Global CO_2 Emissions in 1997

Naikaj Pandya, Amit Gattadahalli, Michael Golas, Austin Pitts

Introduction

In 1960, Charles Keeling in his seminal paper, *The Concentration and Isotopic Abundances of Carbon Dioxide in the Atmosphere* made two notable observations:

1. That a seasonal variation in CO_2 concentrations was observed in the northern hemisphere, corresponding to the activity of land plants

2. That at longer horizons, beyond one year, global concentrations of CO_2 have increased at a rate of 1.3 p.p.m. either from the combustion of fossil fuels or from factors tied to the seasonal variation, exceeding the counteracting oceanic effect removing CO_2 from the atmosphere.

Keeling's analysis was conducted using data obtained from three gas analyzers, equipped to measure carbon dioxide concentrations continuously, located in Antarctica, Hawaii and California.

Our goal, in 1994, is to validate Keeling's observations using data collected during the intervening years, measured using modern optical sensors at higher frequencies and report any observed changes to the rates of accumulated CO_2 in the atmosphere. Using these estimates we plan to extend our study and apply time-series modeling techniques to forecast the trends and variation in expected future CO_2 concentrations to provide bounds on the anticipated levels of CO_2. Since the amount of atmospheric CO_2 carries broad environmental and economic effects, our results are relevant to both environmental and policy researchers as crucial estimates to help guide mitigating courses of action within the appropriate time frames.

CO2 Data

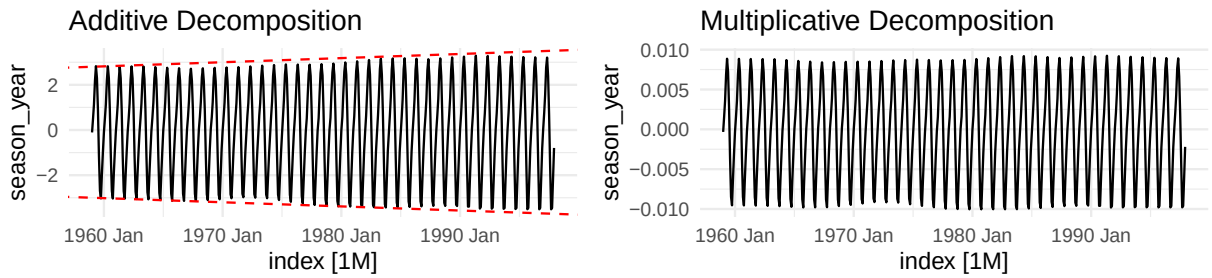
We conduct a timeseries-analysis of atmospheric CO_2 levels using data collected by the NOAA. As stated, the goal is to examine any long-term trends and seasonal fluctuations in CO_2 levels.

The input data sourced from NOAA is collected using a CO_2 analyzer installed at Mauna Loa that uses a technique based on infrared absorption, wherein a sensor measures the magnitude of absorption of light circulating in an optical cavity. Data is collected hourly, daily and monthly, we use the monthly average data for this analysis as our primary interest is devoted to long-term changes in CO_2 levels. An important aspect of the measurements is the ongoing calibrations of the analyzer. The absorption by the instrument depends on the total amount of CO2, therefore the temperature and pressure in the instrument, as well as the flow rate, need to be measured and frequent calibrations performed with reference gas mixtures of known amounts of CO2-in-dry-air. The intake lines are from the top of a 38 m tall tower next to the observatory, to avoid any influence on the measurements by human activities at the observatory. The difference of the ambient air measurements from the reference gas R0 are calculated, and these differences are used to calculate the true fraction CO2.

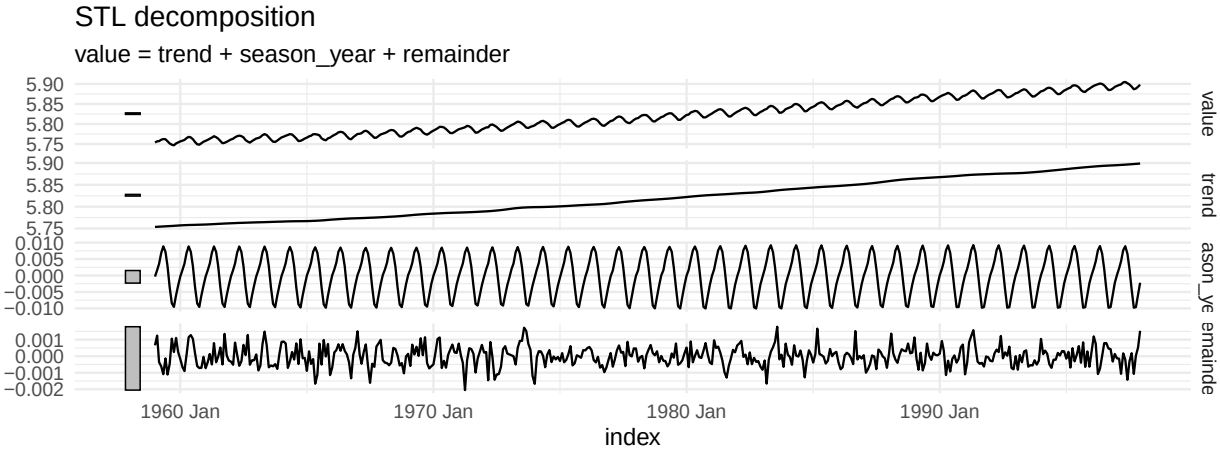
We begin with an exploratory analysis of the data guided by a few general observations apparent from visual inspection of the time-series:

- the data shows variation periodic in time
- the general level increases over time

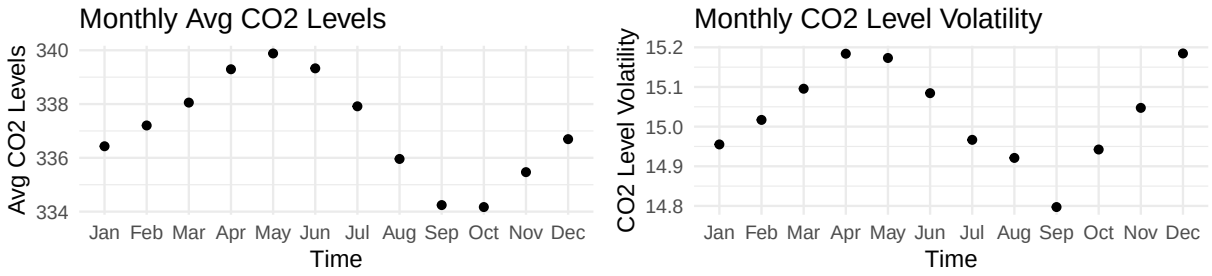
One question is whether the variation remains constant independent of the level of CO_2. Analysis of variation around the trend-cycle reveals a persistent increase in the amplitude of the fluctuations.



As we can see in the figures above, the Additive Decomposition flares outward. This leads us to conclude that the appropriate decomposition of the time-series into Trend, Seasonal, and Residual components is via Multiplicative Decomposition.



Looking at the STL decomposition, although the long run growth rate of co2 is very low, approx. 0.0127402% per year, almost linear at the time-scale of observation, we note that the growth is highly statistically significant.



Finally, observing the month-to-month average CO2 levels and volatility gives us an idea of the seasonality in our data.

Linear Time Trend Model

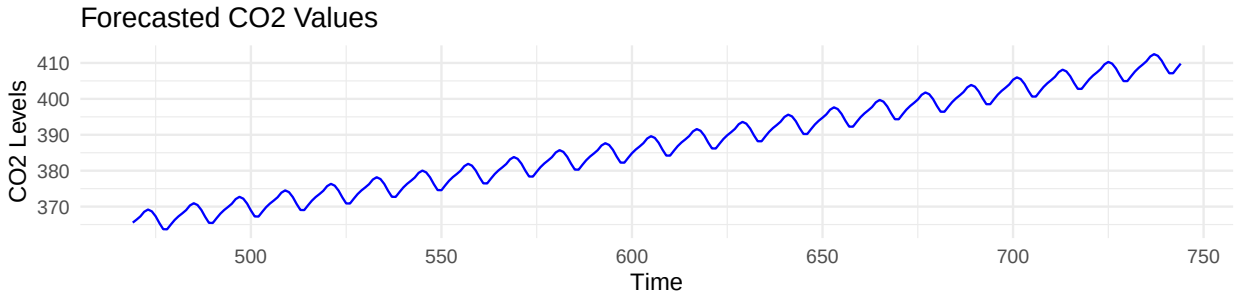
To setup our problem for validation we split our dataset into an in-sample train period spanning years prior to 1998 and post-1998 as the test period.

We fit linear, polynomial and quadratic models. We observe that the using a linear timeseries model the residuals exhibit a positive trend. This is also prevalent for the quadratic and polynomial models, however the magnitude of the trend is reduced. Our final model uses linear, quadratic, exponential and seasonal features.

```
final_model = lm(value~index1+index2+log_index + month,df)
```

In our analysis we observed quality of fit, distribution of residuals and plots showing normality of residuals for each model.

We then used our final linear model to generate forecasts to the year 2022.



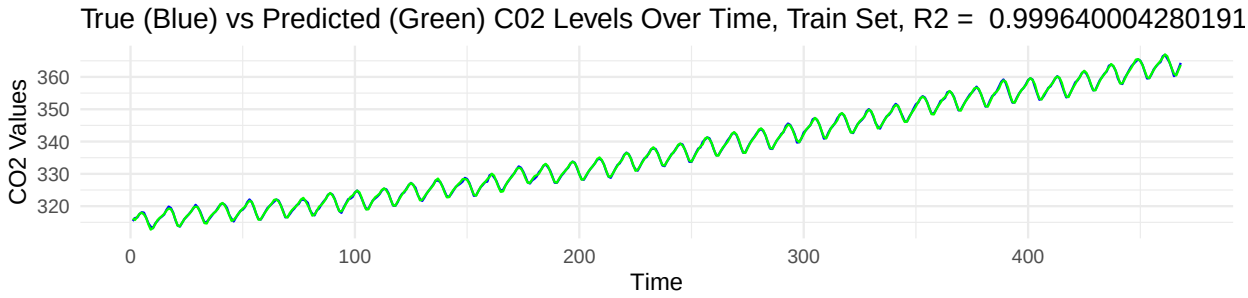
ARIMA Time Series Model

We developed an ARIMA model to fit to the series and generate forecasts to the year 2022. The parameters of our model are chosen by cross-validation. We find that it is necessary to detrend the CO_2 Series such that it is stationary. To achieve this we use a linear model to detrend series as it captures the linear & nonlinear temporal and seasonal trends inherent in the data.

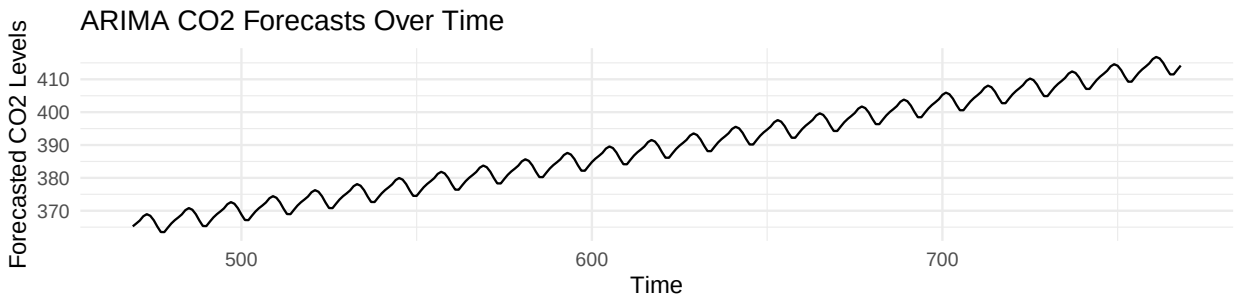
```
df2 = df
df2[['diff_value']] = final_model$residuals
arima_model = arima(df2[['diff_value']], order=c(2,0,0))
```

While there is not enough room to provide all of these plots (see our Appendix if you would like to), in doing this process, we also - checked mean/variance of series over time to visually validate stationarity assumptions - checked ACF/PACF plots - performed a grid search to find P, Q values that minimize in sample BIC - validated that our final model's residuals are white noise and approximately normally distributed

We also used our train and test set to measure our ARIMA model's performance, which turned out well.

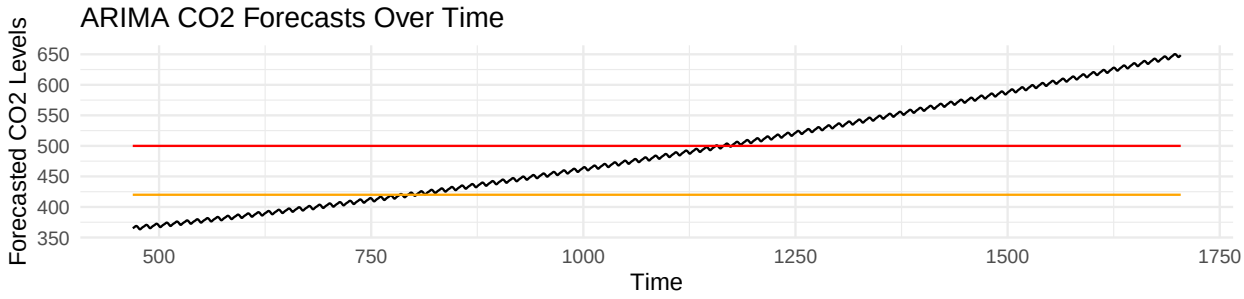


Similar to our linear model, we then used our final ARIMA model to also generate forecasts to the year 2022.



Forecasting Atmospheric CO2 Growth

We generate predictions for when atmospheric CO2 is expected to be at 420ppm and 500 ppm levels for the first and final times. Given errors observed from previous forecasts our hope is that our more modern models which utilize more sophisticated features will provide more accurate results of future CO_2 levels.



Based on this forcasting using our ARIMA model, our predictions are as follows: - First and Final Time at 420, April 2024 - Oct 2026 - First and Final Time at 500, April 2055 - Nov 2056

We are fairly confident that these will be close to accurate predictions based on our analysis of our ARIMA model and its performance, but forecasting so far into the future means that our predictions will likely not be perfect.

Conclusion

TODO - idk if we need this section, feel free to write something or remove the section all-together

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.