Team Control Number

12691

Problem Chosen

B

2022
HiMCM
Summary Sheet

The Identification, Contextualization, and Prediction of Carbon Dioxide Emissions

Climate change has had devastating effects on the world, with the worst still yet to come. Given the trend of climate and temperature time series data, this report is an investigation into the predicted future levels of carbon dioxide (CO_2) concentration and temperature change, and whether the results match with predictions from other sources.

First, to evaluate the statement from the National Oceanographic and Atmospheric Administration (NOAA) that the March 2004 level of CO_2 resulted in the largest 10-year average increase up to that time, we used the given annual data to calculate 10-year average increases. Our results show that the 10-year average increase for March 2004 was one of the highest on record but not the highest, so the result disagrees with the claim made by the NOAA.

To check the prediction from the Organization for Economic Cooperation and Development (OECD) that CO_2 concentrations will reach 685 ppm by 2050, we fit quadratic splines and second differences to describe the past data and used the weighted moving average, divided differences, divided differences by sector and population, and ARIMA models to further describe the trend of the data and predict future values. All the models predicted a concentration of 685 ppm after 2090, so the predictions disagree with the CO_2 level claims. From comparative model analysis, it was determined that the divided differences model was the most accurate.

Next, to predict future land-ocean temperature changes, we visualized the data and created a model based on Bollinger Bands to obtain a general range for future values before getting a predicted value based on the trapezoidal Riemann sum integral. Using the upper Bollinger Band, which was the better predictor, temperature increases of 1.25ºC, 1.50ºC, and 2ºC from the base period of 1951-1980 were predicted to be at 2030, 2040, and 2056, respectively.

Finally, to analyze the relationship between CO_2 concentration and land-ocean temperature since 1959 and predict future values, we used two models: a Pearson correlation and least-squares regression line to show long-term trends, and a Granger causality test and vector autoregression (VAR) model to focus on short-term variation. The long-term model forecasted a growth of approximately 0.01ºC for every ppm increase, while the short-term model forecasted an increase of 7.99 ppm and 0.046ºC for the next 4 years. The limitations of each of the models were also described.

Finally, we conducted a sensitivity analysis on the Bollinger Bands, ARIMA, and VAR models, and the results reflect expectations. The overall strengths and weaknesses of our model are also described. Furthermore, we have completed a non-technical news report which describes our team's findings and recommendations for the future.

Keywords: CO_2, ppm, Divided Differences, Moving Average, ARIMA, VAR, Bollinger Band

November 15, 2022 Team #12691

SCIENTIFIC TODAY

AS CARBON EMISSIONS RISE, THE FUTURE OF OUR PLANET IS INVESTIGATED

The issue of carbon dioxide emissions is not a recent development, nor one that international government officials, corporate leaders, and even common people are unaware of. Yet, as the severity of rising temperature impacts global citizens, it becomes imperative to analyze the future and understand the nuance behind climate change.

Scientists from the Organization for Economic Cooperation and Development (OECD) have made claims regarding future emissions, namely that the carbon dioxide concentration will reach 685 ppm by 2050. In order to test the validity of the statement, Scientific Today hired a research team to investigate data trends, particularly by using yearly ppm data from Mauna Loa, Hawaii between 1959 and 2021.

According to the team, historic ppm concentration values have exhibited oscillations, yet the pronounced trend among the data is that the values are increasing at an exponential rate, with 21-century concentrations growing far faster than those in the mid to late 1900s. Despite this, the team discovered, through the application of various mathematical models, that the prediction by OECD is likely inaccurate. Rather, carbon dioxide concentrations will likely reach 685 ppm by the year 2091, if not later.

Even with these deviations in results, we asked the research team to go one step further and look at temperature changes in degrees celsius from a base period of 1951-1980. Using a model famous to the stock market, Bollinger Bands, the team predicted that temperature is likely to increase 1.25 degrees celsius by 2030, 1.5 by 2040, and 2 by 2056. Finally, the team determined that carbon dioxide concentration and temperature are mathematically correlated and that carbon dioxide concentration is useful in predicting future temperature values, which corroborates the previous notion of carbon dioxide's ability to trap heat emitted by sunlight.

While we may be inclined to dismiss these findings as "tomorrow's problem," such ignorance sets a precedent for global collapse, especially as developing countries are already feeling disproportionate effects. In addition, although we like to place our trust in technology's ability to further decrease per-capita emissions, individuals' actions are vital to changing the course of carbon dioxide's exponential growth. On a personal level, given that electricity and heat is the largest contributor to emissions, we suggest considering the implementation of solar panels or electrically powered vehicles, particularly if you live in a state or province that subsidizes you in a purchase. Alternatively, reducing your consumption of meat or factory-based products, in general, is an easier way to reduce your footprint. On a global level, countries such as the US and China need to be held accountable for their emissions standard, which starts with more emissions limitation compacts such as the Paris Climate Accord. In addition, events such as Russia's Invasion of Ukraine have the tendency to incessantly pour emissions into the atmosphere, meaning that global peace can act as a stepping stone towards a decrease in the adverse impacts.

Contents

1 Introduction

1.1 Background

Climate change is one of the most pressing issues facing modern society. According to the United Nations, it is defined as a long-term shift in temperature and weather patterns [1]. These shifts may be natural, but in the last two centuries, human activity has been the main cause of climate change due to the burning of fossil fuels such as coal, oil, and natural gas. Fossil fuels release carbon dioxide (CO_2), a greenhouse gas that traps heat in Earth's atmosphere, which has led to increased levels of greenhouse gasses and resulted in serious threats to the environment, economies, and public health. Other greenhouse gasses produced, such as methane, further contribute to the magnitude of climate change.

Climate change causes rising sea levels, melting glaciers, warming oceans, and more frequent and intense droughts, storms, and heat waves [2]. These effects wreak havoc on economies and communities around the world. In 2021 alone, the ten most extreme weather events cost over $170 billion in damages [3], and by 2050, it's estimated that the effects of climate change could decrease global economic output by 10 percent [4]. Furthermore, according to the Centers for Disease Control and Prevention, climate change increases the likelihood of respiratory and cardiovascular disease, changes in the prevalence and geographical distribution of food and water-borne illnesses and other infectious diseases, and threats to mental health [5].

However, despite these issues and ongoing action to reduce the effects of climate change, the concentration of CO_2 in the atmosphere continues to grow. In May of 2022, carbon dioxide measured at the Mauna Loa Atmospheric Baseline Observatory peaked at 421 ppm, far higher than pre-industrial levels of around 280 ppm [6]. In a time of urgency to preserve our environment, it's important to understand the consequences of climate change if insufficient action is taken.

1.2 Problem Restatement

With persistent concerns over rising carbon dioxide emissions and temperatures, climate change remains ominous, leaving billions of global citizens uncertain about the living conditions on Earth. Furthermore, the detrimental impacts of climate change have already begun to occur, especially for citizens in countries without the resources to prevent droughts, floods, hurricanes, fires, and other natural disasters. As a result, the analysis required for this problem will help mathematically prove the urgency for resolution, particularly by developing a timeframe and relationship between carbon dioxide emissions and the yearly change in temperature.

To address this issue, the first question asks us to evaluate claims made by climate organizations regarding future concentrations of carbon dioxide in the atmosphere. We will fit various mathematical models to empirical data to evaluate the apparent trends and whether our predictions align with those presented by the organizations. After developing timeframes for carbon dioxide emissions to reach particular concentrations at specified dates, we will comparatively determine which model is the most accurate.

Next, the problem asks us to shift our attention to changes in temperature, through which we will build a model to predict future changes in temperature, once again using empirical data. Finally, we construct and extend a model to determine a potential relationship between carbon dioxide emissions and temperature changes and write a non-technical news article to summarize our key findings and takeaways.

1.3 Our Work

For problem 1, we first looked at the average 10-year CO_2 concentration growth and compared them to conclude whether the March 2004 increase of CO_2 actually resulted in a larger increase than observed over any previous 10-year period. Next, we fit several models, including quadratic splines, weighted moving average, divided differences, divided differences by sector and population, and autoregressive integrated moving average (ARIMA) to describe past and predict future data. We compared the prediction of the models for when CO_2 levels would reach 685 ppm with that of the OECD. Finally, we analyzed the strengths and weaknesses of the models to determine the one that we believed to be the most accurate.

For problem 2, we first visualized the temperature change data. Then, we created two models using Bollinger Bands, which provided the expected range of future temperature change values from which we could further narrow. To analyze the relationship between CO_2 concentrations and temperature changes since 1959, we used both the Pearson correlation method and the Granger Causality test. We then used the results of these two models to create a linear regression and vector autoregression (VAR) model to predict future temperatures as a function of CO_2 ppm changes. We analyzed each model to determine how far into the future it would remain reliable.

Finally, we wrote a one-page non-technical news article for *Scientific Today* to explain and present our findings.

2 Assumptions and Justifications

- Assumption 1: Current trends in carbon dioxide concentrations and temperature changes will continue, and no major climate action or events will affect these trends.

  Justification 1: Although new climate events and actions may cause stagnation or a large decrease in temperature and CO_2 levels, it is difficult to determine if or when they will occur and how large of an effect they will have. Our goal is to model CO_2 concentration and temperature based on historical data and the current climate effort, and including complex hypotheticals will unnecessarily complicate the model.

- Assumption 2: The data taken from Mauna Loa, Hawaii, accurately represent the changes in carbon dioxide concentration and temperature in the rest of the world.

  Justification 2: This assumption is necessary to make conclusions about the effect of climate change overall. The data was taken from the Mauna Loa Observatory, far away from human sources of major pollution and emissions. While CO_2 concentrations and temperature changes will vary from place to place, overall global trends should remain approximately equal throughout locations.

3 Addressing Carbon Dioxide Level Claims

3.1 Largest Ten-Year Average Increase

While the increase in ppm concentration between March 1994 and March 2004 resulted in one of the largest 10-year average increases up to that date, the claim that the increase was the largest is false. The 10-year increase from March 1993 to March 2003 has both a larger absolute and percentage difference. Therefore, the average increase for March 1993 to March 2003 is also larger, since dividing both data by 10 years to get the average increase does not affect their values relative to each other. The comparison is summarized in table 3.1.

10-Year Period	Absolute Difference (ppm)	Percentage Difference
March 1993 - March 2003	18.77	5.2546%
March 1994 - March 2004	18.74	5.2206%

Table 3.1: Comparison of 10-Year Average Growth

3.2 Modeling Past and Future CO_2 Levels

3.2.1 Describing Past Data

In order to describe the historic data beyond its evident exponential trend, we employed two analytic values, second derivatives and clamped quadratic splines, which differentiated between the individualistic and broader sections of the data.

To explore oscillations, we plotted the second derivative of the ppm concentration values using the equation $\frac{d^2g}{dy^2}$, where g is the ppm values and y is the year. The graph of the second derivative values is shown below in figure 3.1.

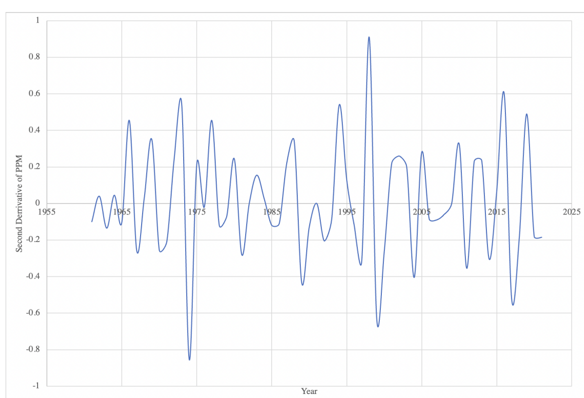


Figure 3.1: Second Derivative Graph

The second derivative of the data incessantly fluctuates between positive and negative, meaning that the data changes concavity rapidly. While we may be quick to attribute this volatility to seasonal fluctuations, the data does not incorporate seasonality, suggesting that the ppm concentrations naturally cycle upward and downward in 3-5 year periods. In addition, the second derivative is mostly stationary, which indicates that a potential fit for the data would be a sinusoidal exponential model ($g(y) = a(sin(by) + my^2 + ny + p)$, with a large period and minute amplitude.

Zooming out from these oscillations, we investigated broader trends using clamped quadratic splines, which ensure that vast differences amongst the earliest and latest values are not inadvertently influenced by a singular model. After choosing nine, seven-year apart data points, we divided them into three groups (Group 1: 1965, 1972, and 1979; Group 2: 1986, 1993, and 2000; Group 3: 2007, 2014, and 2021). Starting with the first group, we applied the following system of equations in order to determine six unknown constants, with the first section of the spline being linear, and the next part quadratic. We used the equations

$$g(1965) = a_1(1965)^2 + b_1(1965) + c_1 \qquad \text{(System 1)}$$
$$g(1972) = a_1(1972)^2 + b_1(1972) + c_1$$
$$g(1972) = a_2(1972)^2 + b_2(1972) + c_2$$
$$g(1979) = a_2(1979)^2 + b_2(1979) + c_2$$
$$0 = 2a_1(1972) + b_1 - (2a_2(1972) + b_2)$$
$$0 = a_1$$

where g is the function for ppm and all a,b, and c values are constant. Using matrix multiplication, we solved for the unknown values and repeated the process for the second and third groups. The quadratic section of our clamped splines is represented below.

$$\text{Group 1: } g(y) = \frac{1}{25}y^2 - \frac{1567}{10}y + \frac{307577}{2}$$
$$\text{Group 2: } g(y) = \frac{29}{490}y^2 - \frac{57461}{245}y + 232704$$
$$\text{Group 3: } g(y) = \frac{57}{980}y^2 - \frac{1136637}{4900}y + 232605$$

From the a-values of these three splines, it may appear that group 2 displayed ppm growth at the fastest rate, yet the fact that $\frac{29}{490}$ is only slightly larger $\frac{57}{980}$ can be attributed to exaggerated growth smaller values tend to portray. Therefore, although it is likely that ppm emissions during the period associated with group 2 likely set the precedent for future growth rates, as time has progressed, the rate at which ppm values have grown has increased. Empirically, this analysis agrees with what was formulated in the late 1900s, as these years saw the boom of Asian carbon emissions, while Western emissions remained high and stable [7].

3.2.2 Weighted Moving Average

As a time series model, the weighted moving average method emphasizes the importance of neoteric data as opposed to distant data by increasing the weight proportional to time. Such a model is appropriate to the situation of predicting carbon concentration, as the industrial system and corporate cultures that contributed to carbon emissions just a few years ago have undergone vast change, and therefore less accurately reflect the trends we predict in the future.

For this model, the data of the previous four years was used as a baseline to predict the next year, with the ppm values of the current year most applicable and therefore receiving the greatest weight, while values from previous years receive less weight. We chose to use the previous four years as one of the well-known climate cycles, El Niño-Southern Oscillation (ENSO), which occurs every 3 to 7 years [8] and indicates the environmental responsiveness to global emissions. The predicted data can be modeled by the following equation

$$\widehat{c_A}(t+1) = \frac{w_{A_1}c(t) + w_{A_2}c(t-1) + w_{A_3}c(t-2) + w_{A_4}c(t-3)}{w_{A_1}+w_{A_2}+w_{A_3}+w_{A_4}} \quad (3.2)$$

where $\widehat{c_A}(t+1)$ is the predicted ppm of the next year, c(t), c(t-1), c(t-2), and c(t-3) are the ppm values from the current year, the previous year, two years prior, and three years prior, respectively, and $w_{A_1} = 4, w_{A_2} = 3, w_{A_3} = 2, w_{A_4} = 1$ are the corresponding weights.

To determine the validity behind the methods, we performed an out-of-sample test, in which we used known data to predict the ppm values of three randomly selected years ranging from 2000 to 2021. The 20-year time span adheres to the principle of valuing recent data higher, especially as differences between predicted and actual values for lower ppm could overestimate the error values. Using the three years outputted by a random number generator (2007, 2015, and 2020), we achieve the following results, in which the predicted values are generally smaller than the actual values.

Year	Actual CO_2 (ppm)	Predicted CO_2 (ppm)
2007	384.02	379.97
2015	401.04	396.54
2020	414.24	409.07

Table 3.2: Weighted Moving Average Out of Sample Test

To account for such discrepancy between our predicted values and the actual values, we incorporated the relative error δ in our predictions for future ppm values. The relative error between the predicted and actual values is calculated as

$$\delta = \left| \frac{\widehat{c_A}(2007)+\widehat{c_A}(2015)+\widehat{c_A}(2020)-(c(2007)+c(2015)+c(2020))}{c(2007)+c(2015)+c(2020)} \right| * 100\% = \left|\frac{1185.58-1199.30}{1199.30}\right| * 100\% = 1.14\%$$

Using the relative error, we can adjust the predicted value by multiplying it by 100% + 1.14% = 101.14%. Therefore, our new predicted value will be $\widehat{c_A}(t+1) * 1.0114$ for our weighted moving average model. This predicted value was used as the c(t) in the equation for the subsequent year, then c(t-1), c(t-2), etc. Employing the aforementioned methods, the ppm predictions for years 2025, 2050, 2075, 2100, and 2109 (when concentration surpasses 685 ppm) are shown in table 3.3.

Year	Predicted CO_2 Levels (ppm)
2025	426.01
2050	490.96
2075	565.80
2100	652.06
2109	686.24

Table 3.3: Predicted CO_2 PPM for Weighted Moving Average

3.2.3 Divided Differences

In contrast to emphasizing more recent data, the divided differences model helps to interpolate a set of numerical data, which accounts for all the historical trends that occur. The end result of the divided differences method is an nth-degree polynomial equation, which we use to predict future values. The steps of the model are as follows:

1. Determine the appropriate degree of the polynomial. To do so, we take increasing numerical divided differences (first, second, third, etc.) until the values are close enough to zero, which is based upon qualitative judgment. We do not want to increase the divided difference infinitely, as it would make the polynomial susceptible to oscillation and sensitive to data error. The divided differences are calculated using the equation presented in figure 3.2.

x	$f(x)$	First Divided Difference	Second Divided Difference	Third Divided Difference
x_0	$f[x_0]$			
		$f[x_0, x_1] = \frac{f[x_1]-f[x_0]}{x_1-x_0}$		
x_1	$f[x_1]$		$f[x_0, x_1, x_2] = \frac{f[x_1,x_2]-f[x_0,x_1]}{x_2-x_0}$	
		$f[x_1, x_2] = \frac{f[x_2]-f[x_1]}{x_2-x_1}$		$f[x_0, x_1, x_2, x_3] = \frac{f[x_1,x_2,x_3]-f[x_0,x_1,x_2]}{x_3-x_0}$
x_2	$f[x_2]$		$f[x_1, x_2, x_3] = \frac{f[x_2,x_3]-f[x_1,x_2]}{x_3-x_1}$	
		$f[x_2, x_3] = \frac{f[x_3]-f[x_2]}{x_3-x_2}$		$f[x_1, x_2, x_3, x_4] = \frac{f[x_2,x_3,x_4]-f[x_1,x_2,x_3]}{x_4-x_1}$
x_3	$f[x_3]$		$f[x_2, x_3, x_4] = \frac{f[x_3,x_4]-f[x_2,x_3]}{x_4-x_2}$	
		$f[x_3, x_4] = \frac{f[x_4]-f[x_3]}{x_4-x_3}$		
x_4	$f[x_4]$			

Figure 3.2: Calculation for Divided Differences [9]

Using the given data for ppm values, we determined that the third divided difference was the most appropriate. Thus, we chose to fit a third-degree polynomial to the data, with the equation

$$g(y) = ay^3 + by^2 + cy + d \qquad (3.3)$$

where g(y) is the ppm, *y* is the year, and *a*, *b*, *c*, and *d* are constants that we must find.

2. After finding the polynomial, determine the constants. Our goal for this model is to minimize S, where

$$S = \sum_{i=1}^{m} [g_i - (ay_i^3 + by_i^2 + cy_i + d)]^2 \qquad (3.4)$$

and m is the total number of years in question. In order to do so, we need to ensure that the partial derivative of each of the variables with respect to S is 0, which is represented by equation 3.5.

$$\frac{\partial S}{\partial a} = \frac{\partial S}{\partial b} = \frac{\partial S}{\partial c} = \frac{\partial S}{\partial d} = 0. \qquad (3.5)$$

Given that g_i must equal $(ay_i^3 + by_i^2 + cy_i + d)$ to achieve partial derivatives of 0, we can set up the following systems of equations to determine our a, b, c, and d values.

$$(\sum y_i^3)a + (\sum y_i^2)b + (\sum y_i)c + md = \sum g \qquad \text{(System 2)}$$

$$(\sum y_i^4)a + (\sum y_i^3)b + (\sum y_i^2)c + (\sum y_i)d = \sum y_i g_i$$

$$(\sum y_i^5)a + (\sum y_i^4)b + (\sum y_i^3)c + (\sum y_i^2)d = \sum y_i^2 g_i$$

$$(\sum y_i^6)a + (\sum y_i^5)b + (\sum y_i^4)c + (\sum y_i^3)d = \sum y_i^3 g_i$$

To ensure greater simplicity of the model, we set the first year of data recorded, 1959 to 0, and increased the y-value of the data by one for each additional year. We then substituted the values from our data to achieve the following systems of equations.

$$3814209a + 81375b + 1953c + 63d = 22512.33$$

$$190694175a + 3814209b + 81375c + 1953d = 731505.83$$

$$9930928833a + 190694175b + 3814209c + 81375d = 31234882.39$$

$$(5.31946E + 11)a + 9930928833b + 190694175c + 3814209d = 1486641434$$

Solving the systems gets us a, b, c, and d values of 0.00004124, 0.009202, 0.9, and 315.054 respectively. Therefore, the empirically cubic model is given by

$$g(y) = 0.00004124y^3 + 0.009202y^2 + 0.9y + 315.054.$$

3. Use the model to predict future ppm values. The same 5-year summary presented in 3.2.1 is given below.

Year	Predicted CO_2 ppm
2025	426.40
2050	504.24
2075	607.66
2091	689.056
2100	740.52

Table 3.4: Predicted CO_2 Values for Cubic Divided Differences Model

3.2.4 Divided Differences By Sector and Population

Although analyzing historical data to predict future trends is useful, such analysis does not take into account two key factors: variation of emissions by sector and population growth. In order to account for these factors, we developed a new model, which draws a correlation between carbon dioxide emissions and carbon dioxide ppm. To begin, we gathered historical data for global, per-capita carbon dioxide emissions by sector from 1990 to 2019 [10] for eight sectors: buildings, electricity and heat, fugitive emissions, industry, aviation and shipping, land-use change and forestry, transport, and manufacturing and construction. In addition, we collected data for both historic and predicted global population growth based on United Nations reports [11]. We noticed in these population trends that the global population is projected to reach a maximum by 2086, but instead of decreasing the population from there, we held the 2086 population value constant for all future years, assuming the human population would remain at the carrying capacity. Next, we used the divided differences method above to extend the per-capita emissions into the future for each of the eight sectors and took their sum to get total per-capita emissions from the year 1990 onward. Finally, we multiplied the total per-capita emissions by the population to get total carbon dioxide emissions, and plotted it against the given ppm values from 1990 to 2021. The graph is shown below in Figure 3.3.

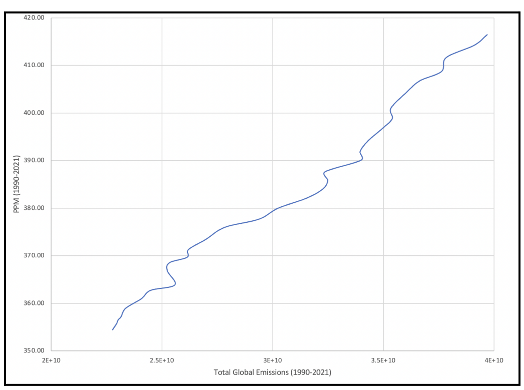


Figure 3.3: Total Global Emissions vs. PPM

Using the values corresponding to this figure, we applied the divided differences model again to achieve a quadratic equation that can be used to predict the future ppm values based on the projected carbon emission values, which is presented below

$$\hat{p}(e) = (5 * 10^{-20})e^2 + (6.7 * 10^{-10})e + 318.69$$

where $\hat{p}(e)$ is the predicted ppm values, and e is the projected carbon emission value. Using this quadratic, and the calculations described above, we achieve the following results for ppm projects, which closely resemble our projection for the weighted moving average model.

Year	Predicted CO_2 PPM
2025	435.45
2050	520.44
2075	601.89
2100	667.46
2107	686.08

Table 3.5: Predicted PPM Values for Sector and Population Model

3.2.5 Autoregressive Integrated Moving Average (ARIMA)

For a non-stationary time series where the mean, variance, and autocorrelation change over time, the ARIMA model is a useful way to understand the data and predict future values because of its ability to apply an initial differencing step to eliminate non-stationarity.

Since ARIMA is a linear regression model and works best when predictors are independent of each other, we first need to differentiate the data. To determine the degree to which the data must be differenced, we apply the Augmented Dickey-Fuller (ADF) test [12], where the null hypothesis is that the time series is non-stationary and the p-value is the level of marginal significance. Only when the p-value is less than 0.05, we reject the null hypothesis and conclude that the time series is stationary. The p-value of the original series is 1.0, so we continue to differentiate. The first difference has a p-value of 0.8307, and the second difference has a p-value of 6.533×10^{-7}, so the null hypothesis is rejected and CO_2 levels are stationary. This conclusion is further supported by the autocorrelation (ACF) graphs in Figure 3.4.

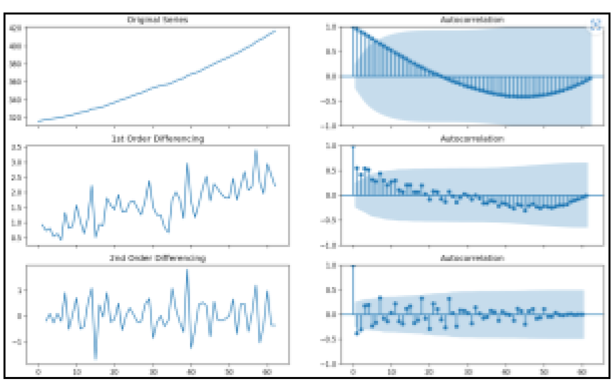


Figure 3.4: CO_2 Concentration ACF Graphs

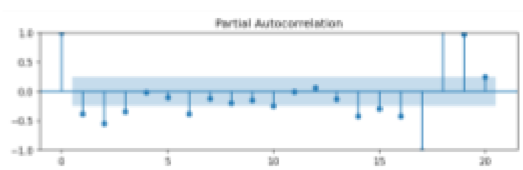


Figure 3.5: PACF Graph for 2nd Difference

After finding the order of differencing and making the data stationary, we build the ARMA(p,q) model where p is the number of autoregressive (AR) or past values used and q is the number of moving average (MA) or past white noise terms. We use the formula

$$y_t = C + \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - ... - \theta_q \varepsilon_{t-q} \quad (3.6)$$

where y is the time series, t is time, C is a constant, ϕ_1, ϕ_2... are AR coefficients, θ_1, θ_2... are MA coefficients, and ε is white noise [13]. Based on the AIC/BIC and ACF/PACF (See Figure 3.5 above) criteria, we choose ARMA(2,1) as our model. Combining the ARMA model with the differencing of degree 2 gives the ARIMA (2,2,1) model. The calculation results are shown in Figure 3.6.

```
                               ARIMA Results
==============================================================================
Dep. Variable:                      y   No. Observations:                   63
Model:                 ARIMA(2, 2, 1)   Log Likelihood                 -41.212
Date:                Fri, 04 Nov 2022   AIC                             90.424
Time:                        12:27:35   BIC                             98.867
Sample:                             0   HQIC                            93.733
                                 - 63
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.1946      0.160     -1.214      0.225      -0.509       0.120
ar.L2         -0.3644      0.170     -2.138      0.032      -0.698      -0.030
ma.L1         -0.5796      0.171     -3.394      0.001      -0.914      -0.245
sigma2         0.2219      0.041      5.356      0.000       0.141       0.303
```

Figure 3.6: ARIMA Results

From the Ljung-Box test, we obtain a p-value of 0.17, so the model is a good fit. Then, we use the ARIMA(2,2,1) to predict. The original data and predicted future values are shown in Figure 3.7 (where year 0 represents 1959) and Table 3.6.

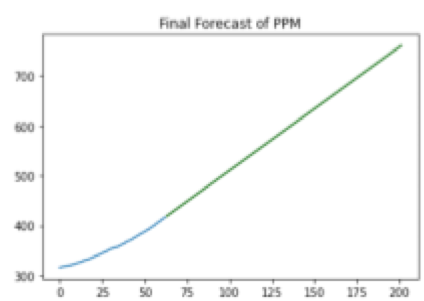


Figure 3.7: ARIMA CO_2 PPM Forecast

Year	Predicted Carbon Concentration (PPM)
2025	426.41
2050	488.35
2075	550.26
2100	612.18
2130	686.48

Table 3.6: ARIMA PPM Prediction Values

The residual plot is shown below in Figure 3.8. Since the residuals are randomly dispersed around the x-axis, the ARIMA model is a good fit.

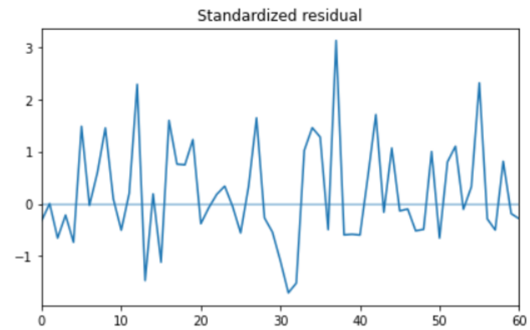


Figure 3.8: Residual Values

According to the model, the CO_2 concentration level will not reach 685 ppm by 2050. Instead, the model forecasts the first year of carbon concentration greater than 685 ppm to be 2130.

3.3 Comparative Model Analysis

Granted that each one of our models staunchly deviated from the prediction that ppm values would reach 685 by the year 2050, we do not believe that the claims and predictions are accurate. Rather, we find that under the assumption no significant action is taken to counteract emission increases, ppm concentration will reach 685 ppm far past the year 2050.

Out of the models from 3.2, the divided differences model appears to be the most accurate model. First, the weighted moving average and ARIMA models, due to their use of previous values to predict future ones, tend to show a much slower exponential increase, which we believe does not predict values as accurately in the future. In addition, these models cannot fully capture historical trends because they only take into account a certain number of lagged values. Although one of the strengths of these models is that they can adjust themselves to fit the data better, this strength turns into a limitation when forecasting in the long term. With the divided differences by sector and population, the future values are based on many theoretical values that were determined through modeling, including the future population and per-capita CO_2 emissions. As a result, although there is more complexity in the model, there is also more likelihood of error, making the values more uncertain.

On the other hand, the divided differences model shows a much faster exponential increase, which aligns more closely with our expectations for carbon concentration growth. However, it also has potential sources of inaccuracy. For example, the model may be overfitting the trend, leading to unwanted oscillations and increased sensitivity to data errors. Nevertheless, it accounts for all historical trends and it minimizes the sum of squared residuals (SSR), leading us to believe that it would be the best in representing the data and predicting future values.

4 Relating CO_2 Emissions and Temperature

4.1 Predicting Future Temperature Changes

4.1.1 Bollinger Bands Model

In the stock market, Bollinger Bands are used to track changes, as they examine where the stock price lies with respect to the upper (pressure) and lower (support) Bollinger Bands, which act as reasonable predictors for the trajectory of the market. Due to the small nature of the values, we found it appropriate to apply to temperature change, with the equation for the upper and lower bands as follows.

$$A_m = \gamma \cdot \ stdev$$

where A_m is the moving average with a rolling window of 4 years, γ is the standard deviation factor, and *stdev* is the standard deviation of each moving average. We chose to use a γ value of 2, as it provides an appropriate range for all the data to fit within. The Bollinger Bands (black) plotted with the change in degrees Celsius (red) relative to the base year 1951-1980 are presented below in Figure 4.1.

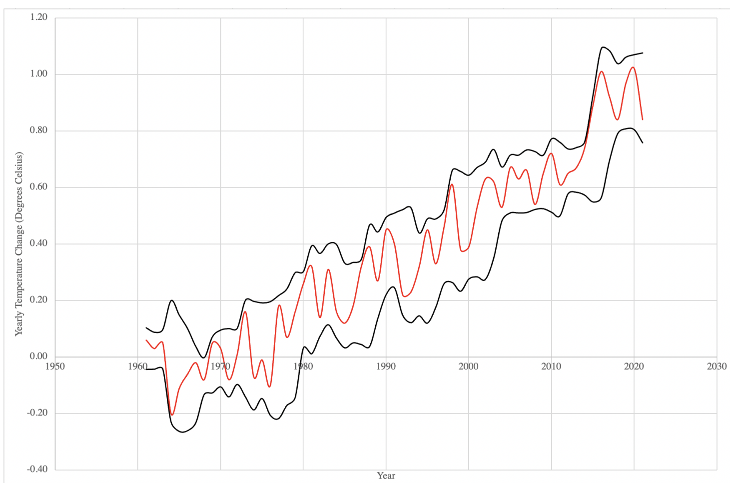


Figure 4.1: Bollinger Bands and Change in Degrees

As the focus of Bollinger bands in business is to analyze the relative position of the stock price in relation to the Bollinger band, we found that two methods, evaluating the absolute distance and evaluating the integral using trapezoidal Riemann sums, were the best to make future predictions for the data. In addition, we decided to further investigate this method by splitting them across the upper and lower bands, as the lower band, due to its smaller values, grows at a rate faster than the upper band.

Before implementing these methods, however, we must extend the Bollinger Bands themselves into the future. To do so, the strategy from the divided differences method was utilized, and the following two quadratic equations were used to model the future Bollinger Band structures.

$$\widehat{ubb}(y) = 0.00011208y^2 - 0.429485y + 411.267 \qquad (4.1)$$

$$\widehat{lbb}(y) = 0.00021471y^2 - 0.838216y + 817.883 \qquad (4.2)$$

where y is the year in question, $\widehat{ubb}(y)$ and $\widehat{lbb}(y)$ are the predicted upper and lower Bollinger bands for years past 2021.

For the absolute difference model using upper Bollinger bands, we first determined the average difference between the upper Bollinger and the change in degrees between the years 1961 to 2021, which was approximately 0.1278996. Then, we subtracted that average from the predicted upper Bollinger band values, which gave us the following equation for future changes in temperature.

$$\hat{u}(y) = \widehat{ubb}(y) - 0.1278996$$

where $\hat{u}(y)$ is the predicted change in temperature for the year in question. We applied a similar process for the lower Bollinger band, except that the average difference of 0.1729816 was between the change in degrees and the lower Bollinger band. Adding the average to the predicted lower Bollinger band values gave us the following equation for future changes in temperature

$$\hat{l}(y) = \widehat{lbb}(y) + 0.1729816$$

where $\hat{l}(y)$ is the predicted change in temperature for the year in question. The results displaying when each of these models surpasses the temperatures change of 1.25, 1.5, and 2 degrees Celsius is shown in table 4.1 below.

Change in Temperature (°C)	Year (UBB method)	Year (LBB Method)
+1.25	2034	2031
+1.5	2043	2038
+2	2060	2051

Table 4.1: Absolute Predicted Change in Temperature

Transitioning to the trapezoidal Riemann sums, we used a similar split process for the upper and lower Bollinger Bands, except instead of using absolute difference, we used trapezoidal Riemann sum integrals. Starting with the upper Bollinger Band, the trapezoidal area between the upper Bollinger Band and the change in degrees was calculated for each set of two subsequent years between 1961 and 2021 using the following equation

$$\int_{y}^{(y+1)} (Upper\ Bollinger\ Band - Change\ in\ Degrees) = \left(\frac{(ubb(y+1) - d(y+1)) + (ubb(y) - d(y))}{2}\right) \qquad (4.3)$$

where (y+1) is the next year and y is the current year, ubb is the value of the upper Bollinger band, and *f* is the change in degrees for the corresponding year. Given that these values had a uniform distribution, we used the average value of these individual integrals, 0.127698, as a constant for all future values. Substituting this value for our integral, we solved for the $d(y + 1)$ value, which gave us the following equation.

$$d(y + 1) = (-1) * ((2 * 0.127698) * (ubb(y) - d(y)) - ubb(y + 1)) \tag{4.4}$$

A similar process was carried out for the lower Bollinger Band, except that the trapezoidal area was found between the change in degrees and lower Bollinger Band for each set of two subsequent years between 1961 and 2021 using the following equation

$$\int_{y}^{(y+1)} (Change\ in\ Degrees - Lower\ Bollinger\ Band) = (\frac{(d(y+1) - lbb(y+1)) + (d(y) - lbb(y))}{2}) \tag{4.5}$$

where lbb is the value of the lower Bollinger Band. The average value of these individual integrals, 0.174323, was used as a constant for all future values, and substituting it for the left-hand side of the previous equation, we solved for the $lbb(y + 1)$ value, which gave us the following equation.

$$d(y + 1)) = (2 * 0.174323) - (d(y) - lbb(y)) + lbb(y + 1) \tag{4.6}$$

In both of these cases, the ubb(y), ubb(y+1), lbb(y), and lbb(y+1) values for years beyond 2021 were taken from equations 4.1 and 4.2. The table below summarizes our findings for both trapezoidal Riemann sum methods.

Change in Temperature (ºC)	Year (UBB method)	Year (LBB Method)
+1.25	2030	2028
+1.5	2040	2034
+2	2056	2048

Table 4.2: Trapezoidal Riemann Sum Change Predictions

Although these four models present reasonable values, we believe that the trapezoidal Riemann sum using the upper Bollinger Band method provides the most accurate data because changes in the upper band are less likely to be exaggerated than they are for the lower band due to the range of values. Furthermore, the trapezoidal method adds complexity to the analysis of the position of the change in degrees, especially as it tends to hug the upper band in the 60-year time frame presented in figure 4.1.

4.2 Analyzing the Relationship Between CO_2 Emissions and Temperatures

4.2.1 Long-Term – Pearson Correlation Coefficient and Linear Regression

Our application of the Pearson Correlation Coefficient investigates the linear correlation between CO_2 concentrations and temperature changes, which is useful for finding the relationship between their growth rates in the long-term. We use the formula for the correlation coefficient *r*, which is

$$r = \frac{n\Sigma xy - \Sigma x \Sigma y}{\sqrt{[n\Sigma x^2 - (\Sigma x)^2][n\Sigma y^2 - (\Sigma y)^2]}} \qquad (4.7)$$

where *n* is the number of terms in the time series, and *x* and *y* are the two variables, which in this case are CO_2 concentration and temperature change, respectively. First, we substitute $n = 63$ since we start from 1959. Then, we substitute values of CO_2 into *x* and degrees Celsius into *y* to obtain $\Sigma y = 22.12$, $\Sigma x = 22,512.33$, $\Sigma y^2 = 14.33$, $\Sigma x^2 = 8,099,762.22$, and $\Sigma xy = 8,483.0381$. Substituting these values into Equation 4.5 gives

$$r = \frac{63(8{,}483.0381) - (22{,}512.33)(22.12)}{\sqrt{[63(8{,}099{,}762.22) - (22{,}512.33)^2[63(14.33) - (22.12)^2]}} \approx 0.961$$

Since the coefficient is positive and very close to 1, there is a strong positive linear correlation between the CO_2 concentration and temperature change. Therefore, the relationship can be modeled by a least squares regression line (LSRL). We use the LSRL equations,

$$\hat{y} = \hat{\beta}_1 x + \hat{\beta}_0 \qquad \text{(System 3)}$$

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} \qquad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$SS_{xx} = \Sigma x^2 - \frac{1}{n}(\Sigma x)^2 \qquad SS_{xy} = \Sigma xy - \frac{1}{n}(\Sigma x)(\Sigma y)$$

where $\hat{\beta}_1$ and $\hat{\beta}_0$ are the slope and y-intercept of the LSRL respectively, $\bar{x}$ and $\bar{y}$ are the means of the *x* and *y* values respectively, and *n* is the number of pairs in the data set. Substituting the same values used for the Pearson correlation model, $n = 63$, $\Sigma x = 22,512.33$, $\Sigma x^2 = 8,099,762.22$, $\Sigma y = 22.12$, $\Sigma y^2 = 14.33$, and $\Sigma xy = 8,483.0381$ gives

$$SS_{xx} = 8,099,762.22 - \frac{1}{63}(22,512.33)^2 = 55238.378$$

$$SS_{xy} = 8,483.0381 - \frac{1}{63}(22,512.33)(22.12) = 578.825$$

$$\hat{\beta}_1 = \frac{578.825}{55238.378} \approx 0.01048$$

Substituting $\bar{x} = 357.34$ and $\bar{y} = 0.351$, we get

$$\hat{\beta}_0 = 0.351 - 0.01048(357.34) \approx -3.39392$$

Thus, the model that relates CO_2 concentration *x* and temperature change *y* is $y = 0.01048x - 3.39392$, and the r^2 value is $0.961^2 \approx 0.924$, meaning very high percentage of the data can be explained by the LSRL.

4.2.2 Short-Term – Granger Causality and Vector Autoregression

Our application of the Granger causality model investigates causality between CO_2 concentrations and temperature changes, which is useful because if the two variables have a causal relationship, one comes before the other in the time series, allowing for the prediction of future values in the short-term to be more accurate [14]. However, to apply the test, the data must first be stationary. From the Augmented Dickey-Fuller Test in 3.2.5, CO_2 concentration is stationary after differentiating twice, and repeating the test with temperature changes gives that it is stationary after differentiating once. However, since the temperature data was differenced one less time, it has one more term, so the first term was removed so only the same years would be compared.

After differentiating, we apply the F-test from the Granger causality test. We start with the null hypothesis that time series x does not Granger cause time series y, or in other words, knowing the value of time series x at a certain lag is not useful for predicting the value of time series y at a later time period [15]. The F-test produces the p-value, and similar to in 3.2.5, a p-value of fewer than 0.05 means we reject the null hypothesis and conclude that time series x Granger causes time series y. We test whether the CO_2 concentration Granger causes the temperature change, and the only number of lags for which the null hypothesis is rejected is at 4 lags, which has a p-value of 0.0438. To ensure that it is not a case of reverse causation, we test if temperature Granger causes CO_2 concentration but obtain a p-value of 0.0688, so we fail to reject the null hypothesis. Therefore, CO_2 concentration Granger causes temperature change with a lag or order of 4. This falls between the expected value of 3 to 7 lags from ENSO in section 3.

With the results from the Granger causality model, we can use the vector autoregression model VAR(p) to model the relationship between CO_2 concentration and temperature change, where p is the order, or lag, of the autoregression [16]. The system of equations used for two time series variables $x_{t,1}$, $x_{t,2}$ is

$$x_{t,1} = C + \phi_{11}x_{t-1,1} + \phi_{12}x_{t-1,2} + \phi_{13}x_{t-2,1} + \phi_{14}x_{t-2,2} + ... + \phi_{1(2p-1)}x_{t-p,2} + \phi_{1(2p)}x_{t-p,2} + \varepsilon_{t,1}$$
$$x_{t,2} = C + \phi_{21}x_{t-1,1} + \phi_{22}x_{t-1,2} + \phi_{23}x_{t-2,1} + \phi_{24}x_{t-2,2} + ... + \phi_{2(2p-1)}x_{t-p,1} + \phi_{1(2p)}x_{t-p,2} + \varepsilon_{t,2}$$
(System 4)

where C is a constant, ϕ_{ab} is a coefficient for time series $x_{t,a}$, and ε is the white noise. From the Granger causality test, we obtain the VAR(4) model. The results can be seen below in Figure 4.2. The bias in temperature predictions testing the last 12 years was -0.0003 and that for ppm was -0.0150, so the model slightly overestimates the values but is still a good fit. [17]

Summary of Regression Results

Model:	VAR		
Method:	OLS		
Date:	Sat, 12, Nov, 2022		
Time:	20:17:07		
No. of Equations:	2.00000	BIC:	-5.78169
Nobs:	57.0000	HQIC:	-6.17613
Log likelihood:	39.4067	FPE:	0.00162615
AIC:	-6.42687	Det(Omega_mle):	0.00121289

Results for equation Degrees C D1

	coefficient	std. error	t-stat	prob
const	0.037695	0.014272	2.641	0.008
L1.Degrees C D1	-0.445070	0.160411	-2.775	0.006
L1.PPM D2	-0.026980	0.035438	-0.761	0.446
L2.Degrees C D1	-0.395085	0.178849	-2.209	0.027
L2.PPM D2	0.013533	0.043113	0.314	0.754
L3.Degrees C D1	-0.412391	0.180325	-2.287	0.022
L3.PPM D2	0.070539	0.040360	1.748	0.081
L4.Degrees C D1	-0.181409	0.171909	-1.055	0.291
L4.PPM D2	0.090098	0.032986	2.731	0.006

Results for equation PPM D2

	coefficient	std. error	t-stat	prob
const	0.123652	0.070862	1.745	0.081
L1.Degrees C D1	0.904093	0.796444	1.135	0.256
L1.PPM D2	-0.865389	0.175952	-4.918	0.000
L2.Degrees C D1	-1.141669	0.887990	-1.286	0.199
L2.PPM D2	-0.599495	0.214059	-2.801	0.005
L3.Degrees C D1	-1.277183	0.895317	-1.427	0.154
L3.PPM D2	-0.149799	0.200390	-0.748	0.455
L4.Degrees C D1	-1.323063	0.853532	-1.550	0.121
L4.PPM D2	0.146276	0.163776	0.893	0.372

Figure 4.2: VAR Model Results

4.3 Extending the Relationship Models

4.3.1 Long-Term Time Series (LTTS) Model - LSRL

With the LSRL from 4.2.1, $y = 0.01048x - 3.39392$, we continue to extend x into the future as our Long-Term Time Series (LTTS) Model. Table 4.3 shows the predictions of change in temperature based on the CO_2 concentration.

Carbon Concentration (ppm)	Temperature Change (ºC)
466.977	+1.5
514.687	+2
562.397	+2.5
685	+3.785

Table 4.3: LTTS Model Predictions

Because our model forecasts long-term growth, we believe that our model will be reliable for up to +2ºC, or around 2060 according to the Divided Differences model in section 3, due to the relatively high r^2 value demonstrating strong linear correlation. However, we believe that after reaching a higher concentration of CO_2, the trend will become more exponential because there is already so much heat being trapped in the atmosphere that the temperature will continue to rise even if the carbon concentration does not. Therefore, we are concerned with the model's ability to predict CO_2 concentrations or land-ocean temperatures far into the future.

4.3.2 Short-Term Time Series (STTS) Model - VAR

With the Vector Autoregression (VAR) model from 4.2.2, we can use the lagged values of both the CO_2 concentration and the temperature changes to predict future values in the short term. The predicted values for both variables when stationary and undifferentiated are shown below in Table 4.4.

Year	Predicted First Difference of Temperature (ºC)	Predicted Temperature Change (ºC)	Predicted Second Difference of CO_2 Concentration (ppm)	Predicted CO_2 Concentration (ppm)
2022	0.098	+0.938	0.176	418.84
2023	0.074	+1.012	0.449	421.67
2024	-0.037	+0.975	-0.249	424.26
2025	0.009	+0.984	-0.016	426.83

Table 4.4: VAR Predicted Values

Because our VAR model is short-term, we believe that it will only be accurate for one lag length, or 4 years. After one lag length, the model no longer has real data to base its forecasts, so both the first difference in temperature change and the second difference in CO_2 concentration converge towards a value between 0.0 and 0.1 (See Figure 4.3). Because the goal of the short-term model is to capture the nuances of the data, it will not be an accurate predictor if the values both converge to a value. The data for both variables are not expected to follow a perfect trend, so oscillation or variability in the predicted values is desired.

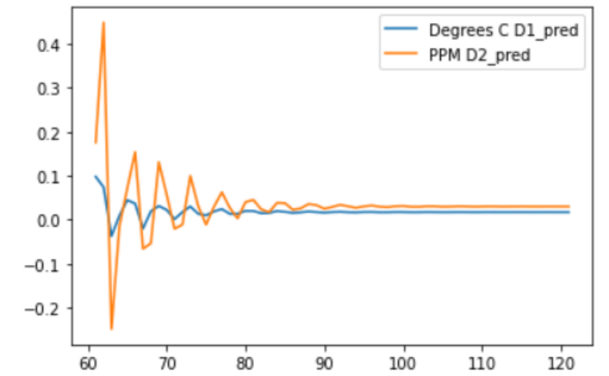


Figure 4.3: VAR Prediction Graph

5 Sensitivity Analysis

Despite the numerous models presented in this paper, we believe that three models in particular, ARIMA, Bollinger Bands, and VAR are important to perform sensitivity analysis towards, as they contain individual, easily-adaptable independent variables.

With ARIMA (2,2,1) from section 3.2.4, the values of p,d, and q were obtained through the optimization of AIC. Changing the value of d will make the data either not stationary or over-stationary, so only the values of p and q will be changed. The effects of different p and q values are shown in Table 5.1.

ARIMA Model (p, d, q)	Predicted CO_2 in 2050 (ppm)	ARIMA Model (p, d, q)	Predicted CO_2 in 2050 (ppm)
(1,2,1)	486.721	(2,2,2)	488.346
(3,2,1)	488.306	(2,2,4)	486.686
(4,2,1)	487.732	(2,2,5)	486.563
(5,2,1)	488.679	(2,2,6)	487.102

Table 5.1: ARIMA Predictions from Changing P and Q Values

Therefore, as p, the number of autoregressive terms, and q, the number of lagged errors, increase, the predicted concentration fluctuates within about 2 ppm of that of the (2,2,1) model. This follows our expectation because the data has random oscillations, and including more autoregressive or moving average terms will change how many terms each oscillation affects. However, because the oscillations are randomly distributed around the x-axis in the second difference of the CO_2 time series, we would expect them to overall not have a large impact on the overall model. Therefore, the ARIMA model is not very sensitive to parameter changes.

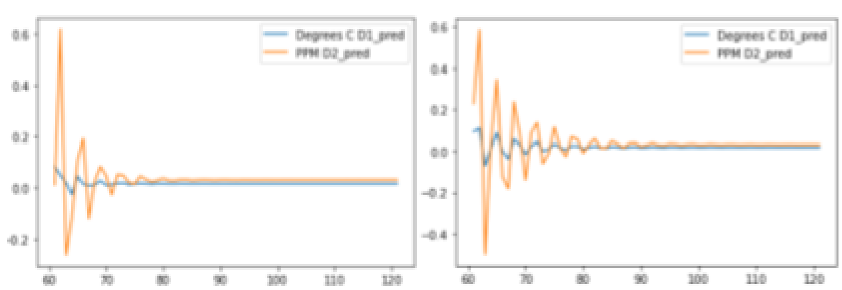
For the Bollinger Bands model from section 4.1.1, we chose to adapt two variables for the trapezoidal Riemann sum using the upper Bollinger Band model: the rolling window of the moving average and the type of model used to extend the band. Changing the rolling window impacts the strength of historic data in predicting future values, and changing the type of model used, in this case, an exponential model instead of a quadratic, affects the growth of the upper band. The results are summarized in table 5.2.

Change in Temperature (°C)	Rolling Window: 4, Quadratic Extension	Rolling Window: 2, Quadratic Extension	Rolling Window: 6, Quadratic Extension	Rolling Window: 4, Exponential Extension
1.25	2030	2026	2030	2030
1.5	2040	2036	2040	2038
2	2056	2052	2058	2052

Table 5.2: Predicted Years from Rolling Window, Extension Type

The results display that as the rolling window is decreased, the change in temperature gets faster, and as it is increased, the change in temperature takes more time to surpass specific thresholds. In addition, an exponential extension of the upper band speeds up the temperature changes in the long run, yet they are similar to the quadratic model in the short run. These outcomes agree with our expectations, as including more historic data tends to decrease the moving average, which draws the upper Bollinger band further down. In addition, an exponential model is expected to show rapid growth in the long run.

With VAR(4) from section 4.3.2, the order p was determined through Granger causality, so changing the value of p would change the Granger-causal relationship between the two variables, including whether they are helpful for forecasting each other’s future values, impacting the forecasts shown below.



VAR(3) VAR(5)

Figure 5.1: VAR Predictions from Changing P Value

Changing p also changes the amount of time that the VAR model can predict before it flattens to a constant value. Therefore, the VAR model is very sensitive to changes to its order.

6 Strengths and Limitations

6.1 Strengths

- **Multiple Methods.** We used multiple models (ARIMA, Divided Differences, Bollinger Bands, etc.) to address each problem, which reflects the broad range and depth of our considerations.
- **Adaptable.** Every model that was used for ppm or temperature can be applied to other sets of data. In addition, as ppm and temperature values are measured in the coming years, the models can incorporate them to increase the accuracy of future predictions.
- **Objectivity.** Our models, using the historic ppm and temperature change data, derive constants from the data itself, making them inherently objective.

6.2 Limitations

- **Specificity of Assumptions.** In the process of modeling, we assumed that trends found in the data would continue into the future. Therefore, we did not take into account potential events that could affect the rate of increase for CO_2 and temperature, such as important climate legislation or war. Including climate action is a fairly reasonable assumption that we could have incorporated into our models.
- **Lack of Total Consensus Among Models.** We used multiple models to address problems, but there was noticeable variability among their predictions. For example, the LSRL differed from the VAR model's prediction for ppm and temperature by a large amount. We could have improved this by conducting a more in-depth error analysis.

7 Conclusion

In the first section of our work, we disputed both claims brought forward regarding the past and future of CO_2 ppm levels by modeling the ways in which historic data correlates to future trends. We predicted that carbon dioxide is set to surpass 685 and 740 ppm by 2091 and 2100, respectively, as per our divided differences model. Then, we went on to analyze the trends with temperature change and discovered that the two variables have a correlational and Granger causational relationship with one another.

Even though our data may not inhibit worry amongst the population, it is important for people to be cognizant of their carbon footprint, especially as it relates to protecting those without the resources to diminish or negate the negative impacts that increasing carbon dioxide levels bring.

8 References

[1] *What Is Climate Change?* (n.d.). United Nations.
https://www.un.org/en/climatechange/what-is-climate-change

[2] *Effects of Climate Change*. (n.d.). World Wild Life. https://www.worldwildlife.org/threats/effects-of-climate-change#:~:text=More%20frequent%20and%20intense%20drought,on%20people%27s%20livelihoods%20and%20communities.&text=As%20climate%20change%20worsens%2C%20dangerous,becoming%20more%20frequent%20or%20severe
[3] Falconer, R. (2021, December 27). Study: *10 biggest climate disasters of 2021 cost $170 billion*. Axios. Retrieved November 13, 2022, from https://www.axios.com/2021/12/27/Climate-disasters-cost-world-billions-2021-study
[4] *The economics of climate change*. (2021, April 22). Swiss Research Institute. https://www.swissre.com/institute/research/topics-and-risk-dialogues/climate-and-natural-catastrophe-risk/expertise-publication-economics-of-climate-change.html
[5] *Climate Effects on Health*. (2022, April 25). Center for Disease Control. https://www.cdc.gov/climateandhealth/effects/default.htm#:~:text=The%20health%20effects%20of%20these,and%20threats%20to%20mental%20health
[6] *Carbon dioxide now more than 50% higher than pre-industrial levels*. (2022, June 3). National Oceanic and Atmospheric Administration. https://www.noaa.gov/news-release/carbon-dioxide-now-more-than-50-higher-than-pre-industrial-levels
[7] Friedrich, J., & Damassa, T. (2014, May 21). *The History of Carbon Dioxide Emissions*. World Resource Institute. https://www.wri.org/insights/history-carbon-dioxide-emissions
[8]U.S. Forest Service. (n.d.). *Natural Climate Cycles*. United States Department of Agriculture. Retrieved November 2, 2022, from https://www.fs.usda.gov/ccrc/education/climate-primer/natural-climate-cycles#:~:text=Interannual%20to%20Decadal%20Climate%20Cycles&text=ENSO%20events%20occur%20every%203,the%20world%20(NASA%202009)
[9] *Numerical Analysis/Divided differences*. (2021, March 31). Wikiversity. https://en.wikiversity.org/wiki/Numerical_Analysis/Divided_differences
[10] Roser, M., & Rodés-Guirao, L. (n.d.). *Future Population Growth*. Our World in Data. https://ourworldindata.org/future-population-growth
[11] Ritchie, H., & Roser, M. (n.d.). *Emissions by sector*. Our World In Data. https://ourworldindata.org/emissions-by-sector
[12] Pathak, P. (2020, October 29). *How to Create an ARIMA Model for Time Series Forecasting in Python*. Analytics Vidhya. Retrieved November 6, 2022, from https://www.analyticsvidhya.com/blog/2020/10/how-to-create-an-arima-model-for-time-series-forecasting-in-python/
[13] Prabhakaran, S. (2021, August 22). *ARIMA Model – Complete Guide to Time Series Forecasting in Python*. Machine Learning +. Retrieved November 4, 2022, from https://www.machinelearningplus.com/time-series/arima-model-time-series-forecasting-python/
[14] Glen, S. (n.d.). *Granger Causality: Definition, Running the Test*. Statistics How To. Retrieved November 12, 2022, from https://www.statisticshowto.com/granger-causality/
[15] Zach. (2021, November 29). *How to Perform a Granger-Causality Test in Python*. Statology. Retrieved November 13, 2022, from https://www.statology.org/granger-causality-test-in-python/
[16] Maitra, S. (2019, October 7). *Forecasting using Granger's Causality and Vector Auto-regressive Model*. Towards Data Science. Retrieved November 10, 2022, from

https://towardsdatascience.com/granger-causality-and-vector-auto-regressive-model-for-time-series-forecasting-3226a64889a6
[17] Brownlee, J. (2017, February 1). *Time Series Forecasting Performance Measures With Python*. Machine Learning Mastery. Retrieved November 8, 2022, from https://machinelearningmastery.com/time-series-forecasting-performance-measures-with-python/

9 Appendix

9.1 ARIMA Python Code

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
df1 = pd.read_csv('2022_HiMCM_Data.csv')
df = df1.iloc[:,[0]]
from statsmodels.tsa.arima.model import ARIMA
import pmdarima as pm
from pylab import rcParams
model = pm.auto_arima(df.PPM, start_p = 1, start_q = 1,
                      test = 'adf',
                      max_p = 5, max_q = 5,
                      m = 1,
                      d = None,
                      seasonal = False,
                      start_P = 0,
                      D = 0,
                      trace = True,
                      error_action = 'ignore',
                      suppress_warnings = True,
                      stepwise = True)
print(model.summary())
model.plot_diagnostics(figsize=(15,9))
plt.show()
n_periods = 129
fc, confint = model.predict(n_periods = n_periods, return_conf_int = True)
index_of_fc = np.arange(len(df.PPM),len(df.PPM) + n_periods)

fc_series = pd.Series(fc, index = index_of_fc)


plt.plot(df.PPM)
plt.plot(fc_series, color = 'darkgreen')
plt.title("Final Forecast of PPM")
plt.show()
```

9.2 VAR Python Code

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import warnings
warnings.filterwarnings('ignore')
%matplotlib inline
df1 = pd.read_csv('HiMCM.csv')
df = df1.iloc[:,[0,1]].dropna()
df.head()
from statsmodels.tsa.stattools import adfuller
n_obs = 12
X_train, X_test = df[0:-n_obs], df[-n_obs:]
x = df.index
y1 = df['PPM D2']
y2 = df['Degrees C D1']
fig, ax1 = plt.subplots(1,1,figsize = (10,6),dpi = 80)
ax1.plot(x,y1,color = 'tab:red')

ax2 = ax1.twinx()
ax2.plot(x,y2,color = 'tab:blue')

ax1.set_xlabel('Year', fontsize = 16)
ax1.tick_params(axis = 'x', rotation = 0, labelsize = 12)
ax1.set_ylabel('PPM', color = 'tab:red', fontsize = 20)
ax1.tick_params(axis = 'y', rotation = 0, labelcolor = 'tab:red')
ax1.grid(alpha = 0.4)

ax2.set_ylabel("Degrees C", color = 'tab:blue', fontsize = 16)
ax2.tick_params(axis = 'y', labelcolor = 'tab:blue')
ax2.set_xticks(np.arange(0,len(x),60))
ax2.set_xticklabels(x[::60],rotation = 90, fontdict={"fontsize":10})
ax2.set_title("PPM and Degrees C", fontsize = 18)
fig.tight_layout()
plt.show()
```

```
import statsmodels.api as sm
from statsmodels.tsa.stattools import grangercausalitytests
from statsmodels.tsa.vector_ar.var_model import VAR
grangercausalitytests(df, maxlag = 19, verbose = False)

mod = VAR(df)
res = mod.fit(4)
print(res.summary())
lag_order = res.k_ar
print(lag_order)
input_data = X_train.values[-lag_order:]
pred = res.forecast(y = input_data, steps = n_obs)
pred = (pd.DataFrame(pred, index = X_test.index, columns = X_test.columns + '_pred'))
pred.plot()
pred1 = []
pred2 = []
expect1 = []
expect2 = []
for forecast in pred['Degrees C D1_pred']:
    pred1.append(forecast)
for forecast in X_test['Degrees C D1']:
    expect1.append(forecast)
Temp_errors = [expect1[i]-pred1[i] for i in range(len(expect1))]
Temp_bias = sum(Temp_errors) * 1.0/len(expect1)
print("Temperature bias: " + str(Temp_bias))
for forecast in pred['PPM D2_pred']:
    pred2.append(forecast)
for forecast in X_test['PPM D2']:
    expect2.append(forecast)
PPM_errors = [expect2[i]-pred2[i] for i in range(len(expect2))]
PPM_bias = sum(PPM_errors) * 1.0/len(expect2)
print("PPM bias: " + str(PPM_bias))
input_data = df.values
pred = res.forecast(y = input_data, steps = 61)
pred = (pd.DataFrame(pred, index = df.index + 61, columns = df.columns + '_pred'))
print(pred)
pred.plot()
```