SEA LEVEL FORECASTING

RORY FREEMAN - CAPSTONE PROJECT REPORT

Contents

Abstract

Rising sea levels pose a significant threat to coastal communities, economies, and infrastructure worldwide. As sea levels continue to rise due to climate change, these threats are becoming more urgent and widespread. One of the key impacts of sea level rise is increased flooding, which can damage or destroy homes, businesses, and critical infrastructure, which in turn can disrupt transportation systems and disrupt supply chains. This means that, in addition to the humanitarian and environmental costs, there are also significant economic reasons for limiting sea level rise.

When the Romans first established a small town on the Thames floodplain, little would have been thought about the effects that rising water levels would have on what would become the UK capital city of London. Nearly 2000 years later, and several of London's boroughs, including Aldgate, Southwark, Westminster and Wandsworth, are at severe risk of flooding in the near future (OpenLearn, n.d.).

A huge amount of money has already been spent on reparations after flooding incidents, and on food defences; most famously after a severe storm surge in 1953 that caused approximately £50m in damages (the equivalent of £1.2b in 2013), resulted in the construction of the Thames Barrier – a retractable wall across the breadth of the Thames at Greenwich that could be raised in case of unusually high tides or future storm surges, at a cost of a further £534m (at 2013 prices) (ThamesClippers, n.d.). As well as the barrier, London also utilises a huge network of walls, embankments, pumps and other defences to defend against flooding (Government), all of which need to be operated and maintained. The cost of upgrading or replacing these defences to be able to cope with the increasing sea levels and more frequent extreme weather events that are occurring as a result of climate change is very large. To ensure maximum value for money, they must be designed to cope not just with the problems being faced today, but also so that they can continue to protect London from future problems. In order to plan these designs economically and accurately, planning authorities need accurate predictions of the oncoming threat.

The aim of this project is to use machine learning techniques to forecast future sea level rise, so that planning teams can make accurate designs and budgets when drawing up London's flood defences of the future.

Data Methods

Collection

First, I needed historical data of my target variable, global mean sea level. Second, I needed historical data of contributing variables that have been proven to contribute to sea level rise. This would increase the accuracy and generalization of any models I created. These key contributors are outlined by NASA as global sea ice levels, ice mass change from the Greenland and Antarctic ice sheets, and thermal expansion of sea water (NASA's Goddard Space Flight Center, n.d.).

I sourced my data from the following reputable, scientific sources to ensure accuracy and transparency:

- Global Mean Sea Level – NASA's Goddard Space Flight Center
- Ice Sheet Mass - NASA
- Sea Ice Extent – National Snow and Ice Data Center
- Sea Temperature – Met Office Hadley Centre

Cleaning

Some of the data was already in .csv format, while some came as .txt or .html, so I first converted everything to .csv and cleaned each dataset individually, as each had different requirements. In order to perform appropriate analysis on the data I also had to make sure each dataset had the same sampling frequency and length. To maintain accuracy as much as possible I reduced each dataset down to the smallest length present across all sets, which begins in April 2002 and ends in November 2022. I also decided to use monthly frequency, so that I didn't have to interpolate a large amount of the data, which could risk reducing accuracy and wouldn't necessarily benefit the model.

I also combined the northern and southern hemisphere sea ice datasets, and the Greenland and Antarctic Ice Sheets datasets, as sea level is not influenced by where the ice-melt is coming from, only the overall quantity!

Once the data was clean and resampled, I combined all the datasets in to one table and exported the clean .csv for analysis.

Exploratory Data Analysis

During my EDA I explored several aspects of the data; distribution (including trend and seasonality), multicollinearity and causality.

All variables showed clear trends in line with scientific reports related to climate change; reductions in both sea and land ice deposits, rising sea levels and increasing sea temperature anomalies.

While all the variables showed a seasonal pattern, the ice sheet mass change data seasonality was very weak, suggesting perhaps that while sea ice is able to melt and refreeze each year, any ice lost from the land based ice deposits isn't reformed, and is lost permanently.

A key finding was that, while looking at causality, I found that sea ice extent didn't contribute towards predicting the target variable. As a result, I decided to remove this data from the dataset before moving on to the modelling phase, as it was very unlikely to contribute any useful insights to the models.

Modelling

I began the modelling phase of the project by creating a very simple baseline model, which took the average reading of the 12-month differenced target variable, and simply extended it across the test set of data. Unsurprisingly, this scored fairly poorly, with an MAPE of almost 600%. However, this provided a baseline with which to compare future models (i.e. any score worse than 600% would mean the model is worse than simply guessing the average value every time).

I then introduced a linear regression model, which relies on the information provided by the exogenous variables to predict the outcome of the target variable. The final version of this model gave an MAPE of approx. 7.5%, a significant improvement on the baseline. However, this model didn't take in to account the information contained within the target variable itself (e.g. using yesterday's value to predict today's).

I therefore implemented a SARIMAX model, which would allow me to create a model that factored in the patterns in the target variable data, as well as the exogenous variables provided, to improve the overall accuracy. The final version of this model achieved an MAPE of approx. 2.7%, a huge improvement on the baseline model, and a very good improvement on the linear regression model.

Even without using the exogenous variables, the SARIMAX model achieved an MAPE of 3.4%. This is useful because, in order to use the model that uses exogenous variables to predict future values, projections of the exogenous values would have to be used. This creates a situation where a projection is based on a projection, and this reduces the reliability of the model significantly. Where this model would be useful is if governing bodies were to set targets for the exogenous variables, and therefore could predict what level the target variable would achieve if those targets were met.

Key Insights

The primary insight to be taken away from this study is that the sea level is rising. Regardless of what people believe the causes are, sea level is rising and the economic risk is rising with it. Long-term strategies must be put in place now to avoid damage in the future.

The damage being done to the main cause of rising sea levels, the melting of the Greenland and Antarctic ice sheets, is not slowing down, nor is it reversing between seasons. People often think that the ice at the earth's poles melts in the summer and refreezes in the winter, and while this is true (to an extent) for sea ice, this is not the case for land based ice deposits. This means that the issues we are experiencing now are the new normal, and any infrastructure needs to be built to adapt to that new normal. Unfortunately, things can only get worse from here, which is why it is so imperative that we put appropriate mitigation strategies in place now.

Bibliography

(n.d.). Retrieved from NASA's Goddard Space Flight Center: https://climate.nasa.gov/vital-signs/sea-level/#:~:text=Sea%20level%20rise%20is%20caused,1993%2C%20as%20observed%20by%20satellites.

Government, U. (n.d.). Retrieved from https://www.gov.uk/government/publications/thames-estuary-2100-te2100/thames-estuary-2100-te2100#:~:text=estuary%2D2100%2Dte2100-,What%20is%20the%20Thames%20Estuary%202100%20Plan%3F,biggest%20challenge%20of%20our%20time.

OpenLearn. (n.d.). Retrieved from https://openlearn.medium.com/will-london-be-able-to-cope-with-rising-sea-levels-da016d43aa6a

ThamesClippers. (n.d.). Retrieved from https://www.thamesclippers.com/blog/thames-barrier