# Global $CO_2$ Emissions in the Present

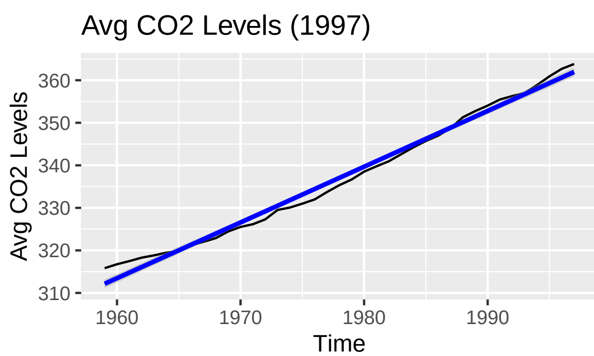
Naikaj Pandya, Amit Gattadahalli, Michael Golas, Austin Pitts

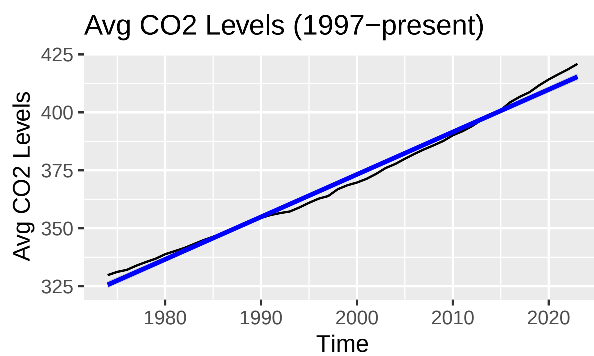
## Introduction

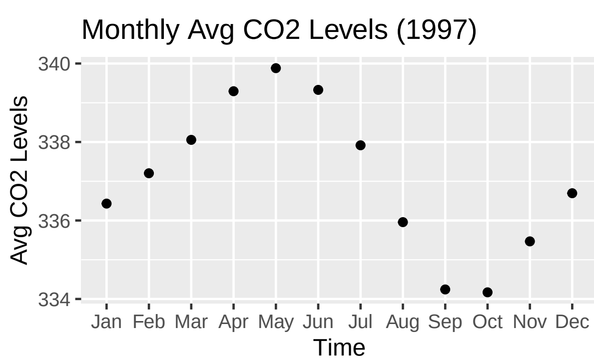
Our goal is to re-evaluate Keeling's observations of accumulated $CO_2$ in the atmosphere using the most upto date data. As of April of 2019, a new CO2 analyzer was installed at Mauna Loa that uses a technique called Cavity Ring-Down Spectroscopy (CRDS). CRDS is based on the measurement of the rate of absorption of light circulating in an optical cavity by comparing the ring down times when the laser is at a wavelength that the CO2 molecule does not absorb, to the ring down time when the laser is at a wavelength that the CO2 molecule does absorb.

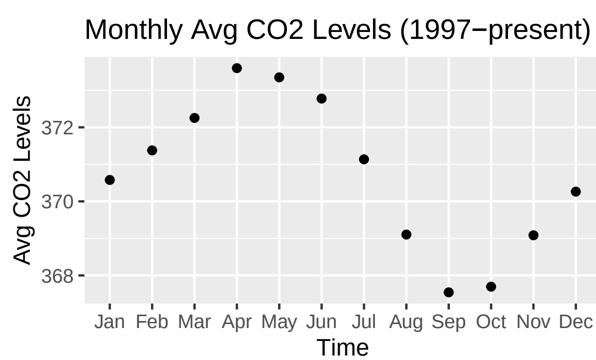
## Create a Modern Data Pipeline for Mona Loa CO2 Data

We establish a modern data pipeline for Mona Loa CO2 data. The data was obtained from the NOAA website, specifically from the CO2 daily data page. Additional columns for analysis were created, such as a log-transformed index and polynomial terms for time. The time series data was converted into a tsibble object for efficient time-series analysis. We then conducted an exploratory data analysis (EDA) to gain insights into the CO2 levels over time. The analysis included visualizations depicting the overall trend, average/standard deviation (SD) of CO2 levels per year, and monthly variations. The first plot illustrated the continuous increase in CO2 levels over the years prior to 1997 then the second plot shows from 1997 to present. The third and forth plots delved into monthly variations, highlighting average CO2 levels and their volatility. This EDA lays the foundation for further analysis and comparison with Keeling's observations from 1997. It also sets the stage for evaluating the performance of earlier models and forecasting future CO2 levels.



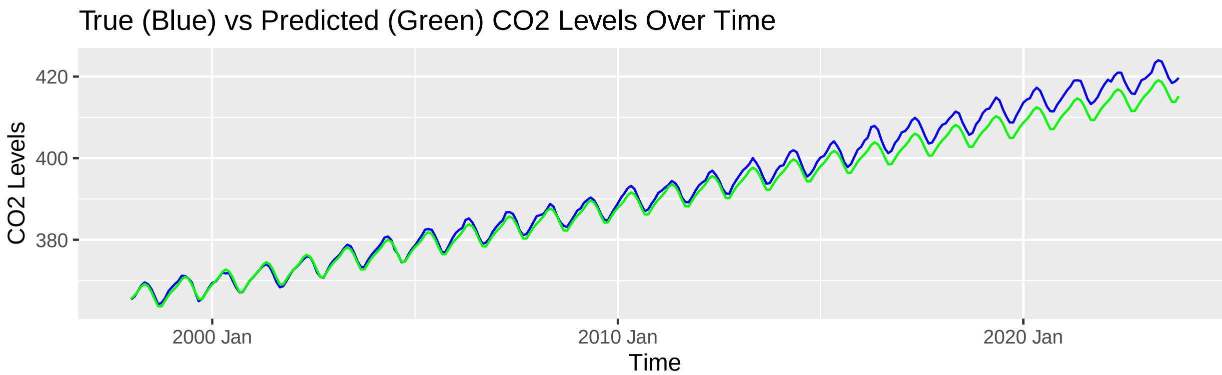



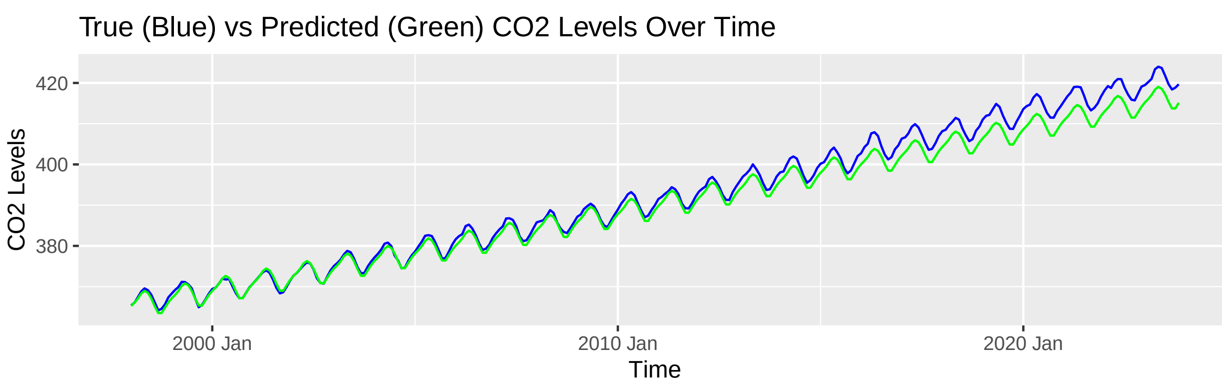
### Compare Linear Model Forecasts Against Realized CO2

In our comparison of realized atmospheric CO2 levels to those predicted by your forecast from a linear time model in 1997 (i.e. "Task 2a"), we can see the results in our graph of True (Blue) vs Predicted (Green) CO2 Levels Over Time. Results stay very consistent at first but as time continues we can see that the true values are increasing at a larger rate than what our predicted values would have estimated.




### Compare ARIMA Model Forecasts Against Realized CO2

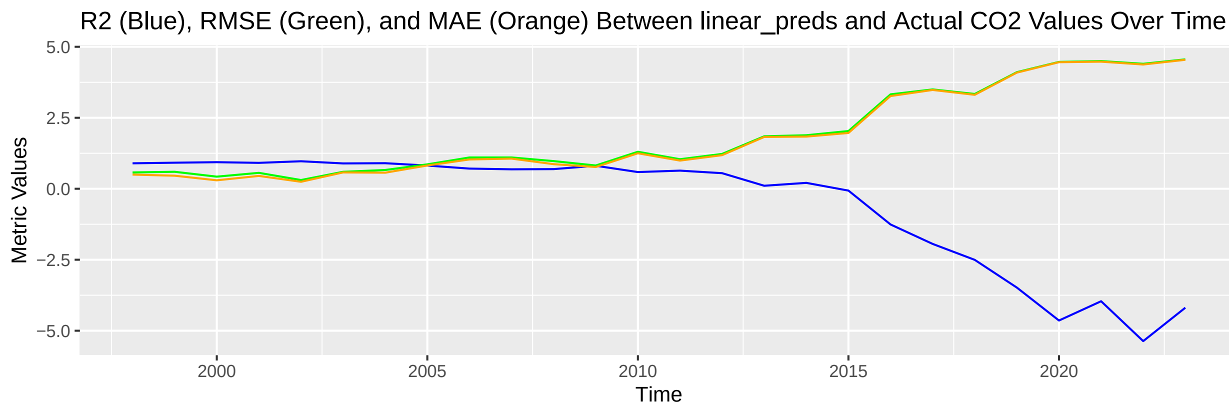
In our comparison of realized atmospheric CO2 levels to those predicted by your forecast from the ARIMA model in 1997 (i.e. "Task 3a"), we can see the results in our graph of True (Blue) vs Predicted (Green) CO2 Levels Over Time. Initially, there is a harmonious alignment between the true (blue) and predicted (green) CO2 levels, indicating that the ARIMA model effectively captured the underlying patterns in the earlier years. However, as time progresses, a noticeable trend emerges: the actual CO2 levels exhibit a more accelerated increase compared to what the ARIMA model predicted. This growing disparity suggests that there are evolving factors or trends influencing atmospheric CO2 concentrations that were not adequately accounted for in the original 1997 ARIMA model. These results are very similar to our previous graph vs our linear time model. Consistent with the broader Keeling Curve's evolution from 1997 to the present, the yearly seasonal patterns and monthly variations in the true values stay consistent. This adherence to historical patterns underscores the persistent nature of the underlying dynamics of atmospheric CO2 concentrations. Furthermore, the observation that the predicted values fall behind the actual values over time hints at the potential influence of a quadratic term in the Keeling Curve. This suggests a more intricate relationship between time and CO2 levels than initially modeled.

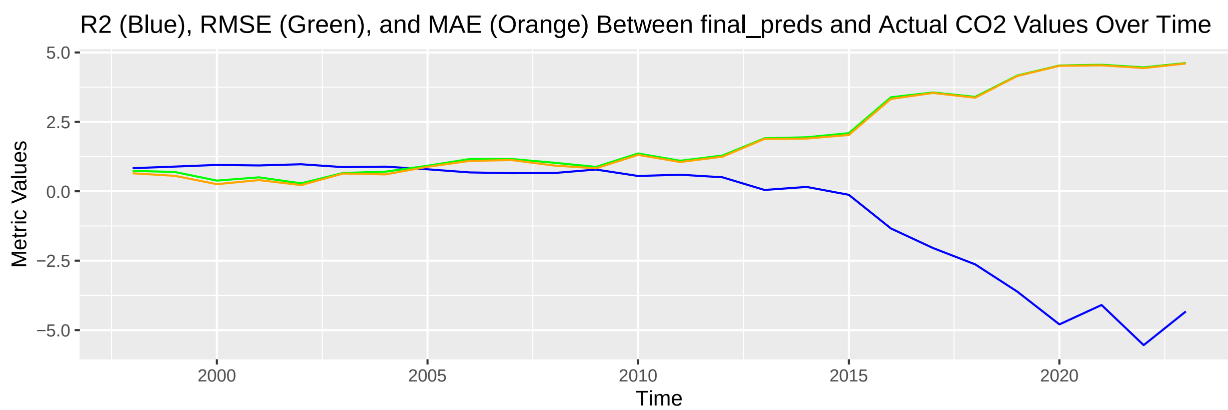



### Evaluate the Performance of 1997 Linear and ARIMA Models

We initially predicted that atmospheric CO2 would cross 420ppm for the first time in Task 4a, but in this stage we calculated the truth and can see than it occurred in April 2022 vs our forecast of April 2024. Our models were close to the truth, being only two years off. This discrepancy suggests a lag in predicting the acceleration of CO2 levels, indicating the complexity of forecasting long-term environmental changes. Now we continue to use the weekly data to generate a month-average series from 1997 to the present (month average series already generated during initial data ingestion), and compare the overall forecasting performance of our models from Parts 2a and 3b over the entire period.

In order to evaluate the performance of our Linear and ARIMA models, we take a look at their R2, RMSE, and MAE values over time. The first plot is the linear model performance and the second plot is the ARIMA model performance.




## Train Best Models on Present Data

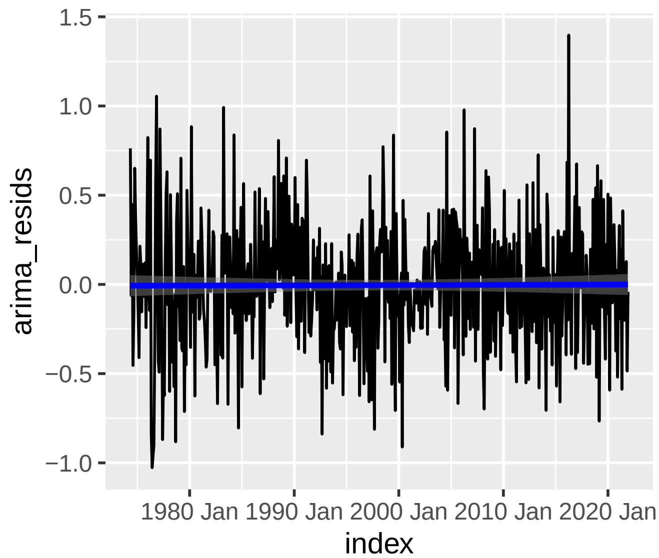
For training our best models on present data, we seasonally adjust the weekly NOAA data, and split both seasonally-adjusted (SA) and non-seasonally-adjusted (NSA) series into training and test sets, using the last two years of observations as the test sets, fitting ARIMA models for both SA and NSA series.

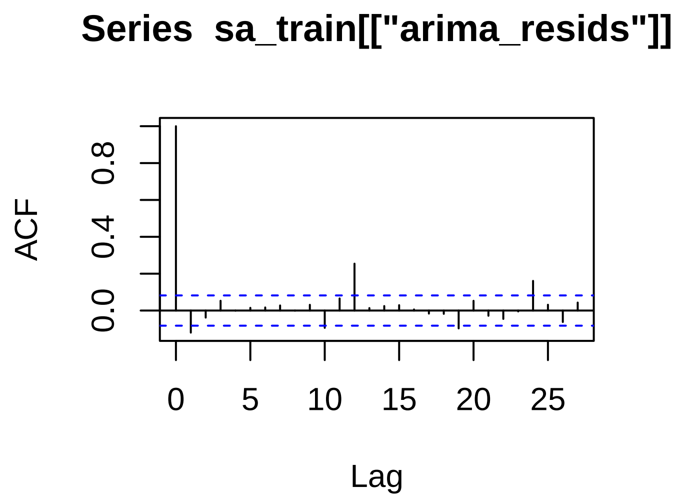
Our process outline was as follows: (if you want to see all of the output please see Appendix)

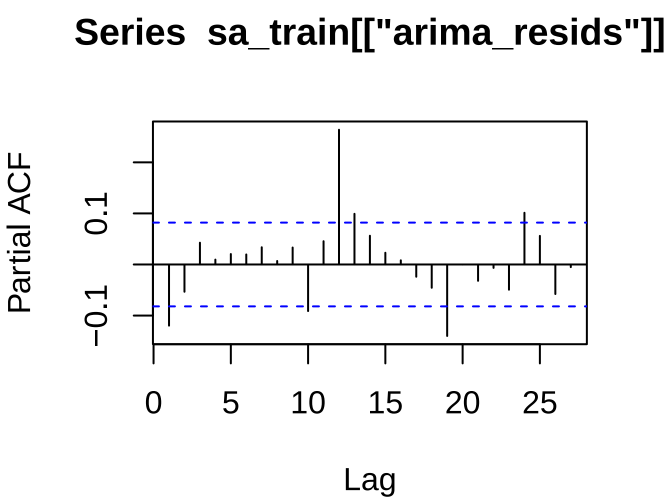
- We started by creating two versions of the CO2 data: one with seasonal adjustments (SA) and one without (NSA). These were then split into training and testing sets for model evaluation.
- The stationarity of the adjusted series (NSA and SA) was visually inspected, and the target variables over time were plotted.
- The autocorrelation and partial autocorrelation functions of the NSA and SA series were plotted to identify potential parameters for ARIMA modeling.
- ARIMA models were fitted to both the NSA and SA series using a grid search approach to identify optimal parameters. The models were evaluated in-sample and pseudo out-of-sample to measure their performance.
- A polynomial time-trend model was fitted to the seasonally-adjusted series, and its performance was compared to the ARIMA model.
- We then retrained the seasonally-adjusted series using a linear model for adjustment and then fit ARIMA models to the adjusted series.

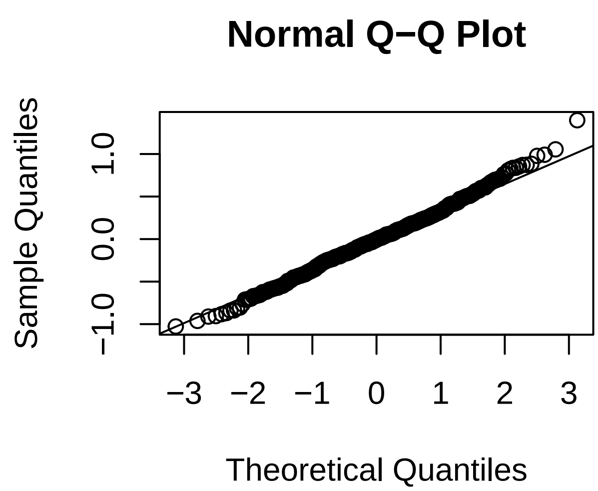
- Fit the polynomial to seasonally adjusted data.
- For our final ARIMA model, we used linear model differencing because it allows us to reconstruct forecasts farther out into the future even when ARIMA model begins generating constant predictions.
- Evaluated residuals and performance on test set.
- Finally, we retried modeling non seasonally adjusted series using linear model to adjust and remade our final ARIMA model.

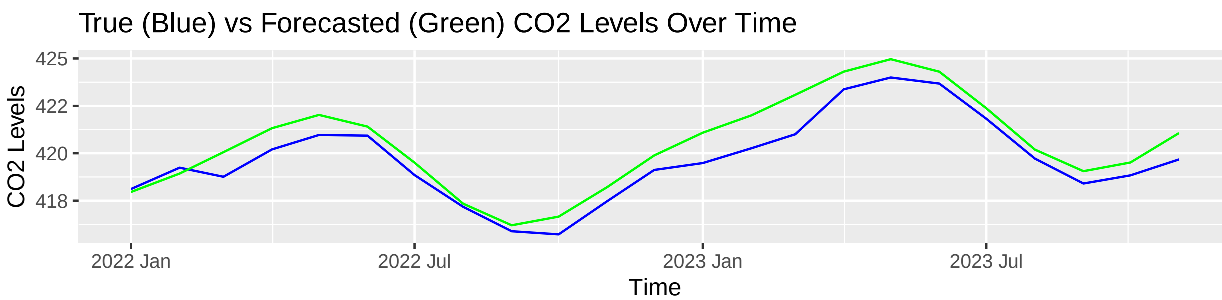
Below are visualizations used to evaluate our final ARIMA model.





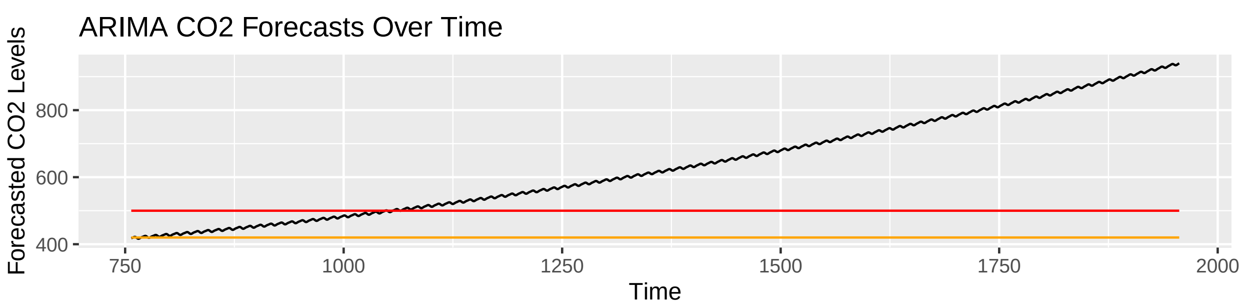

## How Bad Could It Get? (Predictions of the Far Future)

Our code initiates by creating a non-seasonally adjusted subset (NSA) of CO2 data up to the year 2022. A linear model is then fitted to the NSA data, and its residuals are used to train an ARIMA model with parameters (5, 0, 4). We then generate future time points for CO2 prediction from the year 2022 to 2122. The features for the linear model are constructed, including indices, logarithmic transformations, and the month effect. We create a plot visualizing the ARIMA CO2 forecasts over time, including the forecasted CO2 levels, the threshold at 420, and the threshold at 500. And finally we identify the first and final times when the CO2 levels are predicted to cross the 420 ppm and 500ppm thresholds.

Using our non-seasonally adjusted data series, our generated predictions for when atmospheric CO2 is expected to be at 420 ppm and 500 ppm levels for the first and final times is as follows:

- First and Final Time at 420, Mar 2022 - Jul 2022
- First and Final time at 500, Apr 2046 - Jan 2047

Below is our prediction for atmospheric CO2 levels up to the year 2122, if the future keeps with the same pattern historically, we are fairly confident in these estimates.




## Appendix

While our final results are reported here, in our complete notebook (Github Folder: Notebook) we examine alternative models and go into further assessment of the models. The purpose of this background information is to show more of the process in how we reached the conclusions shown above.