Mauna Loa Observatory CO_2 Project

Spring 2021

STOR 556-001

Yiran Li, Xinyue Mei, Cindy Yang, Xinyi Zhang, Olivia Zhu, Yunying Zhu

Introduction

The American scientist Charles D. Keeling monitored carbon dioxide at Mauna Loa Observatory to first alert the world to the possibility of anthropogenic contribution to the "greenhouse effect" and global warming. Our group tries to predict the monthly average of CO_2 level in April, 2021 based on historical data from Mauna Loa Observatory. We use the dataset of monthly mean carbon dioxide from March 2007 to February 2021 (total of 168 months) measured at Mauna Loa Observatory, Hawaii (National Oceanic & Atmospheric Administration, 2021). The carbon dioxide level is measured as the mole fraction in dry air with unit parts per million (ppm).

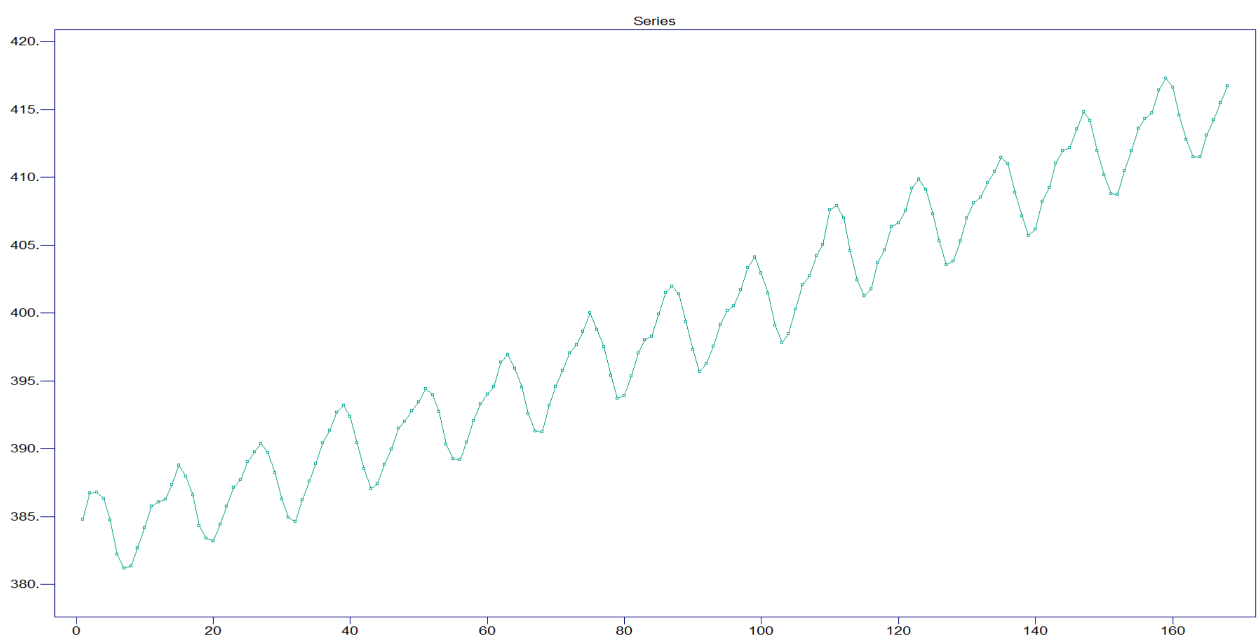

Figure 1: Time series plot on monthly average of CO_2 level 2007.3 - 2021.2

From the plot of the original time series (Figure 1), we can assume that there exists a trend component and a seasonal component. According to the plot, the period is 12 months, so we need to first remove the seasonal component from the original series. We should then fit an appropriate trend to the deseasonalized data and obtain the residuals. Finally, with the seasonal and trend components removed, we can fit the series into a model to obtain IID noise residuals.

Model Estimation

Before doing the actual prediction on the monthly average of CO_2 level for April, 2021, we need to remove seasonality and trend by using classical decomposition (Brockwell & Davis, 2016). From Figure 1, we can clearly detect a seasonal pattern on the monthly average of CO_2 level from Mauna Loa Observatory. Therefore, we want to first estimate the seasonality on this data. By setting x = original time series and using the functions s = season(x, 12) and plotc(s) in RStudio, we can get Figure 2. Here, we are dealing with a 12-month period; this plot shows the estimated seasonality on this dataset.

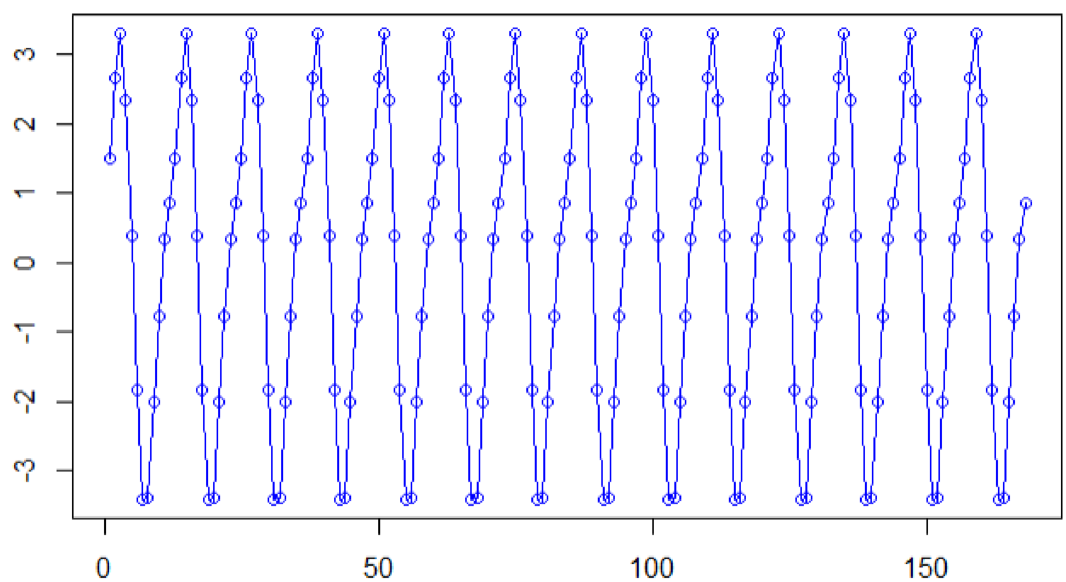

Figure 2: Plot of estimated seasonality of monthly average of CO_2 level

Next, we subtract the seasonality from the original time series. We set resid1 = x - s in RStudio, apply the function plotc(resid1) to visualize the residuals after deseasonalization, and get Figure 3. From this plot, we could clearly detect that there is an upward trend in the deseasonalized data.

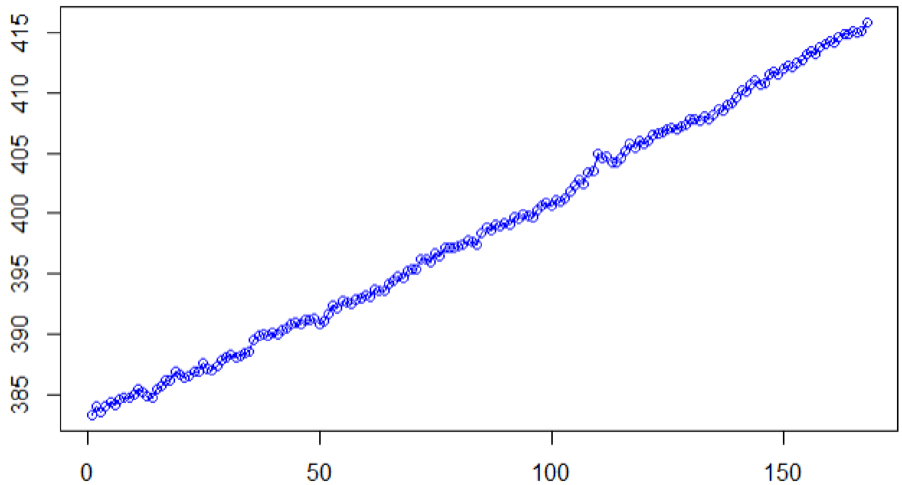

Figure 3: Plot of residuals of the deseasonalized data

The above result leads us to further estimate on the trend of the deseasonalized data. Using the function trend() in RStudio, we set m = trend(resid1, 2) and then plot the estimated trend m, as shown in Figure 4 below. We use a quadratic model for the trend .

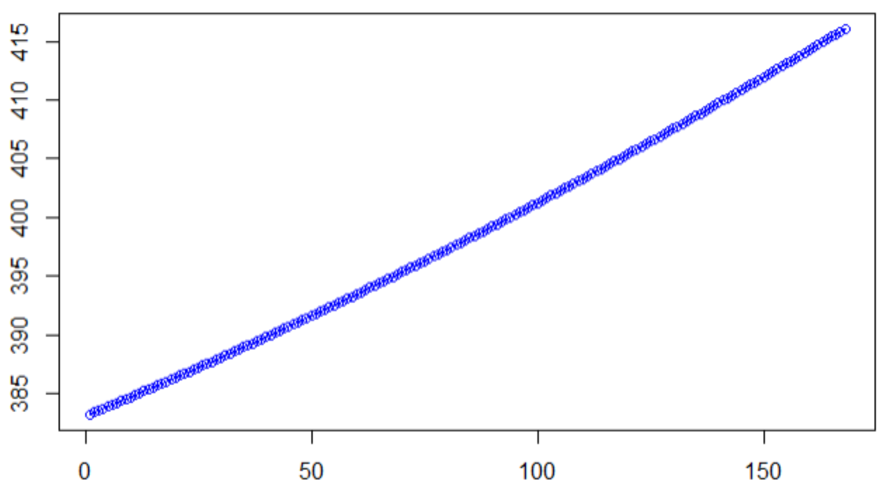

Figure 4: Plot of the estimated trend of the deseasonalized data

Then we take a look at the quadratic model it represents. The results are shown below.

```
Call:
lm(formula = resid1 ~ t + I(t^2))

Coefficients:
(Intercept)            t       I(t^2)
  3.831e+02    1.598e-01    2.168e-04
```

Figure 5: Quadratic trend model on deseasonalized data

By running the code resid2 = x-s-m in RStudio, we deseasonalize and detrend our data to leave out the residuals, resid2. We want resid2 to behave like IID models to demonstrate our model is a good fit to the data. We use plotc(resid2) to plot the data and get Figure 6. The data in Figure 6 does not resemble an IID model well: its mean and variance are not very stable over time. We further use plota(resid2) to generate the residuals ACF and PACF plots, and we get Figure 7. The ACF graph yields a clear left-to-right decaying pattern. At the same time, the PACF has one bar crossing the 95% confidence interval dash line at lag 1, additional to the bar at lag 0. These two properties reflected by the graphs indicate that a AR(1) model could be applied to resid2, so we decide to use Auto Regressive Moving Average(ARMA) model, which includes the AR property, to further model the residuals.

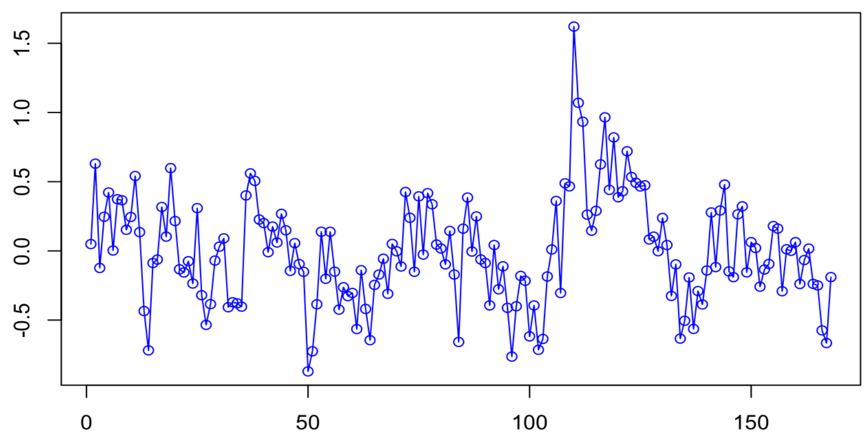

Figure 6: Plot of residuals after removing seasonality and trend

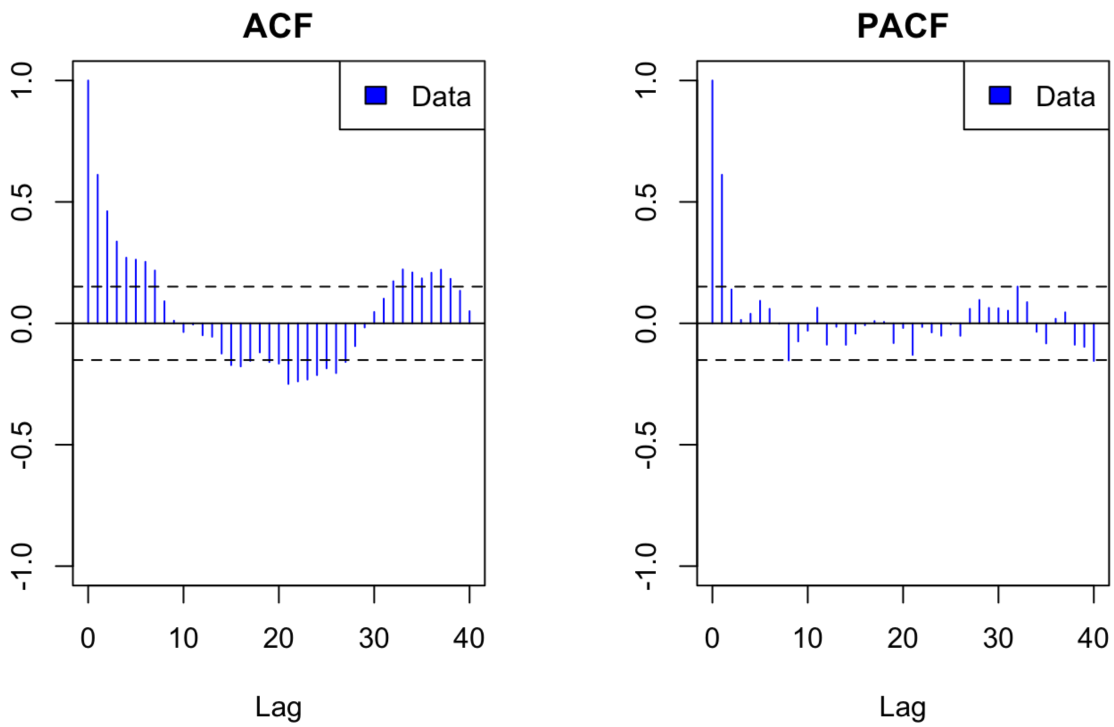

Figure 7: ACF and PACF plots of the residuals after removing seasonality and trend

We utilize ITSM software to auto-fit an ARMA model on the residuals resid2, by setting the max AR Order and MA Order both to be 5. Since the lower value of AICC indicates a better fitting model, the software helps us to find out that ARMA model with parameters (1,1) is the best fit with the lowest AICC, 81.021026. The sample mean of the residuals is $-2.006 * 10^{-13}$. For the final ARMA model, the AR coefficient is 0.7643 and the MA coefficient is -0.2536. The resulting model is written as the following with $Z_t \sim$ WN(0, 0.091150):

$$(X_t + 2.006 * 10^{-13}) - 0.7643 * (X_{t-1} + 2.006 * 10^{-13}) = Z_t - 0.2536 * Z_{t-1}, \text{ or}$$

$$X_t = -2.006 * 10^{-13} + 0.7643 * (X_{t-1} + 2.006 * 10^{-13}) + Z_t - 0.2536 * Z_{t-1}.$$

Model Testing

After fitting the ARMA model, we perform an informal test on sample autocorrelation function plot as shown by Figure 8. We do not see any bar that goes beyond the dash line (95% confidence interval) on the residual ACF plot. On the residual PACF plot, there is only one bar that goes beyond the dash line, which is less than the 5% of the total observations (40*5%=2). Therefore, the residuals after fitting the ARMA model appear to be an IID process, but we need further investigations.

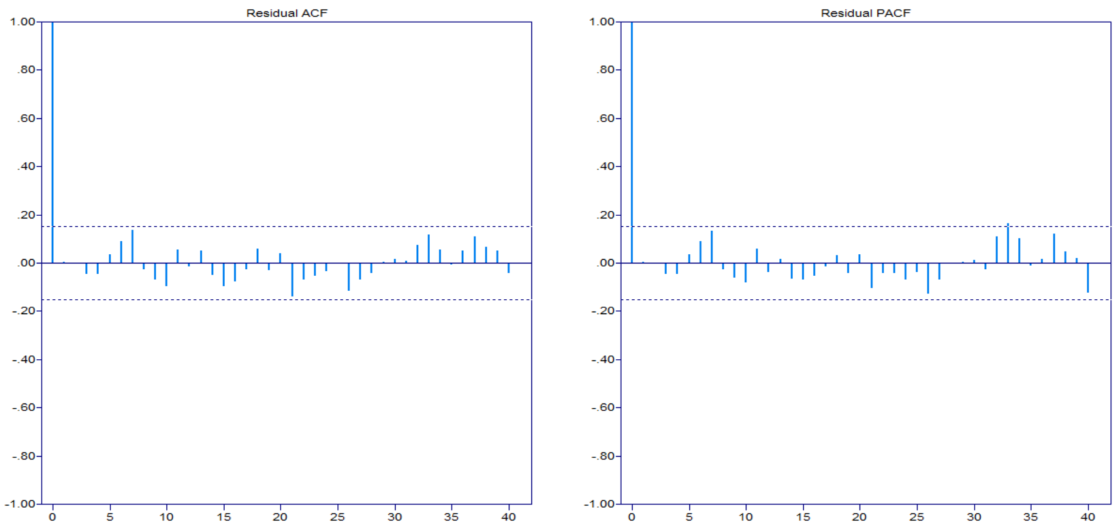

Figure 8: Plot of the final residuals ACF and PACF

To further test whether the final residuals are IID noise, we conduct the Ljung-Box test, McLeod-Li test, turning point test, different sign test, and rank test to test the null hypothesis: residuals are IID noise. The p-values for all five tests are greater than 0.05, so we fail to reject the null hypothesis that the residuals are IID noise.

```
============================================
ITSM::(Tests of randomness on residuals)
============================================

Ljung - Box statistic = 19.375 Chi-Square ( 20 ), p-value = .49758

McLeod - Li statistic = 10.854 Chi-Square ( 22 ), p-value = .97677

# Turning points = .11800E+03~AN(.11067E+03,sd = 5.4355), p-value = .17729

# Diff sign points = 85.000~AN(83.500,sd = 3.7528), p-value = .68937

Rank test statistic = .69190E+04~AN(.70140E+04,sd = .36452E+03), p-value = .79439
```

Figure 9: Tests of randomness

Knowing the final residuals are likely to be IID noise, we now want to test whether the ARMA model makes accurate predictions. We perform predictions on the last three existing observations for December 2020, January 2021, and February 2021 (observation 166, 167, 168) based on the ARMA model constructed from the first 165 observations. As the below table shows, the three actual observations with values of 414.25, 415.52, 416.75 ppm respectively fall in the 95% prediction bounds. These results show that our ARMA model works well for prediction.

Observation#	Actual Observation (ppm)	Prediction (ppm)	Lower 95% Prediction Bound (ppm)	Upper 95% Prediction Bound (ppm)
166	414.25	414.67	414.08	415.26
167	415.52	416.06	415.40	416.73
168	416.75	416.85	416.14	417.55

Figure 10: Table of predictions on last 3 existing observations

Prediction and Conclusion

To predict the monthly average of CO_2 level in April, 2021, we use the ARMA model constructed from all the 168 observations. Using the ITSM software, we find that the predictions for monthly averages of CO_2 levels in March and April, 2021 are 417.59 ppm and 419.05 ppm respectively. The 95% prediction bounds are [417.00, 418.18] ppm for March and [418.39, 419.72] ppm for April, and the 99% prediction bounds are [416.81, 418.37] ppm for March and [418.18, 419.93] ppm for April.

To conclude, we fit the ARMA(1,1) model on the residuals obtained by removing seasonality and trend. The results from the tests of randomness and the predictions of the last 3 existing observations prove the validity and accuracy of our model. We then use this model to make our final prediction for the monthly average of CO_2 level in April, 2021, which is **419.05 ppm**. We are 95% confident that the true monthly average of CO_2 level in April falls between 418.39 ppm and 419.72 ppm; we are 99% confident that the true monthly average of CO_2 level in April falls between 418.18 ppm and 419.93 ppm.

References

Brockwell, P. J., & Davis, R. A. (2016). *Introduction to time series and forecasting*. Switzerland: Springer.

National Oceanic & Atmospheric Administration. (2021). Global Monitoring Laboratory - carbon cycle greenhouse gases. Retrieved March 31, 2021, from http://www.esrl.noaa.gov/gmd/ccgg/trends/.