Time Series Data Analysis Mauna Loa Project

Group 1 - Cora McAnulty and Jinghao Teng

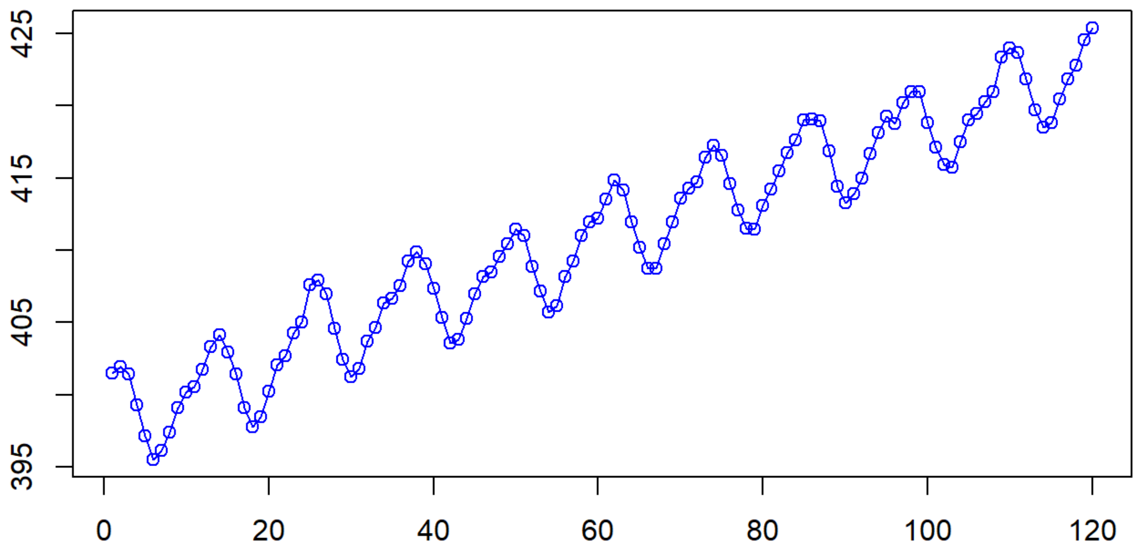
7 April, 2024

Part 1 - Data

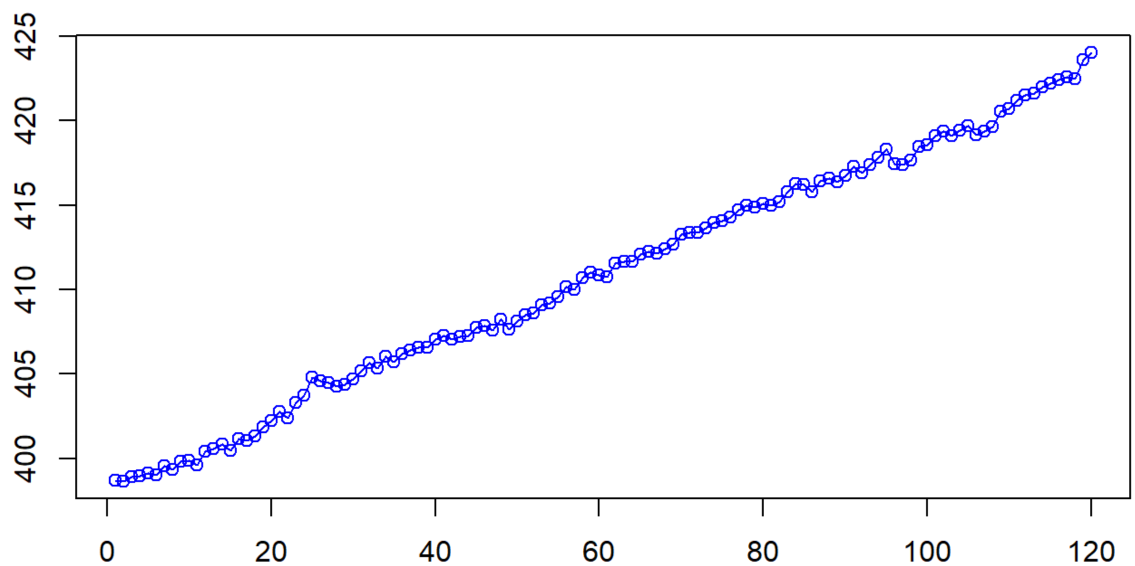
Dr. Charles D. Keeling began collecting data on the monthly average CO2 ppm from the Mauna Loa observatory in Hawaii in the 1950's. This data from one of the world's tallest and most remote places provides an unparalleled source for tracking the growing amount of carbon in our atmosphere. Since Dr. Keeling has been collecting data for so long, it is the longest recorded direct CO2 measurement from the atmosphere. Therefore, the Mauna Loa data is an excellent candidate for modeling with time series methods.

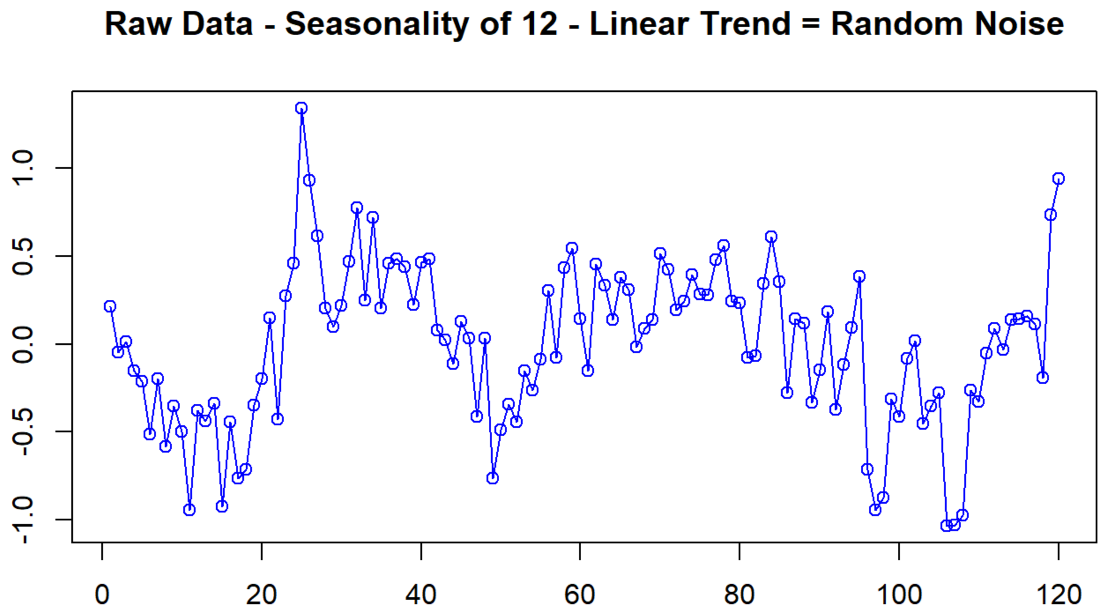
At first glance the data obviously exhibits strong seasonality and trend. The data is from 2014 to 2024, with about ten complete cycles, which leads us to believe it cycles by month. We assumed a period of 12 and eliminated it from the graph, leaving us with a fairly straight line -- the trend. We assumed a linear trend because of the straightness of the line. After eliminating the linear trend, we are left with random noise.

Raw Data

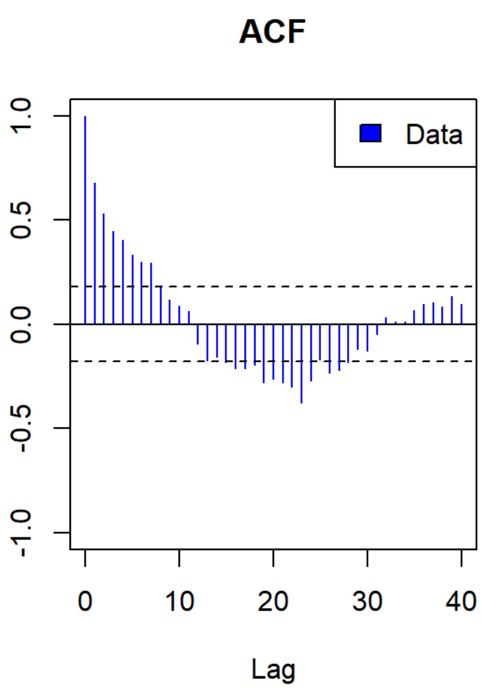


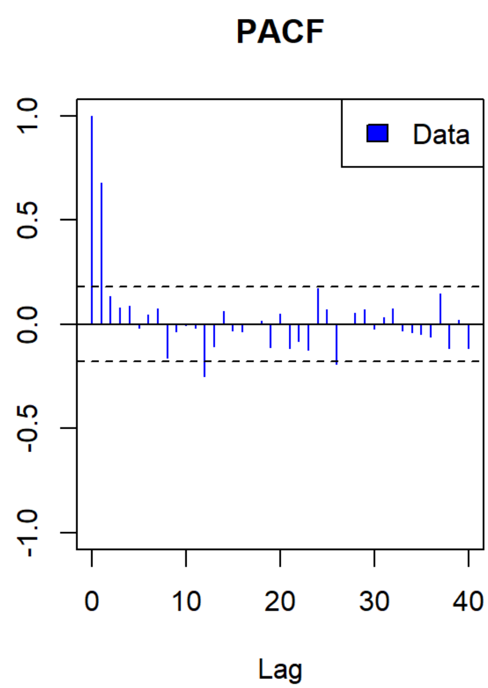
Raw Data - Seasonality of 12





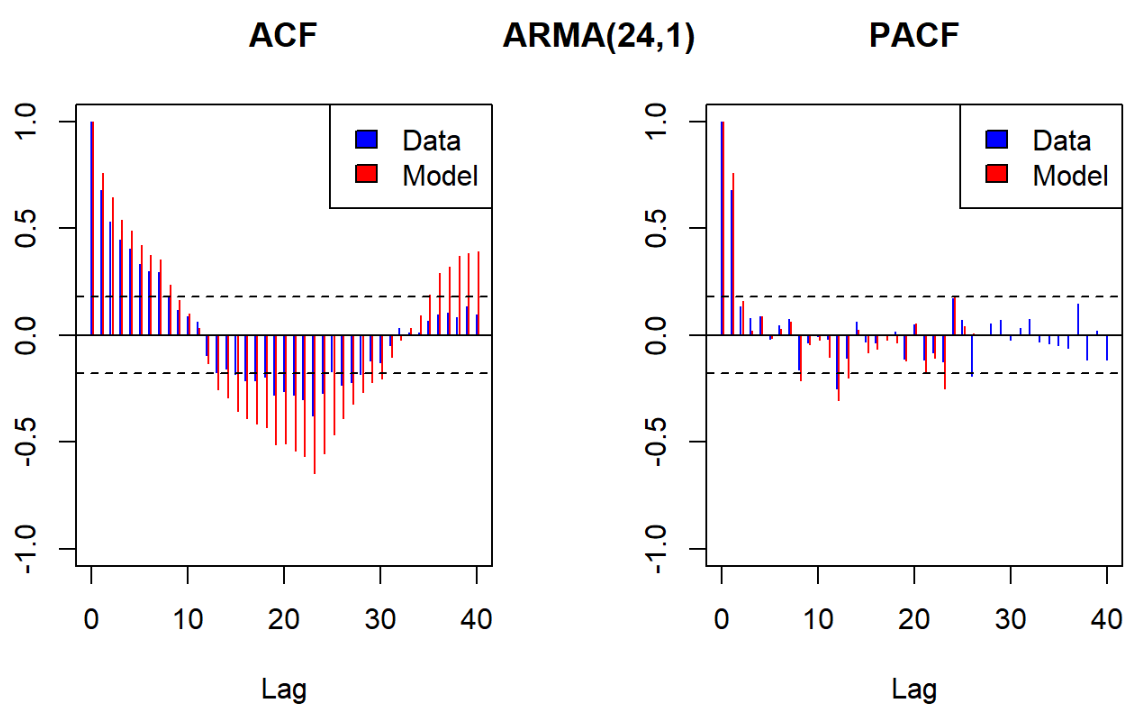
Our random noise's ACF and PACF suggested an ARMA model with a reasonably large p and smaller q. The ACF showed significant autocorrelation up to lag in the mid-20s, and the PACF showed partial autocorrelation up to about lag 1. These plots suggest an ARMA model with a value of p in the 20s and a value of q around 1 or 2.





Part 2 - Model

Based on the ACF and PACF, we chose the ARMA model for prediction. The PACF shows a sharp cut-off after 1, so q might equal 1. We tried multiple p and q values and generated the ACF and PACF for fitted values. Some looked very promising, like ARMA(24, 1) and ARMA(13, 1).

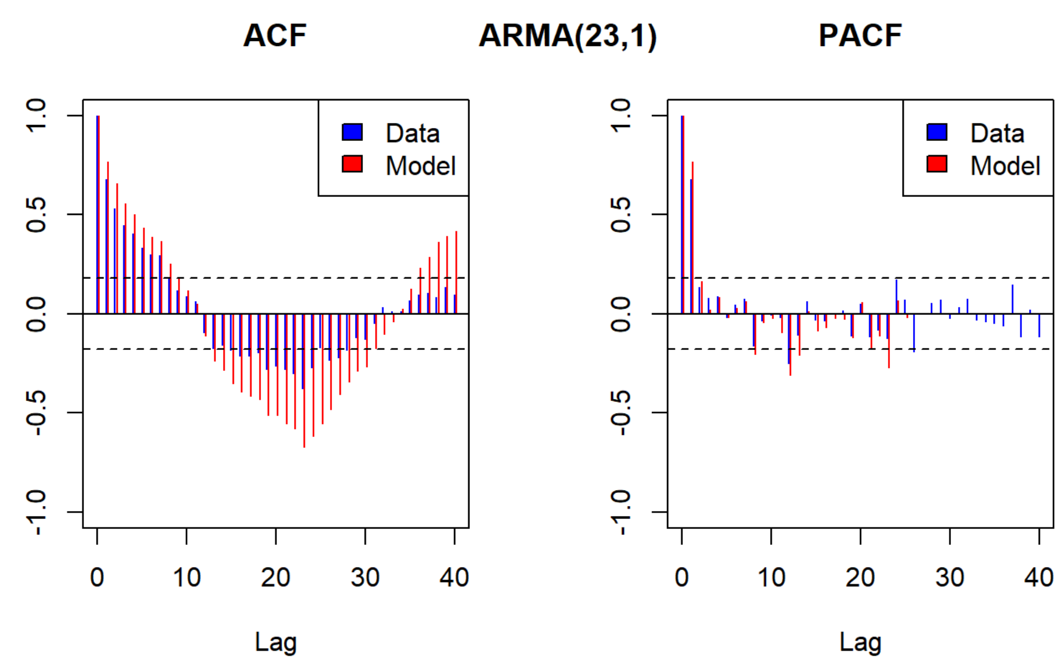


To help us decide which p and q have the best predictive power, we validated our selections using a for loop to calculate the Mean Absolute Error. We created a moving window of observations, each missing the last observation in the window. We then trained our model on this new set and predicted the next month's CO_2 level. We subtracted this from the known CO_2 level for that month to get an error measurement of our prediction. We worked backward from March 2024 and did this to collect 10 errors and calculate a mean absolute error to describe the predictive power of that ARMA model.

To examine the predictive power of various ARMA models, we calculated a matrix of MAE values for a range of p and q combinations. We tested p's from 10 to 30 and q's from 0 to 7, working off of the inferences we made about the data from the ACF and PACF graphs. We found the lowest MAEs in this matrix and created ACF and PACF plots to compare the models' shapes.

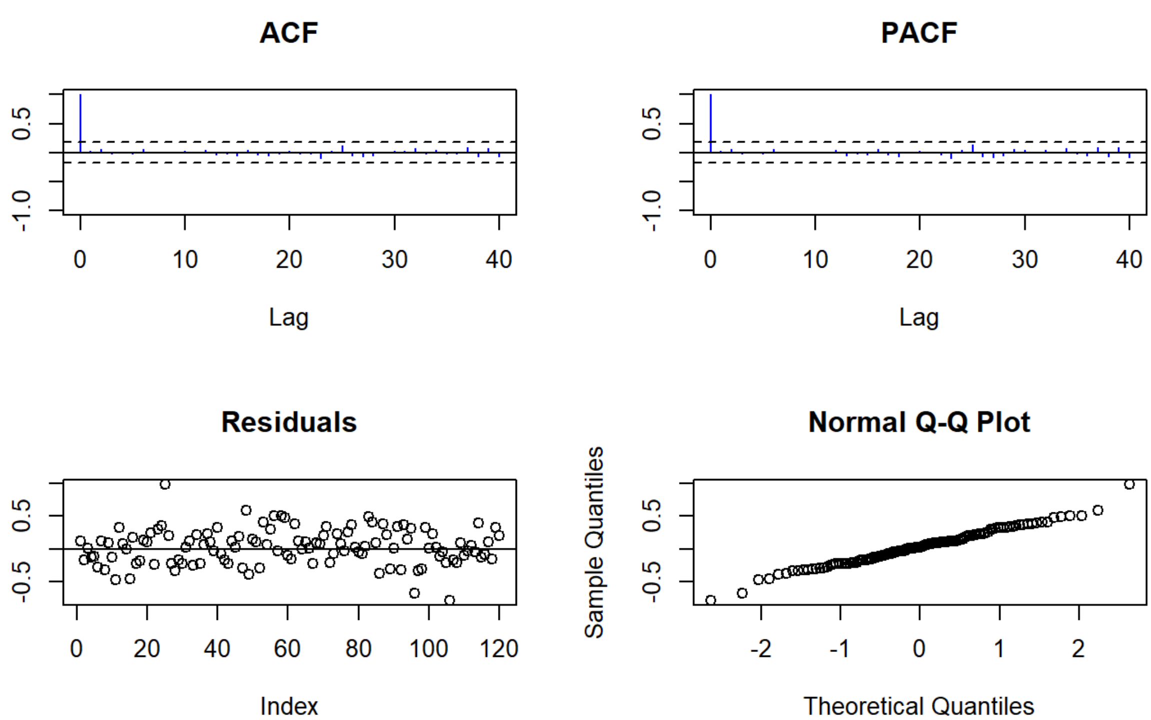
	10	11	12	13	14	15	16	17	18	19
0	0.2823214	0.2828255	0.2751561	0.2567832	0.2698106	0.2612024	NA	0.2592151	0.2654944	0.2670478
1	0.2994485	0.2882611	0.2382191	0.2621647	NA	0.3122310	NA	0.3612851	0.2566001	0.2606700
2	0.2656406	0.2927764	0.2708060	0.2537119	NA	NA	0.3556935	NA	NA	0.2424820
3	NA	0.3164345	0.3483949	0.3363677	NA	0.3334897	NA	0.3610299	NA	NA
4	NA	0.2752944	0.2705583	0.3651182	NA	0.3476394	NA	NA	NA	NA
5	0.2979805	0.2767705	0.3445274	NA	NA	0.3153887	NA	NA	NA	0.3225816
6	0.2918991	0.3357803	0.3457084	0.3195164	NA	NA	NA	NA	NA	0.3110966
7	0.3661015	0.3138622	NA	0.3403584	NA	0.3086773	NA	NA	NA	0.2455935

	20	21	22	23	24	25	26	27	28	29	30
0	0.2699790	0.2529909	0.2469261	0.2153767	0.2193212	NA	NA	NA	NA	NA	NA
1	NA	0.2588811	0.2561347	0.2138917	NA	NA	NA	NA	NA	NA	NA
2	NA	0.2405286	0.2360301	NA	NA	NA	NA	NA	NA	0.2506107	NA
3	NA	NA	0.2690844	NA	NA	NA	NA	NA	NA	NA	NA
4	NA	0.2463997	0.2324488	NA	NA	NA	NA	NA	NA	0.2547092	NA
5	0.3322090	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
6	0.3198681	NA	NA	0.2349651	NA	0.2797442	NA	NA	NA	NA	NA
7	0.2539532	NA	0.2428960	0.2408477	NA	0.2570235	NA	NA	NA	NA	NA



Ultimately, we decided on an ARMA(23,1) model to describe the noise in the Mauna Loa data. We ran some final tests on the model's residuals to further verify the model. The Tests of Randomness failed to reject the null hypothesis. Sample ACF also didn't show evidence against the IID noise assumption. These both showed the effectiveness of the ARMA(23,1) model in capturing the underlying patterns and structures within the Mauna Loa data.

```
Null hypothesis: Residuals are iid noise.
Test                        Distribution Statistic   p-value
Ljung-Box Q                Q ~ chisq(20)      3.12         1
McLeod-Li Q                Q ~ chisq(20)     17.25    0.6366
Turning points T   (T-78.7)/4.6 ~ N(0,1)        82    0.4671
Diff signs S       (S-59.5)/3.2 ~ N(0,1)        56    0.2704
Rank P          (P-3570)/220.4 ~ N(0,1)       3683    0.6082
```



Part 3 - Prediction

Our final prediction for the CO_2 in the atmosphere for the month of April is **427.07** ppm. With 95% confidence intervals we get a lower bound of 426.54 and an upper bound of 427.60. With 99% confidence our lower bound is 426.37 and our upper bound is 427.77.

	95% Confidence Interval	99% Confidence Interval
Lower	426.54	426.37
Upper	427.60	427.77