Global CO_2 Emissions in 1997

By Majid Maki-Nayeri
Vinod Bakthavachalam
Artem Lebedev*

One of the very interesting features of Keeling and colleagues' research is that they were able to evaluate, and re-evaluate the data as new series of measurements were released. This permitted the evaluation of previous models' performance and a much more difficult question: If their models' predictions were "off" was this the result of a failure of the model, or a change in the system? Keywords: Replication, Modern Science

I. Introduction

In this introduction, you can assume that your reader will have just read your 1997 report.

II. Mona Loa CO_2 Data

We first download data Create a data pipeline that starts by reading from the appropriate URL, and ends by saving an object called `co2_present` that is a suitable time series object.

A. Recent Trends in Atmospheric Carbon

Conduct the same EDA on this data. Describe how the Keeling Curve evolved from 1997 to the present, noting where the series seems to be following similar trends to the series that you "evaluated in 1997" and where the series seems to be following different trends. This EDA can use the same, or very similar tools and views as you provided in your 1997 report.

Atmospheric carbon is plotted in Figure 1, and shows relentless increase in atmospheric CO_2 concentration even at the remote Hawaiian island.

III. Models and Forecasts

While these plots might be compelling, it is often challenging to learn the exact nature of a time serires process from only these overview, "time vs. outcome" style of plots. In this section, we present evaluate two classes models to assess which time series model is most appropriate to use.
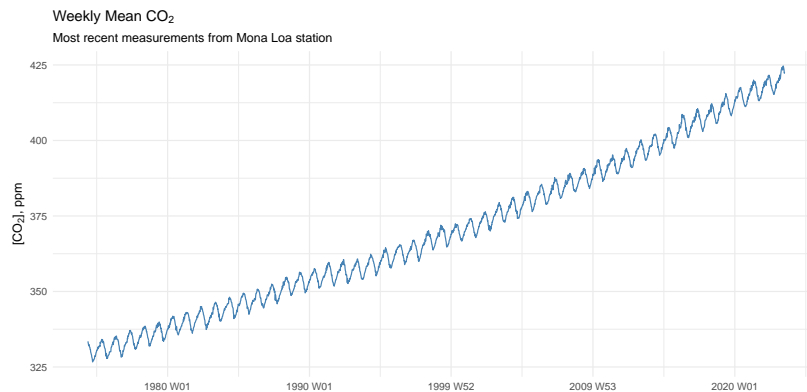
Figure 1. : Recent trend in CO_2 concentration

Note: CO_2 concentration from 1974 onward continues to exhibit ever accelerating growth, with strong seasonal oscillations. These patterns points to the need of de-seasoning and de-trending before further analysis.
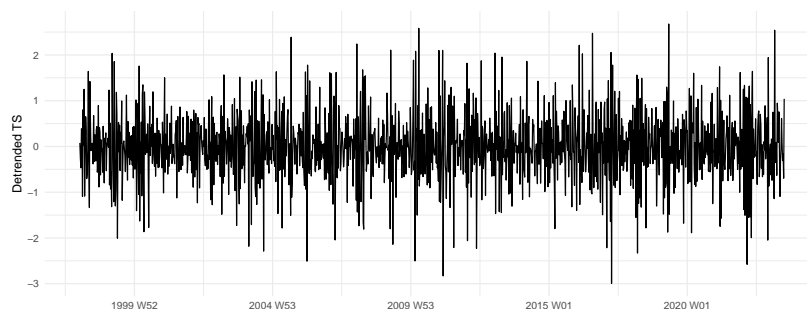

Figure 2. : De-seasoned/-trended series for CO_2 concentration

A.   *Linear Models*

Descriptively compare realized atmospheric CO2 levels to those predicted by your forecast from a linear time model in 1997 (i.e. "Task 2a"). (You do not need to run any formal tests for this task.)

B.   *ARIMA Models*

Descriptively compare realized atmospheric CO2 levels to those predicted by your forecast from the ARIMA model that you fitted in 1997 (i.e. "Task 3a"). Describe how the Keeling Curve evolved from 1997 to the present.
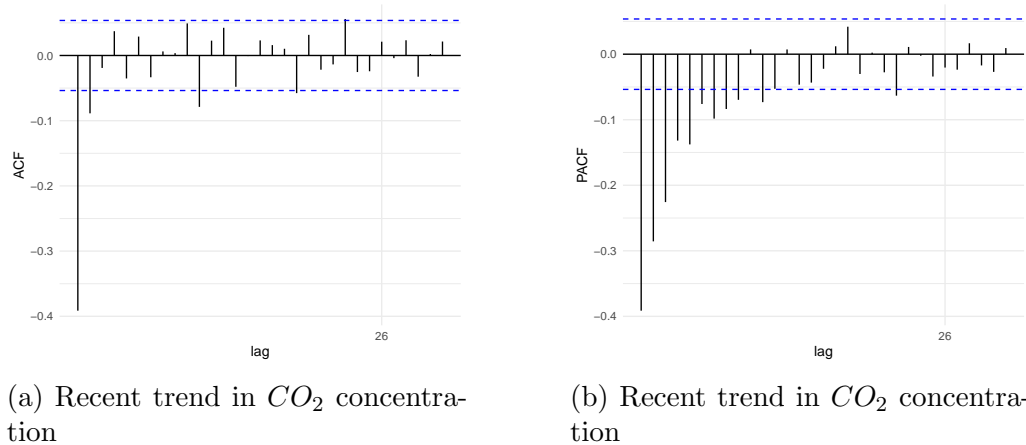
(a) Recent trend in CO_2 concentration

(b) Recent trend in CO_2 concentration

Figure 3. : Diagnostic plots for de-trended series

C. *Performance of 1997 linear and ARIMA models*

In 1997 you made predictions about the first time that CO2 would cross 420 ppm. How close were your models to the truth?

After reflecting on your performance on this threshold-prediction task, continue to use the weekly data to generate a month-average series from 1997 to the present, and compare the overall forecasting performance of your models from Parts 2a and 3b over the entire period. (You should conduct formal tests for this task.)

D. *Best models on present data*

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. For both SA and NSA series, fit ARIMA models using all appropriate steps. Measure and discuss how your models perform in-sample and (psuedo-) out-of-sample, comparing candidate models and explaining your choice. In addition, fit a polynomial time-trend model to the seasonally-adjusted series and compare its performance to that of your ARIMA model.

```
co2_train <- co2_diff %>% filter((week > yearweek("1998-01-01")) &
  (week < yearweek("2021-06-01")))

co2_test <- co2_diff %>% filter(week >= yearweek("2021-06-01"))
```

```
##
```

```
## Call:
## arima(x = co2_train$ppm, order = c(1, 1, 2), seasonal = list(order = c(0, 1,
##      1), period = 52))
##
## Coefficients:
##            ar1       ma1       ma2      sma1
##        0.08409  -0.67320  -0.09333  -0.81686
## s.e.   0.21164   0.21143   0.14559   0.02148
##
## sigma^2 estimated as 0.23119:  log likelihood = -831.02,  aic = 1672.03
```
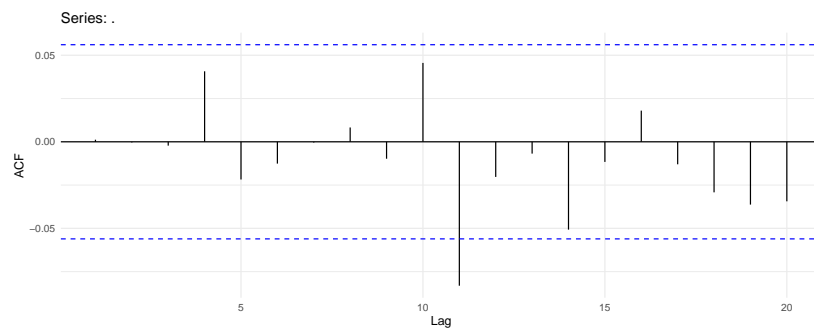

Figure 4. : ACF plot for SARIMA(1.1.1)(0.1.1)[52] residuals

Note: Residuals appear to be white noise

E.   Forecasts: How bad could it get?

With the non-seasonally adjusted data series, generate predictions for when atmospheric CO2 is expected to be at 420 ppm and 500 ppm levels for the first and final times (consider prediction intervals as well as point estimates in your answer). Generate a prediction for atmospheric CO2 levels in the year 2122. How confident are you that these will be accurate predictions?

```
## Don't know how to automatically pick scale for object of type <ts>. Defaulting
## to continuous.
```

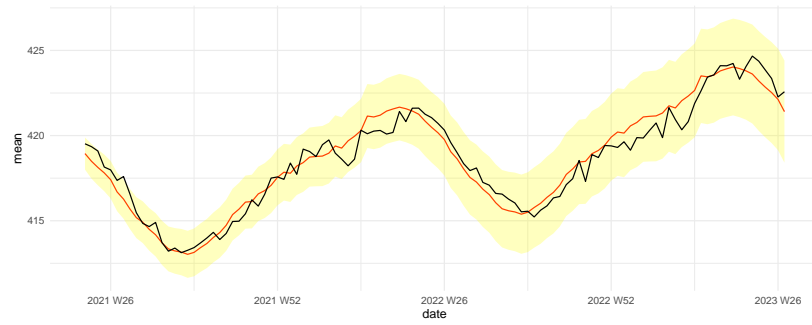## IV.   Conclusions

What to conclude is unclear.

Figure 5. : Forecast for future CO2 concentration

Note: Residuals appear to be white noise

Appendix: Model Robustness

While the most plausible model that we estimate is reported in the main, "Modeling" section, in this appendix to the article we examine alternative models. Here, our intent is to provide a skeptic that does not accept our assessment of this model as an ARIMA of order (1,2,3) an understanding of model forecasts under alternative scenarios.