CSC110 Project Report - Predicting Temperature, Emission, and Deforestation

Hongshou Ge, Jenci Wei, Zheng Luan

December 2020

1 Introduction

Having a glimpse of the current situation of the environment in the world, it is obvious that global warming has become a new threat to mankind. From the shifting weather patterns that threaten food production to the rising sea levels that increase the risk of catastrophic flooding, the impacts of climate change are global in scope and unprecedented in scale. One of the most serious problems of climate change is global warming. After more than a century and a half of industrialization, deforestation, and large-scale agriculture, quantities of greenhouse gases in the atmosphere have risen to record levels not seen in three million years; the concentration of greenhouse gas in the earth's atmosphere is causing the increase in the average global temperature. According to the Fifth Assessment Report from The UN Intergovernmental Panel on Climate Change (IPCC), the average global temperature has increased by 0.85°C from 1880 to 2012 [1]; and in its conclusion, the report claims that climate change is real and human activities are its main cause. Nevertheless, many people still argue for the nonexistence of climate change and treat it as if it is a false alarm created to assert control over the general public. Einstein once said, "Two things are infinite: the universe and human stupidity: and I'm not sure about the universe." Leaving the topic of climate change and global warming as an elephant in the room is an act of ignorance and cowardice. Our group aimed to unveil parts of the chain reaction that ultimately leads to climate change by analyzing real world datasets. **Our question in mind when doing this project is "to what extent do greenhouse gas emission and deforestation contribute to climate change?"** For our project, we found the data of greenhouse gas emission and deforestation and calculated their trends of change; then, we explored the correlation between greenhouse gas emission and deforestation versus the temperature. Since both emission and deforestation are caused by human activity, our research led to the answer to our question. We also developed a prediction of all the variables we studied based on the analysis of the data we collected.

2 Datasets Used

2.1 Adjusted and Homogenized Canadian Climate Data [3]

This dataset contains much detail about the climate in Canada, which includes data measured by all climate stations in Canada for each month. This dataset is retrieved from Environment and Natural Resources Canada, in CSV format. We used columns 9, 10, 11 (or equivalently, columns 8, 9, 10 when counting from 0), where the 9th column is the province, the 10th column is the year and the month where the data is measured, and the 11th column is the temperature recorded in degrees Celsius.

2.2 National Greenhouse Gas Emissions [4]

This dataset contains the yearly emission of greenhouse gas in Canada from 1990 to 2018. This dataset is retrieved from Environment and Natural Resources Canada, in CSV format. We used the entirety of this dataset, which only consists of two columns. The first column is the year and the second column is greenhouse gas emission measured in megatonnes of carbon dioxide (CO_2) equivalent.

2.3 Estimated Area (Hectares) of Annual Deforestation in Canada, by Industrial Sector [5]

This dataset contains the yearly area of deforestation in Canada from 1990 to 2017. This dataset is retrieved from Natural Resources Canada in website format, which we manually converted into csv format. We used columns 1, 6, 7 (or equivalently, columns 0, 5, 6 when counting from 0), where the 1st column is the year, the 6th column is the deforestation area caused by hydroelectric measured in hectares, and the 7th column is the total deforestation area in the given year measured in hectares (i.e. the sum of columns 2, 3, 4, 5, 6).

3 Computational Overview

3.1 Preamble

The temperature mentioned below will be in degrees Celsius unless otherwise specified. The emission mentioned below will be in megatonnes of carbon dioxide (CO_2) equivalent unless otherwise specified. The deforestation area mentioned below will be in hectares unless otherwise specified. Besides the modules we submitted, we also wrote a Python module responsible for visualizing our data on a graph (which are purely for ourselves), and our insights about the graph trends as mentioned below all come from looking at the curves generated by that module.

3.2 Computations

3.2.1 Temperature

Before doing any computations, we created a dataclass named `Temperature`, with instance attributes storing the province, year, month, and the temperature value for each piece of temperature data. To read our dataset for temperature, we wrote a function `read_csv_temp` which converts our csv file to a list of `Temperature`, with a helper function `process_row_temp`, which identifies each column as the corresponding instance attribute of the `Temperature` dataclass. We put a filter in `read_csv_temp` so that it only takes temperature measured in August and September for a reason we would explain shortly. For each province of Canada, we would like to acquire the median temperature every year to represent the temperature of a given year for each province. If we accounted all months (which is what we initially planned to do), the temperature trend we obtain would be inconsistent as it is very likely for there to be more data points in a particular season that shifts the median. We wrote a function `get_yearly_median_temp` that takes a list of `Temperature` (which is all the temperature data for a particular province) and returns a dictionary mapping of year to its corresponding median temperature. We then took the average of the median temperature for all provinces to acquire the representative temperature of Canada (of August and September) for each year, which is the variable `CANADA_MEDIAN` in the main file.

3.2.2 Emission

To read our dataset for emission, we wrote a function `read_csv_emission` which converts our csv file to a dictionary mapping of year to emission. Looking at the trend, we chose to perform a logarithmic regression on our emission data to be able to predict the emission value for any given year. To do so, we utilized the library `scipy` [7] and wrote a function `model_emission` that takes a dictionary mapping of year to emission and returns a tuple of three floats corresponding to parameters a, b, c in equation $y = a\ln(x - b) + c$ where y is the emission in a given year x.

3.2.3 Deforestation

To read our dataset for deforestation, we wrote a function `read_csv_deforestation` which converts our csv file to a dictionary mapping of year to deforestation. We found that there are spikes on the graph of deforestation vs. year, and identified the cause to be the hydroelectric value. On the webpage where our deforestation data is retrieved, it says that "spikes in deforestation occur when large forest areas are flooded during hydroelectric reservoir development" [5]. We then wrote another function `read_csv_deforestation_hydro` that does the same job as `read_csv_deforestation` except that it only takes the deforestation data caused by hydroelectric. After looking at the difference of the two curves, which is the deforestation data without the hydroelectric factor, we chose to perform a reciprocal function regression on our deforestation data to be able to predict the deforestation value for any given year. To do so, we utilized the library `scipy` and wrote a function `model_deforestation` that takes a dictionary mapping of year to deforestation (with the hydroelectric factor removed) and returns a tuple of three floats corresponding to parameters a, b, c in equation $y = \frac{a}{x-b} + c$ where y is the deforestation in a given year x.

3.2.4 Temperature Prediction

Our question in mind when doing this project is "to what extend do greenhouse gas emission and deforestation contribute to climate change?" Since we can already predict the emission and deforestation for any given year at this point, we want to explore the correlation between temperature (as the dependent variable) and emission and deforestation (both as independent variables) so that given the emission and deforestation values, we can obtain the temperature. We decided to perform a multilinear regression on the change of temperature (the following year comparing to the current year) using emission and deforestation. We utilized `scipy` and wrote a function `model_correlation` that takes a tuple of three lists (list of change of temperature, list of emission, and list of deforestation where the indexes of all the lists correspond to ascending year values starting from 1991 all the way to 2017) and returns a tuple of five floats corresponding to parameters a, b, c, d, e in equation $y = a(x_1 - b) + c(x_2 - d) + e$ where y is the temperature with given emission value x_1 and deforestation value x_2.

3.3 Interactive Part

3.3.1 Idea

For our interactive part of this project, we decided to utilize `pygame` [2] to make a game-like simulation that predicts the emission, deforestation, and temperature value of the following year every time the user presses the 'space' key. We made a class `TemperatureGame` for our game where all the instance attributes and methods mentioned below belong to this class. Additionally, the word 'class' used below refer to `TemperatureGame` unless otherwise specified.

3.3.2 Initializing the Class

For our class, we have six instance attributes. The three that do not have to be manually inputted are `emission`, `deforestation`, and `temperature`; all three are dictionary mappings of year to their corresponding value. Their purpose is to keep track of the values the game predicted for each year. The three instance attributes that do need to be manually inputted are `emission_predict`, `deforestation_predict`, and `correlation`, which take the output of `model_emission`, `model_deforestation`, and `model_correlation`, respectively. The game defaults to start at the year 2020 (which cannot be changed by the user), however, the initializer of the class also take the desired starting temperature of this game, which we set to 14 looking at the trend of Canada's temperature in August and September.

3.3.3 Prediction

We wrote three methods responsible for predicting the values of emission, deforestation, and temperature, which are `predict_emission`, `predict_deforestation`, and `predict_temperature`, respectively. The first two takes the current year as the input and predicts the emission and deforestation values for the following year using the formulas mentioned in Section 3.2.2 and Section 3.2.3. Looking at how well the 'best-fit curves' fit the actual data, we decided to add a random value when predicting our emission and deforestation values since real-life data cannot be perfectly smooth. `predict_temperature` takes the predicted emission and deforestation for next year and the temperature of the current year as input, and predict the temperature of the following year using the formula mentioned in Section 3.2.4. We wrote a method `predict_display` that takes the `pygame.Surface`, the updated year value, the emission, deforestation, and temperature values corresponding to the following year as input and display them on the `pygame.Surface`.

3.3.4 Game

Our method for running the simulation game is `run`. It is responsible for creating the 'screen', displaying the background, displaying all the text, calling other methods, and storing game data. Whenever the 'space' key is pressed, time will advance by a year and the new emission, deforestation, and temperature values would be displayed. Since the deforestation value spikes sometimes due to the hydroelectric factor, this game features a random number generator that occasionally increases the deforestation area for one year to account for forest areas flooded during the hydroelectric reservoir development for that particular year.

3.3.5 Game Statistics

When the user closes the game, the Python console will prompt the user for whether they want a statistical graph for the game. If so, the `print_graph` method is called. `print_graph` takes the predicted emission, deforestation, and temperature data each year as generated by the simulation game (which are stored whenever the prediction of values for the new year is displayed while the user is playing the game) and generates three graphs for each variable plotted with respect to time using `plotly` [6].

3.4 New Libraries Used

Our project used `scipy.optimize.curve_fit` [7] to perform various regressions. The function `model_emission` used `curve_fit` to perform logarithmic regression on emission vs. time. The function `model_deforestation` used `curve_fit` to perform reciprocal function regression on deforestation vs. time. The function `model_correlation` used `curve_fit` to perform multilinear regression on temperature vs. emission and deforestation. The interactive part of our project relied on `pygame` [2], which is crucial to our simulation game. The class for our simulation game `TemperatureGame` extensively uses `pygame`. `plotly` [6] is also used to graph the game statistics after one finishes playing the game.

4 Program Instructions

4.1 Setup

Before running our program, the user should have Python 3.8 and the PyCharm IDE installed. To obtain the dataset needed to run our program, the user should go to

`https://tinyurl.com/csc110project`

Or, alternatively, the user should open 'files.txt' and follow the link - both lead to the same website. The user should download all three zip files. After unzipping the files, we ask the user to leave them as folders and put them on the same level as our python files. Additionally, we ask the user to locate 'requirement.txt' on PyCharm and follow the prompt to download the libraries the project uses.

4.2 Running the Program

To run our simulation game, the user should run the file named 'main.py'. It may take **up to 20 seconds** for the computer to perform computations on data, and then the pygame window would pop out:

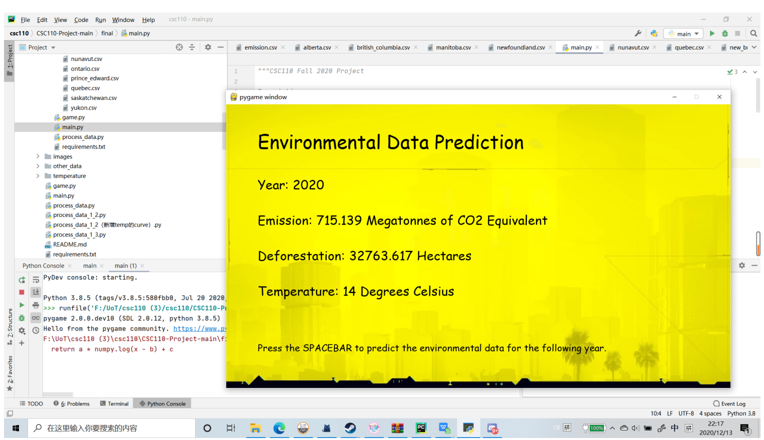

Figure 1: Pygame Window

To predict the environmental data for the following year, the user simply need to press the 'space' key. When the user decides to stop the simulation game, they must close the interface and return to the Python console. If the user would like to see a statistical graph of their game, simply enter 'y' into the Python console after closing the pygame interface. **To be able to see the trend we intended to show, we recommend predicting to the year 2100.** Furthermore, due to our added randomness, different games may be slightly different from each other in terms of the predicted results.

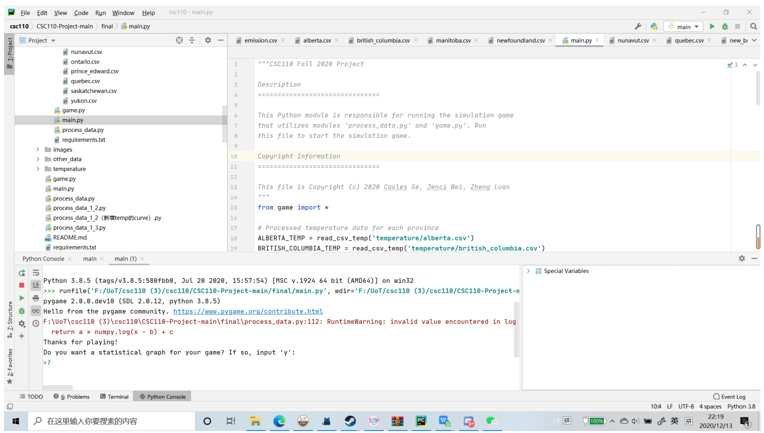

Figure 2: Graph Prompt

Figure 3: Graph

5 Changed Made

5.1 Temperature Data Collected

In our proposal, we were solely thinking about the temperature of Ontario. However, thanks to our TA, we realized that Ontario alone does not reflect the entirety of Canada, which is what our emission and deforestation data are based on. Therefore, we collected temperature data from all 13 provinces of Canada, which may lead to us producing a more accurate correlation between temperature and emission/deforestation.

5.2 Restricting Temperature to Specific Months

In our proposal, our group planned to use the median of annual temperature. Then, we realized that if there are more data points in a particular season, the median would be shifted. To avoid such inaccuracy, we only took the temperature measured in August and September.

5.3 Pygame Interaction

In the original plan, we intended to only find the correlation of the data we are studying. After some discussion within group members, we decided to make our project interactive. Rather than purely presenting the correlation, we want the user to explore such correlation in an interactive way. In our project, we utilized `pygame` which predicts the environmental data for the following year every time the user presses the 'space' bar. Through this, the user has the freedom to choose how far to predict into the future.

5.4 Visualizing Game Data

After the user decides to stop the simulation game and closes the pygame window, the user may ask the game to generate a graph of the game's prediction. This is similar to our original plan, which shows the graph of various data vs. time. This helps our user to understand the impacts of human activity on climate change and shows the consequences of clinging onto the status quo about emission and deforestation.

5.5 Adding Randomness

5.5.1 Emission and Deforestation

Though we have the formula to predict the deforestation (without the impact of hydroelectric, refer to the next section) and the emission for any given year, we acknowledge that the real world does not follow smooth curves since there is a myriad of underlying factors behind any small change in data. Therefore, we assigned randomness in our predictions of emission and deforestation. For simplicity, we used `random.uniform` with a logical range based on how well the curve fits the actual data instead of diving deep into things like covariance.

5.5.2 The Hydroelectric Factor

The dataset we used for deforestation data claims that deforestation area spikes occasionally when large forest areas are flooded during hydroelectric reservoir development in a particular year. However, we cannot tell the frequency of such event happening. Thus, we made a random number generator that occasionally increases the deforestation value for a particular year on top of our predicted value.

5.6 Change of Temperature

We initially decided to predict the absolute temperature given the emission and deforestation values, however, such approach is very naive (i.e. the temperature would not drastically decrease if all the industries suddenly shut down). After some discussion within group members, we decided to predict the change of temperature given the emission and deforestation values and add it to the temperature value of the previous year to obtain our predicted temperature for the following year.

6 Discussion

6.1 Back To Our Question

Our question in mind when doing this project is “to what extent do greenhouse gas emission and deforestation contribute to climate change?”. This question may generalize to “to what extent does human activity contribute to climate change?”. After doing the project, we have an answer to this question. Before the industrial revolution, climate change was never a concern for humans. After the industrial revolution, two things increased: emission and deforestation. Curiously, so did the temperature. Despite the fact that intuition tells us that temperature is correlated with emission and deforestation, there is another voice: the climate change doubters. In our century full of information, it is difficult to identify which one the truth is, but one thing that we may take for granted is that datasets never lie. All the datasets we used in our project comes from Natural Resources Canada, whose job is to record those data without making arbitrary changes; and we do trust them. From such data, we found a positive correlation between temperature and emission/deforestation. In our simulation game interface, we chose to use the Cyberpunk 2077 background as it represents a dystopian future, which coincides with the pessimism of our prediction. By our simulation game, Canada’s median temperature in August and September would go up by 0.5 degrees Celsius in 100 years, and the temperature with respect to year would rise faster and faster, eventually turning our habitat into a giant oven. We acknowledge that our prediction is not extremely accurate, but it does tell us something about our world: given that emission and deforestation follow the same trend in the future, it is inevitable for the temperature to increase. Therefore, we arrive at the conclusion that greenhouse gas emission and deforestation are unarguably factors of global warming, answering our initial question. Since emission and deforestation are subsets of "human activity", we may also say that human activity plays an important role in global warming.

6.2 Obstacles We Encountered

One major obstacle we encountered while doing this project is the lack of knowledge in the area of statistics. Our knowledge of statistics is based on high-school level math, which only allows us to perform regressions with precalculus level functions (such as logarithmic regression). We believe that there are more accurate ways of finding trends, but we are not on that level yet. Additionally, to make our predictions more "real", we arbitrarily added randomness on top of the value outputted by our prediction functions. Despite the fact that the range of our added randomness is based on our investigation of how well the function curve fits the existing data, it is not statistically satisfactory. We see that the curve fitting library we used has the feature of calculating the covariance, but we have no idea what such thing is. With more knowledge in the area of statistics, we would be able to model our data with greater accuracy and make conclusions about the fluctuations based on mathematics.

6.3 Next Steps

In a few years from now, we would be able to look back and see how accurate our predictions are. After we gain more knowledge in the area of statistics in the next few years, we can come back to this project and redesign our algorithm to make our predictions more accurate. Whenever we have free time, we could try to add a third independent variable to our prediction (alongside with emission and deforestation) to find out its relationship with temperature.

6.4 Conclusion

We have successfully answered our question posed before we began our project and sadly proved that climate change is real with real world datasets. If we do not do anything to control our emission and deforestation, the temperature would slowly rise and it would eventually be too late (perhaps in a few generations) for us to turn back. Throughout the process of completing this project, what we learned about Python in this course was applied. We also explored and used libraries outside the scope of this course, such as `scipy`. Though the lack of statistical knowledge reduced the accuracy of our prediction, the trend of temperature as predicted by our program coincides with our intuition. We have shown that emission and deforestation are factors of climate change in our project, and our simulation game may help its users to visualize the foreseeable consequences of emission and deforestation.

References

[1] Intergovermental Panel on Climate Change (IPCC). "Climate Change 2014 Synthesis Report". In: (2014), p. 57. URL: https://archive.ipcc.ch/pdf/assessment-report/ar5/syr/SYR_AR5_FINAL_full_wcover.pdf.

[2] Pygame Community. *Pygame*. https://pygame.org.

[3] Environmental and Natural Resources Canada. *Adjusted and Homogenized Canadian Climate Data (AHCCD)*. https://climate-change.canada.ca/climate-data/#/adjusted-station-data.

[4] Environmental and Natural Resources Canada. *Greenhouse gas emissions*. https://www.canada.ca/en/environment-climate-change/services/environmental-indicators/greenhouse-gas-emissions.html.

[5] Environmental and Natural Resources Canada. *Indicator: Deforestation and afforestation*. https://www.nrcan.gc.ca/our-natural-resources/forests-forestry/state-canadas-forests-report/how-much-forest-does-canada-have/indicator-deforestation-afforestation/16546.

[6] Plotly Technologies Inc. *Collaborative data science*. 2015. URL: https://plot.ly.

[7] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python". In: *Nature Methods* 17 (2020), pp. 261–272. DOI: 10.1038/s41592-019-0686-2.