

CSC110 Final Project: Effects of Melting Glaciers, CO₂, and Global Temperature on Sea Level Rise

Alissa Lozhkin, Ayanaa Rahman, Nada Gamal Eldin, Rachel Xie

December 2020

Introduction

A major concern connected to climate change is sea level rise. According to an article from National Geographic, average sea levels have gone up by about 23 centimetres since 1880, and the current trend is a yearly increase of approximately 3.2 mm (Nunez, 2019). A 2019 report from the National Oceanic and Atmospheric Administration (NOAA) noted that the last twenty years have seen a rapid increase in flooding events along U.S. coasts alone, and that increase is expected to continue (Jet Propulsion Laboratory, 2020). This is problematic because rising seas could lead to coastal areas being flooded, forcing people to flee for higher ground and migrate to new areas. Storms and hurricanes would also be exacerbated, worsening the damage they cause. Additionally, saltwater could leach into aquifers, destroying habitats and leading to huge losses in crop production as freshwater is contaminated (Amadeo, 2020).

We are interested in factors that affect sea level rise, such as glacier melt, ice sheet mass loss, and temperatures rising all over the world (Hennig, 2019). When water heats up, it expands, which causes the oceans to take up more space. While it is natural for glaciers and sea ice melt during the warmer months, they are now melting more than they should each year, and that also adds to sea levels—the natural evaporation and reformation of snow and ice that would normally balance out these effects is less effective as global temperatures increase (Nunez, 2019).

We are interested, additionally, in the relationship between carbon dioxide (CO₂) in the air and sea level rise and whether the two follow similar increasing trends. Due to the fossil fuels we frequently burn for energy, CO₂ levels in the atmosphere are greater than they have ever been in at least the past 800,000 years. As a greenhouse gas, carbon dioxide contributes greatly to trapping heat on Earth, causing global temperatures to rise, which in turn connects greater amounts of CO₂ to rising sea levels (Lindsey, 2020).

Our research question is, **“How are global mean sea levels changing with respect to changes in atmospheric CO₂ concentrations, mean cumulative mass balance of glaciers and global mean temperatures?”**

Datasets

For the processed datasets: Claim ID: Wk5zy38gDYzWCNRw Claim Passcode: H7nSiGA7eXXoBT9C

- Average cumulative mass balance of reference glaciers worldwide: <https://datahub.io/core/glacier-mass-balance#data>
- Monthly CO₂ (this data was also processed and put into an Excel file): [https://www.CO2.earth/monthly-CO₂](https://www.CO2.earth/monthly-CO2)
- Observed land-ocean temperatures: <https://www.bloomberg.com/graphics/2015-whats-warming-the-world/data/observed.csv>
- Global Average Absolute Sea Level Change, 1880-2015: https://www.epa.gov/sites/production/files/2016-08/sea-level_fig-1.csv

The first dataset we used is called “Average cumulative mass balance of reference Glaciers worldwide”. It contains an average cumulative mass of glaciers from 1945-2014 sourced from US EPA and the World Glacier Monitoring Service (WGMS). The values represent the average of all the glaciers that were measured. Negative values indicate a net loss of ice and snow compared with the base year of 1945. Measurements are in meters of water equivalent, which represent changes in the average thickness of a glacier (Datahub, 2015). This dataset is contained within a

csv file. For our purposes, we used the “Mean cumulative mass balance” columns and the “Year” column, focusing only on the time period from 1993 to 2014.

The second dataset that we used is called “Monthly CO₂”. It contains information about the monthly CO₂ levels from March 1958 to September 2018, sourced from the Mauna Loa Observatory. We have inputted this data into an excel file and only included the years 1993 to 2014 (CO₂.earth, 2020). All columns of this dataset were used to produce averages for each year.

The third dataset that we used is called “Observed land-ocean temperatures”. It contains the annual mean and 5-year mean of observed land-ocean temperatures from 1880 to 2014, measured in fahrenheit. The data is sourced by Bloomberg and created through simulation via a supercomputer by NASA’s Goddard Institute for Space Studies (GISS) (Migiliozzo, 2020). The data is contained within a csv file. For our purposes, we used the “Annual_Mean” column and the “Year” column, focusing only on the time period from 1993 to 2014.

The fourth dataset that we used is called “Global Average Absolute Sea Level Change, 1880-2015”. It contains adjusted sea level in inches, from 1880 to 2015. The data is sourced from EPA’s Climate Change Indicators in the United States. The data is contained within a csv file. For our purposes, we used the “NOAA - Adjusted sea level (inches)” column and the “Year” column, focusing only on the time period from 1993 to 2014.

Computational Overview

Statistical analysis functions

We used different statistical calculations to analyse our data. We created a function for each one. There were some built-in functions we could have used from the statistics module, but we preferred to implement them ourselves. These functions include mode and median. We did have to use some built-in functions from the math module, these include: ceil and sqrt. Functions include: sample standard deviations, which calculates and returns the variability of the data. We also calculate r which measures the strength and direction between two quantitative variables. This is implemented using the correlation function. Other functions were implemented to figure out the linear regression: functions to return the slope and y-intercept of the line of best fit. We also implemented an interpolation to return the y values at certain x values in our line of best fit. We implemented an extrapolation function in order to predict future values of the graph. We also calculated the coefficient of determination (r squared) and the root mean squared error, which measures the prediction error for predictions from a linear regression model. Lastly, we created a function (multiple.lin_reg) that performs multiple linear regression on the data using the scikit-learn (sklearn) library. Using sklearn’s LinearRegression model, we can perform multiple linear regression, getting the coefficients and y-intercept of the equation from LinearRegression.coef_, and LinearRegression.intercept_. LinearRegression.fit() was used to ‘train’ the data to produce a final model.

Plotting and curve fitting

Each file was in a csv file or an excel file. Pandas was imported as pd and the functions read_csv_file and read_excel_file were used to read csv files and excel files respectively. Both of these functions returned data frames which are objects associated with pandas. These objects are useful for storing tabular data. Each dataset was processed, and they were processed in different ways. The CO₂ data was processed so that the averages for each year were inputted into the list of x-values. The Sea Level data was processed so that the data from the 5th column was inputted into the list of y-values. The Glacier data and Temperature data were processed so that the data from the 2nd column was inputted into the list of x-values.

The filenames.to.lists function converts each file from their respective format to lists. The create_data_frame function uses pandas to generate dataframes from the data that can then be used to plot the dependent variable against the independent variable. In order to plot them, from matplotlib, plotly was imported as plt and it was used to output the titles on the graphs and to plot the points.

For curve fitting, linear, quadratic, cubic, and exponential functions were used as model functions to create curves of best fit using the curve_fit function from scipy.optimize. The curve_fit function took the model function, the list of x-values, and the list of y-values and outputted the optimal parameters for the function and the covariance of the x- and y-values. For curve-fitting, the covariance was not used. So, for a linear equation, the optimal parameters outputted from the curve_fit function will be floats for m and b. For-loops were then used to generate y-values according to the x-values and the optimal parameters. Matplotlib.plotly was used to generate a scatter plot and a curve of best fit on the same graph. The scatter plot consisted of the actual y-values against the x-values, and the line graph consisted of the calculated y-values against the x-values. The curve equations were generated as strings from the optimal parameters outputted from the curve_fit function.

Visualizing the data

Our program uses Matplotlib to display results in a visual way, because it has “widgets” that allow a user to interact with graphs created by the program. When the `run_program` function is called, it displays several buttons in a new window. These Button objects from `Matplotlib.widgets` can be clicked on to bring up various graphs where sea levels are plotted along the y-axis and one of the three independent variables (CO_2 levels in the atmosphere, glacier mass balance, or global overall temperatures) is plotted along the x-axis. These are graphed using `plt.plot()`. The graphs display curve fits generated using functions described above for the `curve_fitting` module and displays their equations with functions from the `curve.equations` module. Additionally, there is a ‘General Statistical Analysis’ button that will display, in a separate window, the results of performing multiple linear regression, showing the final equation, as well as the coefficients and intercept. Furthermore, this window will show root mean squared error for each independent variable.

Instructions For Dataset Download

We used four libraries for our project: Pandas, Matplotlib, Scikit-learn, and SciPy.

To install `pandas` and `matplotlib`, you first have to install `numpy`. This is because `pandas` and `matplotlib` are built on `numpy`. To install `numpy`, click on ‘file’ in the top-left corner and go to settings. Click on ‘Project: doc file’ and you should see a drop-down menu. In the menu, click on ‘Python Interpreter’. In the bottom-left corner, you can click ‘+’ and search for ‘numpy’. Once you have found it, you can click it and install it. Make sure you install version 1.20.0rc1. You can specify the version by clicking on `numpy` and at the bottom clicking on the specify version box and scrolling through the options.

To install `pandas`, follow the same steps for installing `numpy` only this time searching for ‘pandas’. Make sure you install version 1.1.4.

In order to read excel files, `pandas` uses a library called ‘xlrd’. To install `xlrd`, follow the same steps for installing `numpy` only this time searching for ‘xlrd’. Make sure you install version 1.2.0.

To install `matplotlib`, follow the same steps for installing `numpy` only this time searching for ‘matplotlib’. Make sure you install version 3.3.3.

To install `scipy`, follow the same steps for installing `numpy` only this time searching for ‘scipy’. Make sure you install version 1.5.4.

To install `scikit-learn`, follow the same steps for installing `numpy`, but this time search ‘sklearn’.

All datasets can be downloaded via the links provided in the datasets section. These datasets should be saved in the same folder as where our code is.

After running `main.py`, the user must call the `run_program` function, passing in as arguments the names of the dataset files. The dataset containing data on CO_2 levels must have somewhere in its name ‘ CO_2 ’ (case-sensitive). The user can then expect a new window to appear, displaying a panel of buttons, each labelled with the type of function that will be graphed as a curve fit (exponential, quadratic, cubic, etc.) in a new window when the user clicks on it. Similarly, clicking on the ‘General Statistical Analysis’ button will bring up a new window showing text with the results of the multiple linear regression and root mean squared error.

Note: On the interactive page with the buttons, for the statistical analysis button, it is required for you to make the screen full screen to view all the text. If you close this screen, you should click another button, close the screen, and reopen the statistical analysis in order to view it again.

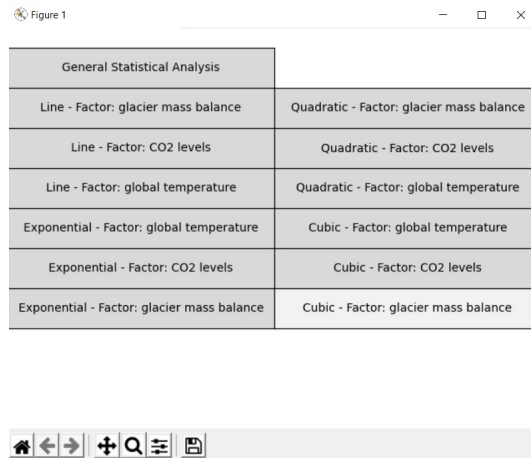


Figure 1: The expected panel of buttons

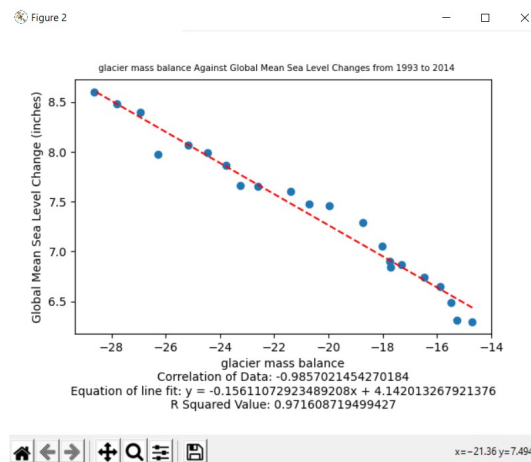


Figure 2: One of the possible graphs produced

Discussion of Changes

We made a few changes from our original project proposal. We decided to forgo the idea of looking into how the polar bear population is affected by the reduction of sea ice due to climate change. This is because we could not find appropriate datasets quantifying the polar bear population. Instead, we decided to look at how sea level is related to the global temperature and glacier mass. There were many datasets on these factors that were appropriate for our use. Additionally, we added more complexity to the mathematical models that we were originally going to implement. We decided to use more statistical models such as multiple linear regression while keeping the main idea of having multiple independent variables and a single dependent variable.

Discussion

By looking at our mathematics models and graphs we can determine how global mean sea levels are changing with respect to changes in atmospheric CO₂ concentrations, mean cumulative mass balance of glaciers and global mean temperatures.

In our multiple linear regression model, since our r-squared value is close to 1, it indicates that most of the variability in global mean sea level change (y) is explained by the regression model.

For our graph showing the relationship between global mean sea level changes and glacier mass balance, since the correlation is very close to -1, it indicates an almost perfect linear relationship. Additionally, since it is negative, this suggests a negative linear association. This means the values of global mean sea level change tends to decrease as glacier mass balance decreases. Since our r-squared value is close to 1, it indicates that most of the variability in

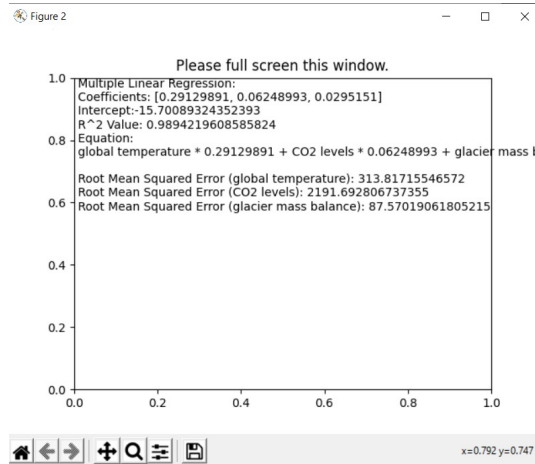


Figure 3: What the statistical analysis window should look like

global mean sea level change (y) is explained by the regression model.

For our graph showing the relationship between global mean sea level changes and CO_2 levels, since the correlation is very close to positive 1, it indicates an almost perfect linear and positive relationship. This means the values of global mean sea level change tends to increase as CO_2 levels increase. Since our r -squared value is close to 1, it indicates that most of the variability in global mean sea level change (y) is explained by the regression model. The positive slope also indicates that the association is positive.

For our graph showing the relationship between global mean sea level changes and global temperatures has a moderate to weak, positive roughly linear association. The correlation is around 0.78, indicating a positive association between the variables, though not as linear as our other graphs. Since our r -squared value is relatively close to 1, it indicates that a significant amount of the variability in sea level change (y) is explained by the regression model.

The RMSE (Root Mean Squared Error) values are large for all of our graphs, which suggests that our regression models are not very accurate. This may be because we did not have a lot of data to work with. Statistical models perform better when more data is used to fit them. The quality and accuracy of our datasets is also a factor, since we used such data to fit our models. A model is only as good as the data it was trained on.

The curve fitting helped us answer our question by allowing us to graph several types of functions and quantitatively see which functions best suited the data. For instance, we saw how the linear and cubic functions fitted sea level changes against glacier mass balance and CO_2 concentrations very well, whereas the quadratic and exponential functions did less so. This shows us that, presently, sea levels are increasing at steady rates against the independent variables with minor fluctuations. We do not know for certain if these variables will maintain this trend in the long-term, though. Through observations, we have established which functions best fit each scatter plot. Therefore, we can describe the change of sea levels with respect to each independent variable in terms of those functions.

Regarding limitations, we were only able to use a subset of the data (1993 to 2014), because these were the only overlapping years between all four of our datasets. If we had used a larger range we would have had more data to work with, which would contribute to the accuracy and quality of our models, since statistical models perform better when more data is used to fit them. This may have reduced our RMSE values significantly. Additionally, our curves of best fit often looked very similar to straight lines, which would suggest a linear relation between global sea levels and each of our independent variables, but we may be misunderstanding the true trends and relationships between these variables because we didn't have enough data to get the full picture of the trend. (For example, if one were to look at a very small section of an exponential function, it might look like a straight line.)

Some next steps might be to add extrapolating and predicting future results. Additionally, we could implement functions to statistically analyze the curves of best fit in a quantitative way.

Regardless of which factors we plotted, sea levels changed in the expected ways. A more negative glacier mass balance corresponded with higher sea levels, which makes sense as negative mass balance for a glacier suggests that it is melting more than it is able to recover from each year. Higher CO_2 levels mean that heat is trapped by the greenhouse gases in the atmosphere, leading to higher sea levels as the Earth heats up. Higher global temperatures lead to rising sea levels as the water expands. All of our graphs and analysis show these things to be true, and the factors we used are directly affected by climate change. We can thus conclude that rising sea levels are related to increased CO_2 , melting glaciers, and increasing global temperatures in a linear way.

References

- Amadeo, K. (2020, April 8). *Rising sea level effects, projections, and solutions*. The Balance. <https://www.thebalance.com/sea-level-rise-and-climate-change-4158037>
- APMonitor.com. (2019, May 2). *Python Nonlinear Regression Curve Fit..* <https://www.youtube.com/watch?v=4vryPwLtjIY>
- Brownlee, J. (2020, November 4). *Curve Fitting with Python*. Machine Learning Mastery. <https://machinelearningmastery.com/curve-fitting-with-python/>
- Daityari, S. (2019, July 10). *How to Plot Charts in Python with Matplotlib*. sitepoint. <https://www.sitepoint.com/plot-charts-python-matplotlib/>
- Datahub. (2015). *Average cumulative mass balance of reference Glaciers worldwide* https://datahub.io/core/glacier-mass-balance#resource-glacier-mass-balance_zip
- CO₂.earth. (2020, November 11). *Monthly CO₂* [https://www.CO₂.earth/monthly-CO₂](https://www.CO2.earth/monthly-CO2)
- Hennig, C. (2019, April 3). *Canada's sea levels are rising — and that's got British Columbians in coastal cities concerned*. CBC. <https://www.cbc.ca/news/canada/british-columbia/climate-change-report-bc-coastal-cities-1.5083449>
- Hunter, J. D. (2007). Matplotlib: A 2D graphic environment. *Computing in Science & Engineering*, 9(3), 90-95. 10.1109/MCSE.2007.55
- Jet Propulsion Laboratory. (2020, August 21). *NASA-led study reveals the causes of sea level rise since 1900*. NASA. <https://www.jpl.nasa.gov/news/news.php?feature=7731>
- Kim, E. (2019, November 18). *Multiple Linear Regression and Visualization in Python*. Pythonic Excursions. https://aegis4048.github.io/multiple_linear_regression_and_visualization_in_python
- Lindsey, R. (2020, August 14). *Climate change: Atmospheric carbon dioxide*. Climate.gov. <https://www.climate.gov/news-features/understanding-climate/climate-change-atmospheric-carbon-dioxide>
- Mckinney, W. (2010). Data structures for statistical computing in Python. *Proceedings of the 9th Python in Science Conference*, 445, 56-61. 10.25080/Majora-92bf1922-00a
- Mester, T. (2020, June 11). *Pandas tutorial 5: Scatter plot with pandas and matplotlib*. data36. <https://data36.com/scatter-plot-pandas-matplotlib/>
- Migiliozzo, B. (2015, June 25). *What's Really Warming the World?* Bloomberg. <https://www.bloomberg.com/graphics/2015-whats-warming-the-world/>
- Nunez, C. (2019, February 19). *Sea level rise, explained*. National Geographic. <https://www.nationalgeographic.com/environment/global-warming/sea-level-rise/>
- Python Tutorial. (n.d). *Read CSV with Pandas: How to excel file using pandas*. <https://pythonbasics.org/read-csv-with-pandas/>
- Yusof, N.. (2020, November 2). *Introduction to Python 3: How to excel file using pandas*. Youtube. <https://www.youtube.com/watch?v=CVjd5T08DjI>