Can we predict sea surface temperatures into the future? What trends are there in sea surface temperatures that are bigger than a year? To answer these questions for the North Pacific Ocean, analysis, filtering, and autoregression models were used. Spectral analysis on 80 years of sea surface temperatures was done first using a discrete fourier transform, which revealed which frequencies most make up the periodic (although noisy) sea surface temperatures. The most prevalent frequencies were determined before a low-pass butterworth filter was applied to reduce the noise of frequencies from periods less than one year. The most prevalent frequencies in the filtered data were determined. Then, many autoregression models were trained on the data using various k/lags. The model with the lowest mse on the testing data was then tasked with predicting sea surface temperatures 20 years into the future.

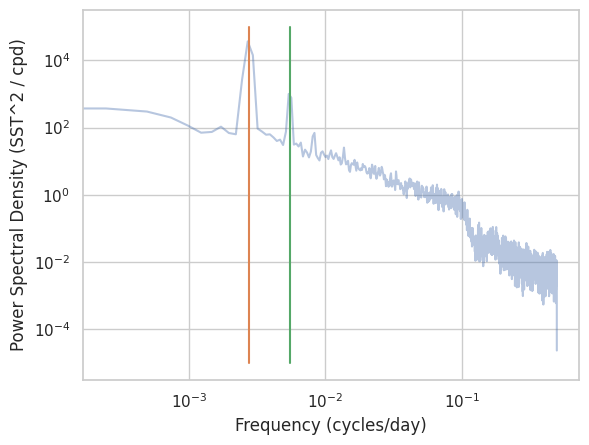


Figure 1

Figure 1 shows the power spectral density of the, which shows which frequencies are most prevalent in the sea surface temperatures, a periodic dataset. The orange line corresponds with the frequency of one over one year. The green line corresponds with the frequency of one over half a year. This means that sea surface temperatures are most periodic about a year, with half a year being the second most prevalent. These are not surprising given how we think of the seasons and the year as cycles.

To investigate trends bigger than a year, a low-pass butterworth was applied with a cutoff frequency of one over one year, meaning noise from smaller periods is reduced. The resulting filtered data was then plotted in a periodogram to investigate which trends bigger than a year were most prevalent.

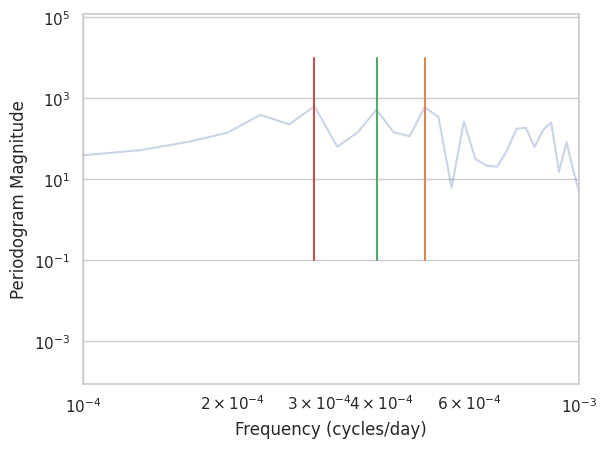


Figure 2

From figure 2, we can see that there are three largest peaks. The most prominent trends then exist at 5.6, 7, and 9.4 years.

For forecasting, many autoregression models were made to attempt to predict the sea surface temperatures 20 years into the future, for which the test data comprised of the last 20 years of the dataset. Each model had a different k value or lag as a hyperparameter. The final model chosen was the one that had the least mean squared error (mse) when used to predict the test data. The best k/lag ended up being just 1, and the model’s mse was 0.03, which is quite small. The model then does a pretty good job of predicting the test data. Figure 3 shows the filtered observational data and the model’s prediction on the test set.

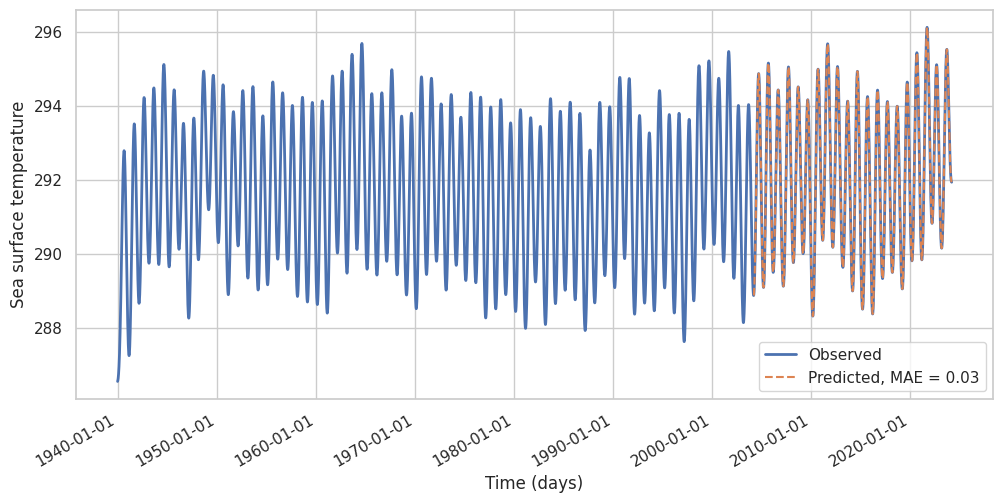


Figure 3

Since the model did a good job predicting the test data, we can have some confidence that it will do well to predict 20 years into the actual future. Figure 4 shows the filtered observational data and the model’s prediction 20 years into the future.

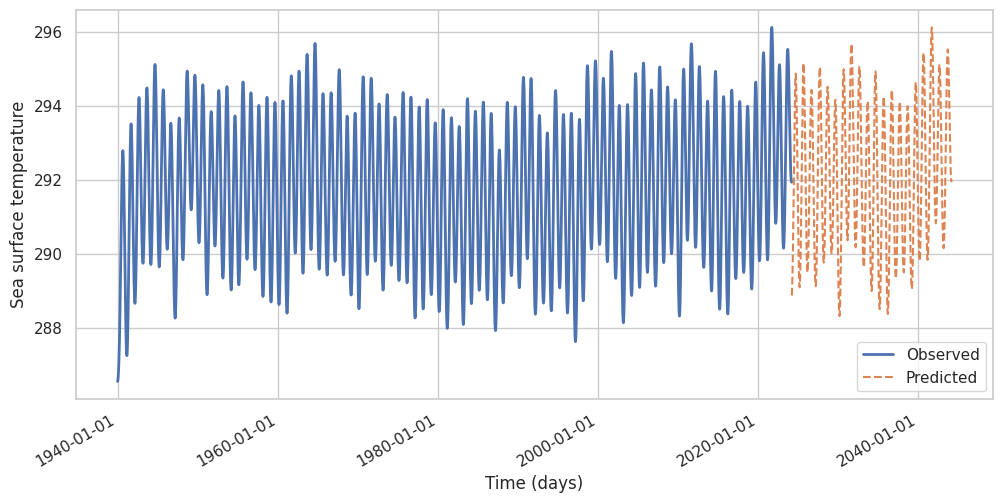


Figure 4

The prediction from Figure 4 doesn't look noticeably bad, and also seems to have slightly higher peaks and troughs than the rest of the data set. This in in-line with what we might expect as a result of climate change. The model appears to be successful in forecasting 20 years into the future.

A year and a half-year are by far the most prevalent frequencies in the North Pacific sea surface temperatures. But by using a butterworth filter to perform lowpass filtering with a cutoff frequency of a 1/year, we were able to identify the prominent trends of 5.6, 7, and 9.4 years with less noise. Additionally, when attempting to predict 20 years into the future, an autoregression model has the lowest mse on the testing data with a k/lag of just 1. It resulted in an mse of 0.03, which is remarkably good. This puts some confidence into the model's predictive power on the actual future, which appears to gently trend upward.