Predicting Sea Levels Using the Actuaries Climate Index

Jennifer Mince, Karl Maier, Wesley Merrick,
Katie Vervack, Toby Duncan, Katie Kessler

April 28th, 2017

Abstract

The main goal of this project was to use the Actuaries Climate Index data set to answer our central question: how do the seasonal components of high temperature, low temperature, heavy rain, and drought affect the seasonal sea levels in North American regions, and by how much? Rising sea level is a threatening force that greatly impacts local and global environments. Many populations live in an area that would be significantly damaged by impending flooding. Destructive storms, coastal flooding, shoreline erosion, and destroyed ecosystems, industries, and infrastructure are all results of rising sea levels. Roads, bridges, power plants, and other important components of human life are not currently equipped to handle this threat.

Using this data, our group could help this societal need by introducing a model for the rate at which the sea levels are increasing, and calculating factors that contribute to the increase in levels. The data set has an Actuaries Climate Index (ACI) created with the components: frequency of temperatures above the 90th percentile (T90), frequency of temperatures below the 10th percentile (T10), maximum rainfall per month in five consecutive days (Rx5), annual maximum consecutive dry days (CDD), frequency of wind speed above the 90th percentile (WP), and sea level changes (SL). The index is an educational tool designed to help inform actuaries, public policymakers, and the public on changes in these measures over recent decades. It is an objective measure of changes in extreme weather and changes in sea level relative to the base period of 1961 through 1990. Using the Anaconda interpreter and Python 3, we created code in linear regression and random forest regression models of each component vs. sea level for each region per season. We found some trends for certain seasons or regions, but no trends over components or regions as a whole. Hopefully, our data analysis can be used to further predict sea-level rise if the Actuaries Climate Index is updated with more years and features to create larger data sets to make it easier to make predictive conclusions.

Introduction

Our group would like to use the Actuaries Climate Index data set to answer the central question: how do the seasonal components of high temperatures, low temperatures, heavy rain, and drought affect the seasonal sea levels in North American regions and by how much?

The research we want to conduct is both important and relevant to the societal need of knowing what causes sea levels to rise and predicting future measurements to be better prepared for future problems. Higher sea levels have a significant impact on both local and global environments. Rising levels can cause destructive storms to move closer inland and as a result, there is more coastal flooding. "In the United States, almost 40 percent of the population lives in relatively high-population-density coastal areas, where sea level plays a role in flooding, shoreline erosion, and hazards from storms" (NOAA). Rising seas threaten the infrastructure used by industries and businesses, such as roads, bridges, power plants, and oil wells. The coastline of the United States is heavily populated and the risk of rising sea levels causes many problems. "Approximately 25 million people live in an area vulnerable to coastal flooding" (EPA). The U.S. economy, transportation, and many ecosystems are threatened by the impending flooding that comes with the rising levels. Integral coastal activities including marine transportation of goods, resource extraction, and tourism help to generate "58% of the national gross domestic product" (EPA). Transportation in the United States is substantially affected by coastal flooding, especially roads. In low-lying communities, the streets are first to be flooded because they are lower than the surrounding land (Titus). The current drainage systems for these roads are not efficient enough to handle increased frequency of flooding due to rising sea levels.

Using this data, our group could help this societal need by introducing a model for the rate at which the sea levels are increasing, and calculating factors that contribute to the increase in levels. The data set has an Actuaries Climate Index (ACI) created with the components: frequency of temperatures above the 90th percentile (T90), frequency of temperatures below the 10th percentile (T10), maximum rainfall per month in five consecutive days (Rx5), annual maximum consecutive dry days (CDD), frequency of wind speed above the 90th percentile (W), and sea level changes (SL). It is an objective measure of changes in extreme weather and changes in sea level relative to the base period of 1961 through 1990. This index is an educational tool designed to help inform actuaries, public policymakers, and the general public on changes in these measures over recent decades (Actuaries Climate Index). Our group will create our own ACI using the seasonal values for T90, T10, Rx5, and CDD to see how they affect the sea level changes.

Existing Conclusions

After researching the various data science studies that were conducted using the Actuaries Climate Index data set, only two other studies were found. One focused on using the data set to build a risk index for areas where economic losses from disasters (i.e.. floods, droughts, etc) would be higher than other areas. They gathered the 90th percentile from the 1961-1990 data period. The study used linear regression on the data sets of SHELDUS data set which consists of economic losses, mortality, and morbidity in the US as well as the Canadian Disaster Database to get the same data for the Canadian Region. The disaster and death data was then compared with various data sets in the Actuaries Climate Index like heat, 5-day precipitation, etc. These data sets were then put in historical impact indexs that ranged from 1 -10. For regions where there were no relationships proxies were used instead. All these indexes were put together to create the ACRI(Actuaries Climate Risk Index). After creating this risk index the a linear regression analysis was performed for all regions, which were then evaluated by hazard and region. The data science research has lead to the possibility of risk models being built in regards to different regions. They also can build models that assess Indirect Supply Chain risks as well as tools to quantify risk with money. This index was then going to be used by insurance companies to help set rates for the regions.

The other was focused on expanding the data set to account for the UK and Europe. Besides these two studies, however, there wasn't any other relevant research that used the ACI data set. Because of this, the search was expanded to research that used just some of the individual data sets that together composed the ACI. Even in these further searches, there were not really any relevant data science studies, but rather lots of papers that simply analyzed the data instead. Overall, after all this research into other data science using the ACI data set, it is safe to conclude that our particular data science study will be answering a question that has not been answered before in another study.

Cleaning the Data

The Actuaries Climate Index data set needed to be cleaned up for our team to effectively and efficiently utilize it for our model. In order to clean up the data, we first needed to understand which variables we wanted to keep as well as determine what each of the variables meant. We decided to keep the temperature above the 90th percentile, below the 10th percentile, consecutive dry days, rainfall maximum for 5 days, and sea level. Our group thought that these variables were crucial to change in sea level. Since the Central Arctic (CAR) and Midwest (MID) regions had no sea level data so we excluded it from the final clean data set. The CAR region consisted of the Northern Territories and Nunavut. The MID territory consisted of Iowa, Illinois, Indiana, Minnesota, Michigan, Missouri, Ohio, Wisconsin. We are trying to predict how sea level is affected and it would not be possible to predict without the data for sea level.

The group excluded the monthly variable sheets in addition to the two above regions because we wanted to focus on the seasonal components. The seasonal components are divided into 4 seasons: Winter (December, January, February), Spring (March, April, May),

Summer (June, July, August), and Fall (September, October, November). Our group decided to use the seasonal data rather than the monthly data because the amount of data there was to process. The data will show a bigger change when split into 4 seasons rather than smaller changes throughout 12 month periods. Now that we had all the data, we needed to add it all into one sheet on excel. Having all the data on one sheet in excel would allow the code to easily import one comma-separated file. When the coding actually started, the one file was more difficult. The one comma-separated file was kept as a master list which was then broken into multiple sheets based on each individual variable. To make the data look cleaner, the original variable names were simplified. For example, the initial variable of ACI_sealevel_seasonal_ALA.csv was then converted to SL ALA. Initially, the datasets columns corresponded to the years while the rows corresponded to the actual variables. The columns and rows were then altered so that the years corresponded to the rows and the variables were in the columns. Excel has a function built in called Transpose, which allowed an easy flip of the rows and columns. The clean data then allowed us to implement each comma-separated value files in our code.

After cleaning up the data, we needed to adjust the ACI average to account for discarding the wind power and sea level data. We left the sea level data out of our ACI average because we want to use this average to help determine sea level. In the data set, each year is broken into 4 seasons. The previous ACI average was computed by taking the average of T90, T10, Rx5, CDD, W, and SL for each of those seasons by region. They did this using standardized anomalies. That is, an anomaly divided by a standard deviation. "The standardized anomaly corresponds to how unusual that season's value is, compared to the reference period mean and standard deviation for that season" (Actuaries Climate Index). They used the mean (T90 - T10 + Rx5 + CDD + W + SL) to calculate the ACI average. The reason for negating the T10 variable is to make the standard deviation 1 for this component, matching all other variables. We used this formula for the ACI average but removed wind power (W) and sea level (SL). This makes our final formula ACI = (T90 - T10 + Rx5 + CDD)/4.

Data Analysis

We soon found out that choosing a machine learning algorithm was more difficult than we thought because there are may different algorithms that we could use. We determined that the k-nearest neighbors regression algorithm would not work with our data. "K-nearest neighbors predicts the class of a data point based on the majority vote k-nearest neighbors" (Ismail). The closer the data points are to each other, the better the prediction value. Since the data points were spread out, we would get a poor prediction value if we used k-nearest neighbors as our predictive model. We then used linear regression and random forest regression on our data to see if there were any positive r-squared, or r^2 values.

Linear regression lets us analyze relationships between two quantitative variables. This involves finding the best-fitting straight line through the data points. This line is called a regression line (Lane). From this, we can determine the r^2 value. R^2 determines how close the data are to the regression line. It is always between 0 and 100%. Generally, the larger the r^2 value, the better the model fits the data (Frost). Random forest regression also uses r^2. Random forests are a "combination of tree predictors where each tree depends on the

values of a random vector..." (Walker). This random vector is obtained separately with the same dispersal for all trees in the forest. Basically, a group of "weak learners" will move together to create a "strong learner".

To begin data analysis on the data set, our group plotted the sea level, which is what we wanted to predict, versus every other variable in our data set. This game us a visual representation of our data, allowing us to choose which machine learning algorithms we wanted to utilize from these plotted scatter graphs. After examination of the data plots, the conclusion was to use linear regression since there were some noticeable trends in the data that looked as though a linear approximation could be made. The code for the plotting is identical to the linear regression plot and will be explained shortly.

When implementing the linear regression on our data, we first needed to read in our data files. Using the pandas library in Python, we read in each of our comma separated value sheets. Pandas is popular with big data implementation because it pulls data into a data frame object instead of pulling data individually by columns and adding them into an array or list. In addition, pandas allowed us to extract data from individual columns for each region using only a single read call on the file. The data frame implementation also allowed our group to easily remove columns, rows, and headers we did not want in our data, such as columns containing season labels. Once the data was imported, we had a set of nested loops that would first iterate over each of the four seasons and then over each region for that season. The linear regression code is then run when the innermost loop collects the data from each subsequent region.

Before doing any of this, however, we first needed to remove the the header columns from our data. In order to do this, we created a new list variable that took the old pandas data frames of sea level and the other variable we used against sea level. Using the pandas function call toList(), the new list variables contained every i+2 column of the data frame as a python list rather than the pandas data frame we initially put the data into. The i columns contained the headers so we added two, making it i+2 to start at the actual data excluding the first two columns which held the year and season labels. Once we had the data for a specific region, we used a list comprehension to assign the data from a single season to that list, effectively flattening the time element of the seasonal data. This was then split into training and testing sets, which were used to fit and test the accuracy of our regression models.

The new arrays of every season were then reshaped using numpy to create array like structures suitable for use in the sklearn linear regression fit() and score() functions. Next, the reshaped data was split into training sets and test sets. The split we used was 80:20. This means 80% of the data goes to the training set and the other 20% goes to the test set. In order to do this we calculated the number of years multiplied by 0.8 (54 years x 0.8) which then gave us the 80% (approximately 43 rows of data) mark. We sliced our list at the 43rd index to separate into our testing and training sets. These slices were done for both of our sea level and variable lists, with the indices corresponding to the year the data was taken from.

Once our data was split into test and training sets, we were able to run the sklearn linear regression algorithm and implement the regression fit for our training sets. From this regression, we calculated the r-squared value of the linear line we had just produced. By doing the r-squared value, it allowed us to understand how well our model is predicted which

will be discussed later in this paper. Sklearns linear regression has a function score() that returns the coefficient of determination r^2 of the prediction. In addition to the r-squared value, a graph was constructed for each region of every season for every variable. The graph gave us another look at how the linear regression line looked over top of the data plots we made earlier.

Plotting the data with the regression line was fairly simple. The library we used to plot was matplotlib. The pyplot package in matplotlib allowed us to format and label out plots so as to distinguish them from one another.

Random Forest Regression is a machine learning algorithm that came out of Bell Labs in the 90s and belongs with ensemble methods. The ensemble learning algorithm generates several different models that take the data and make predictions on it independently of each other. Then once the predictions are finished all the data is brought together and formed into one central prediction. The reason this type of regression is labeled ensemble is due to the fact that it uses an ensemble of decision trees to make its predictions.

These decision trees are then where the random part of the regression comes in, as it produces a random forest of decision trees to be used for prediction. Most of these trees won't be of any use as they are irrelevant to the data, but that is why there is aboral voting. Basically, all the models return a prediction with a value and label linked with it, and at the end of the algorithm all these different models are voted on, and the ones with the most votes are used disregarding the majority, which are not usefully to the problem. The benefit to doing this is that the decision trees that survive the voting process have the better variables, which theoretically means that there will be more relevant results. Along with all these positives, Random Forest Regression does have the tendency to overfit graphs which can result in the data not being as relevant as it typically could be.

The goal of using the Random Forest Regression was to create a better line of best fit. When we did linear regression, we ran into problems with not getting as much correlation between variables as we hoped, which limited are data that we could look over. For this reason, we turned to Random Forest Regression to hopefully solve this problem and get us more data. We setup up the code in Python, trained the sets and ran them. At first it looked like we were getting good data, but soon we realized the one drawback to using random forest was that it builds different random forests each time. The high level of randomness created lots of noise in the data, as one positive correlation on one run could be a very negative correlation on another run.

After realizing this drawback, we reached the conclusion to move away from the Random Forest Regression. We debated the value of the data that it was providing as a good indication that there could be correlation between different graphs, but felt that with all the noise the data was not consistent enough to be used in the project. The data was all over the place and the values we were getting typically were not very high and were mostly negative. This is not to say that Random Forest Regression should be ruled out for future study with this data set, as it has the potential to be particularly useful if the data is aligned well and the other regression algorithms are not providing good results, however, for the sections of the data sets we have been using we opted to use linear regression instead.

Another regression that was considered was Support Vector Regression. Support Vector Regression is an extension of the Support Vector Classification that uses using a subset of the training data in the classifier because the cost function ignores any training data close

to model prediction. We did experiment using it but due to it not offering a comprehensive advantage over linear regression it was not used in the project.

We found slight success when using the method of linear regression. We graphed first each of the following variables (cdd, rx5, t10, t90, and average) versus sea level for each region and season and then ran a linear regression method on each of these graphs. Positive r^2 values were found for some of these graphs but not all of them. We will discuss the largest r^2 values for 'CDD vs SL', 'Rx5 vs SL', 'T10 vs SL', 'T90 vs SL' and 'AVG vs SL'. The largest r^2 value for 'CDD vs SL' was found in winter of region USA with an r^2 value of 0.085 (Figure: 1a). The largest r^2 value for 'RX5 vs SL' was found in spring of region CWP with an r^2 value of 0.602 (Figure: 1b). The largest r^2 value for 'T10 vs SL' was found in winter of region SWP with an r^2 value of 0.19 (Figure: 1c). The largest r^2 value for 'T90 vs SL' was found in winter of region CWP with an r^2 value of 0.081 (Figure: 1d). Lastly, the largest r^2 value for 'AVG vs SL' was found in winter of region CEA an r^2 value of 0.086 (Figure: 1e)

Conclusion

We found small, very specific correlations in the data between certain features and regions that contributed to the rise in sea level.

"Seasonal variations in sea level can be caused by seasonal changes in temperature, seasonal changes in river runoff, or seasonal changes in wind" (Parker). We used seasonal consecutive dry days, rainfall, high temperatures, and low temperatures data to see how much the sea level varied be season. There are some difficulties in predicting future sea level rise. It is important to have data on interannual-to-decadal sea level variation (Parker). The phrase sea level or mean sea level (MSL) is used to describe the water level observations that have been averaged over some period, usually at least a month, so that the shorter period variations have been averaged out and can be viewed as oscillating about this mean sea level. "Sea level is only a mean for a particular time period, and it varies over longer time periods, on a monthly, interannual, and longer basis. This variation in sea level is measured relative to the land. If the land sinks, it will appear that sea level is rising, and likewise if the land rises it will look like sea level is falling. Thus we refer to this as relative sea level" (Parker). These factors may be affecting the data we used through the Actuaries Climate Index that we did not take into account or did not have sufficient information on to correctly predict the changes in sea levels over seasons.

Our group understands that this research is important for the environment and those that may be affected. "Sea level change is a topic of interest to a broad audience for the simple reason that it can influence people directly: around 200 million people live in coastal floodplains" (Milne). The significance and consequence of this topic motivated us to choose the Actuaries Climate Index data set. "it is important to analyze the contribution of these factors to estimate future changes in sea level, intensive studies have been performed. Comparison of the observed sea level rise and its estimation is called the sea level budget. Reducing the gap between these values has been a long challenge in the climate science" (MIMURA). We tried to take on a very important and challenging topic and did not get an exact answer to our overall question.

Although we may not have received the best results, we were able to make some conclu-
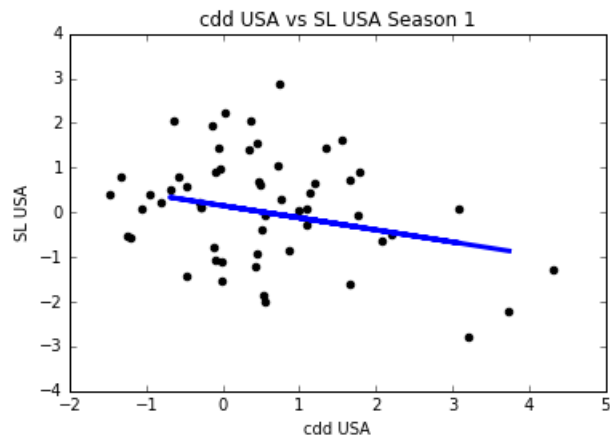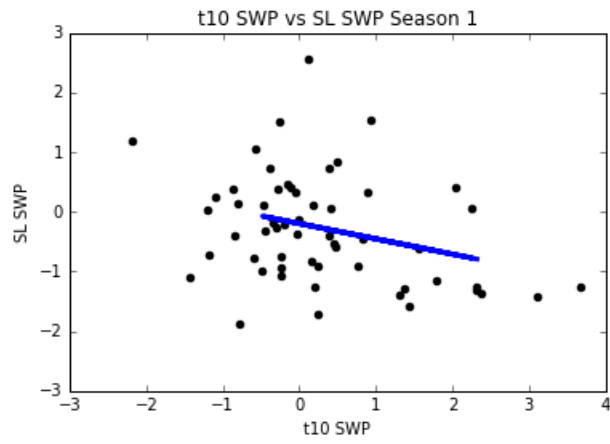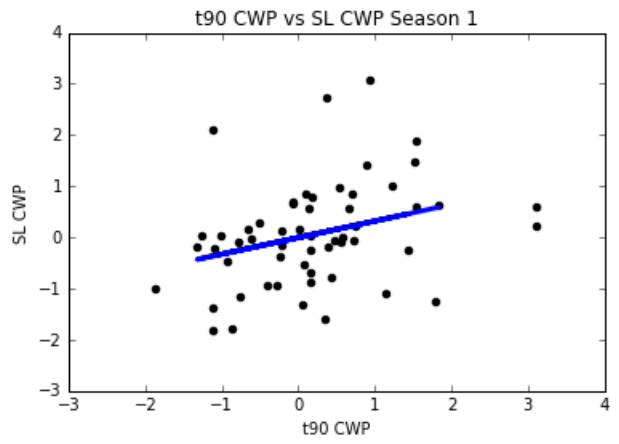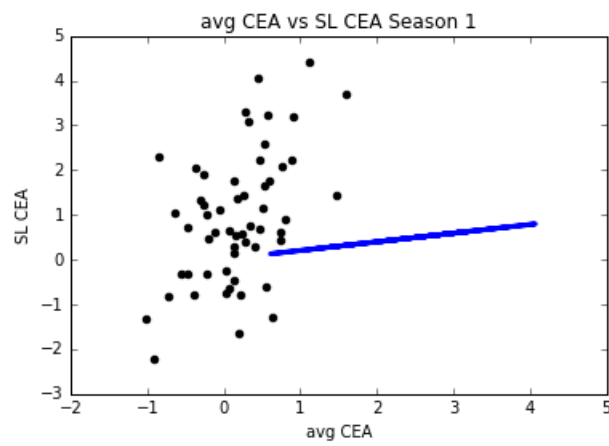
(a) $r^2 = 0.085$
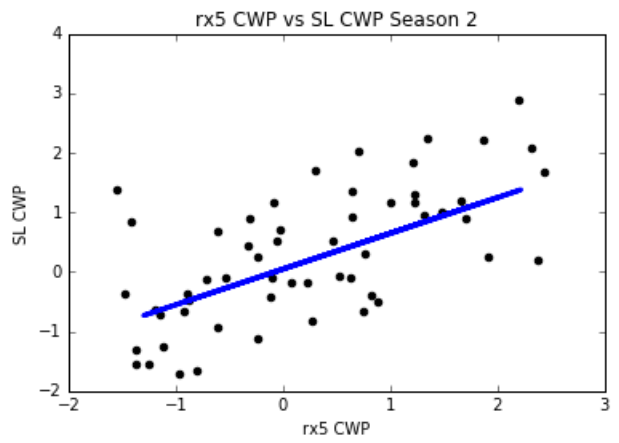
(b) $r^2 = 0.602$

(c) $r^2 = 0.19$

(d) $r^2 = 0.081$

(e) $r^2 = 0.086$

Figure 1: Maximum Linear Regression r^2 Values

sions using the graphs and linear regression we produced. Out of all the graphs we created when performing linear regression, we received five positive r^2 values for 'CDD vs SL', 12 for 'Rx5 vs SL', eight for 'T10 vs SL', three for 'T90 vs SL' and four for 'AVG vs SL'. We have shown the best from each category in Figure 1. The fact that there were many more regions with positive r^2 values for RX5 shows that sea level is affected more by rainy days than dry days, low temperatures, high temperatures, and the average of all of them. Also, the number of positive r^2 values for T10 was relatively high and shows that lower temperature had a large impact on sea level. From this, we can also conclude that T90 had the smallest impact on sea level than any of the other variables that were tested against sea level.

Also, in our graphs, the linear regression line can show how a specific feature affects sea level. For example, the graph seen in Figure: 1b contains a positively sloped regression line. This is true for all but two of the 12 graphs with positive r^2 values. These graphs were for SEA in winter, SWP in winter, USA in winter, USC in winter, CEA in spring, CWP in spring, SPL in spring, SEA in summer, SEA in fall, and USC in fall. We can conclude from this that the more rain a region got in a certain season, the higher the sea level became. Also, we can make conclusions dealing with T10. Of the eight positive r^2 values obtained from 'T10 vs SL', six of them had negative slopes. These graphs were from winter of CWP, winter of SPL, winter of SWP, winter of USA, spring of NWP, and summer of CEA. This shows us that the more low temperature days, the lower the sea level in those regions.

The two variables mentioned above (RX5 and T10) had an impact on a relatively large number of regions. The conclusions we will make for CDD, T90, and AVG will be less impressive because there were so few graphs with positive r^2 values. Although, with each of these variables, all graphs with positive r^2 values follow the same trend. The graph seen in Figure: 1a contains a negatively sloped linear regression line. This is also true with every other 'CDD vs SL' graph that produced a positive r^2 value. All the positive r^2 values graphs for CDD were from SWP in winter, USA in winter, SWP in spring, SEA in fall, and SPL in fall. This shows us that an increase in the number of consecutive dry days results in a decrease in sea level for those regions during those specific seasons. Also, every graph in 'T90 vs SL' with a positive r^2 had a positively sloped linear regression line. These graphs were for CWP in the winter, SWP in winter, and CEA in summer. This shows us that the more high temperature days there were, the higher the sea level became. Lastly, we can discuss how the average impacted sea level. In every graph of 'AVG vs SL' these was a positive linear regression line. These graphs included CWP in spring, SWP in winter, CWP in winter, and CEA in winter. We can conclude that as the average increased, the sea level also increased.

Since we were only able to find trends for certain regions and certain seasons, it was difficult to conclude about how much each variable impacted sea level. We were able to discus whether or not and how a variable impacted sea level but were not able to discus how greatly sea level was impacted by the variables.

"While the spatial variability of sea level change is a complicating factor when making predictions of future changes, it presents a unique opportunity to use observations of sea level changes to understand better the evolution of the climate system in the recent and distant past. This understanding underpins our ability to make accurate predictions of future changes" (Milne). This quote represents some of the challenges we faced trying to predict the changes in sea level during our data analysis.

Future Work

During our work we made several decisions on how we handled our data that resulted in our only using small portions of the total data set in each of the regressions we ran. The largest reduction in data points came without decision to use the seasonal averages exclusively and ignore the monthly data available in the Actuaries Climate Index. While many environmental scientists argue that averages provide a more reliable measurement of most climate factors, we may well have missed opportunities to use that more granular data to perform deeper analyses such as time series manipulation and sliding window data enrichment techniques.

Additionally, we elected to flatten the data by only using measurements from the same season when fitting our regressions. This reduced the variance in climate factor measurements that were caused by seasonal weather shifts, however it also limited our data sets to 54 rows with which to create our training and testing sets. This small number of data points caused a good deal of noise, leading us to consider other techniques that might allow us to leverage larger portions of the total data set should any further work be done.

We could have improved our results by changing the way we split our data into a training and testing set. Currently, our training set is the first 80% of our data. Our data proceeds over time, meaning the first 80% of the data is the earliest data. The last 20% of the data is in the test set. This means that the test set contains data from the most recent years. This could impact our results if it is the case that sea level is changing over time. If this is true, our linear regression algorithm is training on data that is different from the data in the test set. To improve our algorithm, we could have randomly selected the 80% of data for the training set and the 20% of the data for the test set.

Aside from our limited use of the overall data set, we chose to focus on using regression methods, linear regression in particular, and did not employ any classifier type machine learning algorithms in our work. We felt being able to predict values over a continuous range and observing the goodness of fit metrics for our regression lines answered our original question posed in this paper better than results from classifiers would have. That said, classification algorithms make up a large portion of machine learning algorithms available and often offer better performance when properly applied. Future work on this project might well include asking additional questions, some of which could be answered by classifiers such as predicting whether sea level will rise or fall compared to the last month or season. The so-called "sliding-window time-series data enrichment" mentioned earlier in this case would describe a method of using the data within a set number of previous months or seasons (the sliding window) combined with the most recent measurement to predict a rise or fall of sea level. This enables past trends to be accounted for and influence predictions, which could offer improved accuracy over using individual data points to predict class.

While rising vs. falling sea levels are perhaps the most obvious means of labeling our data, many other classification type questions could be posed. The central focus on predicting sea level would likely remain should we explore additional questions, as that represents one of the more impactful environmental factors cataloged by the Actuaries Climate Index. The ability to reliably predict sea levels can serve to help local governments better prepare and plan for the effects of higher sea levels.

Some regions of North America are already planning for these increases in sea level. Gov-

ernments in the Mid-Atlantic region are preparing for sea level rising in Delaware, Maryland, New Jersey, New York, and Virginia with the Mid-Atlantic Regional Council on the Ocean. From the MidAtlantic Governors' Agreement on Ocean Conservation "Prepare the regions coastal communities for the impacts of climate change on ocean and coastal resources. Climate change and its associated impacts threaten to indelibly alter the MidAtlantic region and its resources. Increased coastal hazards, such as flooding and erosion, will threaten existing infrastructure and public health and safety. The widespread nature of this problem also will challenge our efforts to manage human activities across the region" (MARCO). This region has taken statewide initiatives to examine risks and responses to sea level rise. Their main goal was to improve understanding throughout the region.

The projects they implemented are used to acquire and apply data, help their institutions and citizens respond to climate change impacts, such as sea level rise. The important factors of these initiatives and projects are data, outreach, and implementation. Their substantial outreach efforts to address the publics lack of awareness of the fact and consequences of sea level rise. This is exactly what we hoped our data analysis would accomplish. We worked to find correlations in the data set of all regions of North America to help these regions better understand what is happening to sea level. We used this data to create a predictive model for organizations to use as a guide to recognize how sea level is rising and what may cause it to change so rapidly. Using this model, regions like the Mid-Atlantic can better inform their citizens and prepare for the future rise in sea level. Implementation of our model will hopefully build support for the development of sea level rise policy.

A similar project was done in the Florida coastal area by the Florida Oceans and Coastal Council to spread awareness about sea level rise and how it will affect the region. "It provides important information for legislators, policymakers, governmental agencies, and members of the public who are working to address, or who are interested in, issues related to sea level rise in Florida." It is important to ask to what degree sea levels continue to rise and how rapidly, not if they will be affected. They hope to use the data they find on the sea level rise to "seek a thorough understanding of its possible impacts and to provide current and future generations with the information necessary to adjust to higher sea level" (Florida Oceans and Coastal Council). We agree that it is important to look into this research and spread awareness about the rising sea levels.

"It is important for governments and policy makers to be aware of this variability so that appropriate action can be made to plan and implement appropriate mitigatory procedures" (Milne). Hopefully, our data analysis can be used to further predict sea level rise if the Actuaries Climate Index is updated with more years and features to create larger data sets to make it easier to make predictive conclusions. Although we were not able to find any statistically significant correlations, we hope that applying this future work could better the data analysis we conducted in this research. As a group, we thoroughly believe that research into sea level to spread awareness on its rise and its impacts on the environment is important so continued work would greatly benefit society as a whole.

Bibliography

[1] Climate Change Committee, Climate Index Working Group, The American Academy of Actuaries, The Casualty Actuarial Society, The Canadian Institute of Actuaries, and The Society of Actuaries, *Actuaries Climate Index Development and Design*, 2016. Web. 28 Mar. 2017.

[2] Environmental Protection Agency, *Climate Impact on Costal Areas* (06 Oct. 2016). Web. 28 Mar. 2017.

[3] Florida Oceans and Costal Council, *Climate Change and Sea-Level Rise in Florida an Update of the Effects of Climate Change on Florida's Ocean & Costal Resources* (Dec. 2010). Web. 17 Apr. 2017.

[4] Jim Frost, *Regression Analysis: How Do I Interpret R- squared and Assess the Goodness-of-Fit?* (30 Apr.1970). Web. 27 Apr. 2017.

[5] Haydar Ali Ismail, *Learning Data Science: Day 10- Classification, K-Nearest Neighbors, and Cross Validation* (09 Jan. 2017). Web. 24 Apr. 2017.

[6] Stephen Lee Kolk, *An Appraisal of the Actuaries' Climate Risk Index: The Economics Impacts of Sea-Level Rise in Hampton Roads*, May 18, 2016. Paper 7.

[7] David M. Lane, *Introduction to Linear Regression*. N.p. n.d. Web. 24 Apr. 2017.

[8] Mid-Atlantic Regional Council on the Ocean, *MARCO Climate Change and Sea Level Rise Information Exchange* (N.p. Dec. 2010). Web. 17 Apr. 2017.

[9] Glenn Milne, *How the climate drives sea-level changes*. A & G 2008; 49 (2): 2.24-2.28.doi: 10.1111/j.1468-4004.2008.49224.x.

[10] Nobuo MIMURA, *Sea-Level Rise Caused by Climate Change and Its Implications for Society*. Ed. Kiyoshi HORIKAWA. Proceedings of the Japan Academy. Series B, Physical and Biological Sciences 89.7 (2013): 281-301. PMC. Web. 17 Apr. 2017.

[11] Bruce B. Parker, *Sea Level As an Indicator of Climate and Global Change*. Marine Technology Society Journal, Vol. 25, No. 4, 1992. Web. 17 Apr. 2017.

[12] Jim. Titus, *Does Sea Level Rise Matter to Transportation Along the Atlantic Coast? The Potential Impacts of Climate Change on Transportation*. U.S. Department of Transportation, n.d. Web. 29 Mar. 2017.

[13] US Department of Commerce and National Oceanic and Atmospheric Administration, *Is Sea Level Rising?*. NOAA's National Ocean Service. N.p. 27 Oct. 2008. Web. 28. Mar 2017.

[14] Michael Walker, *Random Forests Algorithm*. Data Science Central. N.p. 24 Sept. 2013. Web. 25 Apr. 2017.

[15] Yhat., *Forests in Python*. Random Forests in Python. N.p. 5 June 2013. Web 27 Apr. 2017.