Global CO_2 Emissions in 1997

By James Venter
Ryan Brown
Tara Verma
Zachary Nash*

I. Introduction

A. *Carbon Emissions and the Keeling Curve*

In his 1960 report, Charles Keeling found that atmospheric carbon dioxide (CO_2) is far more variable than other atmospheric gasses, hypothesizing the variation was caused by naturally-occurring and manmade environmental factors (Keeling, 1960). Since then, Keeling has spearheaded efforts towards monitoring CO_2, and chose the Mauna Loa observatory as a representative sample of "baseline" air to proxy global levels as the site is 13,679 feet above sea level and surrounded by lava as opposed to vegetation (Tans, Pieter and Thoning, Kirk, 2020). The seasonal and overarching trends in this data are now widely known as the Keeling Curve, which is the primary subject of our investigation.

Over the last 37 years, the Keeling curve shows a steady upward trend in global CO_2 levels, and our team aims to investigate this phenomenon further. While we do not have a conclusive understanding of the long-term effects of elevated CO_2 levels, there is a growing body of research on the subject that provides initial insights on the impacts to both the natural world and human health. Elevated atmospheric CO_2 impacts decompose ecosystems and soil feedback processes, which in the long term may affect food systems (Jones, T.H., 1998). Recent changes in CO_2 levels are largely attributed to fossil fuel combustion, comparable in magnitude to those of inter-glacial periods, and occur far more rapidly. Many researchers believe increased atmospheric concentrations of greenhouse gasses, like CO_2, will contribute to a rise in global temperatures due to the greenhouse effect. The 1992 best estimate stated that doubling the atmospheric carbon dioxide will lead to a 2.5ºC increase in global temperatures which will impact natural and agricultural ecosystems (Vitousek, 1992). The 1986 Lake Nyos disaster in Cameroon provides initial insights into the medical effects of large-scale CO_2 emissions. When volcanic crater lake Lake Nyos underwent a substantial release of CO_2 gas in 1986, approximately 1700 people died; many of the survivors showed signs of asphyxiant gas exposure and coma states from atmospheric CO_2(Baxter, P.J. and Mfonfu, D., 1989).

As such, we are interested in exploring atmospheric CO_2 levels from 1960 to 1997 to contribute an assessment of recent trends to the literature. Additionally, we built models to illustrate what future levels of atmospheric CO_2 will be if we continue with this trajectory.

II. Measurement and Data

A. *Measuring Atmospheric Carbon Dioxide*

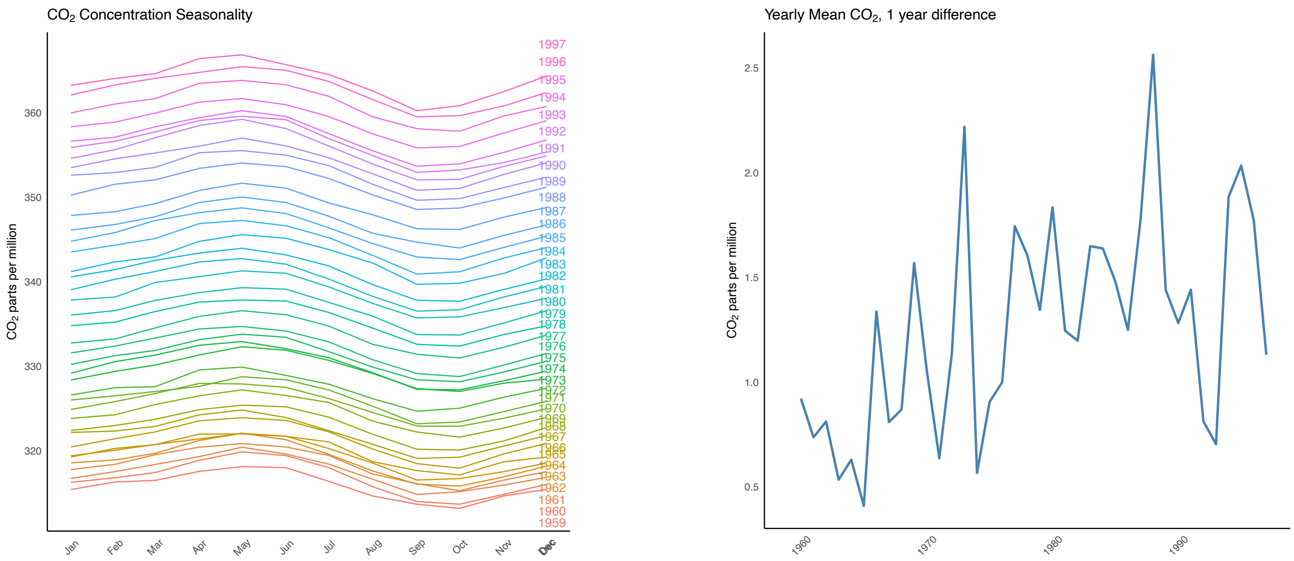
The dataset we are using is the average monthly readings of CO_2 levels collected at the Mauna Loa Observatory in Hawaii, from January 1959 through December 1997. The CO_2 level is reported in parts per million (ppm), which refers to the number of CO_2 molecules in every million molecules of air after removing the water vapor from the sample (Tans, Pieter and Thoning, Kirk, 2020).

B. *Historical Trends in Atmospheric Carbon*

Figure 1a contains the monthly average CO_2 levels by year. The series has a persistent upward trend with the mean increasing from approximately 315ppm in 1960 to 360ppm in 1997. We also observe consistent seasonality across years, where CO_2 levels peak between April and June and hit their lowest point between September and November. This is consistent with vegetation cycles and suggest that any models we build for this series should

* Venter: MIDS, james.venter@ischool.berkeley.edu. Brown: MIDS, rbrown55@ischool.berkeley.edu. Verma: MIDS, tara@ischool.berkeley.edu. Nash: MIDS, zachary.nash@ischool.berkeley.edu.

include a seasonal component. Additionally, the variance, or magnitude of seasonal changes, seems relatively constant throughout the series. Finally, we observe in Figure 1b that the increase of yearly average CO_2 levels is consistent throughout the series, albeit noisy.



(a) Seasonality and trend in CO_2

(b) Change in avg. yearly CO_2

Figure 1. : Seasonality and change in yearly average CO_2

Note: CO_2 levels peak between Apr.-Jun. and bottoms out between Sep.-Nov. each year. The percentage increase in average yearly CO_2 levels is relatively stable.

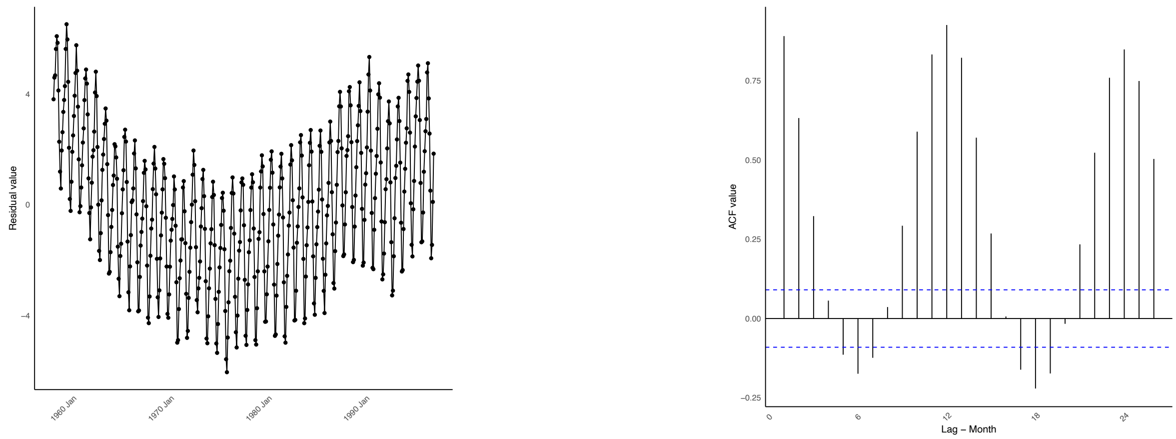
III. Models and Forecasts

A. Linear Time Trend Model

To begin, we fit a model of the form:

$$\mathrm{CO}_{2t} = \beta_0 + \beta_1 trend_t + \epsilon_t \tag{1}$$

The residuals exhibit a non-constant mean and are not normally distributed around zero, as seen in the time series and histogram plots in Figure 2a. Additionally, we observe a strong trend in the seasonal variation in the ACF plot in Figure 2b indicating that the residuals are not stationary.

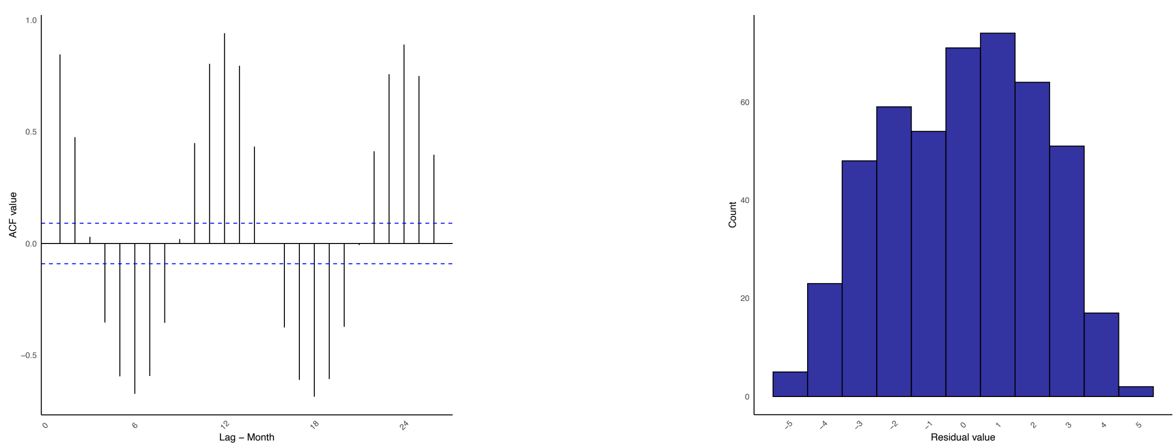


(a) Time series (b) ACF

Figure 2. : Time series and ACF plots of linear time trend model residuals

Note: The residuals exhibit a non-constant mean and are not normally distributed around zero, indicating non-stationarity.

We attempt to address the issue of stationarity by adding a quadratic term to the model. By examining this model's residual plots in Figure 3, we observe that the mean is now much closer to zero and appears to be consistent over time. The ACF plot still shows a seasonal pattern, which we will address by including a seasonal component in our model. Finally, the variance appears to be stable over time, indicating that there is no need to apply a log transformation to the series.



(a) ACF (b) Histogram

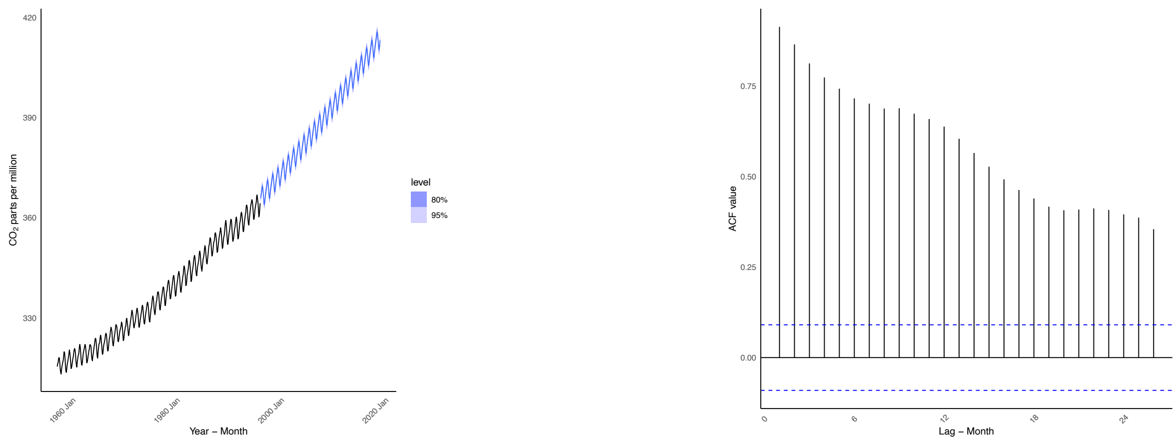
Figure 3. : ACF and histogram of quadratic time trend model residuals

Note: The quadratic term corrects the non-constant mean in the residuals; seasonality still needs to be addressed.

Our final model, used for forecasting, is of the form:

$$CO_{2t} = \beta_0 + \beta_1 trend_t + \beta_2 trend_t^2 + \beta_3 month_2 + ... + \beta_{14} month_{12} + \epsilon_t \tag{2}$$

Figure 4a contains the forecasts from our final time trend model. Although the forecasts appear to have a very narrow confidence interval, the forecast is likely understating the true confidence. The model residuals contain autocorrelation, seen in Figure 4b, indicating that the residuals are not stationary and therefore violate the white noise assumption necessary for us to have confidence in the forecast.



(a) Forecasts through 2020 (b) ACF plot

Note: The forecasts continues the upward trend, capturing the seasonality. Confidence is likely overstated due to autocorrelation in the residuals.

Figure 4. : ACF and forecast plots

A table containing the coefficients for the three time trend models can be found in the appendix.

B. *ARIMA Model and Forecasts*

Next, we will fit an ARIMA model. A KPSS test of the series is not statistically significant with a p-value of 0.01, confirming that it is non-stationary. A KPSS test of the first difference of the series yields a p-value of 0.1, confirming that we will only need to take one difference to make it stationary. We will begin by fitting an ARIMA model using the ARIMA() function to estimate the model with lowest corrected AIC. We will use the form $ARIMA(0,1,0)$ to address the non-stationarity in the series and let the ARIMA function determine if a seasonal difference results in a lower corrected AIC. The function selected a model with the form $ARIMA(0,1,0)(0,1,0)[12]$. This model fits the data fairly well and has an AICc value of 442.01. However, a Ljung-Box test of the residuals is not statistically significant (p-value of 0), indicating that autocorrelation is still present in the residuals. The PACF plot in Figure 5 shows several lags that are statistically significant. It does not display a sharp drop off that would be indicative of an AR process, so we conclude that the model is missing an MA component.

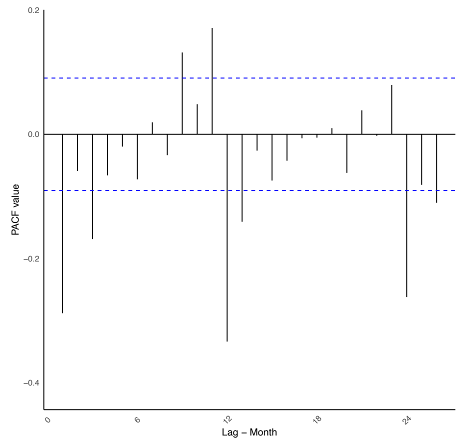
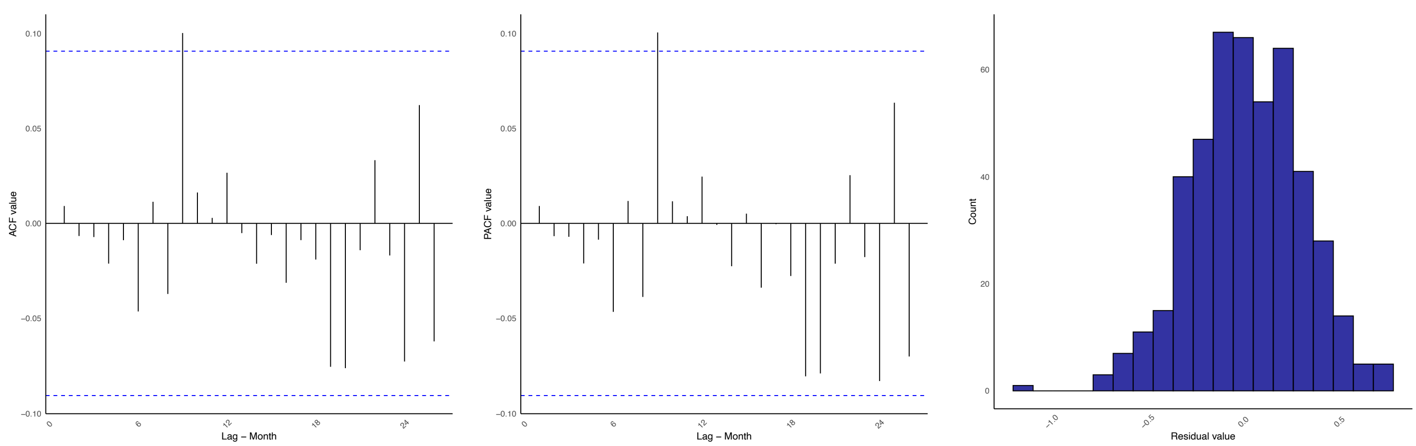


Figure 5. : PACF of ARIMA(0,1,0)(0,1,0)[12] model residuals.

Note: Several lags in the PACF are statistically significant. The shape of the PACF indicates the model is missing an MA component.

We then fit another ARIMA model using the ARIMA() function, again selecting the model with lowest corrected AIC. We kept the difference terms from the prior model fixed and let the ARIMA function select the MA and seasonal MA components with the lowest corrected AIC. The function selected a model with the form $ARIMA(0,1,3)(0,1,1)[12]$. This model has an AICc value of 177, and the Ljung-Box test of the residuals indicates that they do not have any autocorrelation, with a statistic of 7.36 and associated p-value of 0.12. We examined the residuals to ensure the model's adequacy by examining the ACF and PACF plots of the residuals in Figure 6a and Figure 6b, and the histogram of the residuals in Figure 6c. The ACF and PACF plots show no indication of non-stationarity and the histogram shows that the residuals are relatively normally distributed with mean zero, indicating that they satisfy the necessary white noise process assumptions.



(a) ACF (b) PACF (c) Histogram

Figure 6. : Residual diagnostic plots of ARIMA(0,1,3)(0,1,1)[12] model

Note: Only one of the 24 lags is significant in each of the plots. This is within the expected number at a 95% confidence level, indicating that they are stationary. The residuals are normally distributed with mean 0, indicating that the white noise assumptions are satisfied.

We forecast atmospheric CO_2 concentration levels to 2022 using the $ARIMA(0,1,3)(0,1,1)[12]$ model, shown in Figure 7. Our model captures the trend and seasonal patterns in the CO_2 data, providing a reasonable representation of future CO_2 levels. This model assumes historical patterns will continue, but significant changes in global carbon dynamics could affect future predictions. The forecast indicates a continuing increase in CO_2 levels, with clear seasonal fluctuations and widening prediction intervals over time. Our model captures the trend and seasonal patterns in the CO_2 data, providing a reasonable representation of future CO_2 levels. This model assumes historical patterns will continue, but significant changes in global carbon dynamics could affect future predictions.

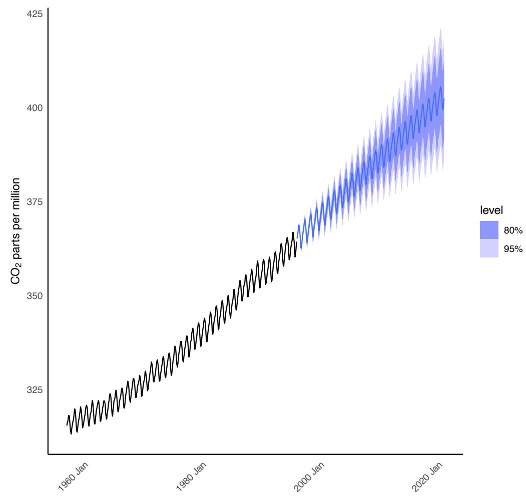


Figure 7. : Forecasts through 2022.

Note: The model captures the trend and seasonal patterns, providing a reasonable representation of future CO_2 levels so long as the trend remains stable.

Additionally, we used our model to produce a longer term forecast of atmospheric CO_2 concentration levels. Figure A2 in the appendix shows this forecast, which extends to 2100. We leveraged this forecast to answer key questions for a longer time horizon. Based on our forecast, CO_2 concentration will reach 420ppm around April 2032 and remain above that level through December 2100, and it will reach 500ppm around April 2084 and remain above that level through December 2100. Both of these are forecasted to persist throughout the time frame after reaching those levels due to the upward trend observed in the series. Finally, we forecast that the CO_2 concentration will be 521.6 at the end of 2100, with a 95% confidence interval of (418.5, 624.7). However, in order for this estimate to be accurate, the trend in atmospheric CO_2 concentration levels would need to remain stable for the next century. Given that this trend is highly impacted by future technological advances and that we have observed a slight deviation in the trend in recent years, it is unlikely to be the same so far into the future. Therefore, we do not have very high confidence in the accuracy of these predictions; however, they can still act as a useful tool to demonstrate what the future could look like if nothing changes in the interim.

IV. Conclusions

Should nothing change, expected concentration levels will increase by nearly 45%, corresponding to an approximate ~1ºC increase in global temperatures which can have potentially devastating impacts on natural and agricultural ecosystems. Additionally, higher atmospheric concentrations of CO_2 could potentially result in more health issues related to asphyxiant gas exposure, and elevated rates of anxiety and panic attacks, which would add an unnecessary burden the healthcare system. However, because these values are projected to occur far in the future based on the assumption that current trends continue, there is still time for us to change the direction. Investing in initiatives to reduce atmospheric concentrations of CO_2 today can help us avoid this scenario in the future.

REFERENCES

Baxter, P.J., Kapila, M., and Mfonfu, D. 1989. "Lake Nyos disaster, Cameroon, 1986: the medical effects of large scale emission of carbon dioxide?" *British Medical Journal*, 298(1437).

Jones, T.H., Thompson, L.J., Lawton, J.H. et al. 1998. "Impacts of Rising Atmospheric Carbon Dioxide on Model Terrestrial Ecosystems." *Science*, 280(5362): 441–443.

Keeling, Charles. 1960. "The concentration and isotopic abundance of carbon dioxide in the atmosphere." *Tellus*, 12(2): 200–203.

Tans, Pieter, and Thoning, Kirk. 2020. "How we measure background CO2 levels on Mauna Loa."

Vitousek, Peter M. 1992. "Global Environmental Change: An Introduction." *Annual Review of Ecology and Systematics*, 23: 1–14.

Appendix: Supplemental ARIMA model plots

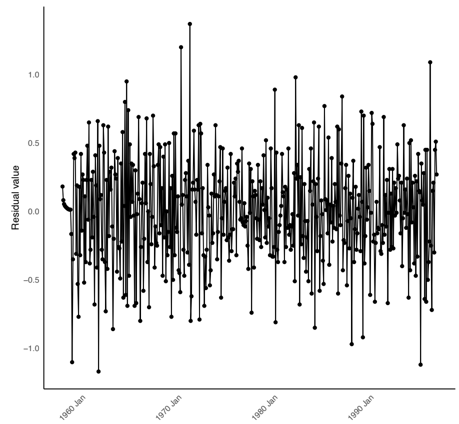


Figure A1. : Time series of initial ARIMA model residuals.

Note: The variance of the residuals is not constant over time, indicating non-stationarity.

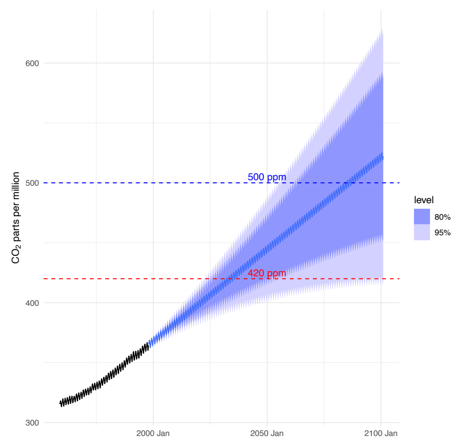


Figure A2. : Forecasts through 2100.

Note: The variance of the forecasts becomes extremely large over time, decreasing confidence in our estimates that are far in the future.

Appendix: Time Trend Model Robustness

	Linear	Quadratic	Quadratic + Seasonal
Intercept	311.50***	314.76***	314.68***
	(0.24)	(0.3)	(0.15)
Trend	0.11***	0.07***	0.07***
	(0.00)	(0.0)	(0.00)
Trendˆ2		0.00***	0.00***
		(0.0)	(0.00)
Month 2			0.66***
			(0.16)
Month 3			1.41***
			(0.16)
Month 4			2.54***
			(0.16)
Month 5			3.02***
			(0.16)
Month 6			2.35***
			(0.16)
Month 7			0.83***
			(0.16)
Month 8			−1.23***
			(0.16)
Month 9			−3.06***
			(0.16)
Month 10			−3.24***
			(0.16)
Month 11			−2.05***
			(0.16)
Month 12			−0.94***
			(0.16)
AIC	904.8	735.4	−286.5
BIC	917.3	752	−224.3
Adj. R-squared	0.969	0.979	0.998
Residual standard error	2.618	2.182	0.724

Revisiting Global CO_2 Emissions

James Venter, Tara Verma, Zach Nash, Ryan Brown

2024-07-16

Abstract

In their initial investigation, the authors analyzed the long-term trends of atmospheric carbon dioxide (CO2) concentrations, focusing on the data collected at the Mauna Loa Observatory from 1960 to 1997. In this paper, we build upon the previous work by evaluating the accuracy of the 1997 forecasts against realized CO2 levels, reassessing the performance of our models, and exploring new forecasting methods using more recent data. Our goal is to enhance our understanding of CO2 trends and improve the reliability of our future projections. Note the most recent data is that provided by NOAA.

Exploratory Data Analysis

To begin, we will review the new data to determine if there have been any significant changes in the behavior of the CO2 series. In this updated analysis, we are considering weekly observations instead of monthly for all new data (post 1997). Despite this change in the frequency of data points, all else equal, we would expect the overall trends observed in the previous paper to remain consistent. By examining the updated dataset, we aim to identify any new patterns or shifts that may have occurred over the extended period.

From the time plot (Figure One A), several key features are immediately evident. Firstly, we observe consistent seasonality throughout the entire series, from 1960 to 2024. We also note a persistent and accelerating upward trend in the series, with the mean increasing from approximately 315ppm in 1960 to over 400ppm by 2024. The transition from monthly to weekly data in 1974 provides a more detailed view of short-term fluctuations, but the overall trend and seasonal patterns remain consistent. Notably, the variance, or magnitude of seasonal changes, appears to increase slightly over time, particularly in the weekly data.

From our ACF plot (Figure One B), we see a very gradual decline, which is strong evidence for the persistent trend within the series. The ACF remains significantly positive even at high lags, indicating a strong long-term correlation structure. We also note the clear scalloping pattern in our ACF plot, which aligns with the seasonality we observed in the time plot. This scalloping pattern is more pronounced in the weekly data, reflecting the finer temporal resolution.

In our PACF plot (Figure One C), we observe a rapid decay after the first few lags. The PACF is nearly 1 at the first lag, with a few subsequent lags showing significant values before quickly diminishing. This pattern suggests that a low-order autoregressive model might be appropriate for capturing the short-term dynamics of the series. The oscillations in the PACF provide further evidence for the strong seasonal component in the data.

Examining the time plot of annualized averages (Figure One D), it's clear that CO2 levels are not only rising but doing so at an accelerating rate. The curve appears to be steepening, especially from the 1990s onward. This acceleration in the growth rate of CO2 concentrations is a critical observation and suggests that the rate of increase is itself increasing over time.

Finally, the seasonality plot (Figure One E) provides a detailed view of the seasonal patterns across all years in the dataset. We note that CO2 levels consistently peak around May and reach their lowest points around September to October. This pattern is consistent across years, although the amplitude of the seasonal cycle appears to increase slightly over time. The consistency of this seasonal pattern, despite the long-term increase

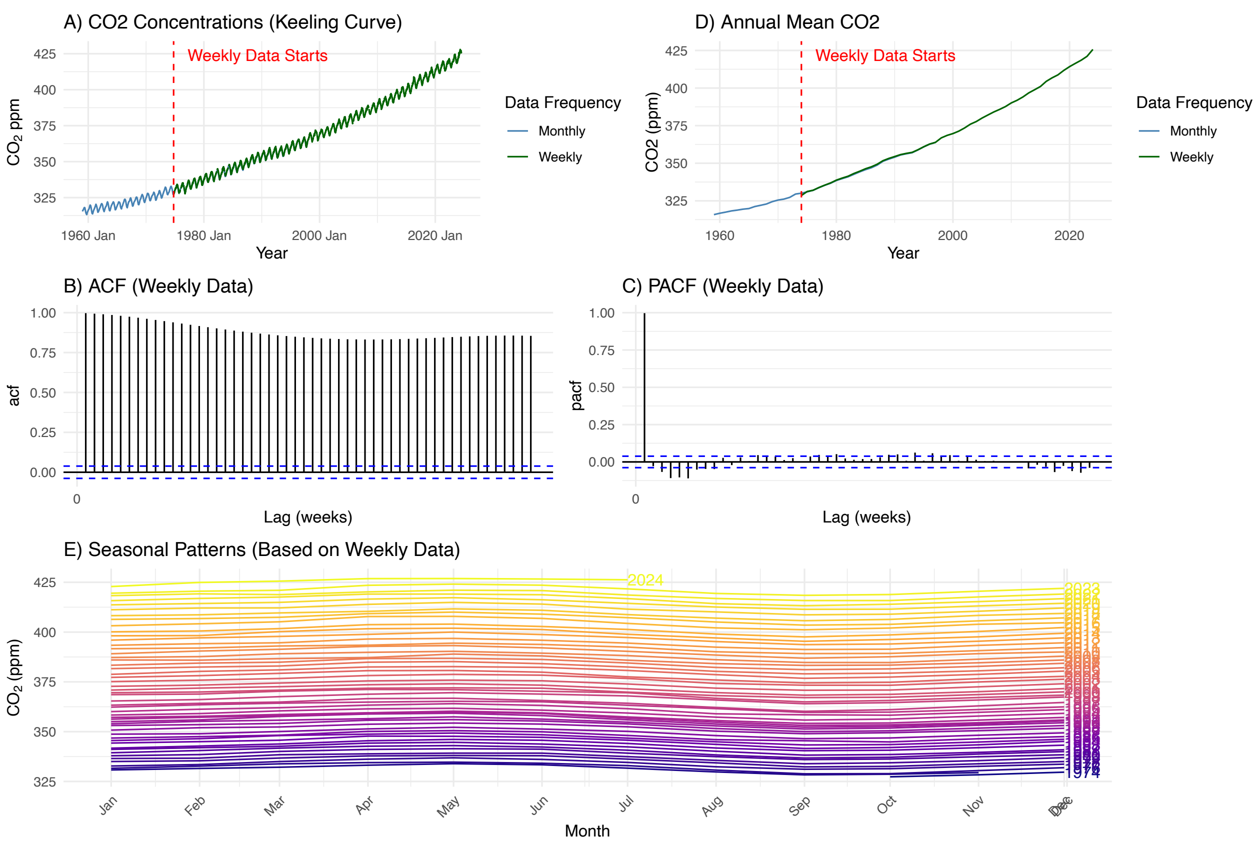


Figure 1: Figure 1: CO2 Concentrations Analysis

in CO2 levels, underscores the importance of considering both trend and seasonality in any models developed for this data.

In the 1997 paper the authors confirmed that the data was not stationary and this appeaser to be the case again in the new weekly data. Ergo, test the original series and its first difference for stationarity using the KPSS (Kwiatkowski-Phillips-Schmidt-Shin) test for stationarity.

The KPSS test for the original series yields a p-value of 0.01, indicating that we reject the null hypothesis of stationarity. For the first-differenced series, the p-value is 0.1, which is greater than our significance level of 0.05. This suggests that first-differencing is indeed necessary to achieve stationarity in our CO2 data. This finding supports our earlier observation of the need for differencing and will inform our modeling approach moving forward.

1997 Model Evaluation

In the next section, we will compare two of the notable models from the original paper to what happened: an ARIMA(0,1,3)(0,1,1)[12] model and a polynomial time trend model with seasonal dummies.

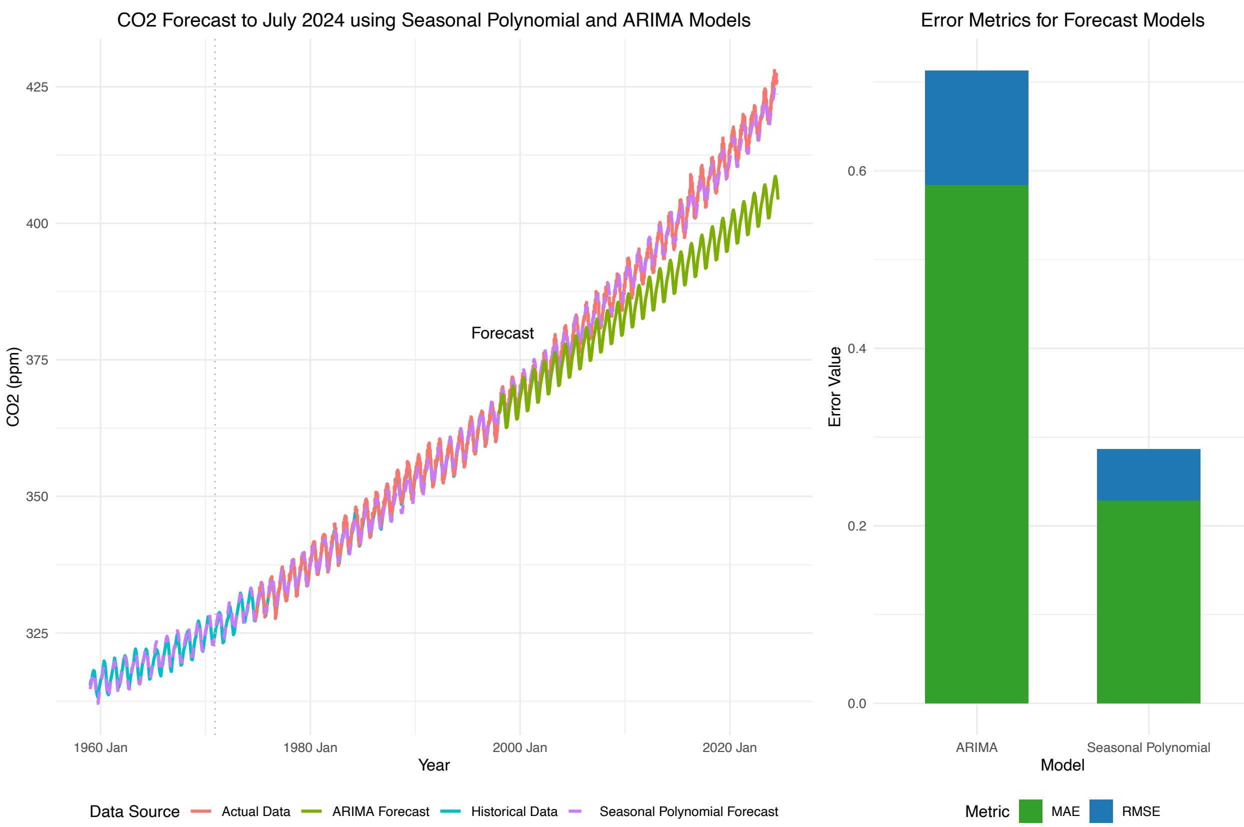


Figure 2: Figure 2: Forecast Comparison

Figure 2 illustrates the CO2 forecast to July 2024 using the aforementioned models, along with the corresponding error metrics for these forecast models. The left plot shows the actual CO2 data from 1960 to 2024 (red line), alongside the forecasts generated by the Seasonal Polynomial (purple line) and ARIMA (green line) models. The right plot provides a comparison of error metrics (MAE and RMSE) for the two models, with the polynomial model showing significantly lower error values compared to the ARIMA model. It seems that the polynomial model was able to capture the accelerating rate of growth, while the ARIMA model was not.

Seasonally Adjust the Data

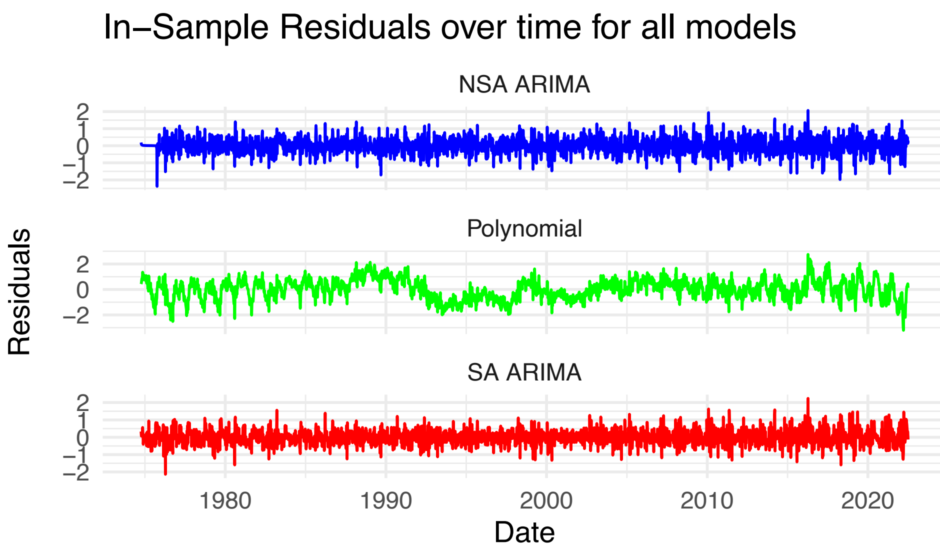
While the polynomial model did well we contend that by seasonally adjusting the data and fitting a model on the new weekly data may result in more reliable forecasts. To test this theory we will fit ARIMA models (using a grid search for lowest BIC) on both seasonally adjusted and raw values and measure performance on both in and out of sample performance. We will also fit a polynomial model on the seasonally adjusted data as a baseline and measure its performance as well. To seasonally adjust the data we use the STL package.

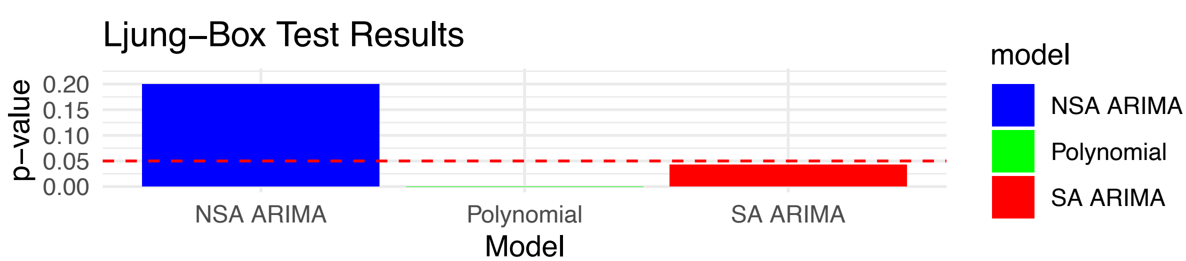
Fit ARIMA Models to NSA and SA Series

Next, we fit ARIMA models to the training sets of both NSA and SA series and evaluate their performance.

The ARIMA model identified for the non seasonally adjusted data is an ARIMA(0,1,2)(2,1,0)[52] and for the seasonally adjusted it is an ARIMA(2,1,2) w/ drift.

Evaluate In-sample and Out-of-sample Performance





The residual plots and Ljung-Box test results highlight differences in the performance of the various model types. The NSA ARIMA model demonstrates consistent residuals over time, with a Ljung-Box test p-value above 0.05, indicating no significant autocorrelation in the residuals.

In contrast, the polynomial model and SA ARIMA models perform poorly. The polynomial residual plot exhibits large, systematic patterns, indicating a failure to capture variations and some longer-term trends in the data. Moreover, both the polynomial and SA ARIMA models have p-values below 0.05 in the Ljung-Box test, signifying significant autocorrelation in the residuals and confirming that they are not randomly distributed.

We now turn to out of sample performance:

Results:

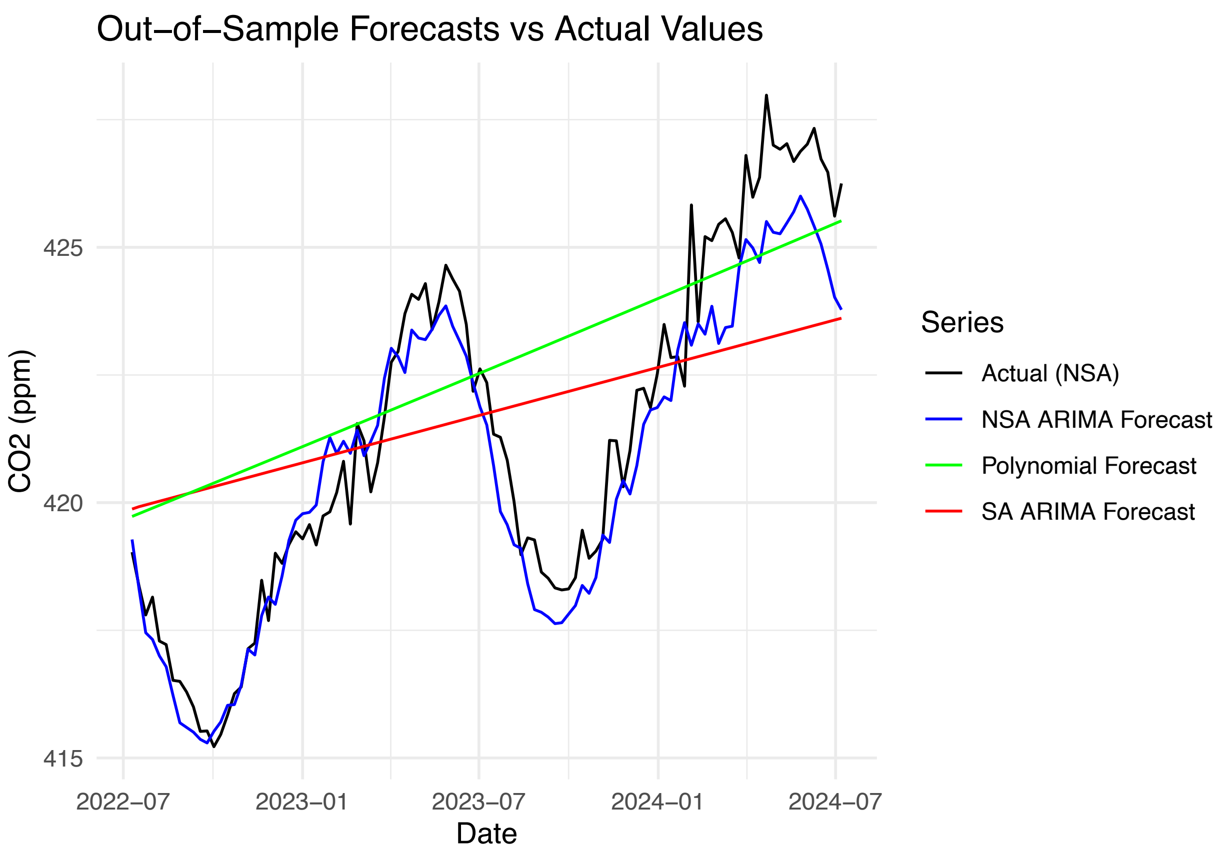


Figure 3: Figure 4: Out of Sample Comparison

Model	MAE	RMSE	MAPE
NSA ARIMA	0.8385329	1.051565	0.1982601
SA ARIMA	1.5425360	1.837126	0.3668053
Polynomial	1.5455302	1.813344	0.3675233

Here we see that while the seasonally adjusted models might be helpful in understanding the long-term trend, the NSA ARIMA model outperforms both the SA ARIMA and Polynomial models in terms of MAE, RMSE, and MAPE.

Where we are going

To look into the future we use the SA ARIMA model and fit it on the entire dataset.

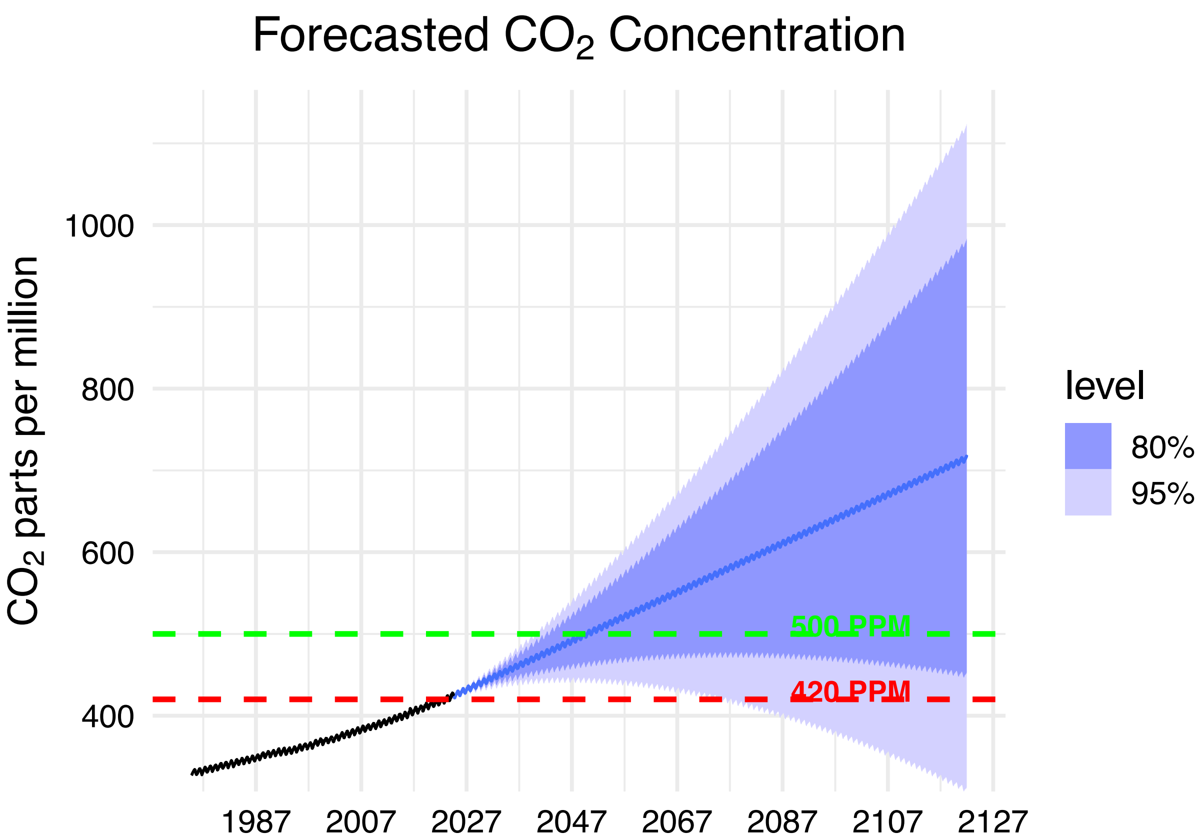


Figure 4: Figure 5: Long Run Forecast

Given the persistent trend and seasonality in the CO2 data, the ARIMA model provides reasonable forecasts within the limits of the model assumptions. However, it's important to acknowledge that long-term forecasts carry inherent uncertainty due to potential changes in emission patterns, technological advancements, and policy interventions. Therefore, while the model provides a structured estimate, the confidence in predictions over such an extended period should be taken with caution.

That said, CO2 concentration crossed 420ppm in April of 2022. Though it may come back down to this level by December 2121 according to this forecast. CO2 is expected to reach 500 ppm for the first time around February 2049. And by the end of the forecast period the model predicts it to be at 719ppm with a 95% CI of 304.86ppm to 1133.14ppm.