Sea Level Rise Prediction Model

Mira Chadha, Kirti Magam, Jake Phelan, Douglas Read

abstract

Sea levels have been gradually rising for many years, and pose a universal concern around the world. The impact of this is felt most strongly in coastal cities such as Boston. This project aims to improve our understanding of how changing factors within the climate can have a long term effect on sea levels, by analyzing sea level data from in and around the Boston area. We aggregated data from various National Oceanic and Atmospheric Administration (NOAA) sources, and conducted multiple tests using the KNN, Random Forest, and Support Vector Machine algorithms to find that a hypertuned Random Forest Model produced the highest accuracy of 99% and MSE of 3.76. We hope to contribute the knowledge that we have gained from this project to the body of research that exists in this area.

introduction

Sea levels have been gradually rising since the late 1800s (Nunez). Greenhouse gasses, melting ice caps, and rising global temperatures have continued to compound their effect upon Earth's oceans, to a degree where such changes now constitute unprecedented yearly rises in sea levels. These sea-level changes pose risks to coastal wildlife, natural ecosystems, as well as coastal cities (Voosen). The effects of such rising sea levels can be found in everything from increased pollution to extreme weather events; in fact, one study found that "between 1963 and 2012, nearly half of all deaths from Atlantic hurricanes were caused by storm surges" (Nunez). Although many coastal cities and conservation experts are planning adaptation measures for long-term sea level rise, it is necessary to predict what future dangers such climate change will bring, as well as evaluate how factors like greenhouse gas emissions can alter these outcomes. There is still work to be done in forecasting how quickly and to what extent the oceans will rise.

Changes in sea levels are a universal concern around the world. While this now-dangerous seawater has always been on the earth, it had been safely frozen in glaciers around Antarctica and Greenland up until the most recent century or so. In fact, the Greenland ice cap's yearly loss has increased from 120 to 300 billion metric tons of ice per year since the 20th century (Lindsey). This rapid increase in glacial melt rates is a result of increased greenhouse gas emissions by humans. These gasses build up in the atmosphere and trap heat, causing the planet's temperature to rise even in regions that have previously always been subject to sub-freezing temperatures. With reduced ice masses at the poles of the earth, the average temperature will only continue to rise as deeper and darker oceans attract more solar radiation and hold a higher specific heat capacity than their previous depths (Lindsey). Some very old buildups of ice which were never meant to melt, called permafrost, have even begun melting and releasing their own stores of greenhouse gasses into the atmosphere, intensifying the vicious cycle of global warming and rising sea levels (University Corporation for Atmospheric Research). The majority of human civilizations were not built to withstand the continually rising sea levels they face today. In fact, about 30% of the US population lives in coastal areas which are highly-likely to suffer in the event of a dramatically-risen sea level (Lindsey). The first step towards combating a problem that harms the public is raising awareness of both the issue itself and its causes, and we hope to do so with this project.

goals & objectives

This project aims to better understand how changing factors within the climate can have a long term effect on sea levels. Specifically, we intend to predict the increasing changes in overall sea levels over a course of a hundred years. Important factors such as historical weather patterns, temperature changes, and previous data concerning sea levels will be examined in order to form predictions for the distant future. This project also intends to account for the effects of changing sea levels by predicting which coastal cities are most likely to face increased risk over time. We can determine how at risk these cities are in comparison to each other to establish which ones will be severely affected by predicted changes in sea levels.

Our main objective is to use this specific climate data available in order to construct a predictive model. This model will be able to analyze past data to show how weather and climate changes have an effect on sea levels. Smaller objectives necessary to achieve our long term goal include collecting the data that will be used in the project, determining what type of predictive model would best fit the data, and adjusting our model in order to make it accurate and precise. We plan to use various National Oceanic and Atmospheric Administration (NOAA) datasets in order to accomplish this.

related work & additional resources

One particular study done in the past related to our project used machine learning techniques to predict the rise in coastal sea levels for certain regions. In the paper "Predicting regional coastal sea level changes with machine learning", Nieves, Radin, and Camps-Valls detail the process to building their machine learning model. In their work, they highlighted the importance of this topic because the changes of climate and weather conditions affect sea levels in different variations based on regional differences. In particular, they wanted to see the influence of the open ocean on the coast. The model they implemented makes near future predictions useful for current decision making (1-3 years) rather than the 30 years we intend to make predictions for. To complete the machine learning techniques necessary, they used a Gaussian process model and recurrent neural networks.

In another study, they focused on the sea level changes on a smaller scale by studying the upper part of an estuary in France. They used machine learning techniques to build multiple regression models as well as using neural networks. When considering inputs, they used both large and small scales to account for the specific location and environment.

To complete the different parts of this project, we will need to utilize multiple libraries. In order to perform the initial exploratory data analysis on our data sources, we will need to access the Pandas library to work with our sources in the data frame format. We may potentially need to use the Numpy library and the scikit-learn library to perform machine learning algorithms for our model that predicts sea levels.

data acquisition & preparation

Data Source I: NOAA Climate Data Online: Sea Level Rise
https://www.star.nesdis.noaa.gov/socd/lsa/SeaLevelRise/

This dataset contains recordings of sea level changes, globally, recorded about 20 times per year. Data is recorded from 4 different satellite missions, with newer missions replacing older ones with time. Some years have only one reading recorded, while some intermediate years have recordings from more than one satellite, as new missions were launched before the older one had completely ended. In order to compensate for this, a column of the average of the readings was added to the .csv file's dataframe during our EDA. This average will be the value considered while training our model. Initial EDA of this dataset, visualized as a preview through a scatter plot relating sea level change to year, displayed a Pearson correlation coefficient of 0.96, so the relationship between sea level and year is very strongly positive.

Data Source II: NOAA Local Climatological Data
https://www.ncei.noaa.gov/cdo-web/datatools/lcd

This dataset contains local climatological data collected in the Boston area. The data is recorded by hour, day, and month, and includes average dew point temperature, average dry bulb temperature, average humidity, average sea level pressure, and precipitation, to name a few. We compiled 30 years of this data for our analysis, and decided that for the purposes of our analysis it made the most sense to use only the monthly and daily data rather than the hourly data. As such, much of our EDA for this dataset included removing all the "empty" rows that contained only hourly data. We also used the methods taught in lectures to clean the data in cases where there was limited daily or monthly data recorded, although these were few and far between. The information gleaned from this dataset will be especially useful in comparison to Data Source I, where it will provide additional context to the trends seen in the sea level rise.

Data Source III: NOAA Global Monitoring Lab: Trends in CO2
https://gml.noaa.gov/ccgg/trends/gl_data.html

This dataset is a map of year and month to the measured atmospheric CO2 in parts per million. There was minimal data preparation that was needed for this source as there were no missing values in the set. Initial analysis showed a clear positive relationship between atmospheric carbon dioxide and time as was expected.

model selection

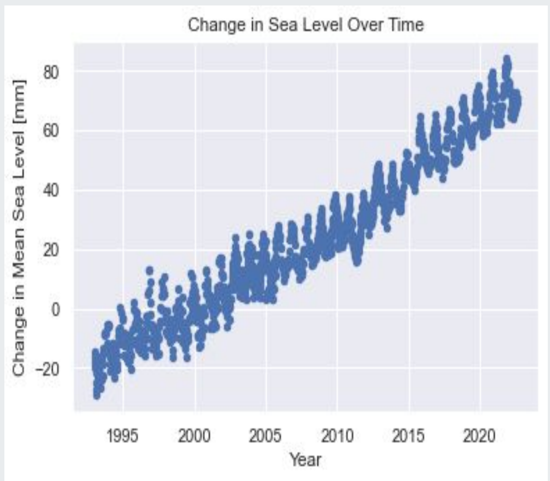
For our model, we used the KNN, Random Forest, and Support Vector Machine algorithms for regression. Some of the variables in our dataset contained categorical data, and so these values had to be encoded before they could be used to train our models. The numerical data did not need to be encoded, but it did have to be normalized at the same step that categorical values are encoded.

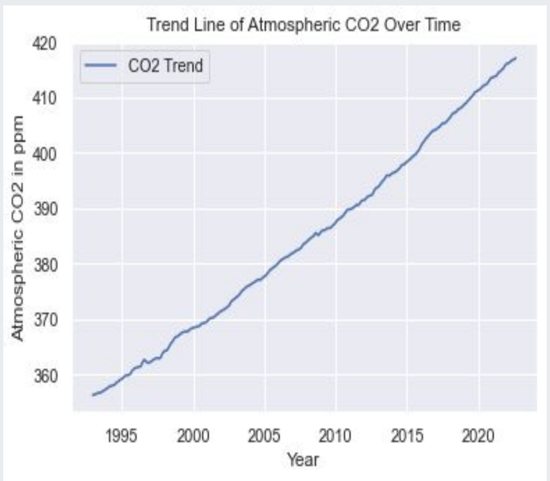
results & evaluation

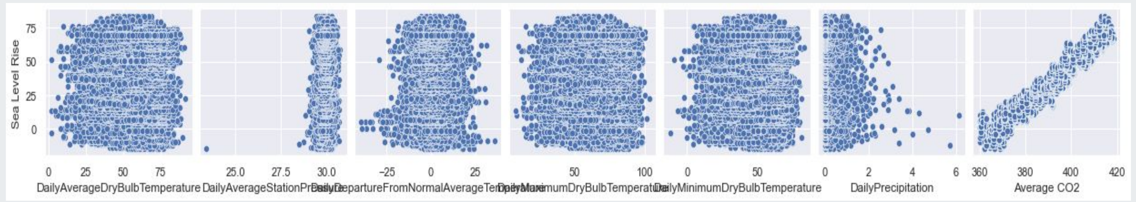
The analysis that we conducted on data around sea level rise confirmed existing conclusions from prior research: namely, that there has been a clear rise in sea levels over time.

Additionally, the data that we collected regarding CO2 tells a similar story: there is a clear positive relationship between atmospheric CO2 and time. Although there are fluctuations in the data on a granular, month-based level, there is a linear and positive trend overall.

We performed additional analysis upon multiple other features, including average bulb temperature, station pressure, and precipitation. While the data around some of these features contained considerable noise, there did appear to be indications that precipitation and temperature have effects upon sea level rise. As such, we proceeded to consider average CO2 levels, average dry bulb temperatures, departures from normal average temperature, and precipitation data in addition to the aforementioned data as we proceeded in our data analysis.







model predictions

For our model, we used the KNN, Random Forest, and Support Vector Machine algorithms for regression. All three algorithms were performed on the data using models that were optimized through hyperparameter tuning using the grid search method. .

KNN Regression: The best nearest neighbors value was k = 24 after tuning. With this parameter set on the model, the accuracy was 92.21%: the second highest of the algorithms. Overall, the accuracy was high, but the MSE was also high as well at 46.98 meaning that the model did not fit the data as best as it could.

Random Forest: The best n_estimators value was 200 after tuning. This model produced an accuracy of 99.31%, the highest of all the algorithms and a significantly lower MSE of 3.73. This signifies that the model fit the data relatively well compared to the other models.

SVM Regression: The best gamma value was 1 and the best C value was 9 after tuning. This model had an accuracy of 91.42% which was the lowest of all of the algorithms and it had the highest MSE of 52.49, meaning that the model did not fit the data as well as the other models did.

Table 1: Accuracy & MSE for hyperparameter tuned models

Model	Accuracy	Mean Squared Error
K-Nearest Neighbors Regression	92.21%	46.98
Random Forest Regression	99.31%	3.73
Support Vector Machine Regression	91.42%	52.49

Table 2: Random Forest Regression Hyperparameter with N_estimators

n_estimators	mean train score	std train score	mean test score	std test score
200	0.999025	0.000012	0.993147	0.000142
400	0.999031	0.000018	0.993115	0.000148
50	0.998933	0.000038	0.992915	0.000296
100	0.998983	0.000020	0.992909	0.000282
25	0.998786	0.000045	0.992731	0.000264

The random forest regression model proved to fit the data the best based on both accuracy and mean squared error values in order to make the best predictions. When further analyzing this model during hyperparameter tuning, it became clear that the 200 n_estimators results in the best scores. Table 2 shows the mean training score as 99.9 and the mean test score as 99.3, with a small difference between the two. This shows that the model does not appear to overfit or underfit the data.

the impacts

The unprecedented and accelerating rise in sea levels poses an extraordinary risk to coastal communities. Unfortunately, prior research focused on the Boston area has revealed that the city is at a particular risk to sea level rise; with much of the city built on landfill over old marsh land, Boston already "experiences some of the worst high-tide flooding in the nation" (Noor, 2022). This could translate into the displacement of thousands of people in the Boston area alone in the relatively near-future. In addition to this, the economic impacts of sea-level rise are far-reaching. To avoid such drastic displacement, billions of dollars of investment would need to be poured into updating infrastructure in the Boston area alone (Carlock, 2022). These costly updates would represent only a small fraction of the infrastructure updates needed around the country and natural disasters that have been increasingly costly due to sea-level rise.

Our project was motivated by these factors, as well as by a goal of reducing the negative impact of sea level rise on our local community. This project has furthered our understanding of the factors that drive sea level rise with more specificity, and contributes to an increasing body of research on understanding sea-level rise.

conclusion

In this project, we were able to thoroughly investigate the effects of climatology and weather data on the rise in sea level through the years. We researched important factors that can explain changes in the overall climate as well as weather data that could capture changes within smaller time frames. The selected features we included in our study to help predict sea levels over time were CO_2 levels, average dry bulb temperatures, departures from normal temperature, and precipitation levels.

By studying the relationships between multiple variables associated with rising sea levels, we hoped to contribute further evidence and analysis of this growing issue regarding climate change and our environment. When conducting our project, the algorithm that produced the best results was random forest regression. After hyperparameter tuning this model so that n_estimators would be set to 200, the accuracy was 99.31%. This emphasizes that the model was able to correctly predict sea levels over time based on how it was trained for the majority of the data. Beyond that, the MSE was 3.73, showing that the model led to small errors when estimating and it proved to be a good fit to the data.

In the future, we can explore more features connected to sea level changes to see if we are missing some important features that should be considered. We can also explore further algorithms for comparison to ensure that we have found the model with the best fit.

references

1. National Centers for Environmental Information (NCEI). (n.d.). Data tools: Local climatological data (LCD). Local Climatological Data (LCD) | Data Tools | Climate Data Online (CDO) | National Climatic Data Center (NCDC). Retrieved December 6, 2022, from https://www.ncei.noaa.gov/cdo-web/datatools/lcd
2. NOAA. (n.d.). Center for Satellite Applications and research - NOAA / NESDIS / star. NOAA / NESDIS / STAR website. Retrieved December 6, 2022, from https://www.star.nesdis.noaa.gov/socd/lsa/SeaLevelRise/
3. University Corporation for Atmospheric Research. "Center for Science Education." Climate and Ice | Center for Science Education, 16 Oct. 2022, https://scied.ucar.edu/learning-zone/climate-change-impacts/climate-and-ice#:~:text=Melting%20ice%20causes%20more%20warming.&text=As%20global%20warming%20causes%20more,the%20heat%20to%20the%20atmosphere.
4. US Department of Commerce. (2005, October 1). Global Monitoring Laboratory - Carbon Cycle Greenhouse Gases. GML. Retrieved December 6, 2022, from https://gml.noaa.gov/ccgg/trends/
5. Carlock, Catherine. "Boston's Sea-Level Rise Will Require a 'Big Dig Level' Solution, City Leaders Say - TheBostonGlobe." *BostonGlobe.com*,TheBostonGlobe,24Sept.2022, https://www.bostonglobe.com/2022/09/23/business/big-dig-level-approach-billions-dollars-are-needed-address-sea-level-rise-city-leaders-say/.
6. Guillou, N., & Chapalain, G. (2021, July 21). Machine learning methods applied to sea level Predictions in the upper part of a tidal estuary. Oceanologia. Retrieved October 30, 2022, from https://www.sciencedirect.com/science/article/pii/S0078323421000634
7. Lindsey, Rebecca. "Climate Change: Global Sea Level." NOAA Climate.gov, 19 Apr. 2022, https://www.climate.gov/news-features/understanding-climate/climate-change-global-sea-level.
8. Nieves, V., Radin, C., & Camps-Valls, G. (2021, April 7). Predicting regional coastal sea level changes with machine learning. Nature News. Retrieved October 30, 2022, from https://www.nature.com/articles/s41598-021-87460-z
9. Noor, Dharna. "Here's What Rising Sea Levels Mean for Boston - The Boston Globe." BostonGlobe.com, The Boston Globe, 22 Feb .2022, https://www.bostonglobe.com/2022/02/16/science/heres-what-rising-sea-levels-mean-boston/.
10. Nunez, Christina, and National Geographic Staff. "Sea Level Rise, Facts and Information." Environment, National Geographic, 16 Feb. 2022, https://www.nationalgeographic.com/environment/article/sea-level-rise-1.
11. Voosen, Paul. "Seas Are Rising Faster than Ever." Science, 18 Nov. 2020, https://www.science.org/content/article/seas-are-rising-faster-ever.