Global Mean Sea Level

Team members

Abhinav Merugu: FB Prophet model and evaluation.

Ancit: ARIMA model and evaluation.

Hamsaraj Tupthi: Data collection and visualization.

Introduction

Global mean sea level has increased rapidly in last century. It has risen about 8-9 inches since 1880, about a third of that has risen in just last 25 years. The rising water level is mostly due to combination of melt water from glaciers and thermal expansion of seawater as it warms. Over last decade, global mean water level is risen by 0.14 inches (3.6 millimeters) per year, which was 2.5 times the average rate of 0.06 inches per year throughout most of the twentieth century. It is estimated by the end of century per year throughout most of the twentieth century, even if the consumption of the greenhouse gas is reduced to an extent. Almost half of the world population live near coastal region. So, it has become important to predict sea level rise with high accuracy and precision.

Problem Definition and Algorithm

Global sea levels are rising and increasing the risk to coastal communities from inundation and erosion. The principal components contributing to global average sea level rise are the melting of land-based snow and ice reserves and the thermal expansion of the ocean water mass. 8 out of world largest cities are near coast. Increased sea level has impact on infrastructure, subways, regional jobs. Sea level plays a role in soil erosion, floods, hazards from storms. Most of early predictions about sea level are done on temperature dataset rather than sea level dataset. Very few researches have been carried out to predict sea level dataset. To our knowledge there is no open source project that examines and predicts sea level changes in consideration with other factors such as CO2 and global temperature.

Algorithm

We have implemented two time series forecasting algorithms ARIMA and FB Prophet.

Design and Implementation

Dataset Collection

We combined three datasets of Global Mean Sea Level Variation, CO2 ppm and Global Temperature Anomalies into one, sources of dataset are CSIRO, NOAA, more information in references.

Dataset Processing

Since our dataset was collected from official government websites, we did not have to deal with null values. We only had to choose right columns and converging dates from different datasets to merge and finally convert date column from string type to date time type.

Models

Facebook's Prophet

Prophet is a procedure for forecasting time series data based on an additive model where non-linear trends are fit with yearly seasonality. It works best with time series that have strong seasonal effects and several seasons of historical data. Prophet is robust to shifts in the trend, and typically handles outliers well.

We use a decomposable time series model (Harvey & Peters 1990) with three main
model components: trend, seasonality, and holidays. They are combined in the following
equation:

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t.$$

Here g(t) is the trend function which models non-periodic changes in the value of the time series, s(t) represents periodic changes (e.g., yearly seasonality), and h(t) represents the effects of holidays which occur on potentially irregular schedules over one or more days, we haven't used this hyper-parameter as we do not consider holidays in our problem context. The error term et represents any idiosyncratic changes which are not accommodated by the model; later we will make the parametric assumption that et is normally distributed.

FB Prophet provides below advantages:

Flexibility: We can easily accommodate seasonality with multiple periods and let the analyst make different assumptions about trends.

Unlike with ARIMA models, the measurements do not need to be regularly spaced, and we do not need to interpolate missing values e.g. from removing outliers.

Fitting is very fast, allowing the analyst to interactively explore many model specifications.

The forecasting model has easily interpretable parameters that can be changed by the analyst to impose assumptions on the forecast. Moreover, analysts typically do have experience with regression and are easily able to extend the model to include new components.

Linear Trend with Change Points

$$g(t) = (k + \mathbf{a}(t)^\intercal \boldsymbol{\delta})t + (m + \mathbf{a}(t)^\intercal \boldsymbol{\gamma}),$$

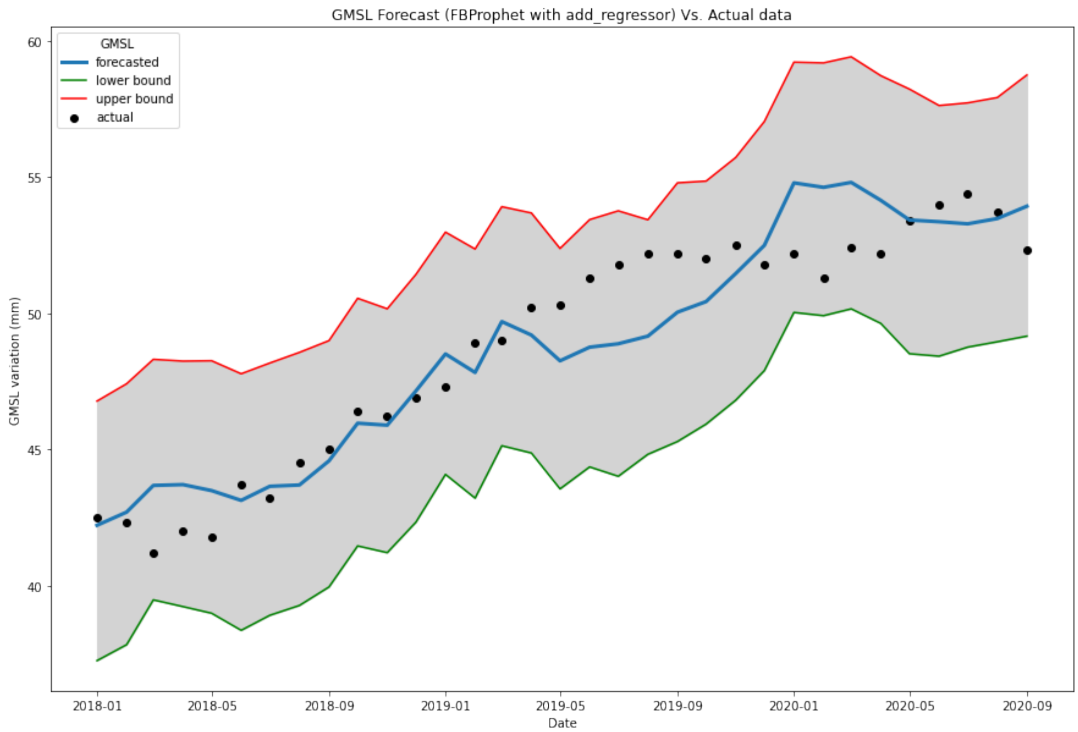
For forecasting problems that do not exhibit saturating growth, a piece-wise constant rate of growth provides a parsimonious and often useful model.

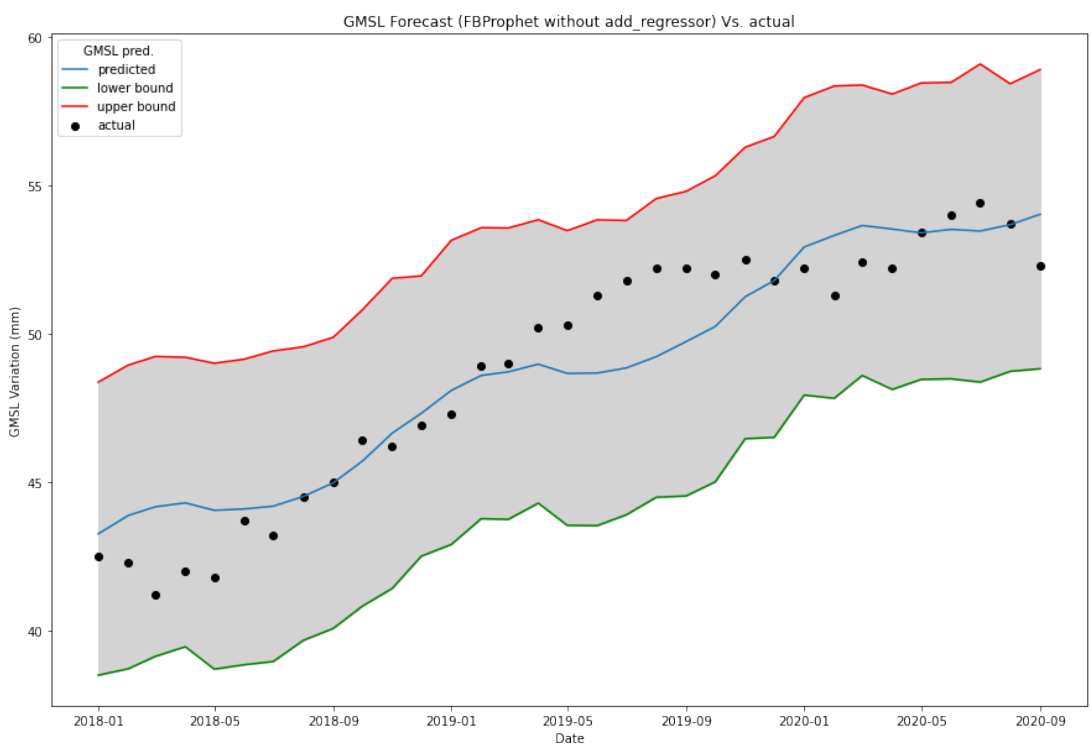
Experiment Setup

Our plan of action for this model is to check if model performs better with or without add_regressor. To implement the model we followed below steps:

1. Rename 'date' to 'ds' and 'gmsl' to 'y'.
2. Split dataset such that there are 300 datapoints in training and rest in testing.
3. Test the model with add_regressor applied for 'co2ppm' and 'gt_anomalies' and plot.
4. Test the model without add_regressor applied for 'co2ppm' and 'gt_anomalies' and plot.
5. Compare how the model predicts while tuning changepoint_prior_scale hyperparameter.
6. After comparing, evaluate the need for predicting regressor values for gmsl with future dates and data by training model with 'co2ppm' and 'gt_anomalies' data separately.
7. Once the final model is generated, evaluate and tune hyperparameters to handle over fitting and under fitting.

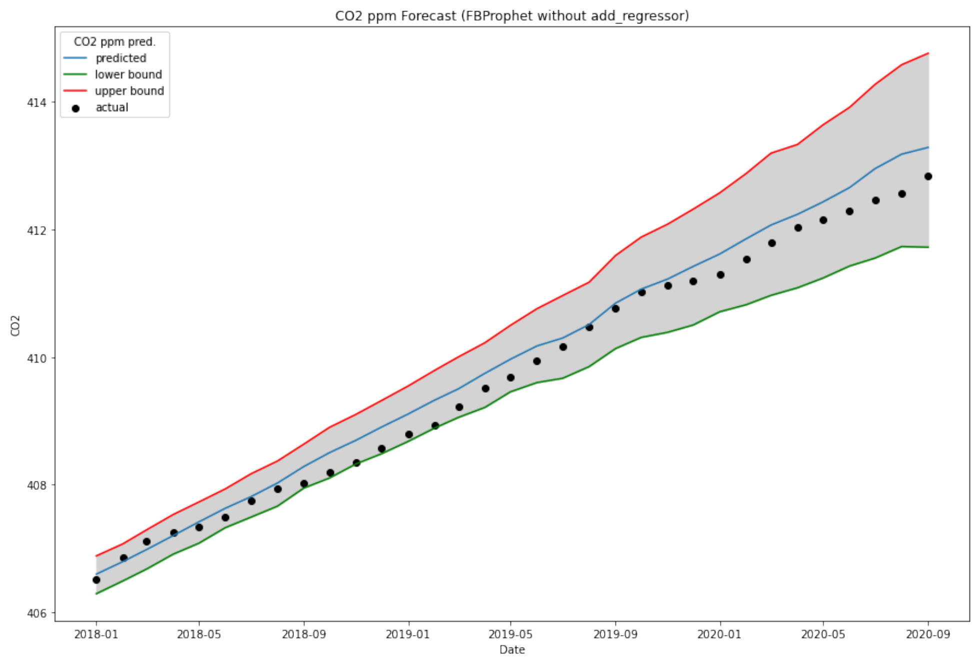
Graphs showing the difference in forecasting with (1) and without (2) add_regressor for 'gmsl' test data with lower and upper bounds for forecast data:

1. 

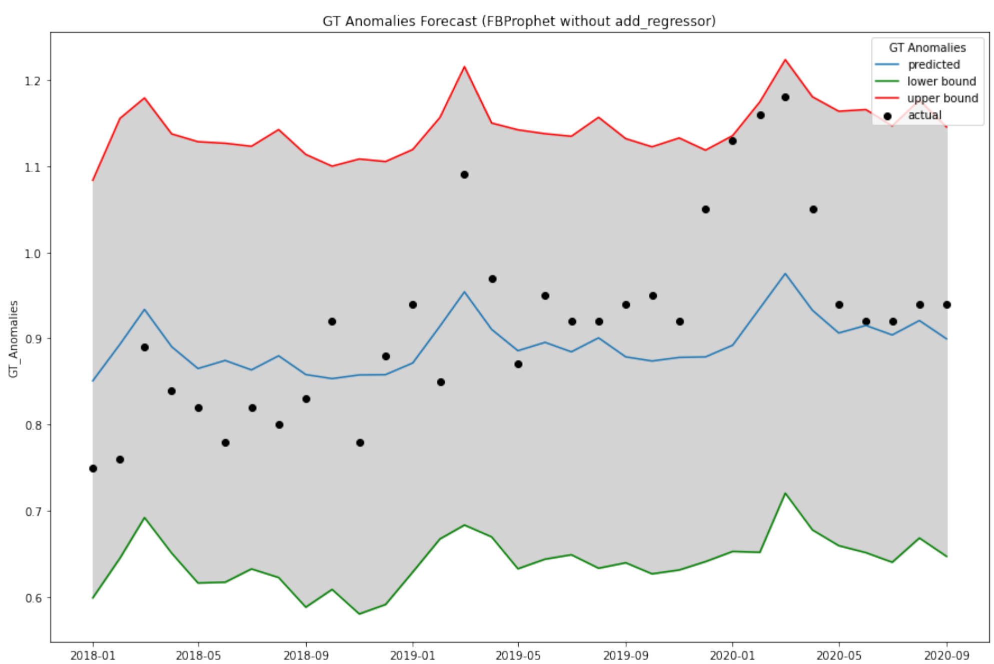
2. 

From the above graphs it is evident that use of add_regressor makes better forecasting, so we generate future dates for 'co2ppm' and 'gt_anomalies' to make an add_regressor model to forecast future dates of 'gmsl'.

1. CO2 ppm

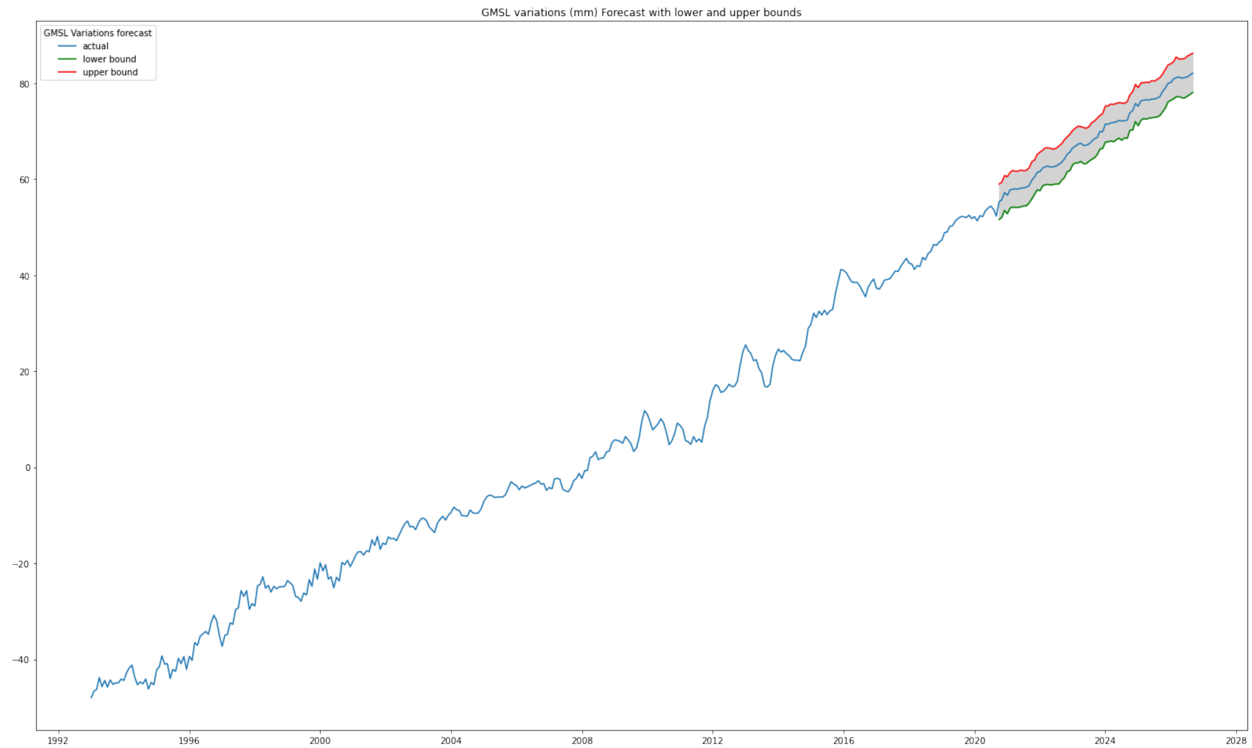


2. Global temperature anomalies



Result

Final plot shows the forecast using above forecast values for add_regressor model



Nature is unpredictable and can sometimes sometimes record abnormal values like from the year 2012 - 2028. Any forecasting model can only help us assume with some confidence that the nature behaves as expected by the model. It also implies that other factors such as CO2 and Temperature anomalies have influences GMSL.

Result Evaluation

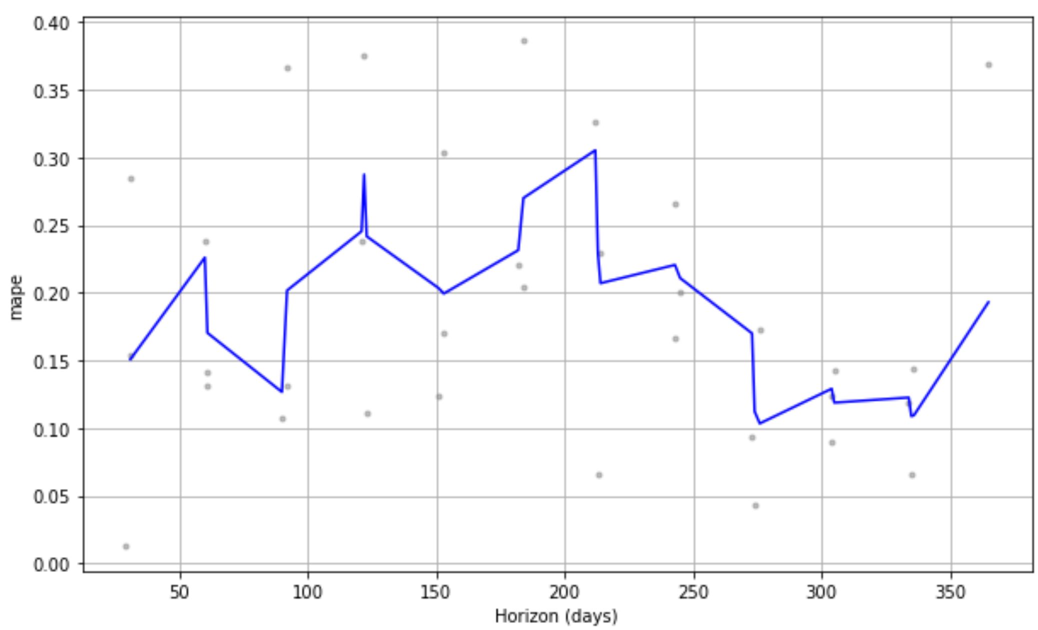
Facebook's Prophet model has built-in model evaluation functions, there is a procedure suggested in the documentation to perform analysis done on historical data:

- Training data (initial): The amount of data set aside for training, the first parameter.

- Cutoff (period): A forecast is made for every observed point between cutoff and cutoff + horizon.
- Horizon: The data set aside for validation. Here we chose 365 days, which is 1 year.

Once the cross validation is done with above parameters, we invoke performance metrics function on the result. Then mean is calculated for all the cutoff value metrics. We can even plot to see each metric graphically.

```
horizon     194 days 02:00:00
mse                  6.209608
rmse                 2.317222
mae                  1.884162
mape                 0.136569
mdape                0.126251
coverage             0.805556
```



Model performs well with regards to historical data analysis, it has ~13% Mean Average Percentage Error, which means that the model makes forecasts with 13% chance of error for each data point which is influenced by the historical data.

Challenges

- Limited dataset
- Limited hyperparameters for tuning.
- Better understanding of under-fitting and overfitting for forecasting.

Future Enhancements

- Find a larger dataset for each of the features.
- Explore how other forecasting models like LSTM and XG Boost performs.
- Find more factors that can be added to the model.

Final Notes

We have divided our code into three different Jupyter notebooks, dataset_prep, fb_prophet and arima to keep the notebook clean.

References

- https://www.cgg.wa.gov.au/live/my-environment/coastal-hazard-risk-management-and-adaptation-planning/the-importance-of-sea-level-rise.aspx
- https://peerj.com/preprints/3190/
- https://facebook.github.io/prophet/
- https://link.springer.com/article/10.1007/s10712-019-09525-z#Par2700
- https://towardsdatascience.com/business-forecasting-with-facebook-prophet-b9aaf7121398
- https://facebook.github.io/prophet/docs/quick_start.html#python-api
- https://podaac.jpl.nasa.gov/MEaSUREs-SSH?tab=background&sections=about%2Bdata%2Bnews%2Bresources
- https://www.digitalocean.com/community/tutorials/a-guide-to-time-series-forecasting-with-prophet-in-python-3
- https://climate.nasa.gov/vital-signs/sea-level/
- https://sealevel.nasa.gov/understanding-sea-level/global-sea-level/overview
- https://www.kaggle.com/datasets/somesh24/sea-level-change
- https://research.csiro.au/slrwavescoast/sea-level/measurements-and-data/sea-level-data/#OurData
- https://www.ncdc.noaa.gov/cag/global/time-series
- https://gml.noaa.gov/ccgg/trends/data.html
- https://ourworldindata.org/co2-and-other-greenhouse-gas-emissions#how-have-global-co2-emissions-changed-over-time
- https://www.climate.gov/news-features/understanding-climate/climate-change-global-sea-level
- https://tidesandcurrents.noaa.gov/sltrends/sltrends.html