Team Control Number

13010

Problem Chosen

B

2022
HiMCM/MidMCM
Summary Sheet

To estimate the patterns in global warming over the last sixty years, our team developed a method that employs statistics and computational approaches to simulate the supplied data on CO2 levels and temperatures for the past and future. We began by evaluating the data with a Python program to calculate the CO2 ppm step sizes for each year. Using this tool, we disproved the notion that the greatest rise in CO2 ppm occurred not from 1994 to 2004, but rather from 1993 to 2003.

We were able to utilize numerous mathematical models to assist model the supplied data and even anticipate what the CO2 levels will be in 2050 and 2100 by looking at this step size and examining several mathematical associations. Using these mathematical associations we were able to use several plotting functions and model our data using various relationships for time with respect to CO2 concentrations.

The first step of the actual model building for both the CO2 concentrations and the temperature levels is by plotting the data that we were provided with in the problem and then using various fits to find which plot had the highest correlation with our data. In addition, we had to complete some additional research outside the problem that allowed us to determine that there was a baseline value before the rise in CO2 and that there should be some point in time when the concentrations stop increasing. While the CO2 has to stop increasing at some point, that point is not modeled in our graph for we have no data on when the human population will reduce CO2 emissions. Based on the research above, we decided that the natural exponent model that we made for the data is most reasonable according to the parameters of the information that our research yielded us and since it provided us with a huge correlation with our existing data to help model additional data points.

In addition to concluding that there is a strong correlation and maybe even causation between the increase in CO2 concentrations and the increase in global temperature levels, we were able to model the data using several different graphs and even found that the claim made by the scientists about how the world will reach 685 ppm of CO2 is false.

Finally, we created a short article made for Scientific Today, representing our findings, along with further information on impact in a grander scope.

Table of Contents

Clarification/Restatement of Problem

Throughout history, the world has been experiencing global warming due to a rise in CO2 and temperature levels. Up until the Industrial Revolution, the CO2 levels in the atmosphere had been fluctuating around 280 parts per million (ppm). With this in mind, scientists have been logging the measurements annually in an attempt to use the data to accurately predict CO2 and temperature levels. To do just this we are tasked with creating models using the existing data and utilizing statistical techniques to create additional data points along various models to help determine what the concentration of CO2 will be in parts per million in specific years.

In addition to being tasked with predicting the rising CO2 levels, we are asked to find a statistical relationship, if any, between land-ocean temperatures and CO2 concentration levels. Such a relationship is key in determining the correlation between rising CO2 levels and the severity of global warming. Analyzing any change in CO2 levels over the years would lead us to determine the step size in the temperature levels over the same years and even allow us to predict future temperature readings based on CO2 concentrations.

As depicted in both tables, a trend can be seen that as the years progress, so do most of the concentrations of CO2 and the temperature levels. Looking at the trends we see in the CO2 concentrations, we are asked to determine the step size and compare the rate of change in the temperatures that are provided in the table to the base period of 1951 - 1980. After creating models for both data tables we are then asked to reflect on our models and test their sensitivity on how accurate they will be far into the future.

After all sorts of modeling and analysis, we then must write a non-technical article that describes our solution in full depth and any possible recommendations that we may have to ensure the future of the upcoming generations.

Assumptions

Assumption 1: CO_2 levels will eventually stop rising at some point in time.

Justification 1: At some point, the levels will reach a maximum because humans are the lead contributors to CO_2 deposition. At some point either due to technological advances or simply due to a decrease in the population the CO_2 levels will stop increasing and maybe even start dipping.

Assumption 2: The baseline of CO_2 levels before the data is given is around 280 parts per million (ppm).

Justification 2: Before the Industrial Revolution the only CO_2 that was produced by the human population was through aerobic respiration and fires. When the human population began mechanizing and undergoing technological developments, all the burning of the fossils fuels caused a hike in the

Assumption 3: There exists some correlation between CO_2 concentrations and temperature levels.

Justification: As shown by numerous reputable sources, the molecule of CO_2 can hold significant energy in its bonds which tends to absorb more energy and directly affect global temperature levels. Colloquially, carbon dioxide is known as one of the greenhouse gasses, which all exhibit similar absorption properties to CO_2.

Assumption 4: Temperatures around the globe will increase but at different rates.

Justification 4: Due to the greenhouse gas effect some regions might drastically feel a change in the climate but overall the heating due to global warming will affect those all around the world through air circulation. The areas that will feel the effect the most will be the landmasses and the polar ice caps such that they absorb the sun’s rays a lot more than the bodies of water and because the main CO_2 emission contributor, humans, live on land.

Assumption 5: Humans are the reason why CO_2 concentration levels are increasing each year.

Justification 5: Over the course of human history the only CO_2 emissions that were released were through bodily processes and environmental causes. When humans began mechanizing, the machines that were created burned the fossil fuels that were found in the Earth which then were released as greenhouse gasses such as CO_2.

Analysis of Problem/Methods

At first glance, the problem seems to be overwhelming, but upon closer analysis, it can be seen that it necessitates simple statistical models. All in all, the entire basis of this problem asks us to use various graphing techniques to model the existing data and determine future data points at the same time as concluding data sets.

Our approach to this problem is to graph the data that is given in as many models as we can and use the various fits to determine which has the highest correlation. We began answering the sub-problems in the order in which they arose such that the problems that we encountered, later on, will utilize the models that we create for the earlier sub-problems.

Using python libraries, various calculators, and numerous tools we were able to graph the given data and utilize various statistical techniques to determine if our models were statistically significant. Through several models, we were able to determine that the most statistically significant models were the exponential and logistic models such that they both help reasonably predict the rise of CO2 levels over the next fifty years.

To determine which graph we would ultimately use for the model, we completed some background research that allowed us to conclude whether to model any of the CO2 data that we were given. The research showed that before the Industrial Revolution the concentration of CO2 was fluctuating around 280 ppm and that eventually humans, the biggest CO2 contributor, would become either technologically advanced enough to decrease CO2 emissions or would start dying off in record numbers which would also decrease emissions.

Even with the data on the various concentrations and levels throughout the years, we simply don’t have enough information to determine when the CO2 levels will begin to decrease. So to determine which model is the most accurate, all should be analyzed with some of the same basic conditions such that they use the same data points and start at the baseline of 280 ppm. As more data points are added and more information is given in terms of outside factors and predictions on what the future may hold, a machine learning algorithm could be used in the future that will accurately model the data and propose when exactly the world will start experiencing a decrease in CO2 concentrations.

Strengths & Weaknesses

Strength:
One strength of the model is that it relies on basic calculations that can be done on the average calculator or Desmos with data coming from basic data from a preexisting reputable source. The model utilizes concepts of regression to develop the fitted lines of the data, something most graphing calculators are capable of.

Strength:
Another strength of the model includes the relatively easy visualization of the model through the regression line. The overall line also provides a holistic representation of the growth of CO2 and temperature far into the future. It allows us to view if CO2 levels are ever stagnant, how fast it's growing, without complex visualization techniques.

Strength:
The regression technique provides a very strong correlation coefficient, and is able to represent the data well.

Weakness:
One weakness of the exponential regression model is that it predicts that the data will progress in the same direction and method for the future data points. It's very unlikely that the CO2 levels will progress in exactly the same manner they did when the Industrial Revolution first began to this point. For one, top contributors of the CO2 levels had no care for the effect of their actions towards the environment. Though such is somewhat unfortunately still true today, it's at a much smaller scale than it once was in the beginning of human industrial development. On top of that, there are many other confounding factors that point to the fact that CO2 growth may not be as consistent and upward to be fitted nicely into a simple exponential equation.

Weakness
The model is only based off of only the past 53 years worth of data, even though the Industrial Revolution as early as the mid-19th century. Currently, the model is based on only the past 50 years, yet humans have had a large impact on CO2 levels for almost another century before. To sufficiently model the growth of CO2 and the relationship humanity has had on those levels over time, more data is simply needed. However, it is understandable that such data may not have been available as CO2 PPM levels were only measured starting from the mid-20th century.

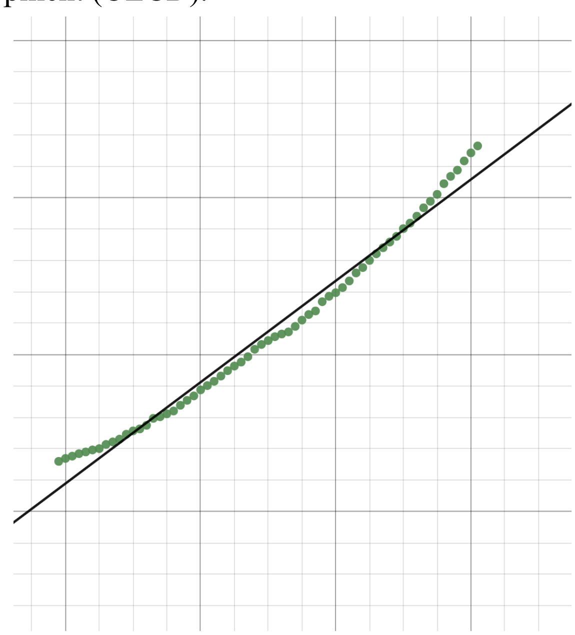
Model/Solution

Problem 1:

The claim that the March 2004 increase was the largest 10-year growth of CO_2 is false. In reality, the largest increase was from 1993 - 2003. Using the CO_2 Data Set 1, we iterated throughout all the 10 year periods before 1994 - 2004 to confirm this assumption. Specifically, 1994-2004 had a 18.74 ppm increase, while 1993 - 2003 had a 18.77 ppm increase. Thus, the larger increase did in fact occur prior to 2004 and makes the claim that the year had the largest increase ever seen before false.

Polynomial Based Modeling:

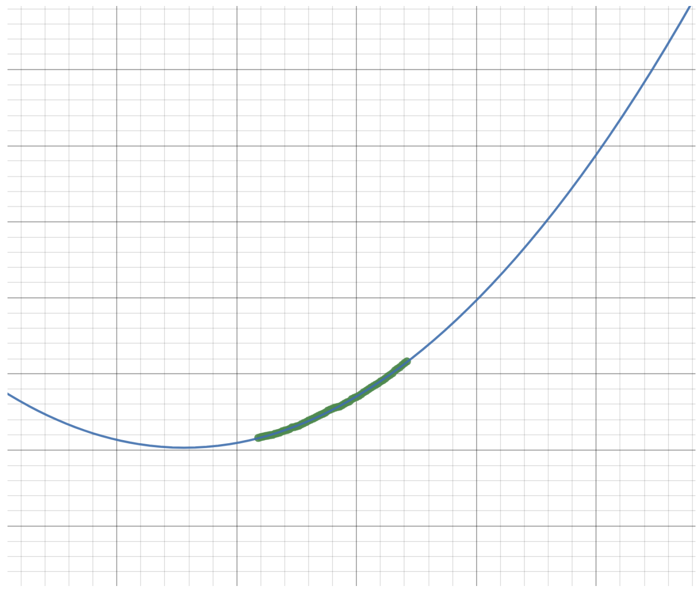
First, we attempted to model the data provided across the time series through various regression methods. On Desmos, we plotted the various data points onto a table and performed a variety of regression models. Such regression methods allowed us to visualize the data across a larger time spectrum. A particular regression type we attempted was a linear regression, as seen in Figure 1 above. It had an equation of y = 1.61404x - 2854.59 and a correlation coefficient of 0.9912. At first glance, the line seems to describe the data in an accurate method, but predicts a CO_2 level of 454 PPM in 205, contrary to the prediction of Organisation for Economic Co-Operations and Development (OECD).



x_1	y_1
1959	315.98
1960	316.91
1961	317.64
1962	318.45
1963	318.99
1964	319.62
1965	320.04
52 more rows Show all	
2018	408.72
2019	411.66

Following the linear, a very similar approach of regression was taken for other regression models. We increased the degree of the x-variable and attempted to model the quadratic regression of the data. Figure 2 represents the quadratic representation of the model below, across the same data points with an equation of $y = 0.013037636x^2 - 50.275753666x$

$+ \ 48771.436440167$



While the higher degree polynomial resulted in a higher precision, it comes with a drawback. As we continued to fit cubic and quartic polynomial models, the data was more accurately represented with a higher correlation coefficient model after model. But such precision resulted in a high volatility. However, all polynomial models created were extremely sensitive. If some data were slightly skewed, the predictions for future CO2 levels would be drastically different. Below shows a table of all the polynomial based modeling we conducted.

Polynomial Based Modeling

Regression Type	Equation	Correlation Coefficient	2050 Prediction	2100 Prediction
Linear	$y = 1.61404x - 2854.59$	0.9912	454 PPM	534 PPM
Quadratic	$y = 0.013037636x^2 - 50.275753666x + 48771.436440167$	0.9997	496 PPM	688 PPM
Cubic	$y = 0.0000412427967x^3 - 0.233181861x^2 + 439.676504254x - 276197.6723$	0.9997	504 PPM	740 PPM

Exponential Based Modeling

Next, we explored a more reasonable exponential approach. Because the increase of carbon emissions is directly linked to human activity, it is reasonable to infer that CO_2 levels are increasing at the same rate as human activity, which is an exponential function. Below are the various exponential fit models we explored.

Regression Type	Equation
Base-10	$y = 36.5404664185937(10^{0.0091970421451999695(x-1959)}) + 280$
Natural	$y = 36.5404657418149e^{0.0211769725580355(x-1959)} + 280$
Arbitrary	$y = 36.54046401713189(1.021402796991871 3^{(x-1959)}) + 280$
Logistic	$y = \frac{413.1615042427866}{1+10.955187676566421e^{-0.026660690147051216(x-1958)}} + 280$

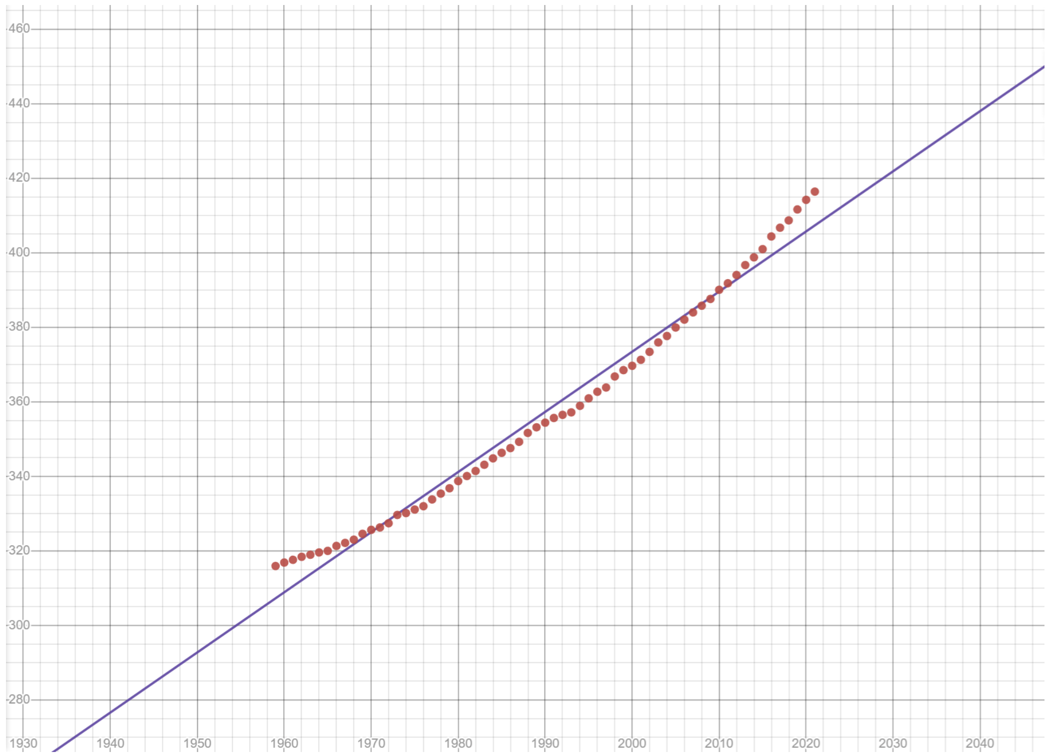
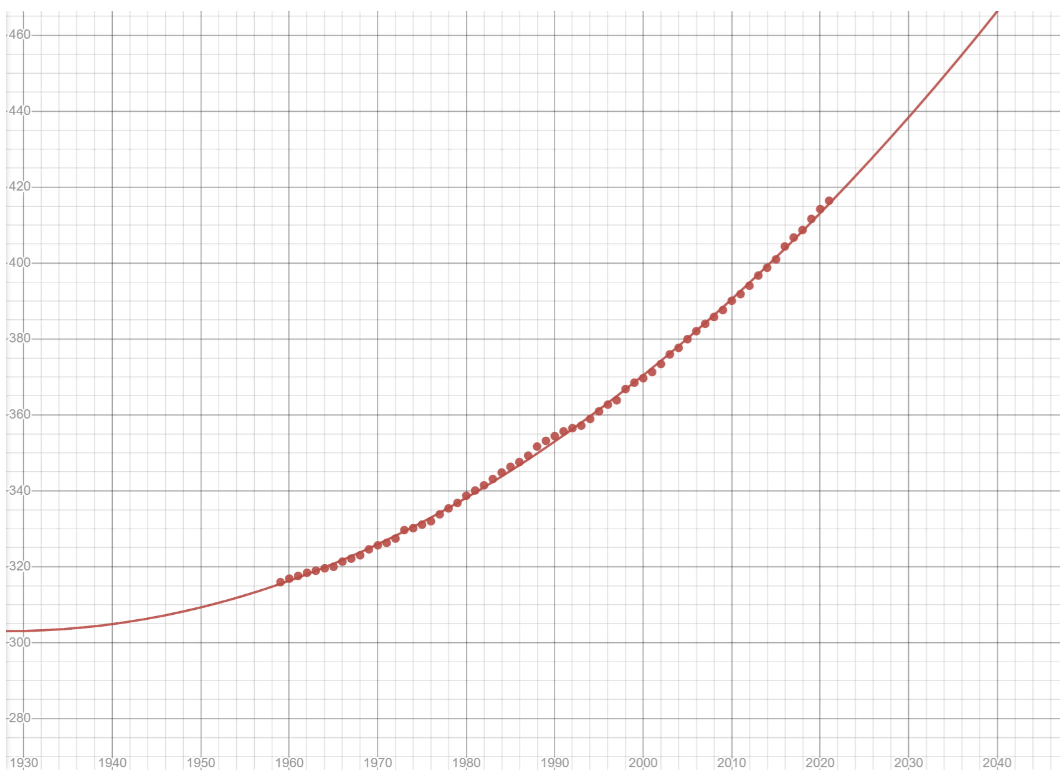
Similarly, solving for each correlation with the given data set, along with finding predictions for 2050 and 2100:

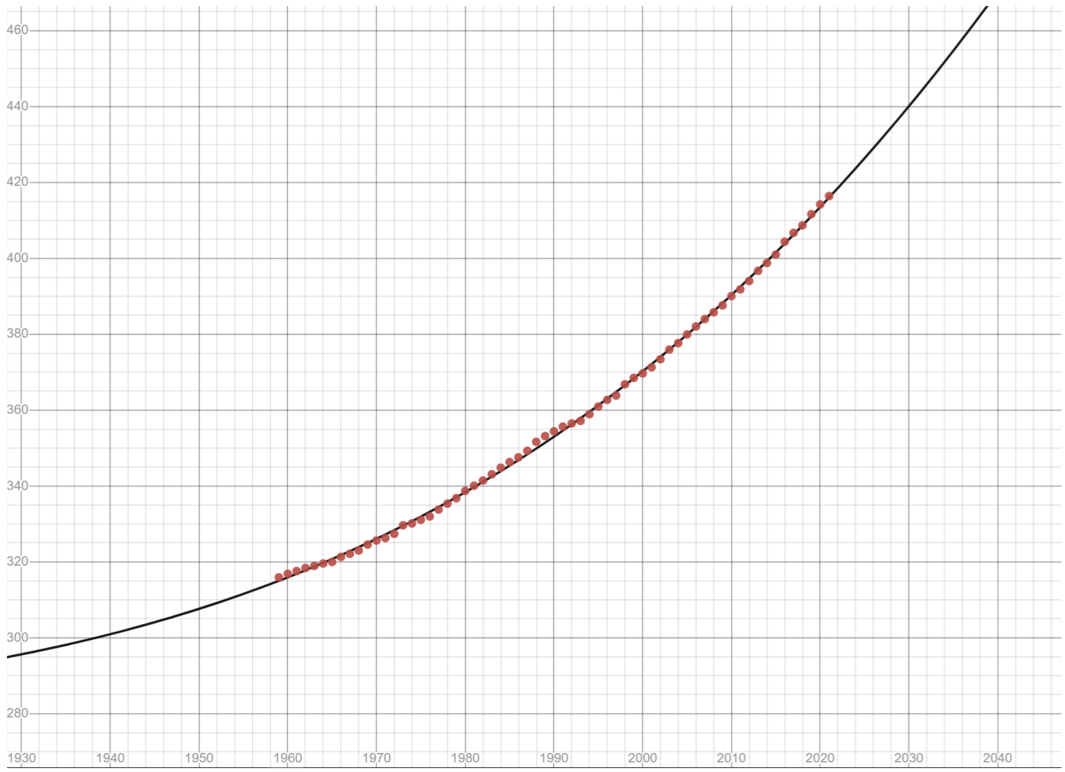
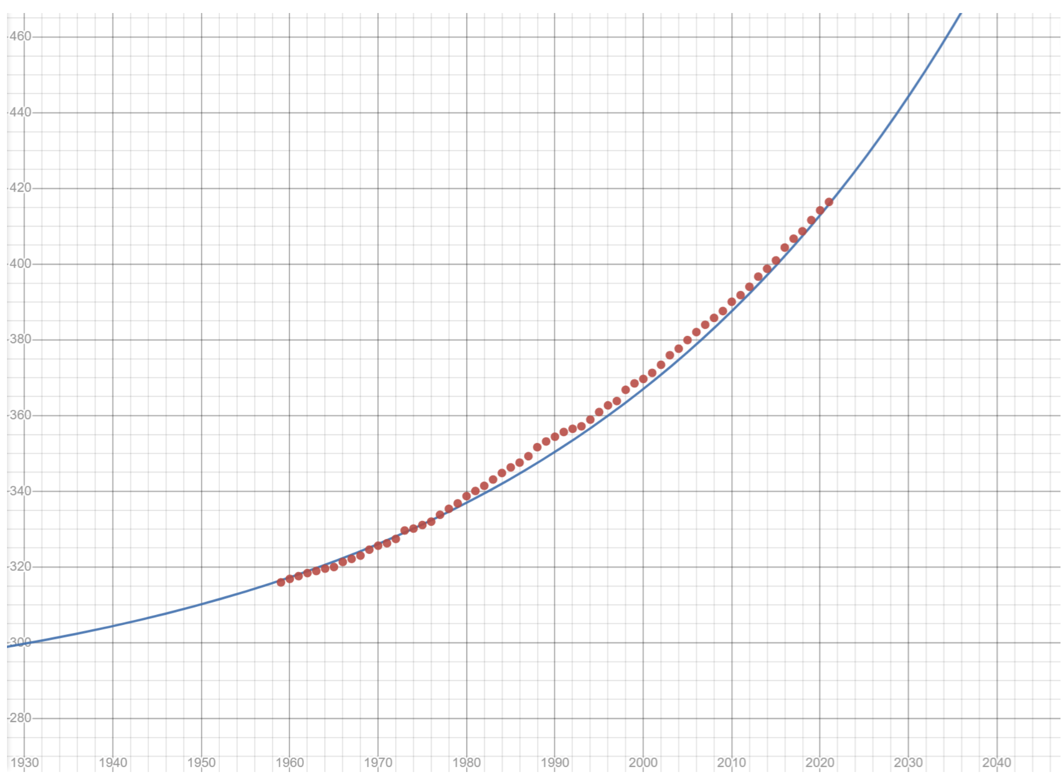
Regression Type	Correlation Coefficient	2050 Prediction	2100 Prediction
Base-10	0.9989	531.018	1003.697
Natural	0.9989	531.018	1003.697
Arbitrary	0.9989	531.018	1003.698
Logistic	0.9997	492.67	610.906

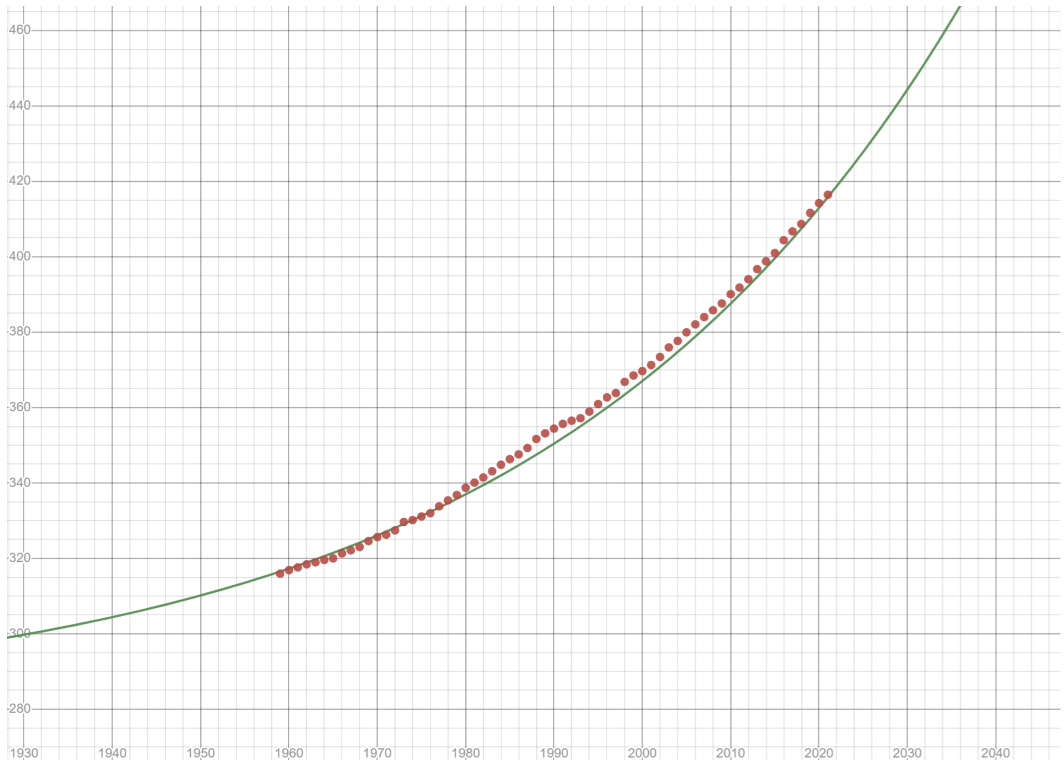
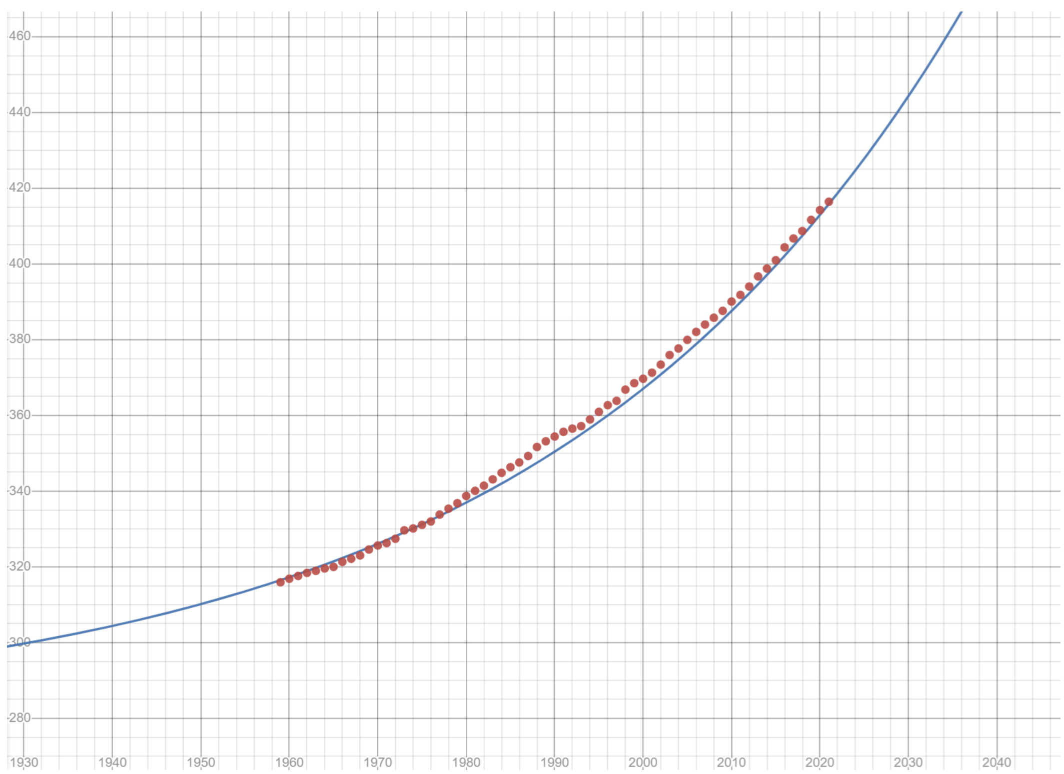
None of our models predicted that carbon dioxide density in the atmosphere would reach 685 PPM by 2050. For each of our models, we predicted the carbon dioxide levels to reach 685 by:

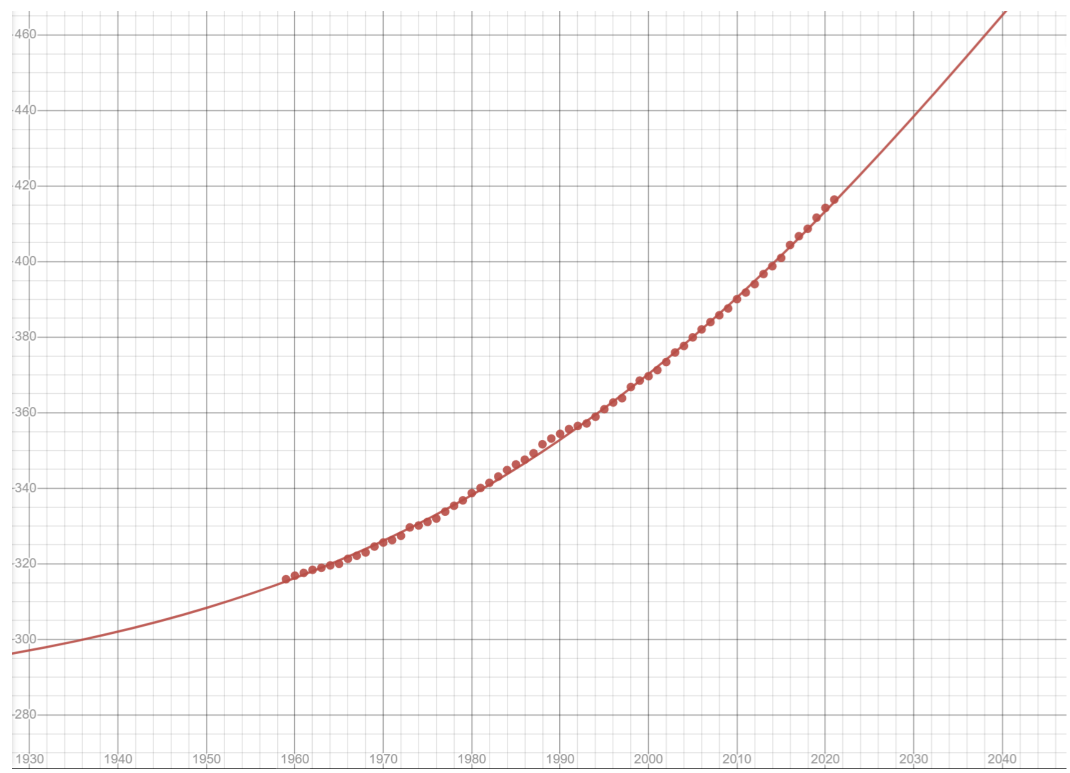
Model	**Year when PPM reaches 685**
Linear	2193
Quadratic	2099
Cubic	2090
Base-10	2072
Natural	2072
Arbitrary	2072
Logistic	2194

Below are the graphic representations of our models fitted against our given data. The x-axis represents the year, and the y-axis represents the PPM of CO_2.

Model	**Graph**
Linear	
Quadratic	

Cubic	
Base-10	

Natural	
Arbitrary	

Logistic	

After looking at the various models we have created, eventually the natural exponent model was chosen, as not only did it represent our data set well, with a relatively high correlation, it also represents a reasonable trend with a 280 PPM baseline, which was what the CO2 levels in the atmosphere fluctuated around before the industrial revolution. Although the logistic curve fit model is reasonable, the value it caps out at, roughly 695 PPM, is not reasonable in the real world. Polynomial models were ruled out initially, and after looking at our model from a holistic approach, we find that the natural exponential model best represents the CO2 increase over time.

Problem 2:

We perform a similar approach as Problem 1 for modeling temperature, and perform curve fits to our data, given a variety of equations.

Regression Type	**Equation**
Linear	$y = 0.016462x - 32.4046$
Quadratic	$y = 0.000174304862x^2 - 0.677097049x + 657.453799$

Cubic	$y = -0.00000336292452x^3 + 0.0202459198x^2 - 40.6075101x + 27135.2338$
Base-10	$y = 1.58287434 * 10^{(0.024183294x - 49)}$
Natural	$y = 3.42688588 * 10^{-12}(e^{0.0133710428x}) - 0.915823587$
Arbitrary	$y = 6.476 * 10^{-40}(1.0457^x)$

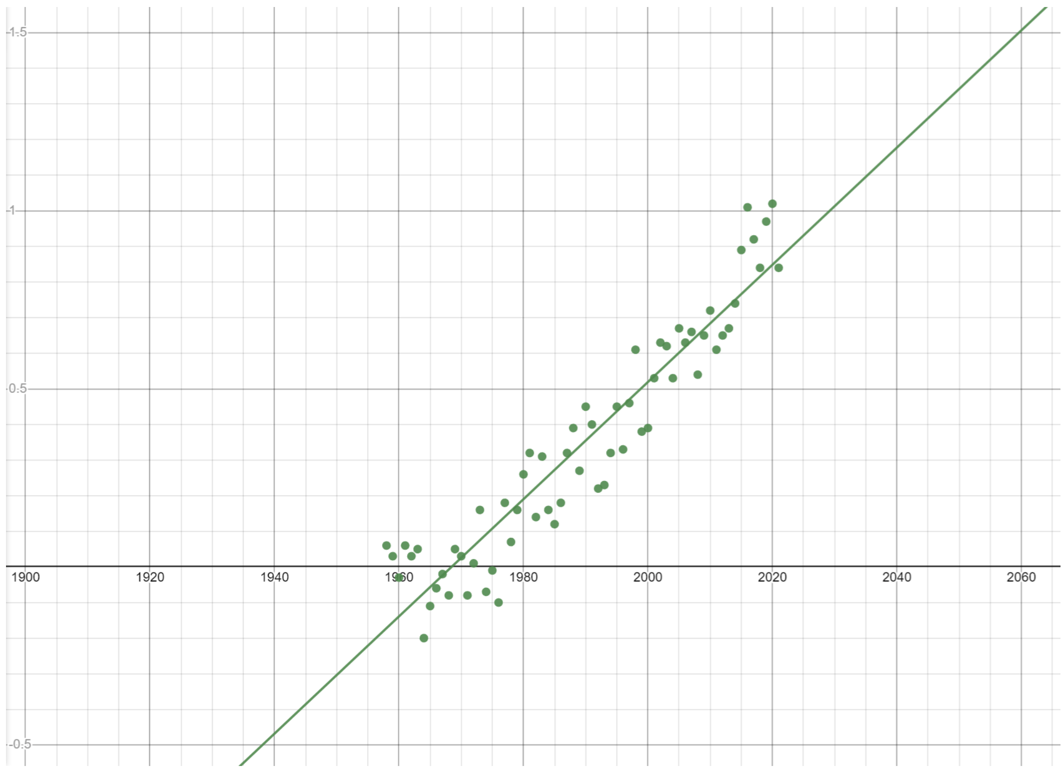
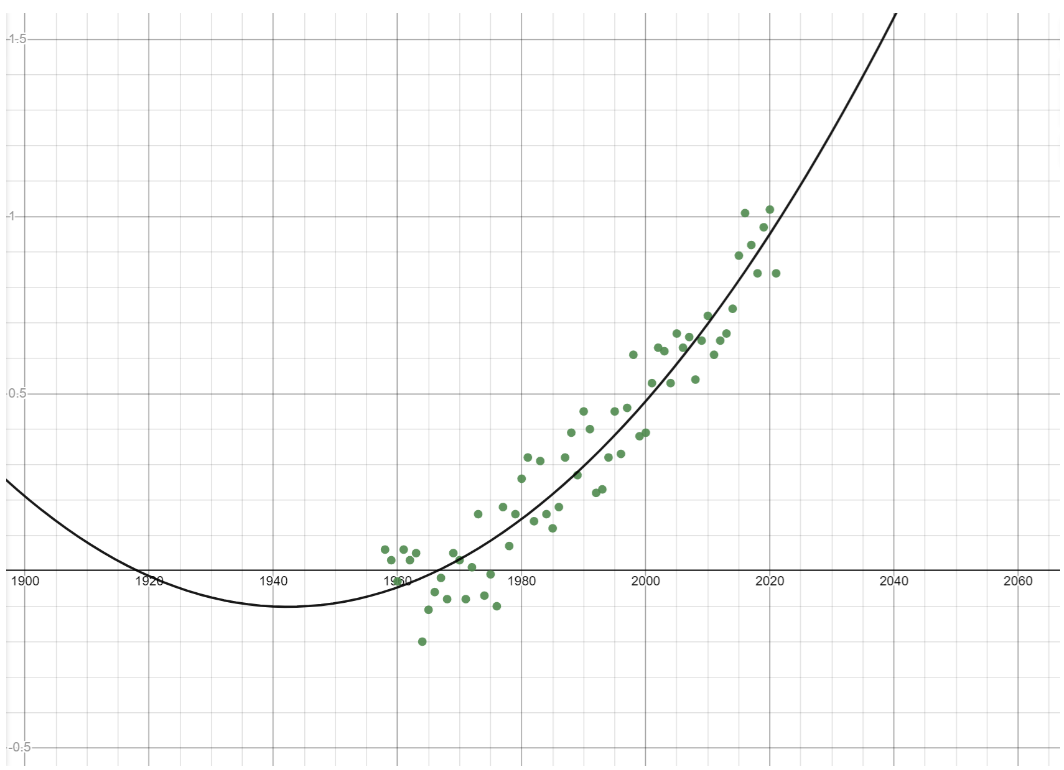
Similarly, we can perform similar analysis to each of the models for temperature. To find the change in temperature, we find some base temperature given the data between 1958 - 1980. We achieve this base temperature by averaging the values, giving us

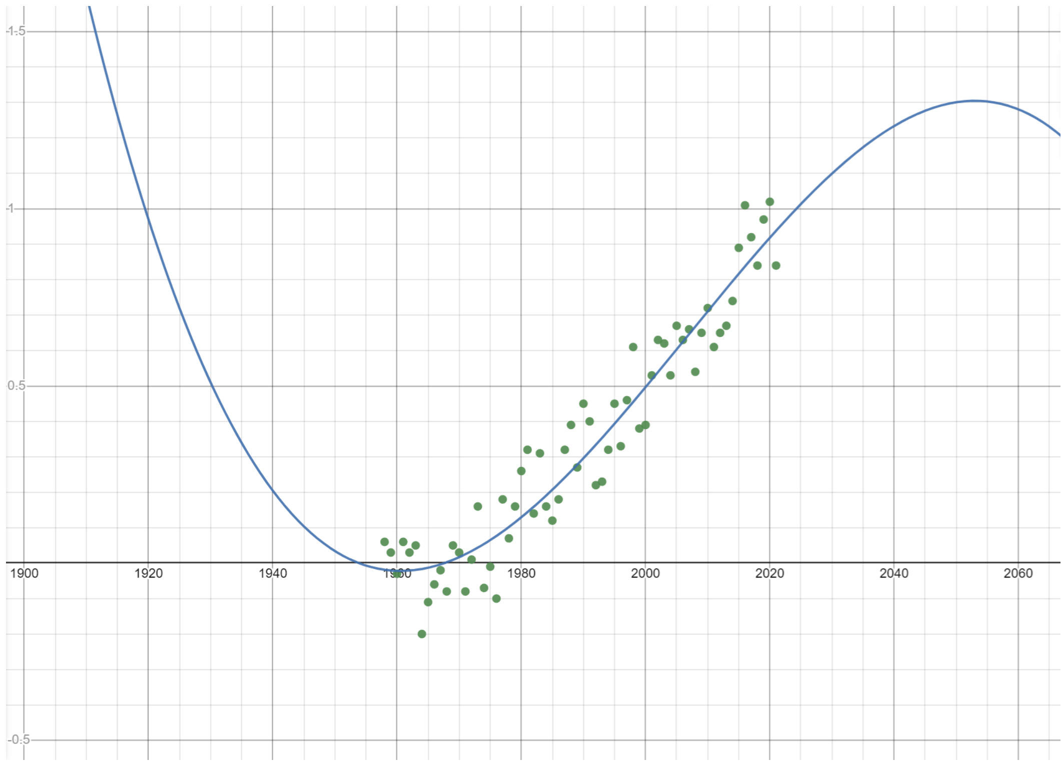
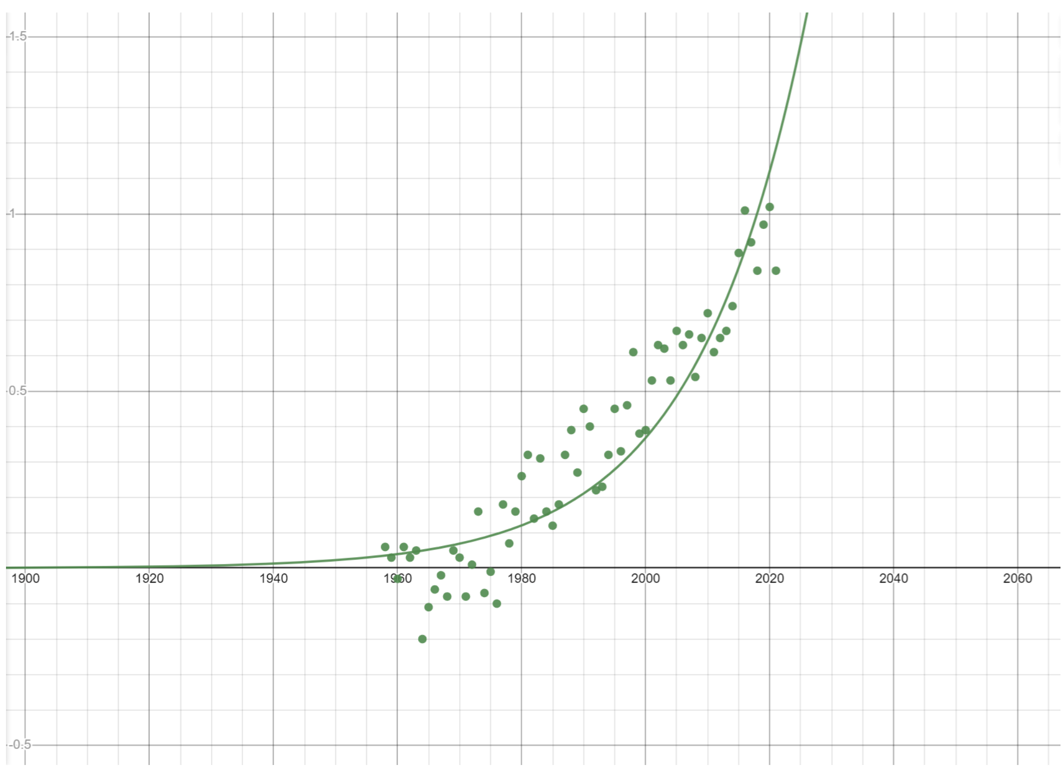
T = 0.02℃

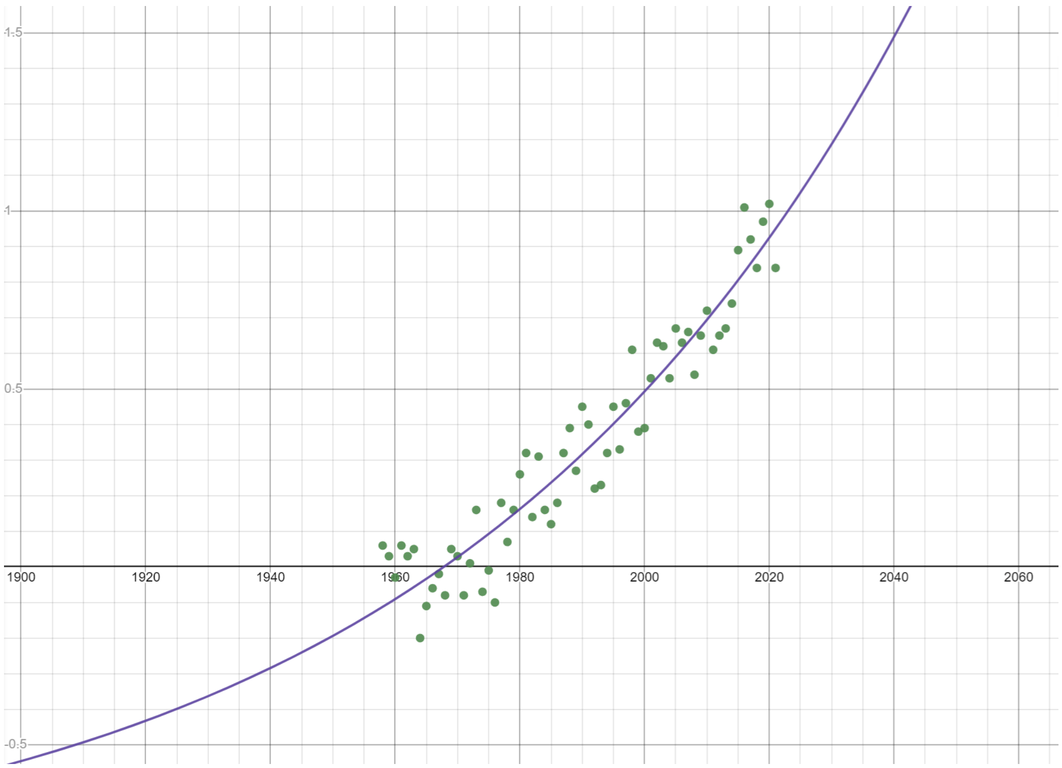
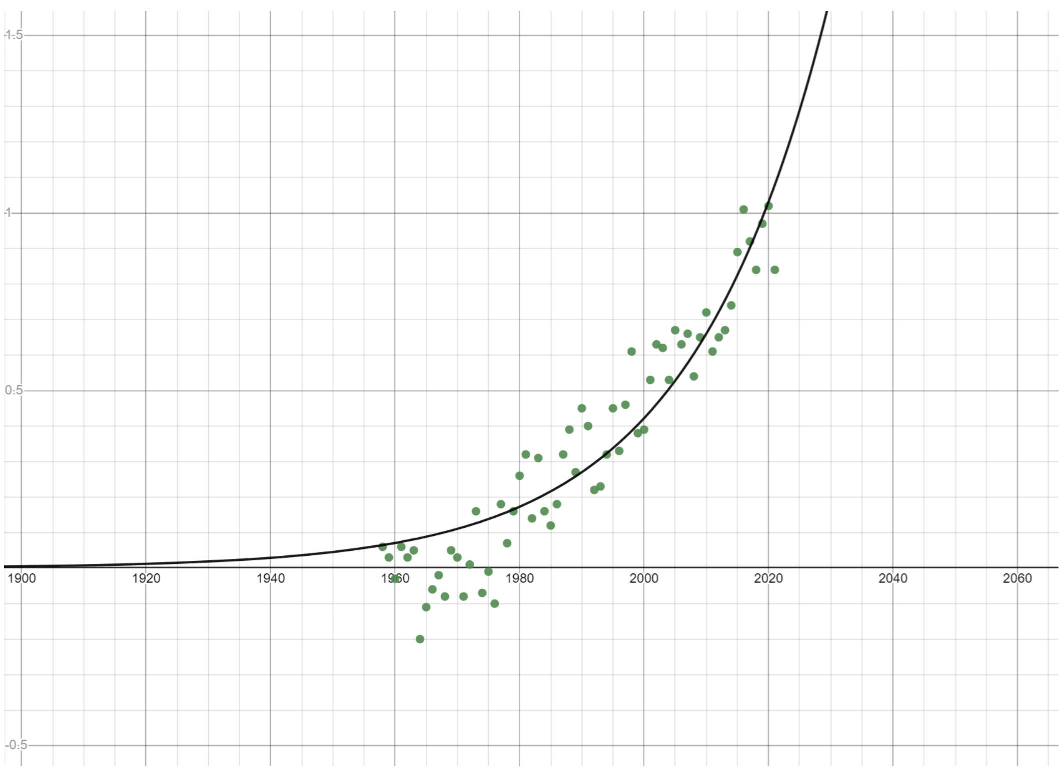
Thus, we can use this value to find when the temperature reaches a certain change based on our various models.

Model	Correlation	Change in 1.25℃ (Year)	Change in 1.50℃ (Year)	Change in 2.00 ℃ (Year)
Linear	0.9438	2045	2060	2091
Quadratic	0.9582	2030	2038	2052
Cubic	0.9595	2044	DNE	DNE
Base-10	0.9211	2022	2025	2030
Natural	0.9558	2032	2040	2054
Arbitrary	0.9359	2024	2028	2035

Below are more graphs modeling our fit lines against the given temperature data:

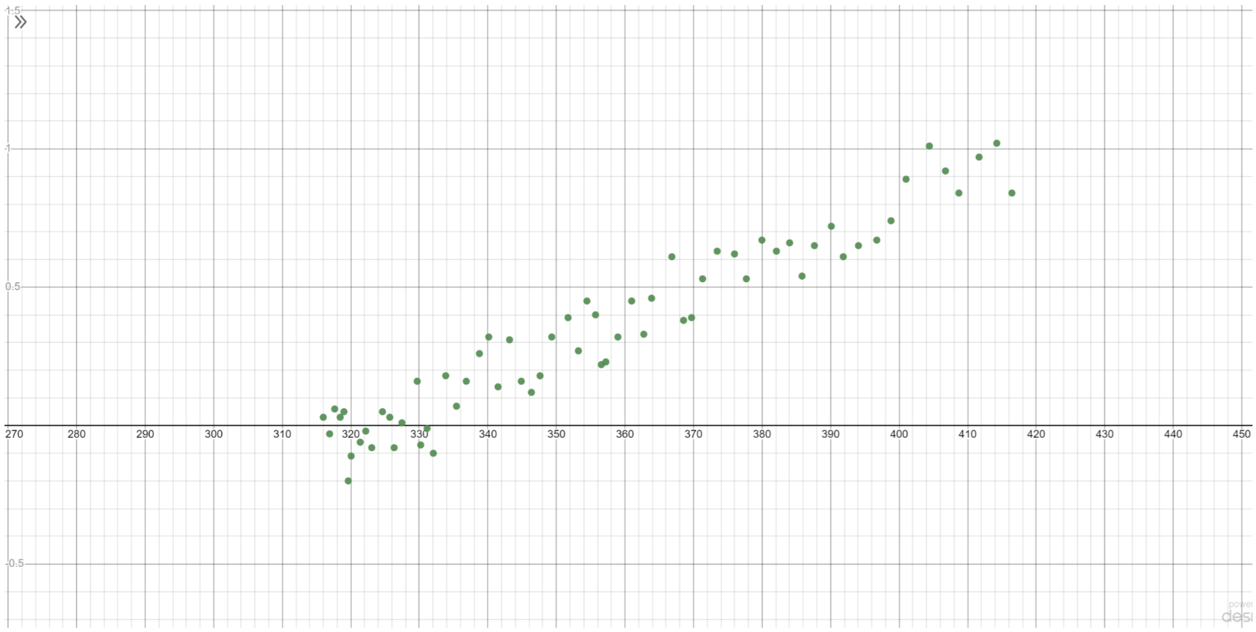
Model	**Graph**
Linear	
Quadratic	

<table>
<tr>
<td>Cubic</td>
<td>
</td>
</tr>
<tr>
<td>Base-10</td>
<td>
</td>
</tr>
</table>

Natural	
Arbitrary	

From this point forward, we again are choosing the natural exponential model, as not only does it have the highest correlation with the given data, it also gives reasonable values for when there is a change in temperature. Our polynomial based models are again eliminated immediately, following the same logic as in Problem 1.

To find a relationship between temperature and CO_2 levels in the atmosphere, we graph the temperature against the PPM data:



With the x-axis representing PPM of CO_2 and the y-axis representing Temperature.

Clearly, since we decided that both employed a natural exponential function to model, we can analyze this scatter plot through a linear sense. As well as this, we omitted the temperature value given in 1958 as there is no corresponding CO_2 level. Modeling this data linearly is simple, employing yet another fit line. Our resulting equation relating PPM of CO_2 and temperature is

$$y = 0.0104766x - 3.39257$$

With its corresponding correlation to the data at 0.9613. This correlation, along with the fact that both temperature and CO_2 levels were modeled exponentially, strongly suggests that there indeed is a relationship between the two variables.

Extending our model into the future, we can explore how reliable our linear model is. For example, when the atmospheric CO_2 levels reach 685 PPM, which our exponential model predicts to happen at 2072, temperatures should reach 3.78 ℃. Using our temperature model though, however, we notice that the temperature only reaches 3.78 ℃ at 2090. From this point, we can scale backwards until our model is reliable, which we found to be until 2052. One major concern with predicting between CO_2 levels and temperature is that although they do directly appear to have a direct relationship, an increase in CO_2 does not affect temperature immediately, as there is some lag in change in CO_2 and change in temperature. This could drastically affect the temperature-CO_2 relationship, as any deviation in the trend of CO_2 would realistically not show up in temperature until after some time. This would change our model to some extent.

Non - Technical Article

Over the past century, humans have only exacerbated the climate crisis. Stemming from the Industrial Revolution, we have been innovating and striving forward, but at what cost? We have been continually building sources of CO2, while deforestation destroys our key sources of CO2 depletion. In fact, according to the Environmental Defense Fund, around 20% of annual global greenhouse gasses (GHG) emissions is contributed to deforestation (Environmental Defense Fund 2015). As such, it's imperative to analyze past emissions trends to gain an understanding of how future CO2 levels may develop. Specifically, we have looked at the data from the scientists from National Oceanographic and Atmospheric Administration (NOAA) and Scripps Institution of Oceanography (SIO). They have collected since 1959 and our team has been able to make several conclusions and predictions regarding CO2 concentrations and temperature levels.

Over the past 50 years worth of data, we were able to analyze periodic trends within the increase of carbon dioxide levels in the atmosphere. For instance, we were able to disprove the largest jump within 10 years of CO2 levels, which happened in 2003, not 2004. Although the difference is very marginal, it's a correction of the NOAA.

Aside from disproving a small assumption made by the NOAA, we also aimed to represent the current data given to us and develop similar regression models. For instance, we first attempted to use many versions of polynomial regression to map out the data to simple equations. These simple equations are key to gaining a better understanding of how CO2 levels & temperature may have developed even prior to the first date of given data, 1959. It also provides us a simple overview of how these same levels may look almost 100 years into the future. However, there are many different models that could be used to fit this data, but after considering volatility of these models, we settled on the natural exponential model. Specifically, we predicted a CO2 level of around 531 PPM by 2050, which is lower than the predicted amount from the OECD. Instead, the model predicts that we will reach the predicted 685 PPM by 2072.

In a similar sense, there exists a direct relationship between temperature and CO2. It's no surprise, as CO2 constitutes almost 76% of total greenhouse gas emissions (Center for Climate and Energy Solutions 2022) and is a major factor for the increased temperatures. Specifically, the data proves this with a statistically significant correlation value. To model temperature itself, we also utilized a natural exponential model for the temperature, which predicts that when 685 PPM is achieved in 2072, we will have increased by 3.87 degrees Celsius. However, this same increase in 3.87 Celsius is only achieved in 2090, according to the temperature model.

References

Carbon dioxide now more than 50% higher than pre-industrial levels. National Oceanic and Atmospheric Administration. (n.d.). Retrieved November 15, 2022, from https://www.noaa.gov/news-release/carbon-dioxide-now-more-than-50-higher-than-pre-industrial-levels#:~:text=Prior%20to%20the%20Industrial%20Revolution,atmosphere%20for%20thousands%20of%20years

Deforestation: Solved via carbon markets? Environmental Defense Fund. (n.d.). Retrieved November 15, 2022, from https://www.edf.org/climate/deforestation-solved-carbon-markets

Desmos

Global emissions. Center for Climate and Energy Solutions. (2022, March 24). Retrieved November 15, 2022, from https://www.c2es.org/content/international-emissions/#:~:text=CO2%20accounts%20for%20about%2076,6%20percent%20to%20global%20emissions

LoggerPro

NASA. (2020, October 12). A degree of concern: Why global temperatures matter – climate change: Vital signs of the planet. NASA. Retrieved November 15, 2022, from https://climate.nasa.gov/news/2865/a-degree-of-concern-why-global-temperatures-matter/#:~:text=Temperatures%20increase%20at%20different%20speeds,regions%20during%20the%20warm%20season.

Python Libraries