Exploring the Relationship Between Carbon Dioxide and the Temperature of the Earth

Bryce Drynan

May 4, 2023

Abstract

My team has found that the National Oceanographic and Atmospheric Association and the Scripps Institution of Oceanography are wrong. We have found that the largest 10 year average increase in PPM happens in 2003, not 2004. We have also found that the Organization for Economic Co-Operation and Development severely overestimated the amount of CO_2 in the atmosphere by 2050. They predict that there will be 685 ppm of CO_2 in 2050, our model does not see that concentration until the year 2133. We have also constructed a model to relate the change in the concentration of CO_2 with the difference in Degrees Celsius from an average of the base period 1951-1980 and forecast future results. We have found that using a VAR model has yielded us the best results, with the temperature difference crossing the $+1.25°C$ threshold in 2031, the $+1.5°C$ threshold in 2039, and the $+2.0°C$ threshold in 2051. We believe that these facts, regardless of the prior institutions being wrong, call for immediate action on CO_2 consumption to curb global warming.

Introduction

Prior to the Industrial Revolution carbon dioxide (CO_2) was consistently around 280 parts per million (ppm). In March of 2004 the atmosphere reached a concentration of 377.7 ppm [1]. This was the largest 10 year average increase of CO_2 in recorded history. According to scientists from the National Oceanographic and Atmospheric Association (NOAA) and Scripps Institution of Oceanography (SIO) the monthly mean CO_2 peaked in March 2022 at 422 ppm. A model constructed by the Organization for Economic Co-Operations and Development (OCED) report predicts a CO_2 level of 685 ppm by the year 2050.

The editors of *Scientific Today* have tasked my team to address the claims made by NOAA and SIO about the current concentration of CO_2 in the atmosphere, as well as OCED's future predictions. We have been provided with two data sets. The first data set is called the CO_2 Data Set 1[2]. It contains the annual month of March averages of CO_2. The data points are expressed as mole fraction in dry air, micromole/mol, abbreviated as ppm (parts per million). This data was collected at the Mauna Loa Observatory, Hawaii, USA through continuous air samples.[Data Credit] The second data set that we have been provided is the Temps Data Set 2[3]. This data set contains the global annual mean surface-air temperature change in degrees Celsius. It is based on land and ocean data compared to the temperature mean of the base period 1951-1980. For example, in 2021 the global land and sea temperature was 0.84°C above temperature mean of the base period 1951-1980.[Data Credit]

In addition to the requests of *Scientific Today* we will also be exploring the relationship between the increase of CO_2 in our atmosphere and its temperature. The field of study this topic falls under is modernly know as the study of global warming. This term was coined in 1975 by a University of Columbia geochemist, Wallace Broeker. It is important to note the difference between global warming and climate change since they are often confused. According to the National Aeronautics and Space Administration (NASA), global warming is the study of the increase in Earth's average surface temperature due to rising levels of greenhouse gasses. Climate change is the long-term change in the the earths climate, or a region of the Earth's climate[4]. This area of study is so important that the President of the United States' Discretionary Budget for the the 2023 Fiscal Year (FY) allocates $44.9 billion to tackle the climate crisis. This is $16.7 billion more that in the 2021 FY, a roughly 50% increase [5].

To complete these objectives the analysis will proceed through X parts:

1. Use a 10 year moving average of the year to year increase to analyze whether the 10 year average temperature increase was largest in 2004.
2. Create a time series model to predict the ppm of CO_2 in 2050 and 2100.
3. Build a model to predict when the average land-ocean temperature will hit +1.25°C, +1.50°C, +2.00°C.
4. Analyze the relationship between CO_2 and surface temperature.
5. Interpret the results and provide my conclusions.

Discussion

To begin we must define certain assumptions that allow our model to be as accurate as possible. We assume that

- The data in the second table Temps Data Set 2 has found the average temperature of the year beginning on March 1^{st} of year x and ending in March 1^{st} of year $x+1$.

- The temperature data was collected from sources world wide, making its use with the CO_2 samples at the Mauna Loa Observatory appropriate and acceptable.

- The data that OCED and NOAA/SIO used to make their predictions is the same.

- Human production of CO_2 increases at them same rate as 1959-2022.

The first person to accurately measure carbon dioxide in the atmosphere was David Keeling [6]. His innovations with continuously operating infrared spectroscopy and the very accurate manometric calibration of reference gasses is much the same technology that we continue to use to this day to measure carbon dioxide[1]. This was the technique that the Mauna Loa Observatory used up until 2019. Since April 2019 a new method called Cavity Ring-Down Spectroscopy (CRDS) is being used, therefore this only affects data points from 2020 on wards. The difference between these two methods is that CRDS looks at the rate of absorption, rather than the magnitude of absorption, of light circling in an optical cavity.[7] The Mauna Loa Observatory has certified that the change in technique for determining CO_2 has not invalidated the prior data, just made the data moving forward more accurate.

In order to look at mean average increase over a 10 year period we need to create a feature in the data for the year to year increase. This feature is known in time series analysis as a lag of size one. It is calculated by taking the difference of the current term minus the previous.

$$d_t = y_t - y_{t-1} \tag{1}$$

In Equation 1, t is the current year, d_t represents the difference at year t, y_t the data point at year t and y_{t-1} data point of the year before. In Figure 1 we can see what the new data looks like. The reasoning for the first term being 0 is because we have no previous term to be able to calculate the difference with. Now we will compute the right aligned moving average for the data. The reason for picking the right aligned moving average is because we are interested in the average of the previous 10 years.

$$ma_{n,t} = \frac{\sum_{i=0}^{n-1} y_{t-i}}{n} \tag{2}$$

Year	PPM	PPM.diff
<int>	<dbl>	<dbl>
1959	315.98	0.00
1960	316.91	0.93
1961	317.64	0.73
1962	318.45	0.81
1963	318.99	0.54
1964	319.62	0.63
1965	320.04	0.42
1966	321.37	1.33
1967	322.18	0.81
1968	323.05	0.87

Figure 1: Example of Data with difference column added as PPM.diff.

Year	PPM	PPM.diff	PPM.10.ma
<int>	<dbl>	<dbl>	<dbl>
1969	324.62	1.57	0.864
1970	325.68	1.06	0.877
1971	326.32	0.64	0.868
1972	327.46	1.14	0.901
1973	329.68	2.22	1.069
1974	330.19	0.51	1.057
1975	331.13	0.94	1.109
1976	332.03	0.90	1.066
1977	333.84	1.81	1.166
1978	335.41	1.57	1.236

Figure 2: Example of Data with 10 yr Moving Averages with results in columns PPM.10.ma

In Equation 2, t represents the year, n is the amount of years prior we want to include in one calculation, $ma_{n,t}$ represents the moving average of size n at year t, and y_{t-i} is the data point at year $t-i$. The reason for the moving average being of size 10 is because of the claim made by NOAA and SIO being change in the 10 year average of the change in PPM.

We can see an example of the results in Figure 2. Note that when computing the moving average we lose data points for 1959-1968. This is because we cannot compute the MA and difference operations for these terms as they require data going out of our domain. From this point we find the largest 10 year average increase. We can see in Figure 3 this occurs in 2003 with an average increase of 1.877 PPM every year over a period of 10 years.

Next we used time series analysis to construct a few models to make predictions with the data we have. In order to do any predictive modeling with time series we first need to use the KPSS unit root test to check if the data is stationary. We will do hypothesis testing to see if the data is

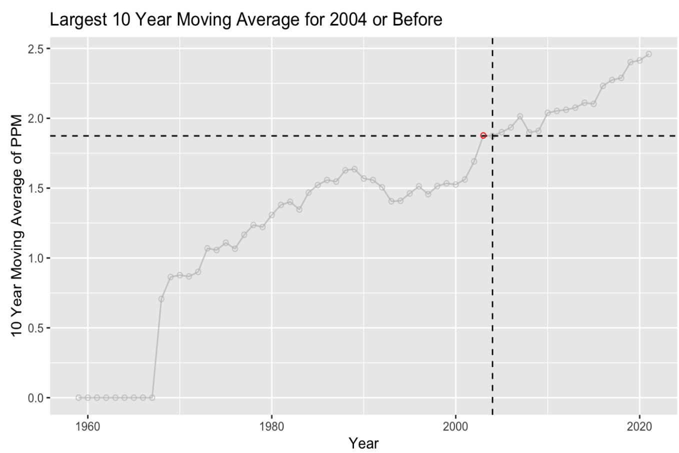

Figure 3: Values Larger Than the 10 Year Moving Average of 2004, Before 2004 (Highlighted in Red)

stationary or not.

$$H_0\text{: The data for PPM is stationary}$$
$$H_a\text{: The data for PPM is not stationary}$$

After computing the KPSS unit root test we got a test statistic of 1.6525 and a p-value $< .01$. Therefore we reject the null hypothesis at any reasonable confidence level and assume that the data is non-stationary. Since we have non-stationary data we will be using models that either make the data stationary or do not need the data to be stationary as a requirement.

The first approach that we will take is to use an exponential smoothing model. We used the RStudio programming language for this analysis. Within RStudio there is a library called forecast which contains a function called "ets()". This function takes an input of time series data. Then the function runs 8 models. These models are identified by a three-character string identifying method using the framework terminology of Hyndman et al. (2002) and Hyndman et al. (2008). The first letter denotes the error type ("A" or "M"); the second letter denotes the trend type ("N","A", or "M"); and the third letter denotes the season type ("N","A", or "M"). In all cases, "N"=none, "A"=additive, "M"=multiplicative and "Z"=automatically selected. Of these models we can specify what identification criteria we want the model to be chosen off of[8]. The first is AIC, which stands for Akaike

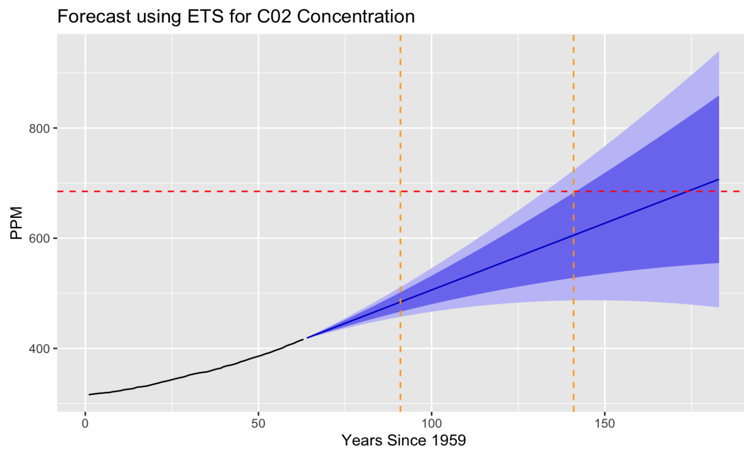

Figure 4: ETS(M,A,N) Forecasting of PPM

Information Criterion. This criterion looks at how well the model fits the data it is generated from. There is also the BIC, which stands for Bayesian Information Criterion. This criterion is derived from the AIC, but uses a Bayesian approach and log-likelihood to determine it. In both cases we will aim to minimize the criterion to determine the best model.

After running the ets function on the data for ppm, with both the AIC and BIC selection criteria we found that they both selected the same exponential model. That model was an ETS(M,A,N). Our model has multiplicative error, additive trend, and no seasonality to it. The model for this is

$$\hat{Y}_t = (l_{t-1} + b_{t-1}) + \alpha(Y_t - l_{t-1} - b_{t-1}) + \beta((l_{t-1} + b_{t-1}) - (l_{t-2} + b_{t-2})) + \epsilon_t \quad (3)$$

Where $\hat{Y}_t$ is the forecast for time t, l_{t-1} is the level estimate for the previous time period, b_{t-1} is the slope (trend) estimate for the previous time period, α is the smoothing parameter for the level estimate, β is the smoothing parameter for the slope estimate Y_t is the actual value at time t, ϵ_t is the error at time t. The model returned initial values of $l_0 = 315.157$, $b_0 = .8251$, $\alpha = .9767$, and $\beta = .213$. The graph in Figure 4 displays the 80% (blue) and 95% (light blue) confidence intervals for the forecasting, as well as the prediction line itself (dark blue). The dotted red line is 685 ppm, the first orange line is when the year is 2050 and the second orange dotted line is when the year is 2100. As we can see graphically in 2100 the ppm prediction is well below OCED's of 685, it is at 605.37 ppm. However, the upper bound

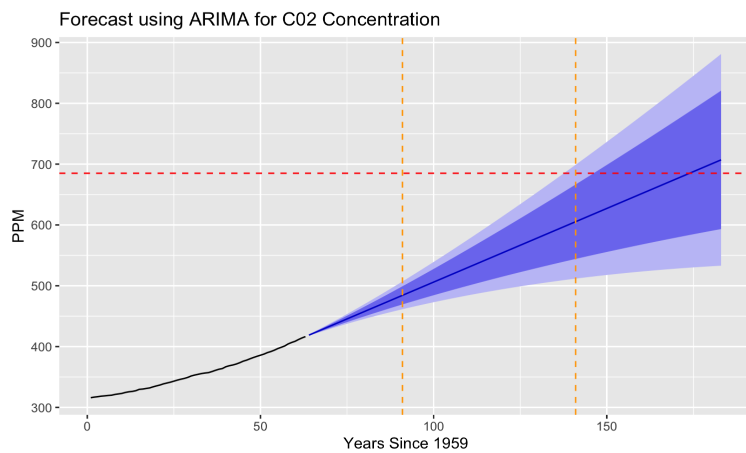

Figure 5: ARIMA(0,2,1) Forecasting of PPM

of the 80% confidence interval is at 682.63 ppm, almost to 685 ppm. In the year 2050 the prediction is at 487.27 ppm.

The next model that we looked at was an ARIMA model. This is an acronym for Auto-Regressive Integrated Moving Average model. This model will look at three parameters, p,d, and q. "p" is the amount of auto-regressive terms we want to look at, d is how many times we take the difference of the terms to make them stationary, and q is the amount of moving average terms we will use. We used the auto.arima() function in r to process our results. The identification criteria for this model is the same as the exponential smoothing model that we performed. We are looking to minimize either our AIC or BIC. The auto.arima() function will assess the data given to it, then run an iterative process, changing each parameter from 0 to 5, then analyze and looks for the lowest BIC or AIC, whichever is specified. After running our CO2 data through both AIC and BIC criterion, we noticed that they both returned the same model, an ARIMA(0,2,1), meaning that it has 2 difference steps and 1 moving average term. The graph in Figure 5 displays the 80% (blue) and 95% (light blue) confidence intervals for the forecasting, as well as the prediction line itself (dark blue). The dotted red line is 685ppm, the first orange line is when the year is 2050 and the second orange dotted line is when the year is 2100. This model reaches 685 ppm in the year 2133 at 685.23 ppm. In the year 2100 we should have a concentration of 605.32 ppm in the atmosphere and in 2050 we should see a concentration of 484.25

ppm.

If we look at the size of the confidence intervals (cone) in the ETS model we can notice that it is significantly larger as we get further into the forecast. This leads me to believe that the ARIMA model would perform better in the long run. However when looking at the mean average error (MAE) of the ETS model it is lower than the ARIMA, .3621 and .3706 respectively. Since the MAE is an unbiased parameter and more robust we will use it to choose the ETS model. Additionally the ETS model is largely considered more simple than an ARIMA model and that is another reason that we will choose the ETS model.

Moving on to looking at the temperature of the earth we will be applying the same ARIMA and ETS models as above. First we will look at the results of the ETS model. If we decide to set our criteria to AIC we get the model in Figure 6. If the criteria is BIC we get Figure 7. For the AIC criterion an ETS model of (A,A,N) was produced, meaning additive error, additive trend and no seasonality. For the BIC model it was a (A,N,N), meaning additive error, no trend and no seasonality. The BIC model is more commonly known as a simple exponential smoothing model. In order to pick the correct model we looked to see if the base data was stationary. If it is then we will choose the BIC model, if it is not we will choose the AIC model. We preform the same test as we did for the CO_2 data, the KPSS unit root test.

H_0: The data for Degrees.C is stationary
H_a: The data for Degrees.C is not stationary

After running the test we got a test statistic of 1.59 and a p-value$< .01$. This means we reject the null hypothesis and the data is not stationary. We therefore picked the model that incorporates the trend of the data or the ETS model that used AIC as its identification criterion. Next we run the auto.arima() function on the data. After running the function it returned two different models for the AIC and BIC criterion. The AIC model is an ARIMA(0,1,2) with drift (Figure 8). The addition of a drift term indicates that using just the data points does not allow the model to fully fit the data. A positive drift, which we achieved indicates a long term overall trend increase. The parameters returned from the model indicates that there are 0 auto-regressive terms, 1 difference taken and 2 moving average terms. The BIC model is an ARIMA(0,1,1) with drift (Figure 9). The parameters indicate that the model has 0 auto-regressive terms, 1 difference was taken and there is 1 moving average term. When looking at the error measures, mean

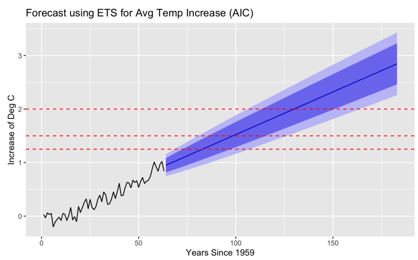

Figure 6: ETS(A,A,N) for Temperature with AIC Criterion

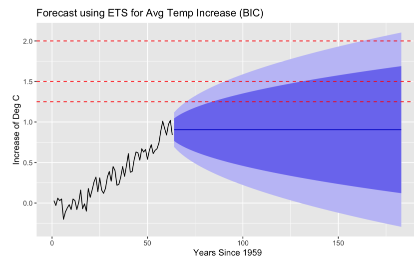

Figure 7: ETS(A,N,N) for Temperature with BIC Criterion

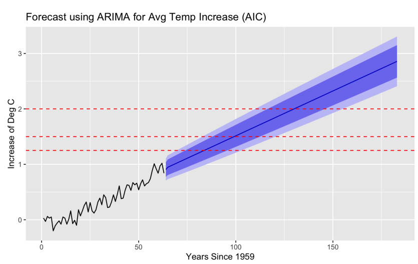

Figure 8: ARIMA(0,1,2) for Temperature with AIC Criterion

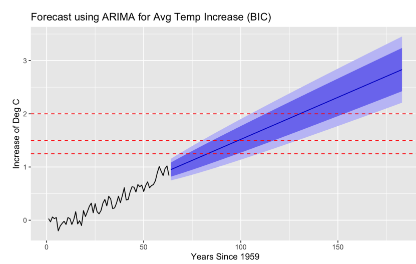

Figure 9: ARIMA(0,1,1) for Temperature with BIC Criterion

error (MA), root mean squared error (RMSE), mean average error (MAE), and mean percentage error (MPE), the AIC performs better, it only looses out to the BIC on the mean average percentage error (MAPE). Additionally the confidence intervals long term are narrower in the AIC model versus the BIC Providing a smaller possible window for the true value at time t is more promising than a wider one at the same significance levels.

Since the readability of the models is again similar as it was in the CO_2 case we will look to the error measures to compare the models. The ARIMA(0,1,2) (Figure 8) outperforms the ETS(A,A,N) model for Degrees.C because in RMSE, MAE, MPE, and MAPE the ARIMA model performs marginally, but consistently better, we select it to be our forecasting model. From this model we predict that we will cross the $+1.25°C$ threshold in 2042, the $+1.5°C$ threshold in 2058, and the $+2.0°C$ threshold in 2089.

After looking at the two models independently we wanted to see what, if any impact and correlation CO_2 has on the change in Degrees.C to the

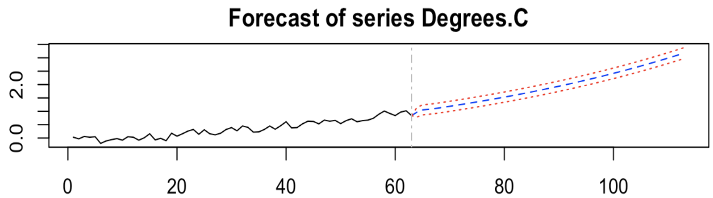

Figure 10: VAR(3) that regresses CO_2 PPM onto the Change in Degrees Celsius

base period. To do this we will analyze the data with a Vector Auto-Regressive Model (VAR). This model will return data similar to a Ordinary Least Squares regression. We used the vars library in r to compute the following results. within the vars library is a function called VARselect(). We used this function to calculate the optimal lag necessary for the best model. When run on the data it returns a value of 3, meaning the model will look at the previous 3 terms to get a result. Then we fit the data to the model using the VAR() function, then plotted the results and forecasting using the predict() function from r, the results are in Figure 10. The formula that describes the model and some of its attributes are in Figure 11. We can see the correlation of the PPM and Degrees.C vectors at a lag of 3 by applying a cross correlation function. This returns Pearson correlation coefficient of .82, which means the data is highly positively correlated.

Conclusion

NOAA and SIO have made the claim that 2004 saw the largest 10 year average increase in history up until that point. According to the model that my team and I have constructed we believe that the largest 10 year average increase happened in 2003 instead of 2004. This means that we believe they are a year late with their prediction. We also believe that the OECD are vastly over estimating their results. We are seeing a level of 487.27 ppm in 2050 and then 605.37 ppm in 2100. We do not hit OECD's forecast amount until the year 2133 with a level of 685.23 ppm. In both situations we see that the organizations are wrong. We acknowledge the fact that the assumption that human consumption of CO_2 remaining constant and not increasing decreases the accuracy of our model. However, when looking at

```
Estimation results for equation Degrees.C:
==========================================
Degrees.C = PPM.l1 + Degrees.C.l1 + PPM.l2 + Degrees.C.l2 + PPM.l3 + Degrees.C.l3 + const

             Estimate Std. Error t value Pr(>|t|)
PPM.l1       -0.03203    0.03870  -0.828    0.412
Degrees.C.l1  0.25716    0.17220   1.493    0.141
PPM.l2        0.05124    0.05095   1.006    0.319
Degrees.C.l2 -0.20163    0.17219  -1.171    0.247
PPM.l3       -0.00736    0.03350  -0.220    0.827
Degrees.C.l3 -0.09801    0.16100  -0.609    0.545
const        -3.80321    0.79308  -4.795 1.36e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1


Residual standard error: 0.09337 on 53 degrees of freedom
Multiple R-Squared: 0.9256,     Adjusted R-squared: 0.9172
F-statistic: 109.9 on 6 and 53 DF,  p-value: < 2.2e-16
```

Figure 11: VAR(3) that regresses CO_2 PPM onto the Change in Degrees Celsius Parameters and Features

the model of Tans (2008), our numbers for CO_2 are roughly in line with this model as well. When looking at the models we created to forecast the future temperature we had two distinctly different options. The first option is the ARIMA model and the second is the VAR(3) model. We will choose the VAR(3) model because when we applied a cross correlation the value it returned of .82 implies that the data is highly correlated. Thus give us a more informed model we select the VAR(3) model for Degrees.C. This model tells us that the temperature difference will cross the $+1.25°C$ threshold in 2031, the $+1.5°C$ threshold in 2039, and the $+2.0°C$ threshold in 2051.

Recommendations

Based on the findings of our models we believe that NOAA and the SIO should revise their model and adjust their claim to being 2003. We also believe that the OECD need to revise their model. With our model as well as the model, Scenario A and B, in Tans' paper we believe that the OECD has grossly overestimated the amount of CO_2 in our atmosphere[6]. We also believe that there is a highly positive correlation between the amount of CO_2 in our atmosphere and the increase in temperature relative to the base period 1951-1980. Increasing global surface and ocean temperatures can have many negative effects on the climate. Those effects include changing the climate of regions, this could put farmers out of business if they are no longer able to grow their native foods and could also result in mass extinction of animals in the climate. The increase in temperature could also melt our ice caps faster increasing the sea level. This could put many people and animals out of their homes in our life time if this issue is not addressed. A common claim by global warming deniers is that the earth will self regulate and this is all part of some seasonal process of warming and cooling. Their statement is true, we have seen net CO_2 emissions by the planet dropping as well as the ocean's absorption increase[6]. However, up until the industrial age humanity's burning of fossil fuels was negligible, providing no more impact than a forest fire. Then in the Industrial Age our burning of coal to run our factories, then gasoline to run our transportation and machinery, have exponentially grow to the point were CO_2 is playing a more than significant role in any model. We need to act now, if do not our world could heat to a crisp and we could kill our ocean life with CO_2 poisoning.

References

[1] NOAA Earth System Research Laboratory, "Trends in atmospheric carbon dioxide." https://gml.noaa.gov/ccgg/trends/data.html, 2022.

[2] "Gml data," tech. rep., National Oceanographic and Atmospheric Association and Scripps Institution of Oceanography.

[3] National Aeronautics and Space Administration Goddard Institute for Space Studies, "Gistemp dataset," tech. rep.

[4] E. Conway. https://www.nasa.gov/topics/earth/features/ climate_by_any_other_name.html, 2008.

[5] W. H. Staff. https://www.whitehouse.gov/omb/briefing-room/2022/03/28/president-bidens-fy-2023-budget-reduces-energy-costs-combats-the-climate-crisis-and-advances-environmental-justice/: :text=The%20President's%20Budget%20invests %20a,increase%20of%20nearly%2060%20percent., 2023.

[6] P. Tans, "An accounting of th observed increase in oceanic and atmospheric co_2 and an outlook for the future," *Oceanography*, vol. 22, no. 4, pp. 26–35, 2009.

[7] P. Tans and K. Thoning. https://gml.noaa.gov/ccgg/about/co2$_measurements.html$, 2020.

[8] http://freerangestats.info/blog/2016/11/27/ets-friends: :text=In%20all%20cases%2C%20%E2%80%9CN%E2%80%9D, multiplicative%20errors%2C%20and%20so%20on.