Present Global CO_2 Emissions

Meral Basit, Alex Hubbard, Mohamed Bakr

Contents

1 Background

In the 1997 report, we explored the trend and variability in atmospheric CO_2 levels using polynomial and ARIMA models. Our analysis indicated a significant upward trend in CO_2 levels, with an accelerating rate of increase. In this report, we aim to re-evaluate those models and their predictions using the latest data. The central question we are addressing is:

> *Have the previous models accurately predicted current CO_2 levels?*

1.0.1 Null Hypothesis

2 Measurement and Data

2.1 Measuring Atmospheric Carbon

Since our last report, the volcano near the research center has erupted. Therefore the measurements from Dec. 2022 to July 4, 2023 are from the Maunakea Observatories, which are just over 20 miles north of the original observatory. Additionally, there is a note that the last several months worth of data is "preliminary" and therefore could be revised. Furthermore, the provided data consists of weekly averages, and we will need to calculate monthly averages from 1997 to the present in order to compare with the forecast data from the 1997 report.

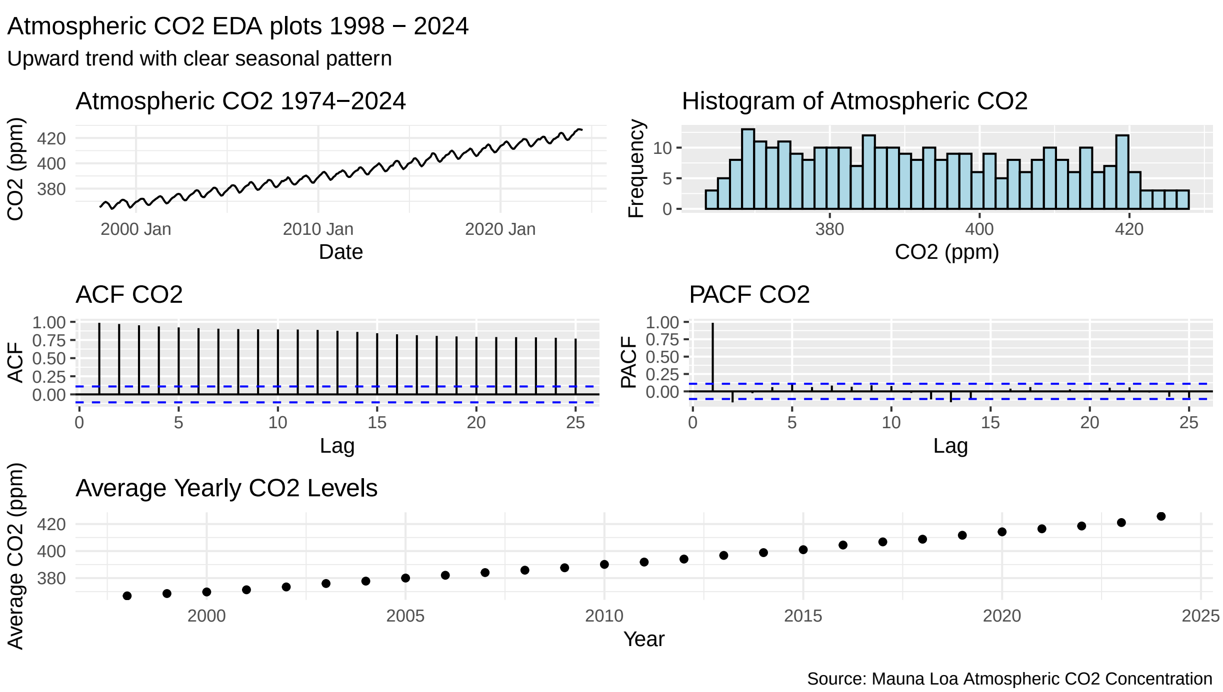

Figure 1: Atmospheric CO2 EDA standard plots

2.2 Historical vs Present Trends in Atmospheric Carbon

We see that the atmospheric CO_2 levels continued to have a strong upward linear trend since 1997, as seen in the bottom plot of average yearly CO_2 in Figure 1. It appears that this linear trend very slightly increased in slope after the year 2000. We also see that the distribution of values is fairly wide from the histogram, with a slight right skew. We also see that the ACF tails off very slowly while the PACF drops off after lag 1 but still has few lags above the segificance level. This indicates that there may be some unit roots. As this timeseries is a continuation of our previous time series, we know that this data is also not stationary.

3 Old Models Comparisons

3.1 Linear Model

Our polynomial model did a pretty good job at predicting CO_2 up to 2024. It is missing the peaks and valleys, but it looks to capture the average yearly increase in CO_2 levels as shown in Figure 2.

3.2 ARIMA Model

In the 1997 report, we used an ARIMA(3,1,0)(2,1,0)[12] model to forecast atmospheric CO_2 levels. Both the ARIMA model forecast and the the realized CO_2 show an upward trend with seasonal pattern. However, the model predicted lower values than the realized values in the long term as shown in Figure 2.

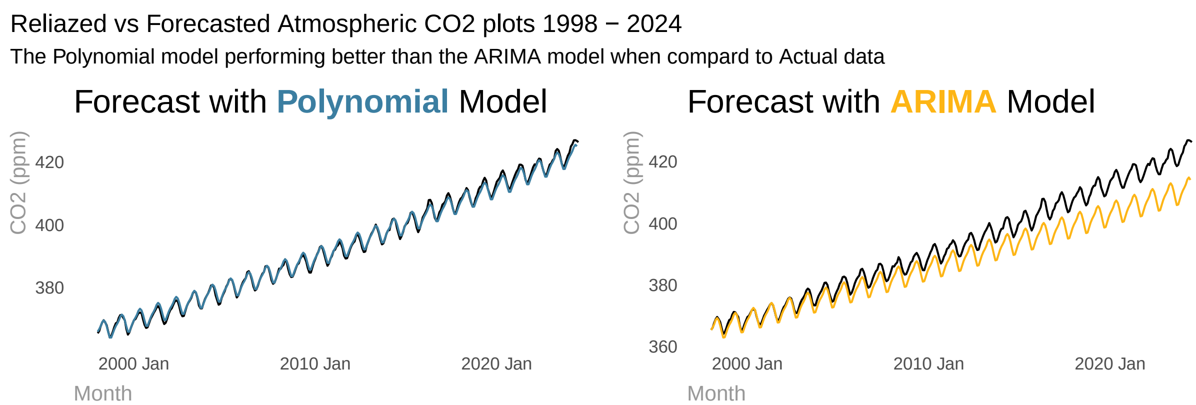

Figure 2: Comparing Realized Atmospheric CO2 Levels with Polynomial and ARIMA Models

3.3 Linear vs ARIMA

In 1997, we predicted that CO_2 levels would reach 420 PPM by March 2025 using our ARIMA model and 2022 May using the Polynomial model. However, Actual CO_2 levels reached 420 PPM by 2022 Apr.

Our ARIMA model with log transformation produces RMSE = 7.184, while the Polynomial model produces an RMSE = 3.48. Considering the descriptive analysis, threshold-prediction results, and

Table 1: ARIMA Models Comparison - Seasonally Adjusted CO2 values

.model	sigma2	log_lik	AIC	AICc	BIC
arima_fit_log.search	0.00	1624.8	-3238	-3237	-3216
arima_fit.search	0.11	-88.3	189	189	211

Table 2: ARIMA Models Comparison - non-Seasonally Adjusted CO2 values

.model	sigma2	log_lik	AIC	AICc	BIC
arima_fit_log.search	0.00	1538	-3067	-3066	-3049
arima_fit.search	0.13	-105	225	225	250

the RMSE comparisons, we can conclude that despite the fact that the ARIMA model performed better with the 1997 data, the Polynomial model outperformed the ARIMA model in the long-term forecast.

4 New Model

4.1 ARIMA Model

We made two copies of the `NOAA` data and seasonally adjusted one of them. After that, we split both datasets (the seasonally adjusted (SA) and the non-seasonally adjusted. (NSA)) into training and test sets, using the last two years of observations as the test sets.

We used the KPSS test to determine whether the SA and the NSA data are stationary. For both series, the first test yields a p value of 0.01, leading us to reject the null hypothesis, indicating that the data is not stationary. After taking one difference, the p value for both series rose to 0.1, and we failed to reject the null hypothesis, suggesting that both datasets are stationary after one difference.

Based on the EDA performed earlier, it was challenging to estimate teh p and q terms for the ARIMA model just by looking at the `ACF` and `PACF` plots. We estimated two non-stepwise ARIMA models for each set one with log transformation and another without the log transformation. Using the information criteria `AICc` in Table 1, we see that the best SA model was an ARIMA(5,1,0)(1,0,0)[52] with a log transformation. This model has five `AR` terms and is first differenced. It also has one seasonal `AR` term with a period of 52 weeks. The best NSA model was an ARIMA(0,1,0)(0,0,1)[52] with a log tranformation. This model is first differenced and has one seasonal `MA` term where the period is 52 weeks. We selected both of these models because they had the lowest `AICc`. We will examine the residuals to see if they resemble white noise.

The top performing models for SA and NSA data both has residuals that rejected the null hypothesis of the Ljung-Box test, which indicates that they do not have white noise residuals. Therefore, we selected the models with the second lowest `AICc`, which have residuals that follow a normal distribution, and appear to be white noise in their `ACF` plots. Additionally, they both fail to reject the null hypothesis of the Ljung-Box test, indicating that the residuals exhibit no autocorrelation for 10 lags and can be regarded as white noise (SA p-value = 0.23, NSA p-value = 0.48).

The superior model for the SA data is an ARIMA(5,1,0)(1,0,0)[52], which has five `AR` terms, first differencing, and one seasonal `AR` term with a period of 52 weeks. The superior model for the NSA

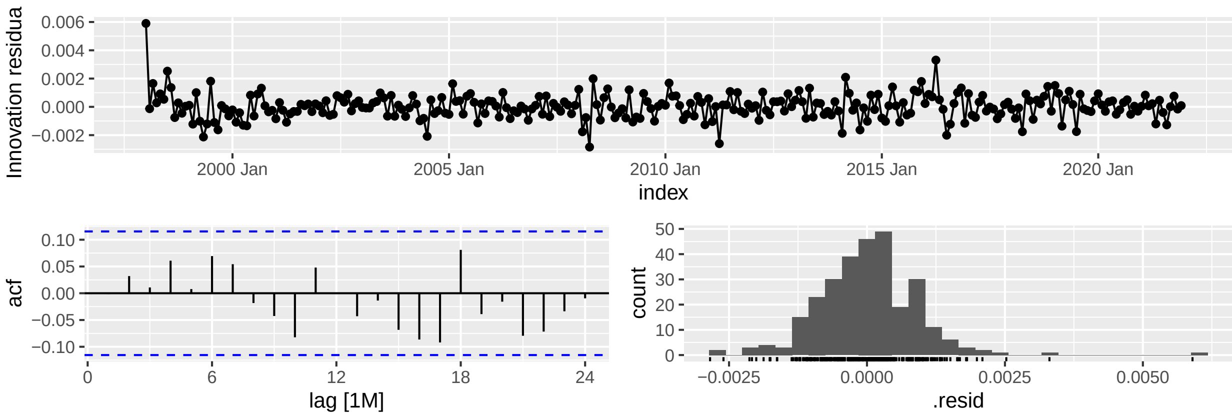

Figure 3: SA Trained ARIMA(,,)(,,)[12] Model Residuals

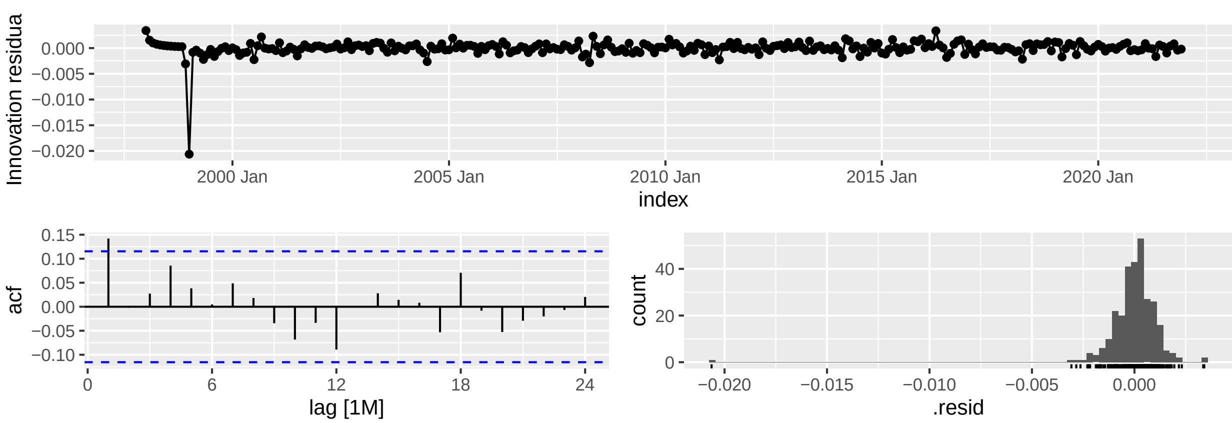

Figure 4: NSA Trained ARIMA(,,)(,,)[12] Model Residuals

data is an ARIMA(1,1,4)(0,0,1)[52], which has one `AR` term, first differencing, four `MA` terms, and one seasonal `MA` term with a period of 52.

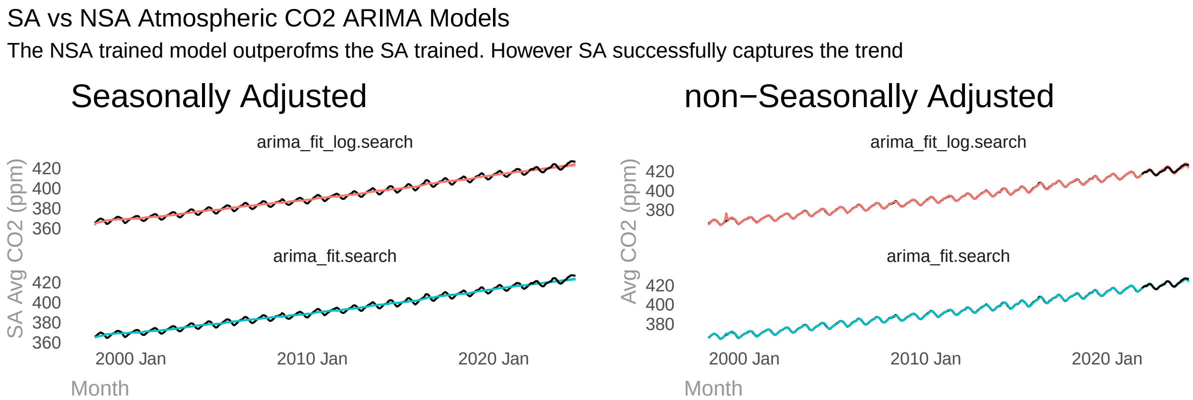

Figure 5: Comparing SA vs NSA data sets ARIMA models perfomance

The selected NSA trained model produced RMSE = 0.71 which outperforms the selected SA trained model that produced RMSE = 0.76

4.2 Polynomial Model

We estimated a Polynomial model of the form: $CO2_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \epsilon_t$ and trained it on the SA data. Based on the model fit results, the estimated coefficient $\beta_1 = 0.14$ indicates that the CO_2 levels increase by ≈ 0.14 units per month which is lower than the rate of ≈ 0.0674 we estimated in the 1997 report. The p-value of the time index is < 0.05 which suggests that the coefficient is statistically significant. We reject the null hypothesis that the coefficient $\beta_1 = 0$. that also provides evidence that the CO_2 levels continues to have an upward linear trend. The estimated quadratic term coefficient $\beta_2 = 0.000141$. The positive coefficient suggests that the rate of increase in CO_2 levels is accelerating at a higher rate than the model estimated in 1997 report. The p-value of the time index is < 0.05 which suggests that the coefficient is statistically significant. We reject the null hypothesis that the coefficient $\beta_2 = 0$.

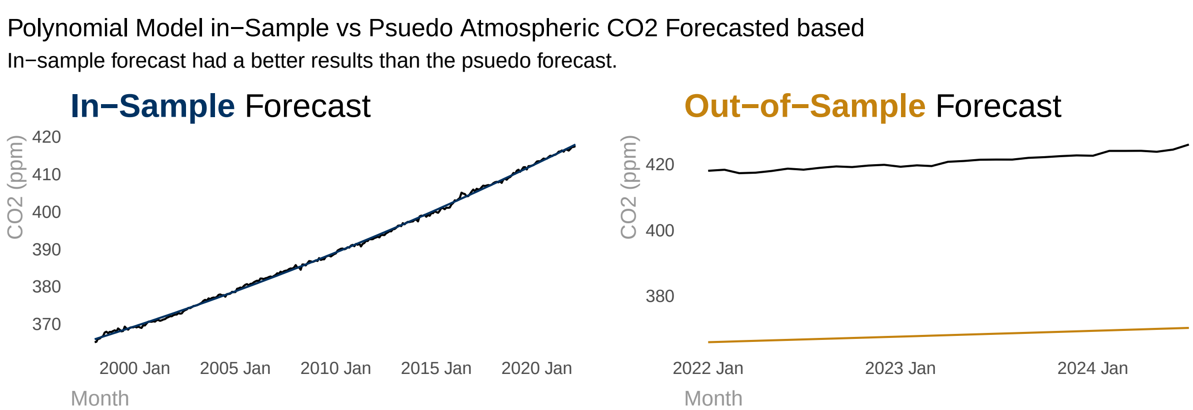

Figure 6: SA trained Ploynomial Model - in-Sample vs Psuedo

Our polynomial model performs well in sample, but appears to underpredict out of sample. The polynomial model produced RMSE = 52.77 which is much higher than the ARIMA model out of sample RMSE = 0.76.

5 Conclusion