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 Mauna 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. The second set of plots focused on yearly averages and SDs, providing a clearer picture of trends and volatilities. The third set of 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 here we calculate 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.

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.

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.