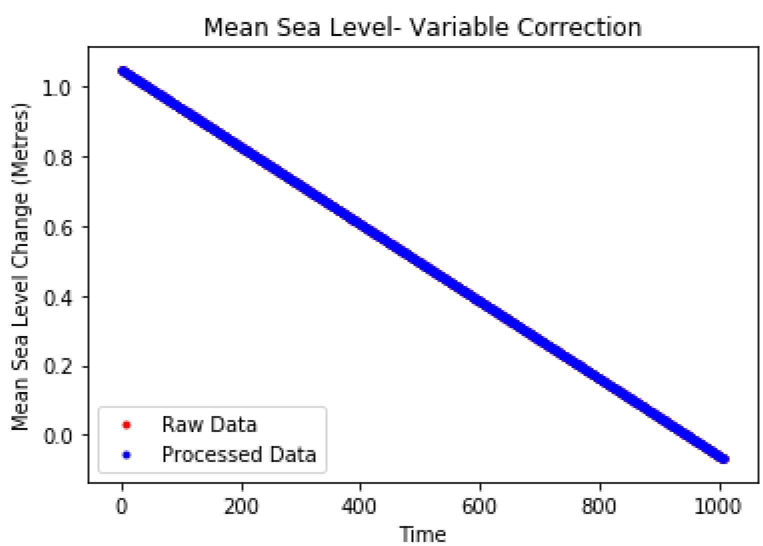

Figure 10: Mean sea level correction for Alaska shows no difference in the values for this dataset as the values have been cleaned before being updated on the website.

Description of Data Quality Checking Results:

For the Guam dataset 37 data points were removed due to the presence of NA values and no values were removed due to the range check or confidence check. For the Alaska dataset no values were removed for any checks

It is possible that the datasets were cleaned before they were uploaded on the NOAA website. However, it is always a good practice to ensure that the dataset is processed for any data quality errors before working with it.

Summary of Descriptive Statistics:

Since this project looks at mean sea level data and compares the data and forecasts between two recording stations, the following metrics were calculated and compared for each station:

1. Maximum mean sea level
2. Minimum mean sea level
3. Average mean sea level
4. Median mean sea level
5. Percentage of months where mean sea level exceeds twice the median
6. Percentage of months where mean sea level exceeds twice the mean
7. Lowest average mean sea level for 12 months

The following data was found for each station:

Guam dataset:

Description	Statistic
Maximum mean sea level	0.248
Minimum mean sea level	-0.431
Average mean sea level	0.021
Median mean sea level	0.018
Percentage of months where mean sea level exceeds twice the median	42.331 %
Percentage of months where mean sea level exceeds twice the mean	40.736 %
Lowest average mean sea level for 12 months	-0.202

Alaska dataset:

Description	Statistic
Maximum mean sea level	1.050
Minimum mean sea level	-0.069
Average mean sea level	0.490
Median mean sea level	0.490
Percentage of months where mean sea level exceeds twice the median	6.244%
Percentage of months where mean sea level exceeds twice the mean	6.244 %
Lowest average mean sea level for 12 months	-0.063

ARIMA forecasting:

Using the cleaned data sets the station data was used to forecast the mean sea level changes for each of the stations. For this project a large number of datapoints have been chosen to show the forecasting for the datasets.

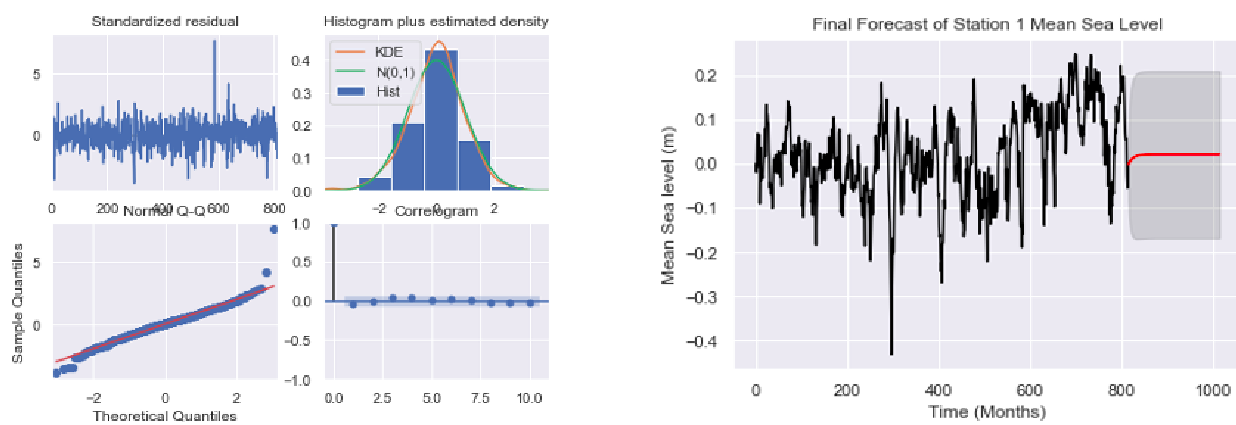

Figure 11: ARIMA forecasting for Station 1 (Guam) shows that the most of the remainders used for this forecasting were uniform and they follow a normal distribution. The forecasting shows a slight increase in mean sea level. Additional points have been forecasted to highlight the forecasting portion of the project.

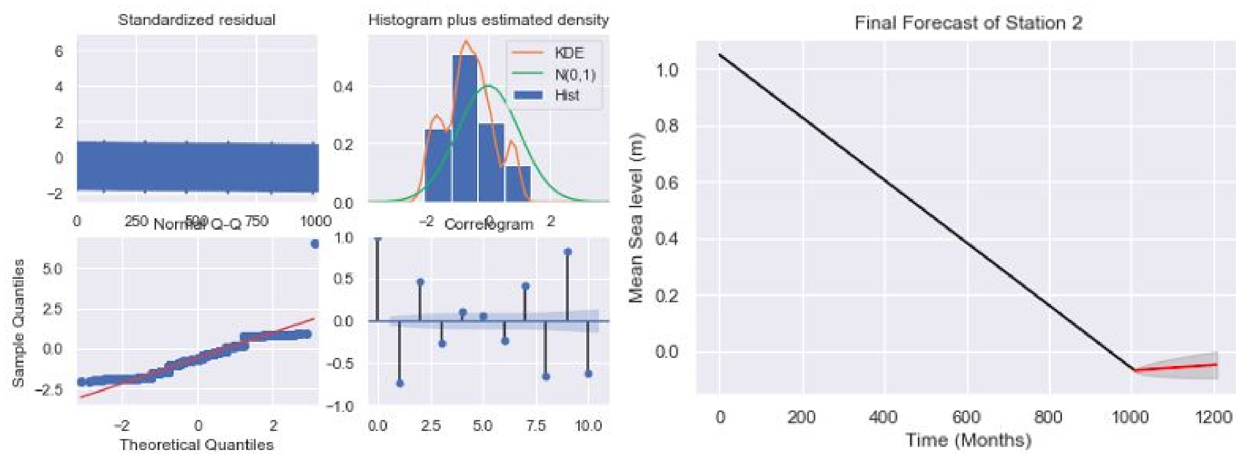

Figure 12: ARIMA forecasting for Station 2 (Alaska) shows that most of the remainders were not uniform for this forecasting as they are negatively skewed which affects the forecasting done. The forecast is unable to show a clear trend before it levels out.

This project provides the above graphs as a result for comparing the forecasts of the sea level change for two recording stations. This should be treated as a starting point for analysis on changes in mean sea level.

Overview of data and script files provided

This submission includes the following documents:

- “Project-code.ipnyb”: This file contains the code used to generate all the results and forecasting done in this project.
- “1630000_meantrend.csv”: This file contains the Guam dataset used to generate all the plots and figures for this project. This file is directly used for developing the project code.
- “9452210_meantrend.csv”: This file contains the Alaska dataset used to generate all the plots and figures for this project. This file is directly used for developing the project code.
- “IntermediateResults.png”: This image shows the data for the two stations after data quality checking has been performed on both stations
- “Finalresult-station1.png”: This image shows the ARIMA forecasting for station 1 (Guam for this test)
- “Finalresult-station2.png”: This image shows the ARIMA forecasting for station 2 (Alaska for this test)