From Sea to Shining Sea

M3 Challenge

- 2017 -

By: Team 8800

February 26, 2017

1 Executive Summary

It has become more clear as time goes on that climate change has continued to have an increasingly immense impact on a parks resources and visitor experience. While climate change can continue on infinitely, unfortunately funding to maintain every park simply cannot and it is vital that funds are allocated accurately in precise proportions.

To begin with, we began our tackling of the problem by inspecting, organizing, and producing coherent data sets and charts that allowed us to observe common trends. We compiled monthly data for MSL into yearly data to create a more clear chart and noticed quickly that chunks of data were missing. Using differences between consecutive rates of change and calculating averages of those rates of change, we were able to fill in the missing data values with our most accurate point estimator being the average rates of change of the values we were given. After implementing the data in charts, we noticed that the distribution followed a roughly normal distribution without any true outliers so we utilized normal statistical analysis and created a 95 percent confidence interval for the rates of change of sea level for each park given. Alongside factoring in surrounding parks rates of change of sea level in a 500 mile radius, we deemed that a high risk in sea level rates of change would be indicated by a stronger difference between the respective parks sea level rate of change and its surrounding parks, and vice versa.

Our goal was to create a method to find a single climate vulnerability index score that will be ranked in a continuum. Our mathematical model is based on susceptibility factors concurrently with non-climate stressors formatted into multiple spreadsheets. We categorized our scores into present circumstances, susceptible risks in a climate, adaptive capacity, and certainty based on a range of -18 to 90. This ranking will take into account future predictions of 50 years based on adaptive capacity which tells us how an environment acclimates as it grows. Lower ranges will portray safe habitats while higher ranges represent critical habitats. Our main objective is to help the current management of NPS units make imminent decisions that are based off of expert opinion and justifiable information.

Finally, we decided to analyze the visitor populations of each park. After implementing the data of each park for visitor population in a basic chart, we again noticed that it did not follow a linear trend and we decided to then apply polynomial regression from Microsoft Excel to get a rough estimated equation, and then once more utilizing machine learning polynomial regression in the programming language, Python, alongside the lowest possible least squares regression value in order to optimize the accuracy of the curve. In order to precisely allocate funding to different parks, we utilized a weighted mean formula in which the estimated future visitor population and our CVI (climate vulnerability index) were factored in, in order return data on which of the parks will be in the most dire need of funding. Based on this, we found that Padre Island national park will require the most funding followed by Kenai Fjords, Olympic national park, Cape Hatteras, and Acadia will accordingly require the least funding.

2 Introduction

Due to our largely industrialized world with millions of inhabitants, we find ourselves faced with problems such as climate change with the large amount of carbon dioxide emissions and other harmful things we, as humans, do to our environment. These changes to our environment lead

to the warming of the atmosphere and a largely warmer planet, which can lead to things such as the rising of sea levels and the melting of polar regions which have a profound effect on our environment and raise many concerns. The National Park Service (NPS), is tasked with preserving our national parks, and this task becomes difficult as the climate changes and different parts of the environment begin to change. With rising sea levels, the parks on the coasts of the United States become at risk of drastically changing, which would be a variation and failure of preservation of the park. There are also considerable effects on the actual environment of the park from climate change, from wildlife to overall health of the environment, there can be some immense changes in the environment of the park, along with the interest in the park. Different geographical events can lead to shifts in interest in the park with varying levels of visiting. The interest in the park and amount of visitors can have a large effect on the budgeting and spending for a park. The climate change has a massive effect on the decisions and actions that the National Park Service make in the future.

3 Problem Restatement

With the rising problem of climate change in modern society, we were tasked with creating many models to create and classify different problems that the National Park Service may face.

1. Create a model to sort 5 different national parks into 3 different categories based on their level of sea level change for the future, including all of the factors for climate change and previous data on sea level changes.

2. Develop a model to give a score to the parks based on the vulnerability of the park in relation to the risk it is at from climate change. Create a model that works for over 50 years.

3. Create a model to properly finance the national parks based on the vulnerability scores and interest in the park with past data on visitors to the park. With this data, give a model that will predict the changes to the park and advise where the National Park Service should allocate their money.

4 Low, Medium, or High?

4.1 Completing the Dataset

The method for calculating the missing data values contains assumptions and the following calculations

Assumption: Rate of change of sea level change is constant throughout one year per location.
Justification: There are too many factors and random variables in finding the rate of change of the rate of change of sea level.

Assumption: Seasonal changes are not a factor in sea level change
Justification: The national weather service graphs for seasonal changes does not show high variation.

Assumption: Padre Island cannot be filled in accurately with data
Justification: Our method of calculating empty values will result in high variability for the Padre Island.

A is the number of available data points. We set A as 12 because that is the number of months in a year.

$\boldsymbol{m}$ stands for month. This helps us identify the specific month we have selected with the subscript of $\boldsymbol{n}$

$$\sum_{n=1}^{A/2} m_n - m_{n+1} = X \tag{1}$$

$$\frac{X}{A} = R \tag{2}$$

$\boldsymbol{R}$ is the rate of change of sea level trend per year

4.2 Safe to Use Normal?

The data set provided that contains data for sea level trends has more than 30 samples per location. This satisfies the central limit theorem that requires only 30 samples. Since we are dealing with means, the central limit theorem is the correct rule to use.[**10**]

4.3 The Next 10-20 Years

To judge the mean rate of change for sea level we constructed a 95 percent confidence interval for every location.
Since we have population data we decided it was safe to use a one-sample $\boldsymbol{z}$ interval. This will allow us to see a range of values that can estimate the true population mean of sea level increase values over the next 10-50 years.

$$Interval = \bar{x} \pm z^{\star}\frac{\sigma}{\sqrt{n}} \tag{3}$$

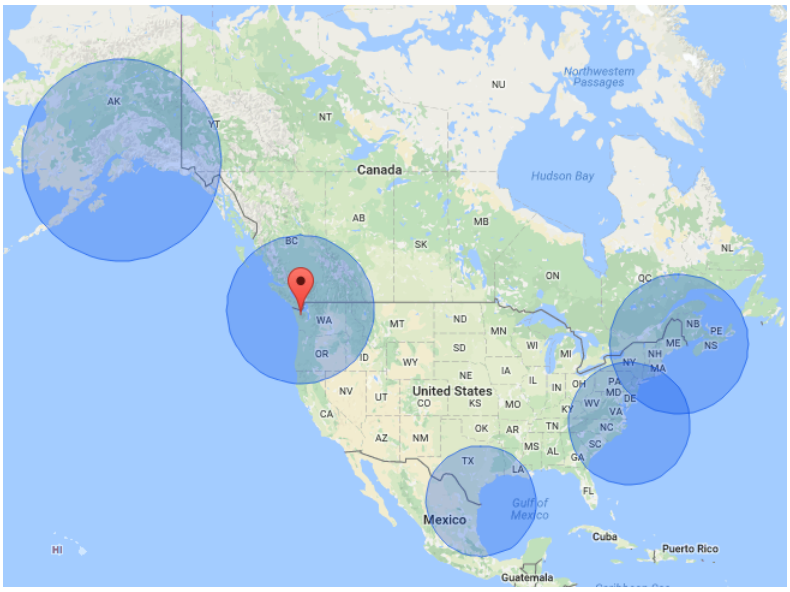
Confidence Intervals			
Park Name	Point Estimator	Margin of Error	Interval
Olympic	-0.00044	0.02497	(-0.0253, 0.02453)
Cape Hatteras	0.726918	0.35535	(0.37156, 1.0823)
Acadia	0.546208	0.20471	(0.34149, 0.7509)
Padre Island	0.628383	0.52201	(0.10638, 1.1504)
Kenai Fjord	-0.72792	0.31122	(-1.039, -0.4167)

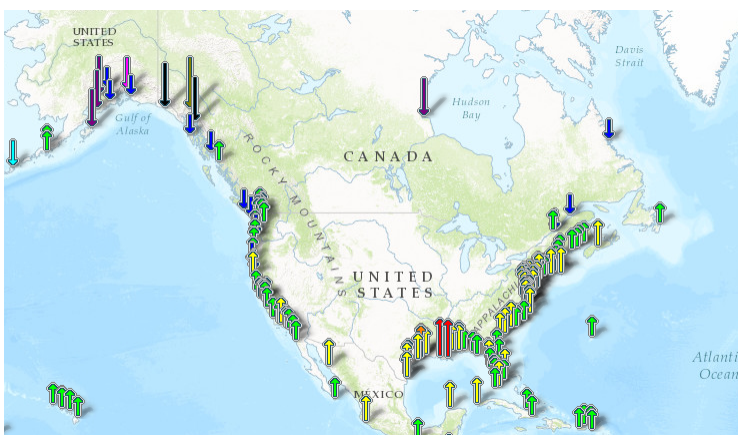
4.4 High, Medium, and Low

Our conclusions for what parks are classified as high, medium, and low risk ratings for sea level change.

Risk Level Ratings	
Park Name	Rating
Olympic	Medium
Cape Hatteras	Medium
Acadia	Low
Padre Island	High
Kenai Fjord	Low

Our confidence interval values and surrounding park sites allow us to jump to the conclusion stated in the table. The surrounding station sites show that these geographical areas are similar to the parks selected for the problem. By using a radius of 500 miles, we can classify the clusters near the park site.**[5]**[**6**]





Overlaying the images allows us to see this clustering classification.

4.5 Beyond

Our model will hold true for the next 10 to 50 years, where carbon dioxide emission and glacial melting trends will not increase or vary much. But the model will not prove to be accurate 50 to 100 years from now, due to too many unknown variables. As time goes on, the amount of carbon dioxide being emitted on a yearly basis could change drastically with the development of new technology or new environmental policies that would limit or monitor carbon dioxide emission. The change in levels of carbon dioxide in the atmosphere would change the rate of thermal expansion of the oceans along with the rate that the polar regions melt at.**[1]** Based on the uncertainty of how carbon dioxide levels will change. The melting of the polar regions has increased incredibly, and it would be very

hard to predict if the ice would keep melting at the same rate, or increase exponentially in 100 years.

We do not have any data concerning the shift of elevation by Alaska and the rest of North America due to Alaska becoming lighter from the melting glaciers.[**2**] Depending on how much mass of ice melts off of Alaska, the tilt due to the alleviation of mass would drastically change and skew sea level changes for all of the locations in North America. There is a lot of uncertainty with the effects of melting ice caps, to the point where the data from the past 27 years where sea level increase is 2.5 times the rate from 1900 to 1990. The trend is too underdeveloped to come up with an accurate model for the far future. These factors that we cannot change will make either keep the rates similar to where they are now, or drastically change the sea level change, which would change the classification for the different locations.[**3**]

5 Is the Coast Clear?

5.1 Criteria

The assignment of giving a climate vulnerability score to a coastal unit is dependent upon several geological and physical processes. We can measure climate-related events and their effects within an index. This index is formatted into a spreadsheet tool aggregating data of climate change impacts and environmental stressors. Taking various mathematical models, we will predict the vulnerability in coastal regions. This information can be directly applied for the next 50 years and help current management and conservation decisions of the NPS.

5.2 Assumptions

Assumption: Our model is established on the expectation that fields of higher environmental quality can hold higher abundance of indigenous species, and that losses in habitat conditions contributed to species decline.
Justification: An increase in prosperous land mass gives enough space for species composition and structure to flourish, but when there is habitat disturbance species lack the nutrients to endeavor the circumstances.
Assumption: We can assume that multiple climate change exposures result in higher vulnerability over time.
Justification: When a habitat has overlapping issues of exposure it has a higher chance of becoming bereft of life because multiple disturbances will accelerate the rate of life decline.
Assumption: An average score would not naturally compare the difference between a high or low score.
Justification: An average rank of the foreseen feedback would propose a neutral rank. The most applicable score should follow the so-called worse case response.

5.3 Developing the Model

Our tool is a spreadsheet-based device that will rank coastal regions based on climate change exposures (temperature increase, precipitation change, sea level rise, and extreme climate events) and non-climate stressors (invasive species, nutrients, sedimentation, and contamination toxicity) that portray the persistence of a habitat. We will also take in factors of direct sensitivity of the environment to climate alterations, the ongoing circumstances of the habitat, and instinctive and

anthropogenic conditions that influence adaptive capacity. All of these components will factor into our final numerical vulnerability score.

5.4 Scoring

The scores assigned for current environmental factors, non-climate stressor interactions with climate exposure, adaptive capacity and certainty relative to these scores gets ranked in our spreadsheet. The determined response criteria is classified into three categories: Present circumstances, susceptible-risk, adaptive capacity, levels of confidence.

Present circumstances: This score is calculated to gain the relative healthiness of environmental region preceding the influence of stress of extraneous exposures. How particular units in this environment react to imminent cases will depends on whether the habitat has already been arbitrated to other non-climate stressors. We must take into account the direct climate effects that mightve occurred to which the surroundings have been once inveigled.

Susceptible risks: The explicit effects of alternations to the climate and the foreseen communications of the environment with the non-climate stressors are scored here. This risk is attributed to sensitivity into our six categories (direct effects, invasive species, nutrients, sedimentation, erosion, and environmental toxicity) and climate exposures.

Adaptive Capacity: Instinctive traits of a habitat that allow an environment to adjust a growing climate is scored in this section. Adaptive capacity is based on degree of fragmentation, barrier of migration, recovery, and human response.[**7**]

Level of Agreement:

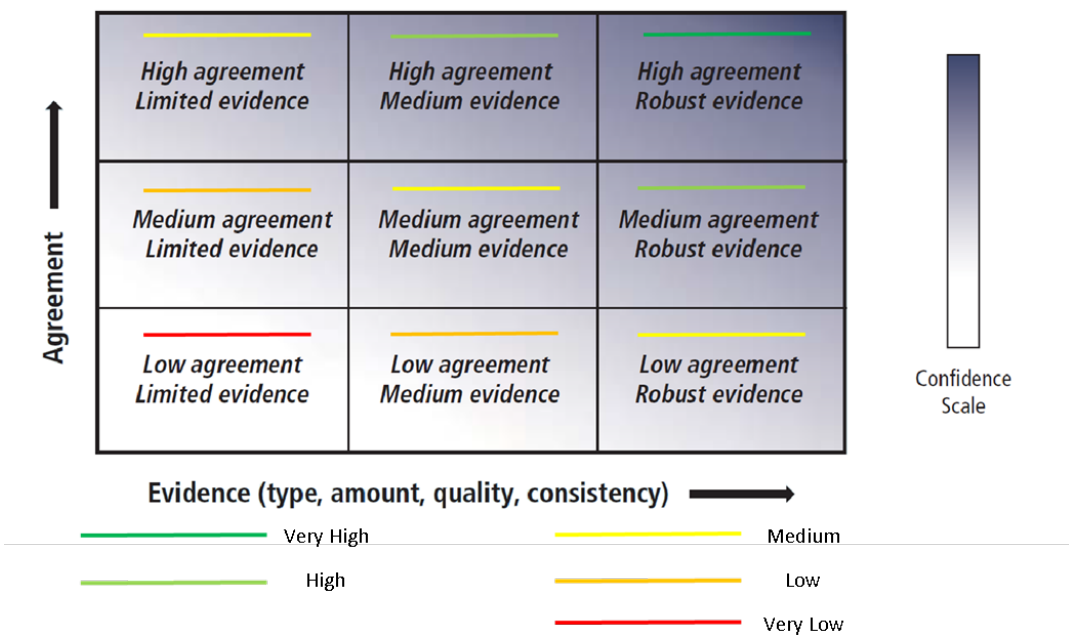


Figure 1: Confidence levels are a combination of level of agreement and evidence. There are five levels, shown with colours. (IPCC 2013)

Category	Points	Description
Present Circumstances	**0**	Habitat is not impacted
	2	Habitat is impacted to limited degree
	5	Habitat is currently moderately affected
	10	Habitat is severely impacted
Susceptible risks	**-2**	Habitat may benefit
	0	No anticipated in habitat found
	2	Habitat will be impaired to a limited degree
	5	Habitat persistence will be limited
	10	Habitat will wasted
Adaptive Capacity	**0**	Severe hindrances to habitat persistence or diffusion of the habitat that isn't sufficient
	2	Moderate hindrances to habitat persistence or diffusion of the habitat that isn't sufficient
	5	No hindrances to habitat persistence or diffusion of the habitat that isn't sufficient
Level of agreement	**0**	No direct support is found to support score
	1	Low evidence that is inconclusive
	2	Medium evidence that is suggestive to be on expert opinion
	3	High: moderate evidence that includes several sources
	4	Very high: Strong evidence with a high consensus and established theories.

5.5 Process of Scoring

Based on our mathematic model shown later a user will apply a score for each category dependent upon the results of each model. The best way to conclude accurate results is to base your data and calculations to use off of expert opinion and official NPS management. Also, take your scores into consideration of multiple stressor interactions happening at once. For example, when perception lowers concurrently during sedimentation salinity increases, vegetation growth decreases, and sediment production decreases. Keep in mind of your vulnerability score when these circumstances occur.

Our model will be split into three spreadsheets to get an overall vulnerability score for a coastal region. The first spreadsheet will accumulate the score susceptibility to risk depending on categories seen below. This basic framework will be used to aggregate the data from all the categories for any coastal region. The second spreadsheet will calculate the score of adaptive capacity which tells us how well that region adapts to supplementary, outside factors. The third spreadsheet will add all the scores and present an overall vulnerability. This number will be placed on an index of high to low where -18 presents a non-vulnerable region and 90 presents a very vulnerable region.

5.6 Sub-categories of Climate Change and Environmental Factors

Susceptibility factors: Continuous Environment Effects Invasive Species Nutrients Environmental Toxicity All these factors are concurrently scored with climate factors or represent on the spreadsheet as risks. The risks are current conditions, CO2 levels, temperature, sea levels, and extreme climate events. Next is breaking down the Adaptive Capacity spreadsheet. We need to take in adaptive capacity to see how much can a coastal region change before it renders wasted. The Adaptive Capacity is broken down into degree of fragmentation, barrier of migration, recovery, and human response. Remember all of these are scored in adjacent to certainty and how supported each score is

based on evidence. The lower the certainty of score is the less value the data accounts for. Certainty is one of the main deciding factors if the data is valuable or credible to use for predictions.

(INPUT ANY SUSCEPTIBILITY FACTOR)

(RISKS)	**-2**	**0**	**2**	**5**	**10**
Current Conditions		No Change	A shift in species composition and phenology no significant losses	A significant change to synergy due to recent climate change	A phenologic disconnect resulted in loss of habitat
Increase in CO2	Habitat will benefit from elevated CO2	No change	Community composition will alter habit structure	Community composition will severely alter habit structure	Foundation species will be displaced due to changes
Increase in Temperature	Habitat will benefit due from increase in growing season length	No change	Changes in temp. will mildly reduce species' reproduction	Habitat functions alter temp increasing mortality rates	Habitat cannot persist under alternated disturbance
Change in Sea Level	Habitat will benefit e.g. increase area due to inundation	No Change	Community species will be limited in access to freshwater	Frequency inundation will cause stress to habitat and affect interactions	Habitat will be completely inundated
Increase in Extreme Climate	Habitat will benefit to species that progress	No Change	Foundation species too slow to regenerate	Critical species will be affected by disturbance	Habitat is not likely to recover from major disturbance

(Changes)	**0**	**2**	**5**
Degree of fragmentation	Sufficient enough to prevent habitat adaption across habitat's original extent	Not likely to influence more than habitat's current extent	Not a impendent to habit persistence
Barrier of migration	Barriers border the current habitat are lily to be greatly impaired	Barriers border the current habitat are likely to be somewhat impaired	Effective barriers the cause shifts do not exist for habitat
Recovery	Foundational species do not have recovery	Foundation species have modest recovery	Foundation species have access to recover at good rates
Human Response	Response is neutral or negative	Response is moderate	Adaptive capacity response is high

5.7 Our Assessment

Climate Vulnerability Index (-18, 90)	
Park Name	Index Value
Olympic	55
Cape Hatteras	16
Acadia	16
Padre Island	9
Kenai Fjord	29

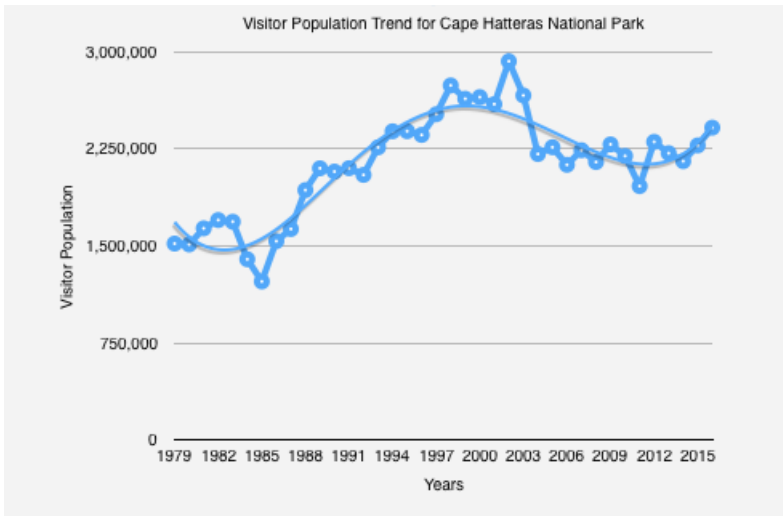
These specific values for the climate vulnerability of each national park were computed by utilizing the Susceptibility Factor Table in the Sub-categories section.

6 Where Will the Money Go?

6.1 Modeling the Future Visitors

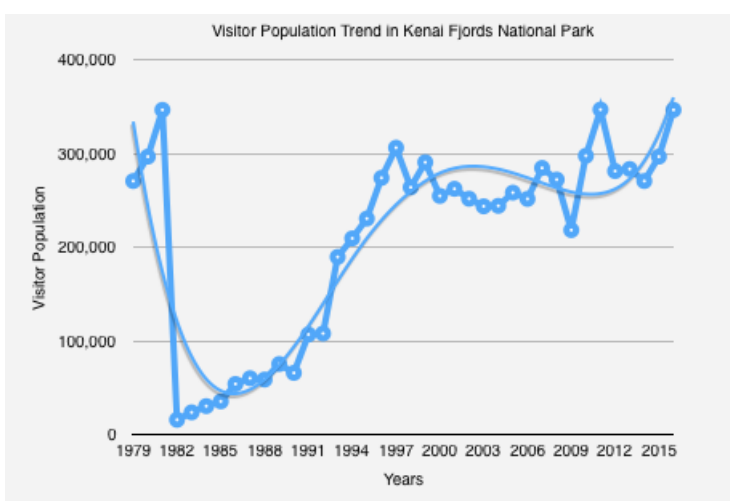
The visitor data provided gives us a solid baseline to develop two polynomial regression models. The first one we constructed uses data fitting techniques from Microsoft Excel.

Cape Hatteras National Park:



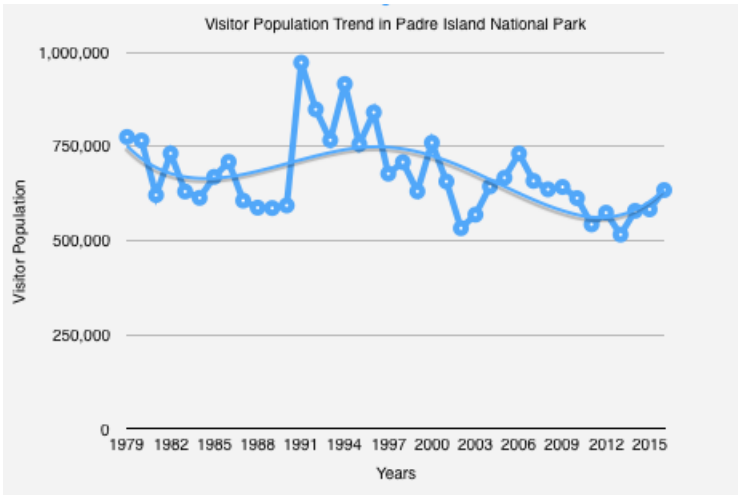
$$y = 16.708x^4 - 1321.5x^3 + 31997x^2 - 209489x + 1.862E6 \tag{4}$$

Kenai Fjords National Park:



$$y = 4.9577x^4 - 427.92x^3 + 12304x^2 - 12390x + 446019 \tag{5}$$

Padre Island National Park:

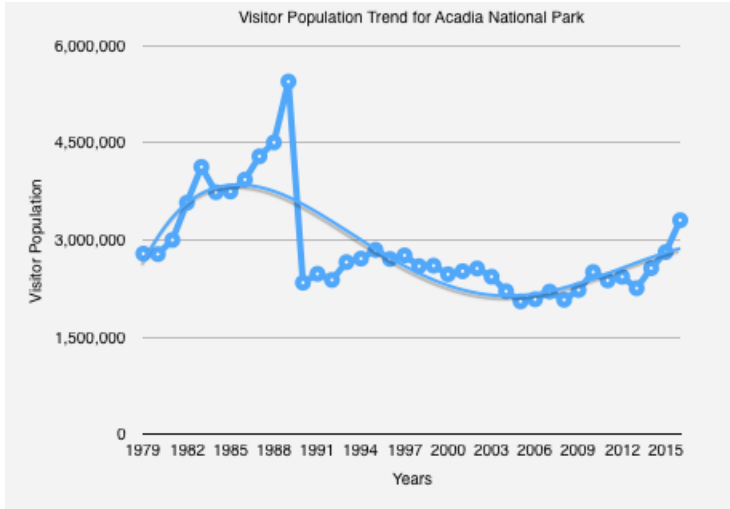


$$y = 3.6719x^4 - 283.06x^3 + 6771.5x^2 - 54372x + 799401 \tag{6}$$

Olympic National Park:

$$y = 11.937x^4 - 847.84x^3 + 16528x^2 - 27627x + 2.208E6 \tag{7}$$

Acadia National Park:



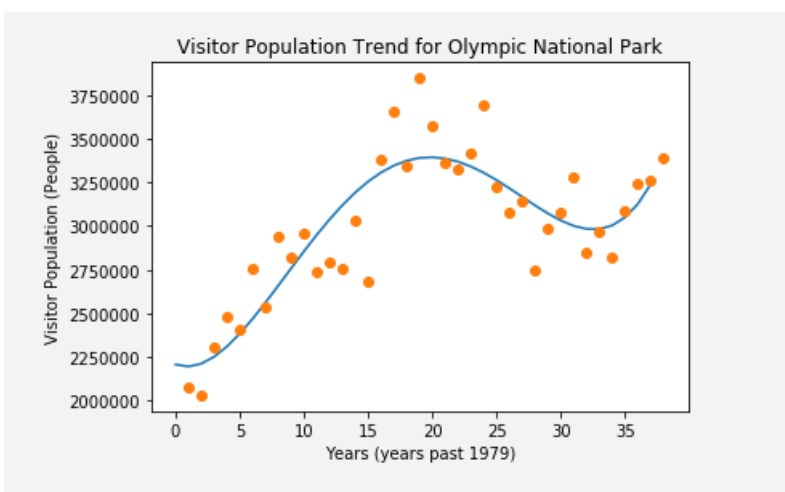
$$y = 16.66x^4 + 1648x^3 - 52032x^2 + 529739x + 2.181E6 \tag{8}$$

We want to predict how many visitors will be at each park for the Excel Model. Plugging 86 for $\boldsymbol{x}$. This allows us to compute an estimate for the amount of visitors in 2065 which is 48 years from 2017.

Visitors Predicted in 2065 - Excel Model	
Park Name	Visitors
Olympic	235,760,000
Cape Hatteras	2,938,900,000
Acadia	1,622,400,000
Padre Island	67,019,000
Kenai Fjord	89,390,000

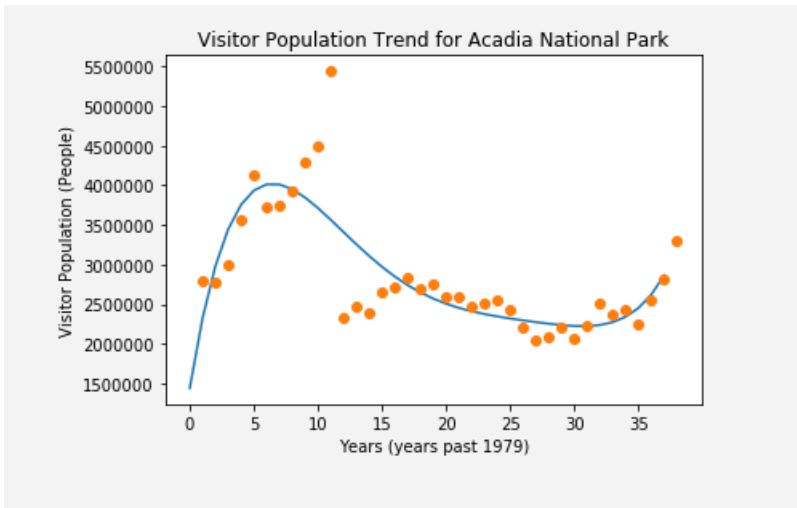
Our second model focuses on a Python machine learning approach using polynomial regression. We used MatPlotLib and Numpy.[**4**][**9**]

Olympic National Park:



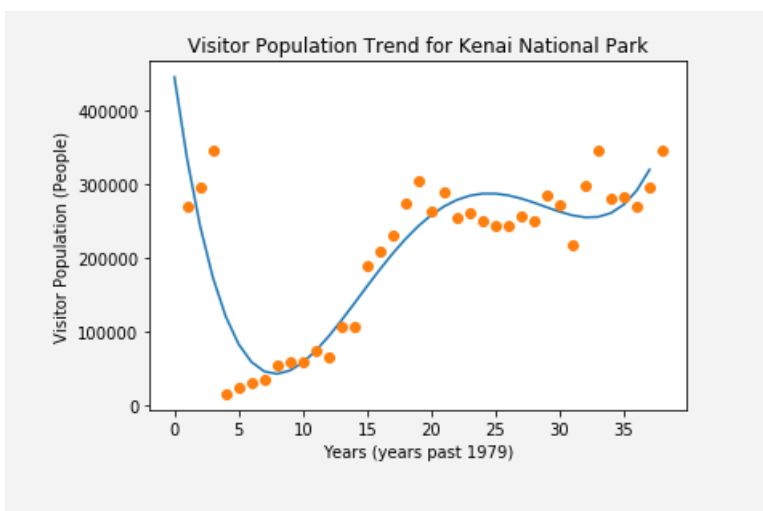
$$y = 11.9370453246x^4 - 847.835432395x^3 + 16527.9484832x^2 - 27627.4841072x + 2208359.27281 \quad (9)$$

Acadia National Park:



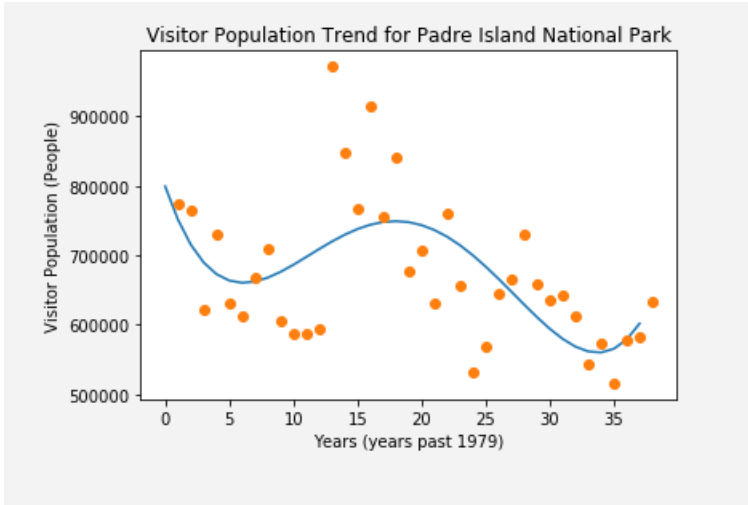
$$y = 1.6127727x^5 - 173.90531x^4 + 7136.7939x^3 - 133956.381x^2 + 1010830.591x + 1442227.521 \quad (10)$$

Kenai Fjords National Park:



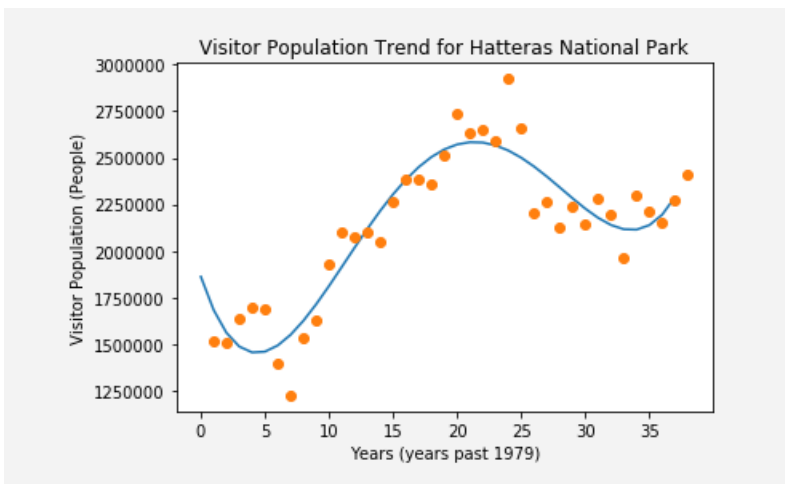
$$y = 4.95774753307x^4 - 427.9176298x^3 + 12303.5142258x^2 - 123929.77747x + 446019.033317 \quad (11)$$

Padre Island National Park:



$$y = 3.67189419596x^4 - 283.058068536x^3 + 6771.538599x^2 - 54372.0144735x + 799400.910384 \quad (12)$$

Cape Hatteras National Park:



$$y = 16.7080517217x^4 - 1321.504822x^3 + 31996.82369x^2 - 209488.66575x + 1861658.75333 \quad (13)$$

Minimizing the Mean Squared Error for Polynomial Regression:

$$E = \sum_{j=0}^{k} |p(x_j) - y_j|^2$$

This equation is a common measure of the quality of our estimation. This was integrated into our polynomial regression model.

6.2 Visitors in the Future

Visitors Predicted in 2065 - Python Model	
Park Name	Visitors
Olympic	235,768,409
Cape Hatteras	293,887,134
Acadia	1,711,189,121
Padre Island	67,020,534
Kenai Fjord	79,798,109

6.3 Who Needs the Money?

We decided to use the Python model output because it has minimization of squared error integrated and it has the most realistic population values.

Using a weighted mean, we can assign a score to every park that factors in their population and Climate Vulnerability Index(CVI). The parks with high weighted means do not need funding since they are given funds by the admission fee to the park. Parks with low weighted means mean that they need funding since they are not getting large visitor influx. The population was given more weight since that is what is giving a park funding.

v is the amount of visitors for a certain park. ***cvi*** is the Climate Vulnerability Index for a certain park.

$$(0.6)(v) + (0.4)(cvi) \tag{14}$$

6.4 Ranking in the Future

Weighted Mean Values	
Park Name	Weighted Values
Olympic	1.414610674E8
Cape Hatteras	1.763322868E8
Acadia	1.026713479E9
Padre Island	4.0212324E7
Kenai Fjord	4.7878877E7

Padre Island requires the most funding. It has the lowest weighted value meaning that it has low visitors and a high CVI value. This makes sense because Part 1 of our model shows Padre Island with a **high** rating for sea level risk.

7 Strengths and Weaknesses

7.1 Part 1

Strengths: For our model in part 1 our strengths were in being able to determine a set range of values of the rate of change of the sea level in the five national parks listed. This confidence interval for each park allowed us to provide a valid rating of risk by looking at the surrounding parks around the listed ones.

Weaknesses: Our model is weak in the fact that it does not take into effect the tilt of the tectonic plates from the shift in weight as Alaska becomes lighter, along with not knowing how the levels of carbon dioxide emission will change. There are too many factors with different policies that may be put in place in the future which might limit carbon dioxide emission due to it being a pressing matter around now, or even the development of greener, more efficient technology that would reduce the overall carbon dioxide emission.

7.2 Part 2

Strengths: The table model we were able to conclude off of in the model in part 2 was strong in that it provided acknowledgement of the many factors that can act on the parks in the future.

Weaknesses: However, its weakness is that it is not specific in the way each category's points can be determined. Only with true research, we were able to conclude climate vulnerability indexes of the various national parks.

7.3 Part 3

Strengths: The biggest strength of our models in part 3 was the machine learning aspect we were able to utilize from libraries within Python. Being able to model the visitor data of each of parks with Python on the computer gave us a more exact answer to the actual outcome in 50 years than just solely polynomial regression could.

Weaknesses: However to fully rank the parks in priority for funding the variable of the climate vulnerability index is a weakness. This was not an exact index for every park like the total populations almost were, thus leaving the rankings to be slightly skewed.

8 Conclusion and Future Direction

Essentially, it was firstly observed that an extremely accurate measurement of whether a rate of change of sea level is high, low, or neutral can be calculated and concluded based upon similar rate of change readings of surrounding parks in a justified radius. National parks near a body of water (sea) in each others surroundings seem to follow similar patterns of rise and are in direct relation with each other. Furthermore, we concluded with 95 percent confidence that Padre Island will be the national park that will be in the most dire need of funding in the upcoming future of 10, 20, and 50 years, followed by Kenai Fjords, Olympic national park, Cape Hatteras, and Acadia.

1. To improve our model for problem one, we could've taken into effect the change in elevation by the melting of Alaska to get the shift over time of the tectonic plate, while also getting more data on the trends in carbon dioxide emission over the years and also more data on factors that could affect this in the future and how these emissions would change the thermal expansion of water and the melting of polar ice regions.

2. To improve problem 2, we could have added more data to aggregate a mathematical model to show patterns of each climate change factors. I could also explain more info on how each susceptible factor coincides with the risk factors. we would've also could compare how susceptible risk relates to vulnerability rates.

3. To improve this model we could of collected more information on the particular climate change factors that affect the interest in the park along with different costs to maintaining and preserving the park. We also could use our data to better find the perfect degree polynomial to use for our graphs.

References

[1] http://www.cbsnews.com/news/alaska-glaciers-sending-75-billion-tons-of-water-into-sea-each-year/

[2] http://io9.gizmodo.com/why-are-sea-levels-dropping-in-places-closest-to-the-me-1684599241

[3] http://science.howstuffworks.com/environmental/earth/geophysics/question473.htm

[4] https://github.com/eckels/m3challenge/blob/master/problem3.ipynb

[5] http://obeattie.github.io/gmaps-radius/?lat=51.500358lng=-0.125506z=10u=mir=5

[6] https://tidesandcurrents.noaa.gov/sltrends/sltrends.html

[7] http://www.vims.edu/cbnerr/resources/CCVATCH_Guidance_Final.pdf

[8] https://www.ipcc.ch/pdf/supporting-material/uncertainty-guidance-note.pdf

[9] https://wch.github.io/latexsheet/latexsheet-1.png

[10] https://mym3challenge.siam.org/node/30133

9 Software Used

1. Latex
2. Apple Numbers
3. Excel
4. Google Drive
5. MatPlotLib
6. Numpy
7. IPython Environment