Biology paper

Carbon dioxide and Global warming

By: Zekai Yu, Haoyi Cai, Zihao Yuan, Qixin Zhu

Step into the industrial Era, with the development of technology and the life quality of people, people had more influence on the natural world than the sum of the one people made before 1750. One essential change agent in the environment people made worldwide is the amount of CO2 (carbon dioxide) people emitted into the atmosphere. CO2 emission is the most well-known and concerning consequence of industrialization for most people. Due to the identity of CO2 as a greenhouse gas, many people are concerned the high emission of CO2 would lead to global warming.

In this report, we would first made graphic analysis, then we would draw a conclusion based on our model. Furthermore, we plot data on PPM growth compared to last year (ΔP) to illustrate our idea more clearly. We concluded that the increase in PPM in 2004 did not directly lead to the largest increase over any previous 10-year period, but year 2004 indicates the time it has since CO2 increases more 1.5 PPM per year. From these findings, we began to build various models of PPM, QR and ARIMA, to achieve the effect of changing trends. From the QR model, we predict that the PPM may reach 496.8 in 2050.

In addition, we have released a lot of trend charts with different situations. Later, we began to focus on the adaptation of the quadratic function and show future trends by the LSM model, the least squares method, and using data from 1951 to 2021, which shows that when the temperature changes by 1.25 °C, 1.5 °C, 2 °C corresponds to the years in the quadratic function, they are 2030, 2038 and 2052. To figure out if there is a relationship between temperature and CO2 concentration, we plot the data for these two values on the plot to see if the trends of the two values are close. With this method, we see that the trends in temperature and CO2 are very similar, so they are likely to have co-related, and these data and graphs show that PPM has a significant effect on temperature.

The article

In 1896, Swedish scientist Svante Arrhenius warned that carbon dioxide (CO_2) emissions could contribute to global warming. In 1988, the United Nations Environment Program and the World Meteorological Organization established the Intergovernmental Panel on Climate Change (IPCC), and subsequently, the IPCC released the "Climate Change Assessment Report" that the global climate is warming, of which CO_2 is the most important factor causing global warming, and if control measures are not taken, global warming will cause ecological and environmental disasters. Since the release of the IPCC report, there has been great controversy in the academic community about global warming and whether CO_2 is the main cause of climate warming.

After the second industrial revolution, with the great increase in production levels, the world's population increased dramatically. Large-scale production has led to excessive carbon dioxide emissions, which makes it impossible for the earth to self-regulate to consume human carbon dioxide emissions, so this leads to the greenhouse effect, which creates this environmental problem. In recent years, the continuous excessive carbon dioxide emissions have continued to carry out carbon dioxide emissions such as the current one, and the impact of the greenhouse effect on the environment will increase even more significantly in the future.

Therefore, we began to analyze and study the trend of carbon dioxide on the greenhouse effect. In the beginning, we had an assumption that emissions would not surge significantly in 2004. Thus, we plot the data on PPM growth compared to last year. From these charts, PPM is likely to reach 496.8 in 2050.

Overall, the relationship between CO_2 concentration and the temperature is strong. Higher CO_2 would lead to Higher temperature. For the sake of our world, our future, we need to pay attention and considered plans for decrease CO_2 emission in order to mitigate temperature rise.

Table of Contents

1.Introduction

1.1 Background

Before the Industrial Revolution, atmospheric carbon dioxide (CO_2) remained around 280 parts per million (ppm). In March 2004, the concentration of CO_2 in the atmosphere reached 377.7 ppm.
Along with the development of society, people more and more realize the importance of the environment, but people's life today is hard to leave those industrial products that influence the environment more than ever people made before the Second industrialization. Of all the great changes people made on Earth, CO_2 is the most well-known one. People usually related it to global warming. Many countries made claims about their plan to reduce CO_2 emissions. However, the information about CO_2 may be too abstract for many people to know. In this report, we wanted to analyze the data collected to some simple and understandable model for people to understand the trend of CO_2 emission, why is it important, and how it would be in the future. Prior to the Industrial Revolution, carbon dioxide (CO_2) in the atmosphere was consistently around 280 parts per million (ppm). The concentration of CO_2 in the atmosphere reached 377.7 ppm in March of 2004, resulting in the largest 10-year average increase up to that time. According to scientists from the National Oceanographic and Atmospheric Administration (NOAA) and Scripps Institution of Oceanography (SIO), the monthly mean CO_2 concentration level peaked at 421 ppm in May 2022. An Organization for Economic Co-Operations and Development (OECD) report predicts a CO_2 level of 685 ppm by 2050. For a greenhouse gas like CO_2, a high concentration will likely aggravate global warming.
The increase in global temperature would lead to the sea level coming up, animal extinction, and extreme weather, hurting both humans and nature.

Although the harm of high CO_2 concentration is very obvious, there are still some people who cannot perceive the emergency our world meets. Thus, our team want to make models to illustrate the future CO_2 concentration and justify the relation between CO_2 and temperature.
We obtain data on CO_2 concentrations and temperature changes from 1959 to 2022.Then, our team want to find out the problem between Carbon dioxide and global warming.

1.2 Problem Restatement

Our group will create graphs to study and describe past and predicted future levels of carbon dioxide in the atmosphere, and we will use these models to predict future trends in the greenhouse effect and the relationship between the greenhouse effect and carbon dioxide. Finally, we will analyze our model and prepare an article explaining our team's findings and possible future recommendations.

Problem 1: Building models conclude and predicted CO_2 concentration in the atmosphere. Answer the question that whether the claim shown in the stem is true. Analyzing the credibility of each model we made.

1a. We need to build a statistical model to conclude data and figure out whether the March 2004 increase in CO_2 leads to the greatest increase among any previous 10-year period according to the overall pattern of CO_2 emission

1b. We need to make predictions of the future trend of CO_2 based on different models.

1c. Using a prediction model to find the PPM value in the year 2100. Figuring whether the CO_2 concentration level will reach 685 PPm by the year 2050, if not the year 2050, which is the year the CO_2 PPM level reached 685 PPM.

1d. Evaluating our model to find the most accurate one.

Problem 2: First find the relationship between the concentration of CO_2 in the atmosphere and the temperature change. Analyze is based on my formal work in Problem 1, and the dataset of land-ocean temperature provided.

2a. We need to build the prediction model for the temperature change by the given data, and specifically figure out when would the average temperature have increased by 1.25°C, 1.50°C, and 2°C, compared to the base period of 1951-1980.

2b. Construct a model to figure out and justify whether there is a relationship between CO2 concentration and the land-ocean temperature. If there is a relation between the

2c. Based on the model built in 2b which is the model explains the relationship of CO2 and the temperature to predict future temperature change. Evaluate how far the model is still reliable. Writing down the possible cause of error in prediction for our model.

2.Preliminary Information

2.1 Assumptions

1: Other greenhouse gas provided little influence on temperature change compared to CO2.
-There are other greenhouse gases such as CH4 also exist in the atmosphere, it is hard for us to analyze all greenhouse gas together in this report.
2: There is little temperature change caused by the geographical activity of Earth
-The amount of CO2 in the atmosphere and the average temperature on Earth are not static because of geographical activity on Earth, but the geographical change often happened over a very long period calculating in thousands of years. It is reasonable for us to ignore geographical activity's influence on temperature or CO2 percentage in the atmosphere.
3: The change of CO2 would follow its original pattern.
-There might be some new technology developed in the future that can decrease the amount of CO2 in the atmosphere, but we cannot predict whether it would happen based on present information.

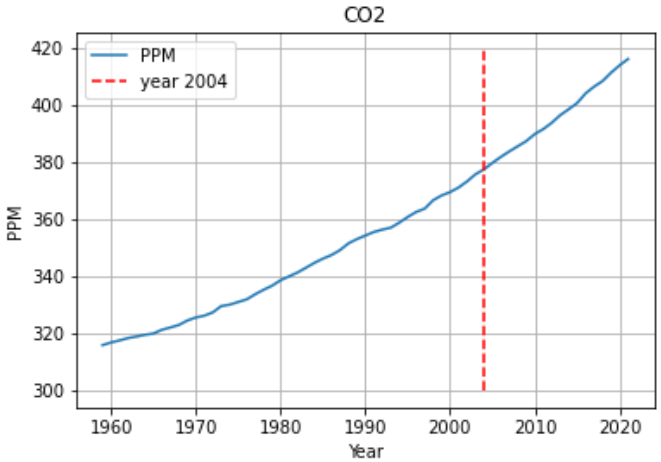
2.2 Variable Chart

meanings	symbols	units
CO2 amount (PPM)	C	0.001‰
Change of CO2 amount	ΔC	0.001‰
Temperature change	ΔT	°C
Temperature	T	°C
Distance from points	D	/

3.Formating models

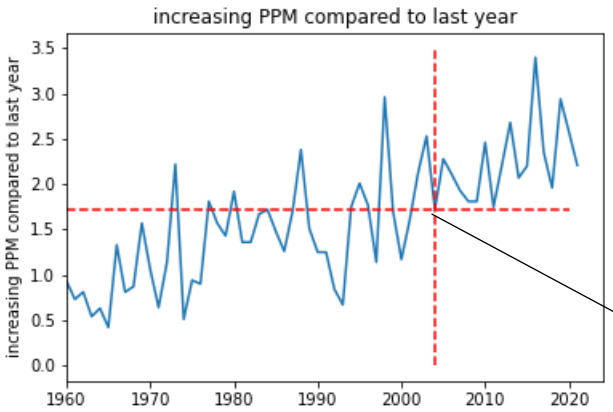
3.1 Problem 1(a)

Our Conclusion: We disagree with this statement that the March 2004 increase of CO2 resulted in a larger increase than observed over any previous 10-year period.



3.1.1 Graphic Analysis

First, we plot the information of CO2 emission compared to time on the graph to illustrate the data to give us a better understanding of what model we should use. Based on this raw graph we already have a hypothesis that the year 2004 doesn't have a big surge in emissions according to the bland slope of the trend.



Year 2004
ΔC 1.72

$$\Delta C = Cx - Cx\text{-}1 (x=2,3,4...)$$

To figure out whether the year 2004 lead to the biggest burst of CO2 emission compared to any previous 10 years more clearly. We plot the data of PPM increase compared to the last year (ΔC). In the third graph indicating the increase of PPM compared to the last year, we could see that the increase of the year 2004 compared to last year is actually smaller than the amount of last year compared to the last year. Thus, the increase in PPM in 2004 had not directly resulted in a larger increase than observed over any previous 10-year period. Nonetheless, the year 2004 indicated the time that each year led to more than 1.5 PPM of CO2 after that.

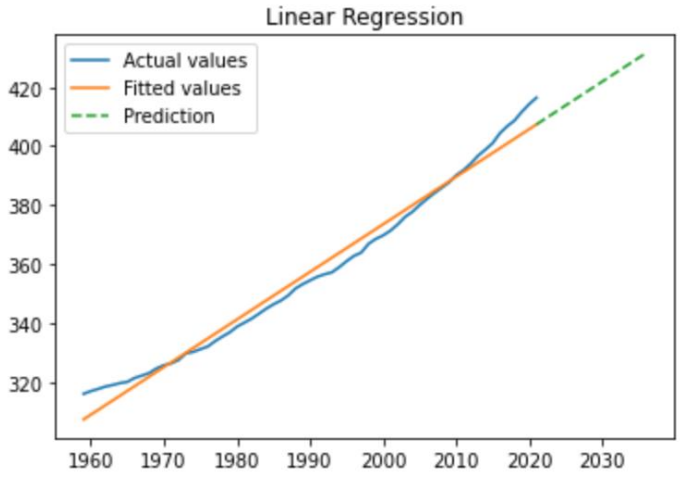
For predicting the future trend of CO2 emission, we need to first regress our data by a specific model based upon the feature of data, and then predict by the feature of the model we use.

Our thought is that the scale of ΔC is change through time with some pattern, and the graph of the change of CO2 is fluctuant through an upward trend. We believed this is because over a period of time there is increased productivity and increased demand for products because of new-coming technology and product which lead to the increased production process that emits CO2. The year after 2000 has a new area of market of Internet and relevant product, which may be the cause of the increased scale of increase of CO2 after 2000.

3.2 Problem1(b) Building various perdition model of PPM

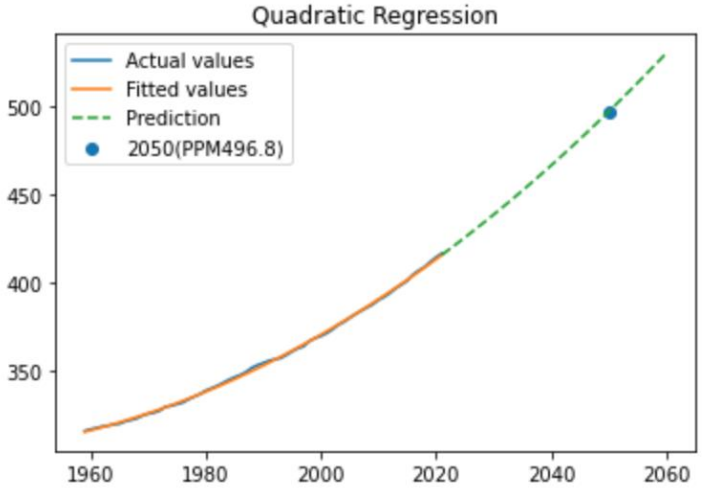
3.2.1 Regression

The first model we made is built by the regression. This model is often used to determine the best-fit curve of polynomial function for a bunch of data. The principle of this model is to find the curve where the quadratic sum of residual error reaches a minimum.



Function: y=1.614x+ 307.303

Standard Deviation: 3.921195995641181



Function: $y=0.013x^2 +0.806x +315.522$

Standard Deviation:0.720

It is clear that the data is more fit in the quadratic regression.

3.2.2 ARIMA

The other model we tried to use is the ARIMA (Autoregression Integrated Moving Average) model. This model is normally used in predicting those data that have different periods of the tendency of value change but with an overall trend, such as the data of the stock market.

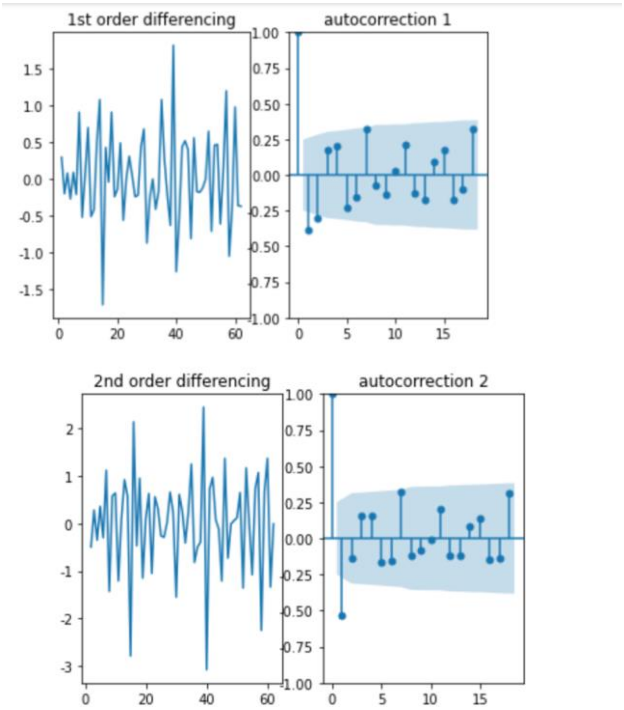
An autoregressive integrated moving average model is a form of regression analysis that gauges the strength of one dependent variable relative to other changing variables. The model's goal is to predict future securities or financial market moves by examining the differences between values in the series instead of through actual values.
We wanted to use this model to predict the future ΔC and calculated the future PPM based on it.

Our team used The ARIMA model to solve the problem. It can be understood by outlining each of its components, as follows:
（1）Autoregression (AR): A model that shows a changing variable that regresses on its own lag or previous value.

（2）Integration (I): Represents the difference in the original observation to allow the time series to become stationary (that is, the data value is replaced by the difference between the data value and the previous value).

（3）Moving average (MA): Contains the dependencies between observations and residuals of the moving average model applied to lagging observations.

To use ARIMA, first we need to find out the time need for the overall dataset to be stationary, allow us to make further conclusion by computer function. To find out how many time the dataset need to be stationary, we need to use difference operation.



In those two graphs above we could see that the autocorrection part in first order difference is similar to the on in second order difference. Also, the data of those autocorrection is arrayed thoroughly between the positive and negative. Thus, the data is likely to be stationary.

To be sure of whether the data is stationary we running ADF method, the result is written below:

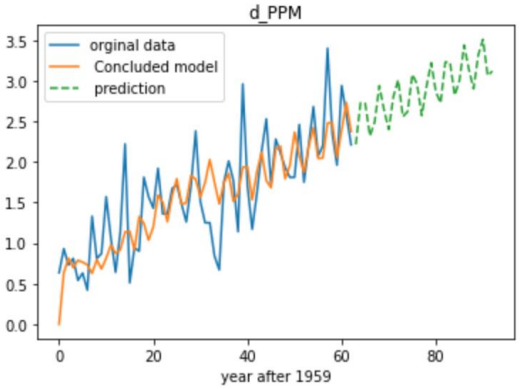
First order difference	
Test Statistic Value	-5.931
p-value	2.3868e-07
Lags Used	5
Number of Observations Used	56
Critical Value	{'1%': -3.553, '5%': -2.915, '10%': -2.595}

Second order difference Test Statistic Value: p-value: Lags Used: Number of Observations Used: Critical Value:	
Test Statistic Value	-5.0607
p-value	1.6804e-05
Lags Used	10
Number of Observations Used	50
Critical Value	{'1%': -3.5685, '5%': -2.9214, '10%': -2.5987}

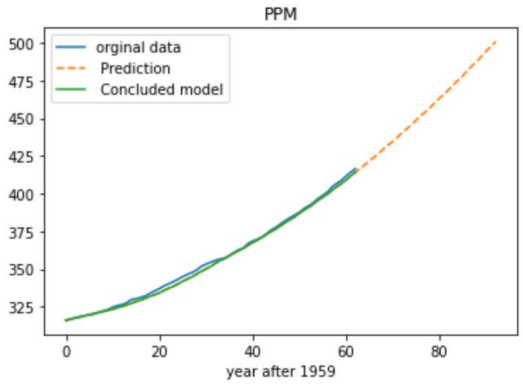
And running ADF test based on the data we got from difference. To know if the data is stationary, we need to see the p-value of data from the box above. The p-value less than 0.05 indicate that the data is stationary when it is differenced X times at the corresponding to the X order difference. All the p-values after the differencing are much less than 0.05, there is no need to differentiated the dataset to make it stationary. By the other word, the dataset is originally stationary at that situation.
From above, we know that this data is already stationary, which means the model we are going to make is simplified to an ARMA model.

The data would conclude like
$C_t = \beta_0 + \beta_1 C_{t-1} + \beta_2 C_{t-2} + \cdots + \beta_p C_{t-p} + \varepsilon_t + \alpha_1 \varepsilon_t$,
β is weight of data, ε is the white noise of data.

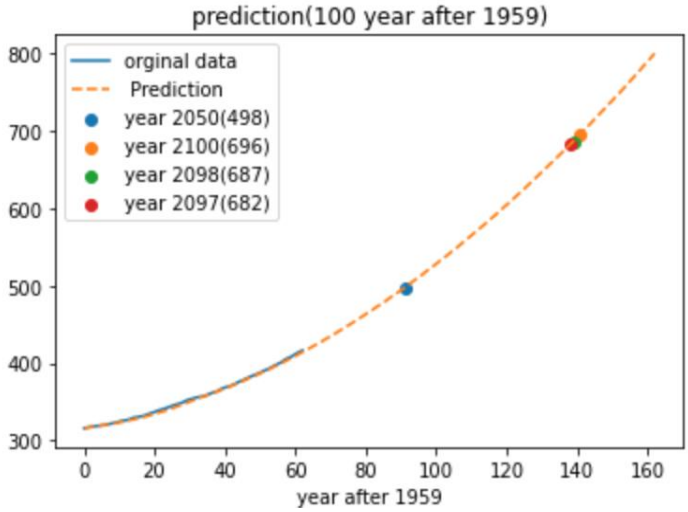
There would be a lot of possible models.
Then we need to use the variance between our concluded models to the original dataset to figure out the credibility of our model. The model we found had the least variance difference from the dataset is this one.



d_PPM is the prediction graph of ARIMA of ΔC, we could see ΔC is increase over time with a little flection.



PPM is the graph of perdition based on Δ C predicted by the ARIMA graph, the graph is obvious non-linear, and seems similar to the one we get thorough quadratic regression



3.3 Problem 1(c)

	Regression model		ARIMA model
PPM in year 2050	496.8		497.6
PPM in year 2100	688		696.3
PPM reach 685:	Around 2099-2100		Around 2097-2098
Standard Deviation of according value	0.720		0.4233 (ΔC)
Standard Deviation of the graph of PPM	0.720		1.7663

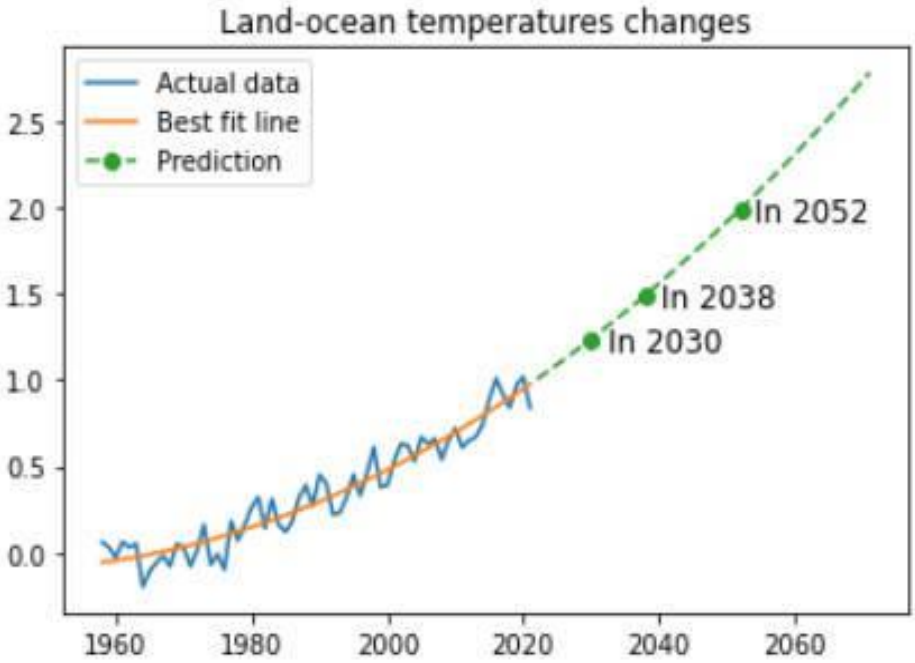
Specific analysis: Both models have a special non-linear increase in the amount of growth that increases over time. We found that the predicted results were not as exaggerated as the title gave 2050 in question. The standard deviation of according value of ARIMA is less than the one of PPM. In this case, ARIMA is more fit in its own predict value, the change of PPM overall time. However, the difference of Standard deviation of PPM value and the value generated according to models suggested that the value regression model is more fit to the given data. This is because the method of determine regression is find the best-fit curve with least standard deviation.

3.4 Problem 1(d)

The model using ARIMA is more accurate because it has a less standard variance compared to the model built by LSM, considering that the model built ARIMA has less standard variance on the value it considered than the LSM one.

Also, when using the ARIMA model, we also exercised ADP to simplify the model, and for the LSM model, we have no way to effectively verify the conclusions reached.

3.5 Problem 2(a)



We continue to use the LSM model, namely the least square method, and use the data from 1951 to 2021 to fit a suitable quadratic function and show the future trend and it shows that when the temperature changes by 1.25℃, 1.5℃,2℃ corresponds to the years in the quadratic function.

The quadratic function that we figured out is

$$0.00017430486180484227x^2 - 0.6770970487844706x + 657.4537993880336$$

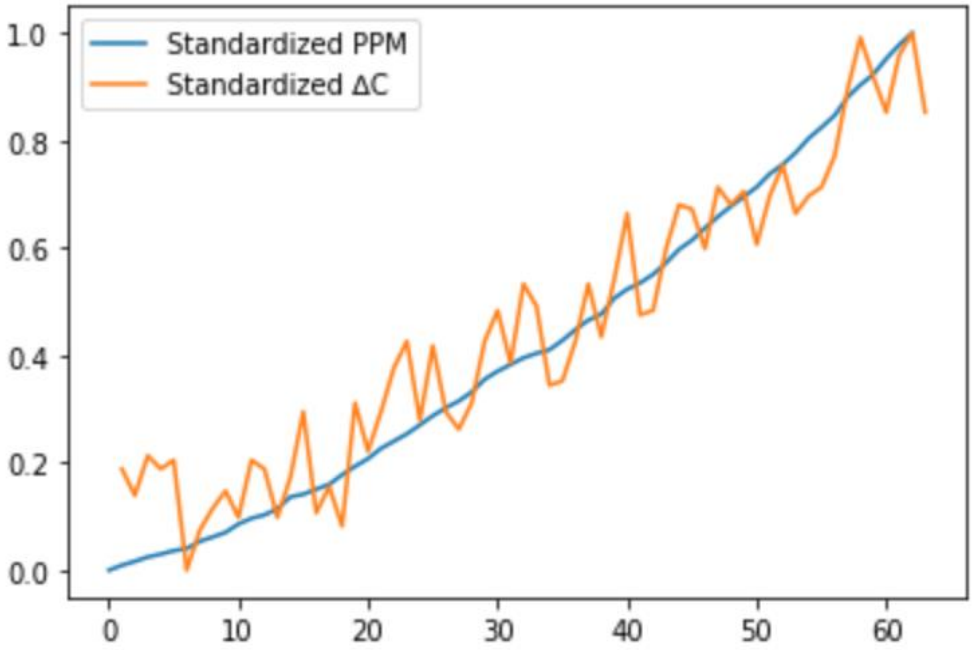
and we can make y=1.25, 1.5, 2 respectively and we can figure out x is 2030，2038，2052.

3.6 Problem 2(b)

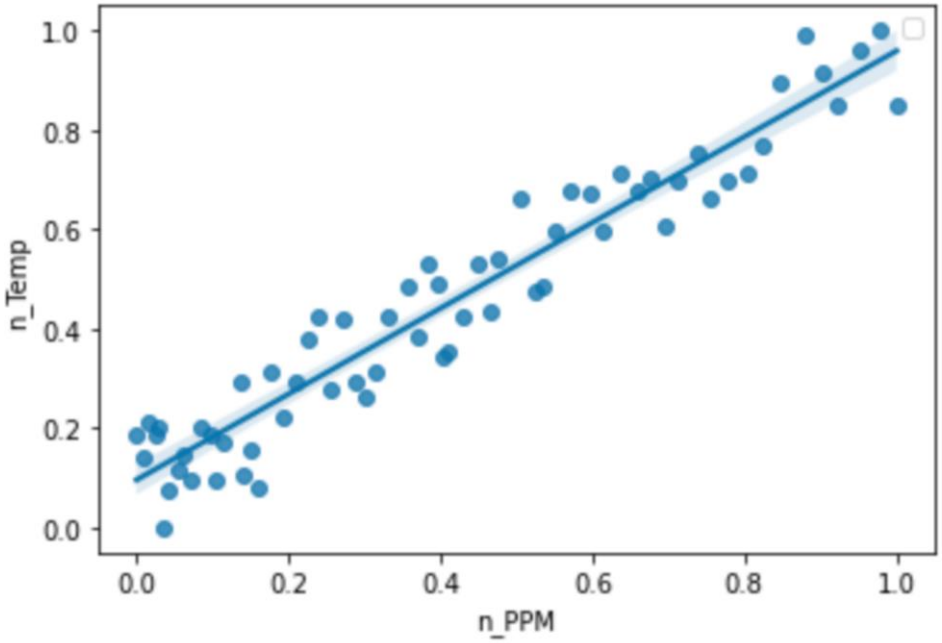
Our team figured out the relationship between the temperature and the CO2 concentration. To figure out if there is relationship between the temperature and the CO2 concentration, we need plot data of these two values on the graph, to see if the trend of the 2 value is close

First, we need to standardize those two kinds of data because they have different unit and scale.

The way we standardized data is: $C_s = \frac{C_i - C_{min}}{C_{max} - C_{min}}$ $T_s = \frac{T_i - T_{min}}{T_{max} - T_{min}}$



Through the graph above, we could see that the trend of temperature and CO2 is close, so they are likely to have relationship. Then we would plot the standardized PPM value as x-component and standardized ΔC value as y-component in a same graph to ensure our guess.



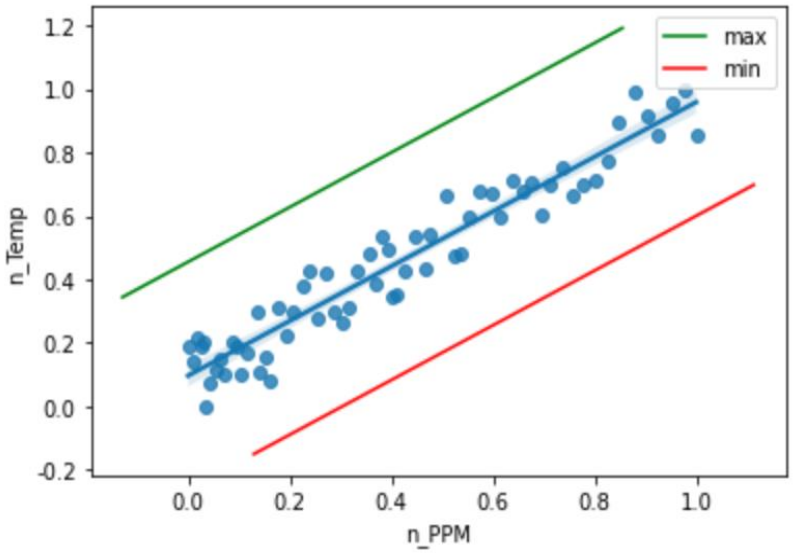
y=0.86x+0.097

Standard Deviation:0.1305

The best-fit line of this graph is linear and the standard deviation of this graph is low, suggesting that the PPM has a directly influence on the temperature.

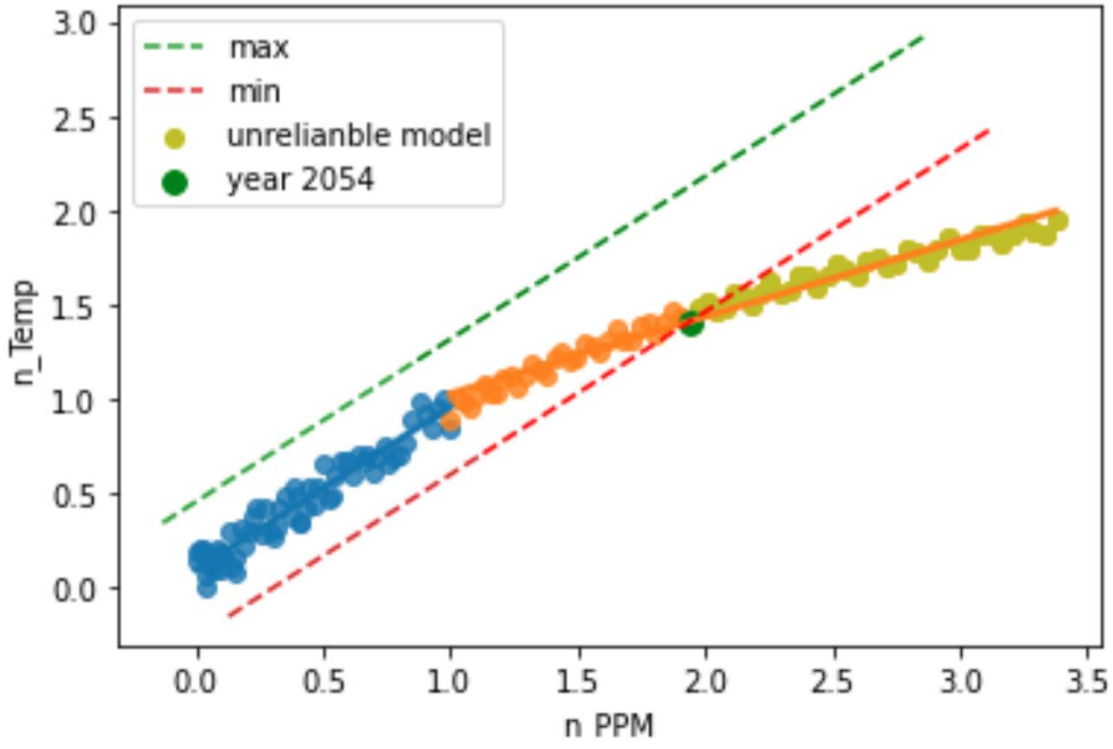
3.7 Problem 2(c)

To evaluate the credibility of our model, we need to figure out a range that our model is credible.
We can get the range of credible value based on the standard PPM value vs. standard ΔC value. We find out the furthest point from the best-fit line, find its distance from best fit-line, D. Then we determined that D is the distance of credible range from the continuous of the best-fit line.



The graph above shows the standard PPM and ΔC values, including the credible range(max,min).

D is equal to: 0.279 in this case



We plot our prediction of PPM and temperature change in to the graph, the prediction values are also standardized by the method mentioned above.

From the graph the data until 2054 are in the max-min range, and the straight lines connected by these data are roughly parallel to the min and max lines. But in 2054, the data changes dramatically, forming a dividing line that deviates the data beyond 2054 from the max-min line we could see that the graph is at first reached out of the credible range at the year 2054.

Thus, our model is likely to be credible before 2054, but less credible after it.

4.Strengths and Weaknesses

4.1 Strengths

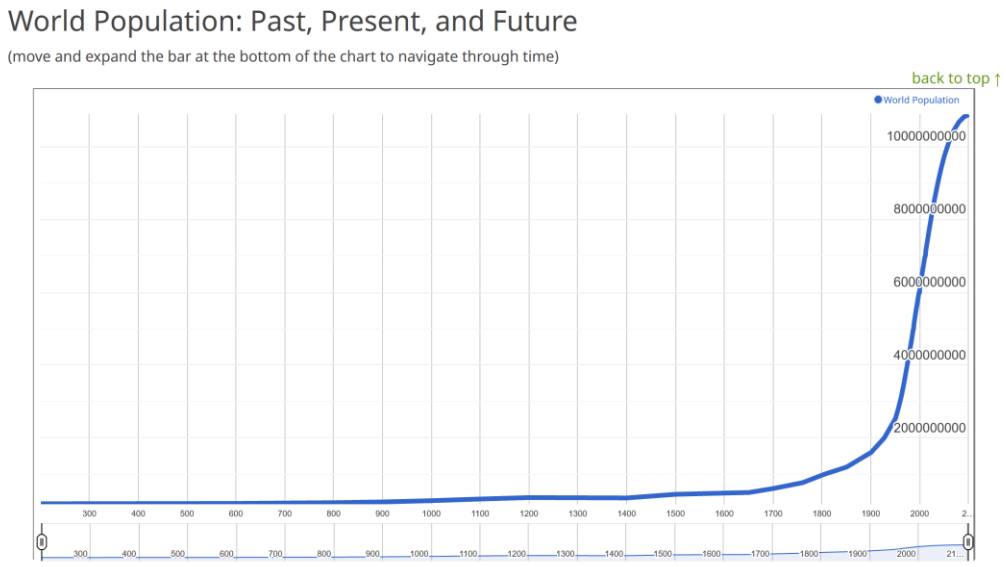
1. We considering the fluctuation of CO2 increase, which made the model be able to more in line with reality. For example, we put all the valid data into the model, on the one hand, we respect the authenticity of the experiment and retain all the data that may affect the results, on the other hand, these data help us make better predictions, according to this data, we know the fluctuation curve of the data, and we can also know the error of our curve.
2.In order to study the problem more holistically, our group considered more comprehensive factors. For example, we took into account that the global population would be largely related to carbon emissions, so we looked for data. Based on authoritative data, we summarized the carbon emissions situation, and then conducted a deeper study through the data given by the title.
3.We looked at a lot of data, such as the factors that affect the greenhouse effect of carbon dioxide, the growth of the global population, the data of the world's carbon emissions, and then analyzed. For example, when we look at the huge increase in the world's population after 1990 due to the end of the second industrial revolution, which confirms the increase in world carbon emissions, it also confirms the chart predictions we made

4.2 Weaknesses

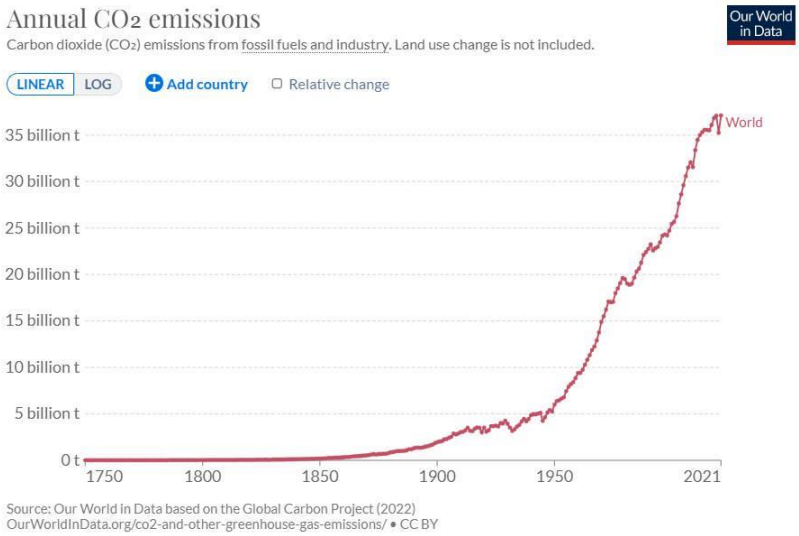
1. Our group knows that many countries will pay attention to environmental pollution, but we have not considered the impact of developed technologies on CO2 emissions concerns and temperature. In fact, we believe that there are still objective factors that have not been taken into account.
2. Our predictions for the data are biased towards the considerations of the supervisor, and do not specifically consider the continuous deterioration of the environment. After reviewing the data, we found that many countries have taken environmental pollution (including carbon emissions) seriously, so we believe that environmental pollution will be generally visible in the future.

5.Space to Improve

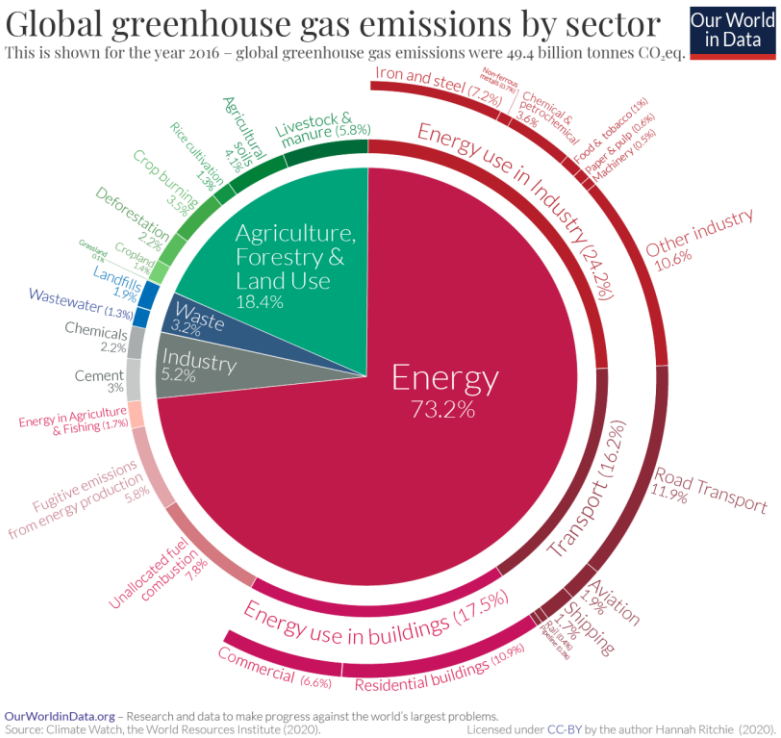
There are some other really important factors for temperature and CO2 concentration that we haven't included in out model, such as the world population, the emission of other greenhouse gases like methane Our group found that by constantly combining the display data, such as the world population trend chart placed below, we found that the population has passed the stage of substantial growth and is already in a period of population growth, so in this case, many countries can be more effective in controlling the greenhouse effect caused by carbon dioxide emissions in this situation, and this also helps the country to better manage environmental problems, and we hope that the conclusions and data reached by the group can help the management of the environment. I also look forward to a better nature!



In addition, our group also found the world carbon emission situation map and energy share map (shown below), these data are from the official website (links are in the literature), so they also help us better understand this environmental problem, but also most of the verification of the reliability of our model at the moment, as mentioned earlier, we maintain a positive attitude, so when the world begins to focus on reducing carbon emissions, the trend data predicted by our group will be greater than the data at that time. Conversely, the trend data predicted by our group will be larger than the data at that time.



The chart graphs annual CO2 emissions from before about 1750 to the present. From this, the line graph graphs the extreme increase in annual CO2 emissions since 1950. In recent years, we can know that although this data is high, it has plateaued. This is related to the world's population base and productivity levels.



This pie chart of global greenhouse gas emissions by industry visually shows that energy consumption accounts for the majority.

6.Reference

[1] https://www.investopedia.com/terms/a/autoregressive-integrated-moving-average-arima.asp
[2]https://www.investopedia.com/terms/a/autoregressive-integrated-moving-average-arima.asp
[3]https://www.statisticshowto.com/adf-augmented-dickey-fuller-test/
[4]https://wenku.baidu.com/view/1c252a020740be1e650e9a21.html?fr=aladdin664466&ind=1&_wkts_=1668231212650&bdQuery=最小二乘法
[5]https://reformship.github.io/pages/1capacity/1model/模型%20 预测模型 27%20 时间序列-非平稳时间序列%20(1).html
[6]https://www.worldometers.info/world-population/
[7]https://ourworldindata.org/emissions-by-sector
[8]https://ourworldindata.org/co2-emissions
[9]https://www.co2.earth/global-co2-emissions#
[10]https://ourworldindata.org/co2-emissions
[11]https://www.co2.earth/historical-co2-datasets