Talia Arauzo
CE 203 Surface Water Hydrology

Time Series Analysis of SST

I recently watched the Netflix documentary "Seaspiracy" and it left me shook. The amount of destruction and plunder we humans have incurred on our oceans (not to mention planet) is astounding. Covering heavy topics from whale and dolphin hunting to overfishing to corruption to modern slavery, the documentary aims to highlight some key injustices occurring in our marine environments. Although there is much (questionable?) controversy surrounding the film's statistics and representations, there was one indisputable implication that struck a chord with me. The ocean is the planet's largest carbon sink and as ocean temperatures continue to rise, the ocean is losing its ability to sequester carbon. For reference, losing just one percent of the ocean's carbon stores is equivalent to releasing emissions from 97 million cars. Thus, the ocean losing its carbon sequestration capabilities is catastrophic for all life on earth.

After learning this devastating information, I became curious as to the magnitude, speed, and consequences of warming ocean temperatures. Rising sea surface temperatures (SST) have a wide range of impacts besides affecting carbon sequestration abilities, including ecological, human health, and economic impacts. Ecologically, rising SST can alter species distributions, food webs, predator prey relationships, and reproductive timing, to name a few. However not only is marine life impacted but human life. Warming ocean temperatures will lead to sea level rise, causing erosion, flooding, salt water contamination, beach usage reassessment, etc. These frightful outcomes will incur severe economic impacts as well, as infrastructure will become damaged, many will be forced to relocate, fishing practices and markets will be disrupted or terminated, etc. Thus, rising ocean temperatures is a crucial issue and I will explore such through a time series analysis of SST.

Utilizing NOAA's public COBE SST dataset, I acquired the monthly mean SST anomalies from the climatology for a one degree latitude by one degree longitude global grid. To focus on local impacts and simplify my research, I chose to concentrate my analysis on the dataset's closest location to Berkeley: Half Moon Bay, located at a latitude of 37.5 and longitude of -122.5 (coordinates consistent with the dataset's nomenclature). Using the data from this locale, which contains recordings from 1880 until 2017, I built three time series forecasting models and then with the best performing model, forecasted ten years into the future.

As the dataset contains many years worth of data, I decided to segment it in order to test and train the models. Attempting to omit sampling discrepancies and utilize recordings acquired with modern technology, I selected a recent twenty seven years of Half Moon Bay data as my training set, from January 1980 to December 2006. My test set consisted of a recent ten years of Half Moon Bay data, from January 2007 to December 2017. I partitioned my data in this manner so that the training and test data would be used to build and validate the models respectively. Figure 1 displays the test and training datasets.

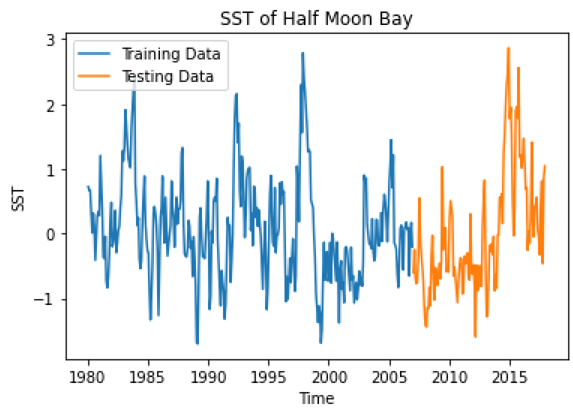

Figure 1: Test and Training Data

To preface, time series forecasting is used to forecast a measured quantity at points in the future based on historical, consecutive observations. Often, time series data is decomposed into components of trend and seasonality. Trend is the increasing or decreasing behavior of a series over time whereas seasonality is the repeating patterns or cycles of behavior over time. Trend and seasonality are described as either additive or multiplicative, where additive indicates a somewhat constant change over time and multiplicative indicates an increasingly exaggerated change over time. Figure 2 displays the varying combinations of trend and seasonality.

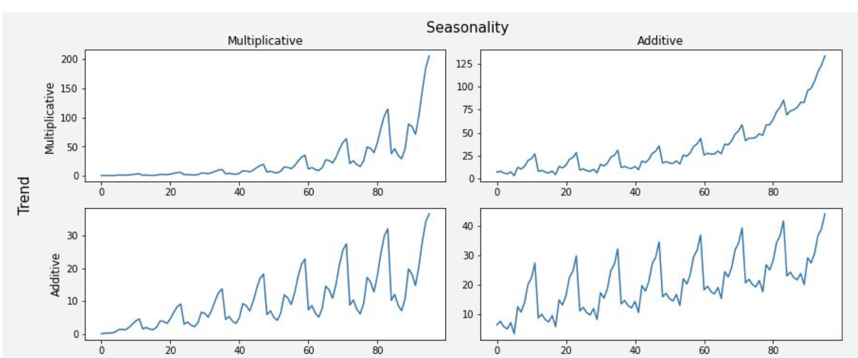

Figure 2: Seasonality and Trend Combinations

Understanding these basics of time series data informed my model selection. As the COBE SST mean anomaly data for Half Moon Bay displayed additive seasonality, I decided to utilize models that best represent such: triple exponential smoothing, SARIMA, and Prophet. The triple exponential smoothing model, which I implemented from the statsmodels library, bases predictions on a weighted sum of past observations, with exponentially decreasing weights for past observations. Unlike single and double exponential smoothing, triple exponential smoothing incorporates support for seasonality to univariate time series data. The SARIMA model, which I also implemented from the statsmodels library, bases predictions on a weighted sum of past observations, differencing, and moving averages. SARIMA stands for seasonal ARIMA, or AutoRegressive Integrated Moving Average, thus indicating the model's ability to take into account seasonal components. The last model, Prophet, which I implemented from Facebook's Prophet library, predicts non-linear trends that experience seasonality and holiday effects. The most complex of the three models, the Prophet model is best for data with strong seasonal effects and several seasons of historical data (hint hint).

In order to determine which of these three models performed best, I assessed each model with the following four metrics: R^2, MAE, MAPE, and MSE. The coefficient of determination, or R^2, is the proportion of the variance in the dependent variable that is predictable from the independent variable; in simpler terms, it represents the model's goodness of fit to the actual data. MAE, also known as the mean absolute error, is an arithmetic average of the absolute errors, where the errors are the difference between the actual values and the model's predicted values; this metric indicates how big of an error you can expect on average from the model. The third metric, mean absolute percentage error (MAPE), measures the accuracy of a forecasting model and highlights its relative error. Finally, the mean squared error, or MSE, represents the average of the squares of the errors; as such, MSE emphasizes large errors or forecasting outliers produced by the model. Figures 3 displays the equations used to calculate the four metrics. However, the metrics were largely implemented via the sklearn metrics library.

$$\mathrm{R}^2 = 1 - \frac{SS_{res}}{SS_{tot}} \qquad \mathrm{MAE} = \frac{\sum_{i=0}^{n} |y_i - \hat{y}_i|}{n} \qquad \mathrm{MAPE} = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \qquad \mathrm{MSE} = \frac{\sum_{i=0}^{n} (y_i - \hat{y}_i)^2}{n}$$

Figure 3: Metrics Formulae

The results of the trained triple exponential smoothing, SARIMA, and Prophet models are displayed in Figures 4-6 respectively and the corresponding metrics in Table 1. Overall, the Prophet model performed best, despite visible objections. Due to the model's complexity and propensity toward minimizing error, the Prophet model output conservative (yet accurate) predictions.

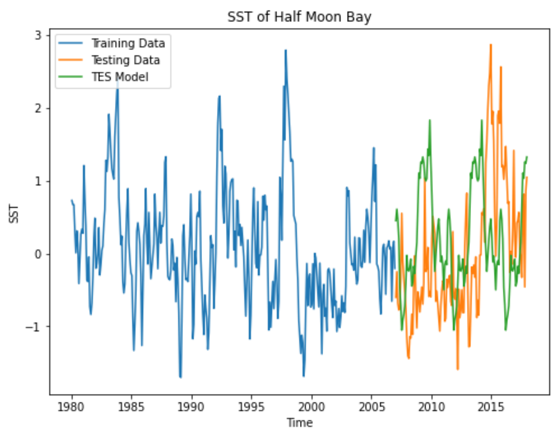

Figure 4: Triple Exponential Smoothing Model

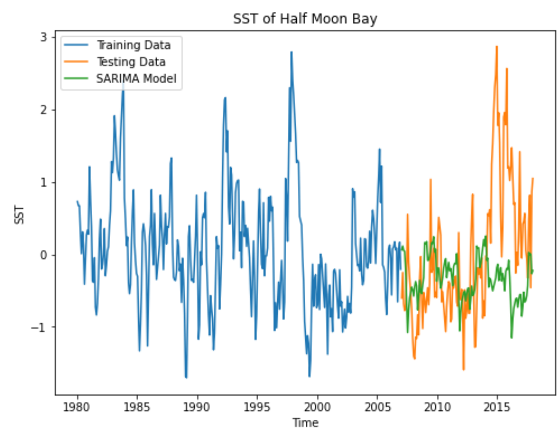

Figure 5: SARIMA Model

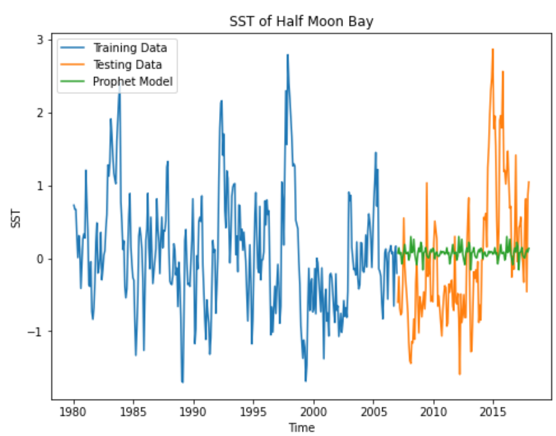

Figure 6: Prophet Model

Model	R^2	MAE	MAPE	MSE
Triple Exponential Smoothing	-0.654	1.0049	316.2973	1.4152
SARIMA	-0.2889	0.8003	186.4918	1.1029
Prophet	-0.0078	0.7561	114.3487	0.8624

Table 1: Model Metrics

As the Prophet model performed best, I utilized Facebook's model to predict SST mean monthly anomalies for ten years into the future passed the dataset, or January 2018 to December 2028. Figure 7 displays the aforementioned predictions.

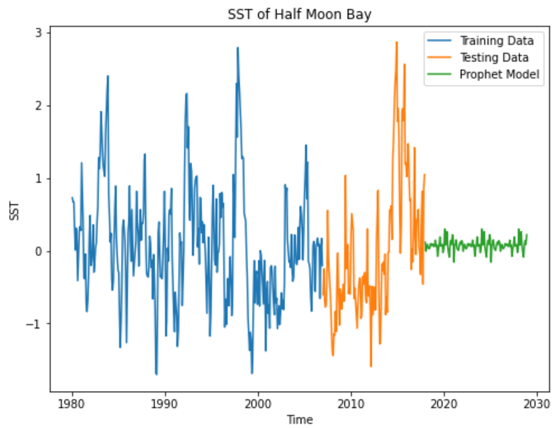

Figure 7: Prophet Model Future Predictions

Similar to the validation phase, the Prophet model outputs very conservative predictions for future SST values. Thus, it lacks the extreme seasonality spikes, which for the purposes of my analysis, are actually quite insightful. To correct for this, next steps would be to include predictable peak dates in the model formulation to simulate holiday spikes. However, the predictions do maintain an average value of slightly greater than zero, indicating an increasing SST mean monthly anomaly from the climatology, thus confirming rising ocean temperatures in Half Moon Bay.

As Half Moon Bay is an infamous coastal city centered around agriculture, seafood, and surfing, rising SST can have severe impacts on the community. Agriculture can be in danger of saltwater intrusion, flooding, and loss of production quality due to the effects of warming SST. Further realities include loss of fish species and fishing markets, as well as changing wave breaks, sizes, shapes, and densities. As Half Moon Bay is home to the Mavericks – a world-renowned winter surface break that boasts some of the biggest waves in the world – alterations to the waves could potentially harm wave quality and the associated touristic attraction of the town. Overall, rising SST can spark a myriad of negative consequences and it is our responsibility to utilize our resources and mitigate/adapt to our new realities as best as possible.