Global CO_2 Emissions in 1997

By Abbas Siddiqui
Matthew Mollerus
Eric Jung*

Global attention is turning toward the consequences of human-actions in our environmental system. The IPCC has been in existence and studying these trends for more than ten years, and has released its second assessment report in 1995. In this report, the IPCC notes that the balance of the evidence suggests that human-actions play a role in the changing climate. Although, there is little political will to change this activity, neither have global progressive and conservative politicians broken into clear partisan camps. Here, we assess data from the Mona Loa observatory to describe and predict global CO_2 concentrations under several possible scenarios. What we find, when we run the analysis, is concerning.
Keywords: Replication, Modern Science

I. Introduction

In light of the global interest in CO_2 emissions and the impact it may have on accelerating climate change we investigate Global CO_2 patterns in order to help with forecasting and describing global CO_2 concentration. We hope that through better understanding the process of CO_2 emissions we can craft better legislation, develop better metrics for measuring global progress on tackling CO_2 emissions, and appropriately bring attention to any alarming changes. Our goal is to bring awareness to CO_2 emissions trends over approximately the last 40 years, and to warn the public of the ramifications of such trends continuing on into the future. Furthermore, we are interested in weighing the impacts of technology and policies upon the levels of atmospheric CO_2.

II. Background

A. Carbon Emissions

It is only recently that the international community has decided that rising greenhouse gas emissions, including gases like CO2, are a harrowing threat to our planet by causing changes in climate, including but not necessarily limited to global warming. However, we've known about this problem for years. In 1938,

* Siddiqui: UC Berkeley, School of Information, asiddiqui6@ischool.berkeley.edu. Mollerus: , mjmollerus@ischool.berkeley.edu. Jung: , erijung@ischool.berkeley.edu. The authors would like to thank their instructors from MIDS 271.

scientists discovered that global temperature was rising [1]. In 1958 the Keeling curve was published and marked the first attribution of global temp increases to CO_2 [2]. Later, in 1988 the Intergovernmental Panel on Climate Change assisted in making climate research accessible for policymakers. Just 3 years ago, the United Nations Framework Convention on Climate Change (UNFCCC) was the first international treaty supported by 197 countries designed to address greenhouse gas emissions such as CO_2. However, despite the new international effort CO_2 levels continue to climb, thus prompting further investigations into the impact of CO_2 emissions.

III. Research Question

1. What type of model best estimates the CO_2 process?

IV. Measurement and Data

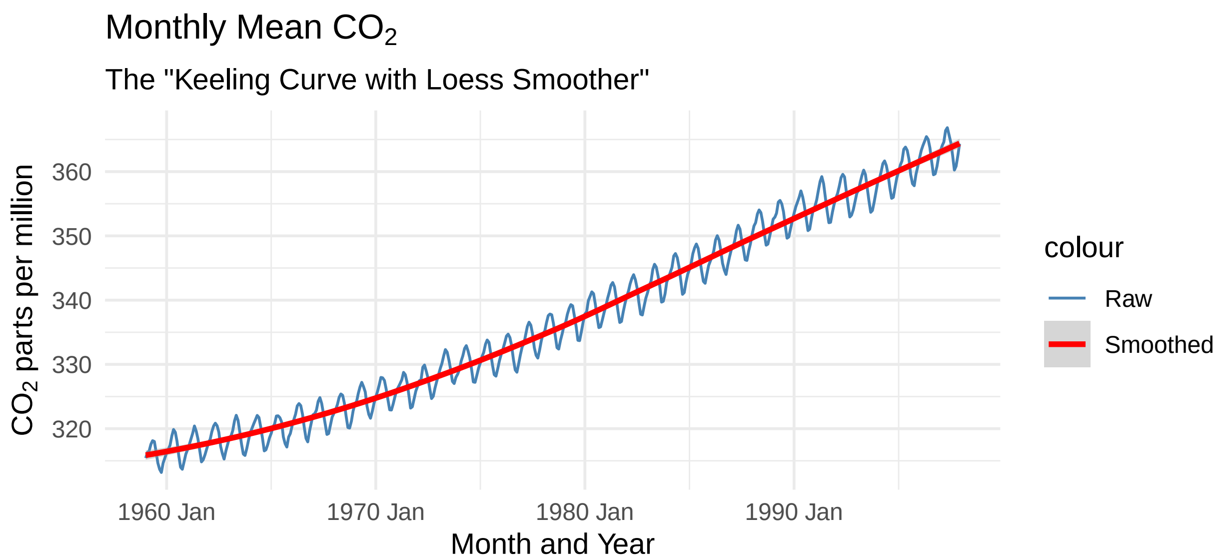
A. *Measuring Atmospheric Carbon*

In this study, we rely on the data collected at the Mauna Loa Observatory in Hawaii, which sits on the side of the Mauna Loa volcano approximately 3400 meters above sea level. At its high altitude and relatively isolated location, this station is in an apt location for collecting atmospheric samples representative of Earth's atmosphere.

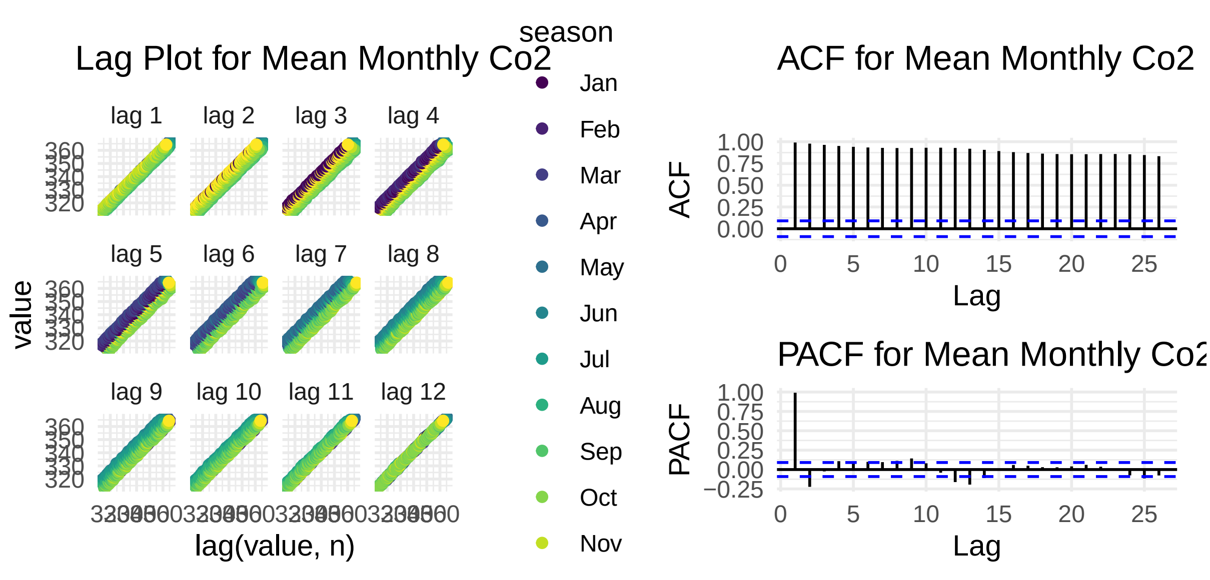
B. *Historical Trends in Atmospheric Carbon*

Looking at the data we will be looking for trends in the mean CO_2 values by month over the span of nearly 40 years (1959 - 1997). In the following EDA we hope to discover key trends to help inform our subsequent modeling efforts.

We first plot the data we have.



Monthly average atmospheric carbon is plotted in the time series plot, and shows some worrying trends. We see a consistent upwards trend in CO_2 part per million over time. Even more, a careful examination of the time series plot suggests some worrying trends as the curve appears potentially quadratic, indicating quadratic growth of CO_2 concentration over time. We also examine the ACF and PACF plots.



Looking at ACF and PACF plot likely indicates MA(1) process with a seasonal component, likely of lag 12 because it's a yearly process with monthly data as shown in the lag plot. However, given that the data is non-stationary we believe an ARIMA model would be most appropriate.

V. Models and Forecasts

While these plots might be compelling, it is often challenging to learn the exact nature of a time series process from only these overview, "time vs. outcome" style of plots. In this section, we present three classes of models to assess which class of time series model is most appropriate to use.

A. Linear Models

To begin, we fit a linear model of the form:

$$\mathrm{CO}_2 = \phi_0 + \phi_1 t + \epsilon_{eit} \tag{1}$$

We estimate best fitting parameters on this model in the following way,

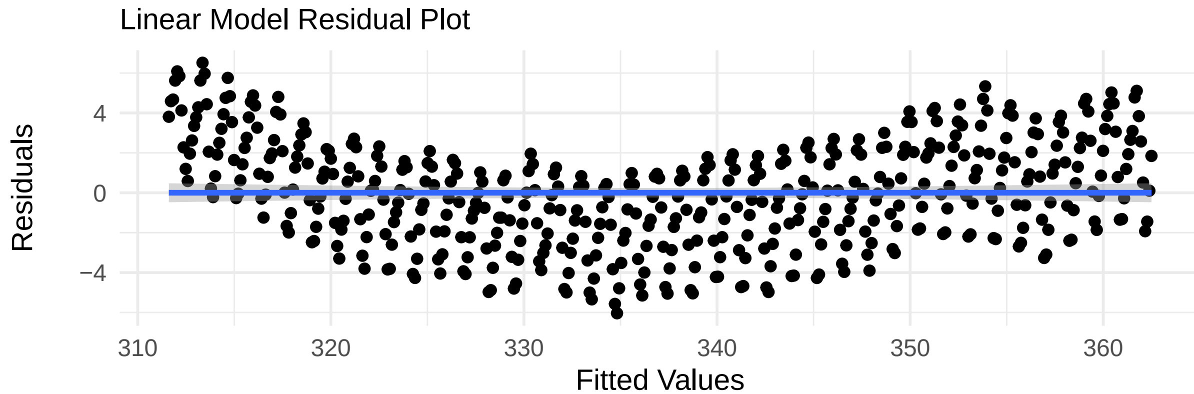
```
linear_model <- co2_tsib %>%
  model(
    linear_trend = TSLM(value ~ trend())
  )
```

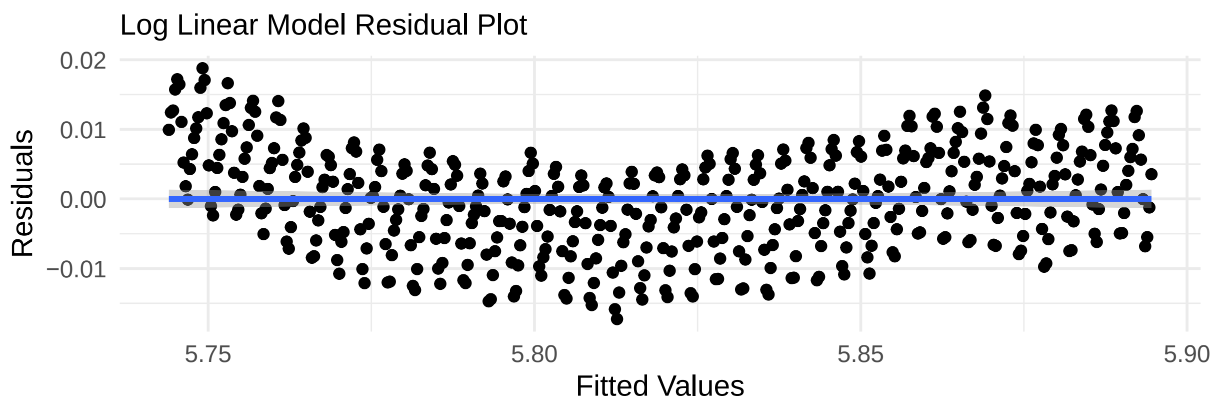
For completeness, we also estimate a log-linear model of the form:

$$\log(CO_2) = \phi_0 + \phi_1 t + \epsilon_{eit} \tag{2}$$

```
log_linear_model <- log_co2_tsib %>%
  model(
    log_linear_trend = TSLM(log_value ~ trend())
  )
```

We examine the residuals





However, the log transformation did not appear to have much impact as we were still lacking a seasonal component. We didn't have a clear exponential trend or heteroskedastic/exploding residuals in our initial linear model, so it makes sense that this has little effect.

B. *Quadratic model*

We compare this linear model to a quadratic model of the form:

$$CO_2 = \phi_0 + \phi_1 t + \phi_2 t^2 + \epsilon_{eit} \tag{3}$$

We estimate the best fitting parameters on this model in the following way,

```
quadratic_model <- co2_tsib %>%
  model(
    quadratic_trend = TSLM(value ~ trend() + I(trend()^2))
  )
```

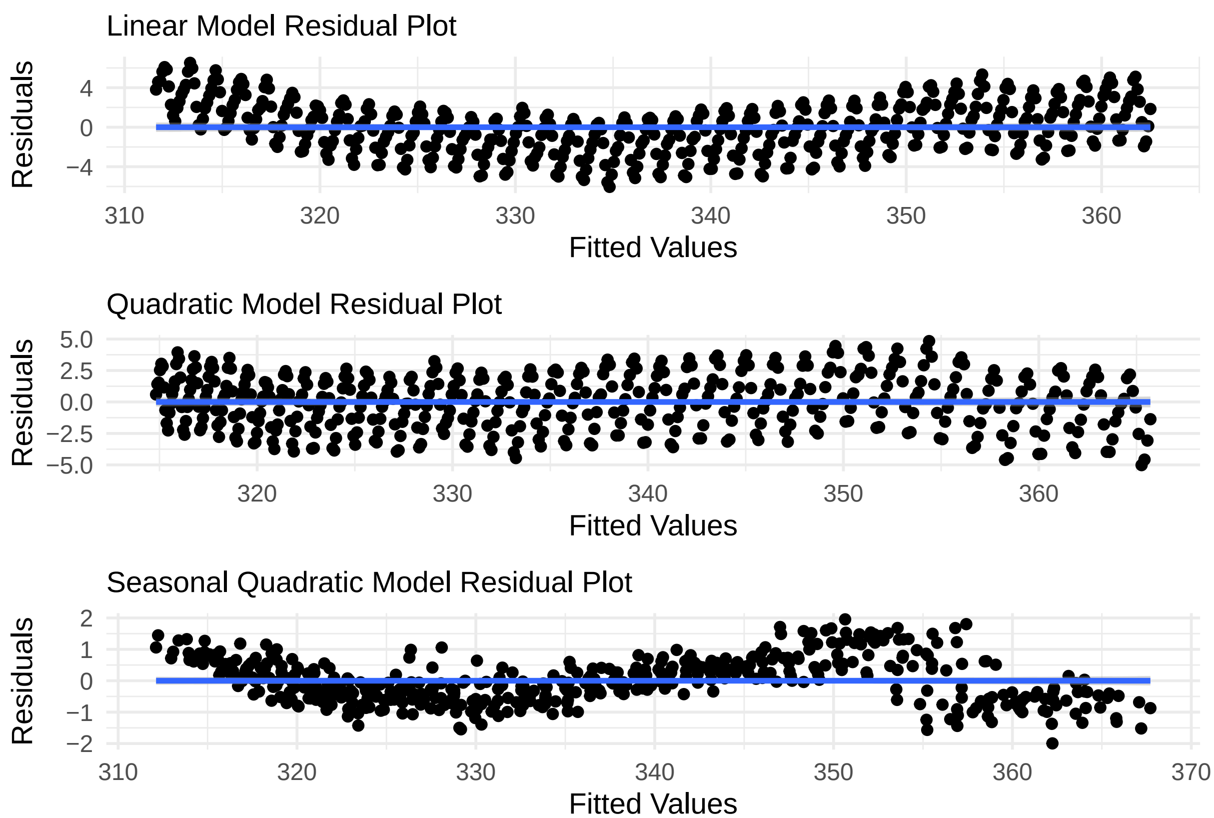
Finally, we fit a quadratic trend model with seasonal dummys of the form:

$$\text{CO}_2 = \phi_0 + \phi_1 t + \phi_2 t^2 + \text{Season} + \epsilon_{eit} \tag{4}$$

We estimate this model as below:

```
quadratic_seasonal_model <- co2_tsib %>%
  model(
    quadratic_seasonal_trend = TSLM(value ~ trend() + I(trend()^2) + season())
  )
```

We compare the residuals of the linear model to our quadratic models



The quadratic model also appears to have slightly more normally distributed residuals relative to the linear model, but there is still clear evidence of seasonal effects not being captured. Albeit both the linear and quadratic models show these seasonal artifacts. We see that the seasonal quadratic model performs best relative to the other models here.

C. ARIMA Models

We recognize that the above models are simple and are time deterministic. We build up to a ARIMA model by examining the ACF, PACF, and lag plots of the data. Recall that from ACF and PACF plots we found that the series was likely indicated a MA(1) process with a seasonal component, likely of lag 12 because it's a yearly process with monthly data as shown in the Lag plots. Additionally, the series appears to be non-stationary and thus an ARIMA model would likely be best for this data. Consequently, let's use autoARIMA to select the lowest by AIC.

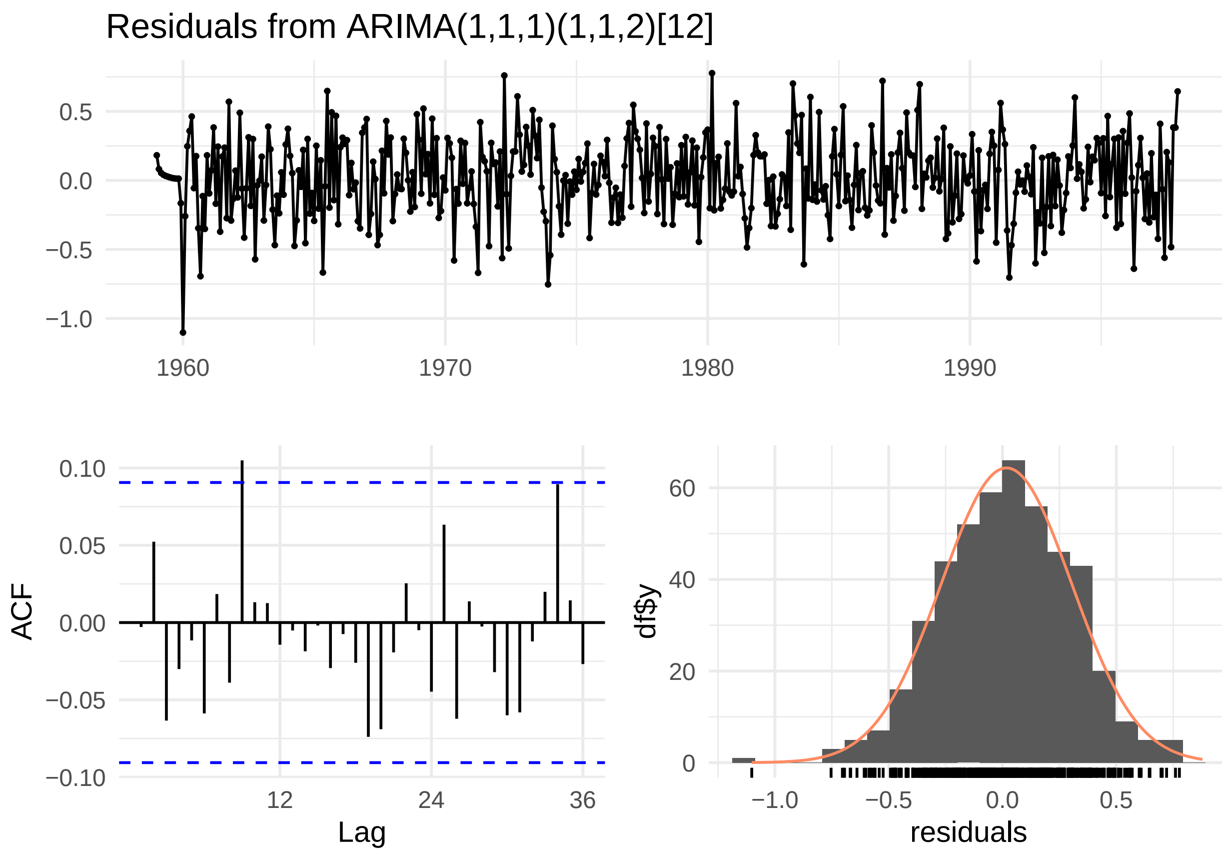
We fit our ARIMA model as follows:

```r
best_arima_model <- auto.arima(co2_ts, seasonal = TRUE)
```

We note that the final ARIMA model chosen is a ARIMA(1,1,1)(1,1,2)[12]. This means this model is of the form:

$$(1-\phi_1 B)(1-\Phi_1 B^1 2)(1-B)(1-B^1 2)Y_t = (1+\theta_1 B)(1+\Theta_1 B^1 2+\Theta_2 B^2 4)\epsilon_t \tag{5}$$

where Y_t is CO_2 ppm at time t, B is the backshift operator, and ϵ_t is the error terms at time t.




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(1,1,1)(1,1,2)[12]
## Q* = 19, df = 19, p-value = 0.4
##
## Model df: 5.   Total lags used: 24
```

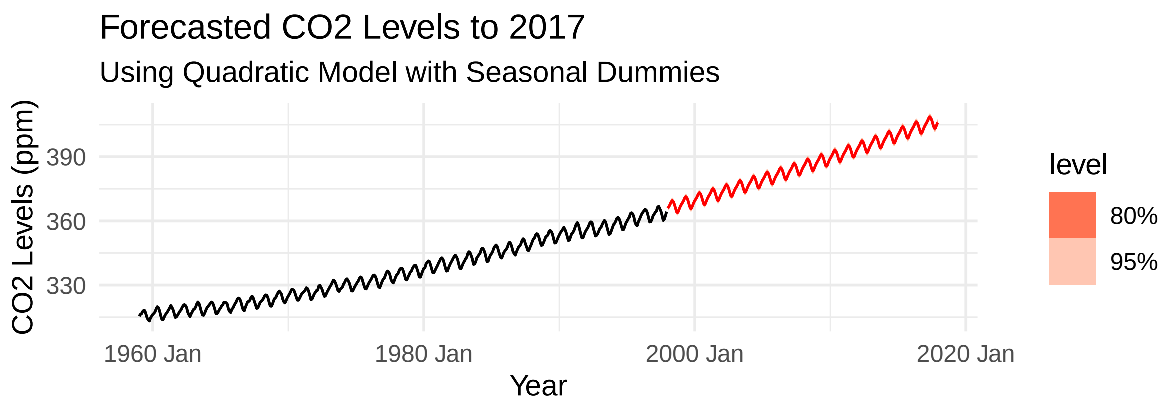
Ljung-Box test indicates no serial correlation, and the residual plots look normally distributed. The ARIMA model appears to be well fit.

D. Forecasts

Finally, we use two of our estimated models to forecast for future periods.

First, we forecast using the quadratic trend with seasonal dummies model for a 20 year projection.

Second, we also use the estimated ARIMA model for projecting until 2022.



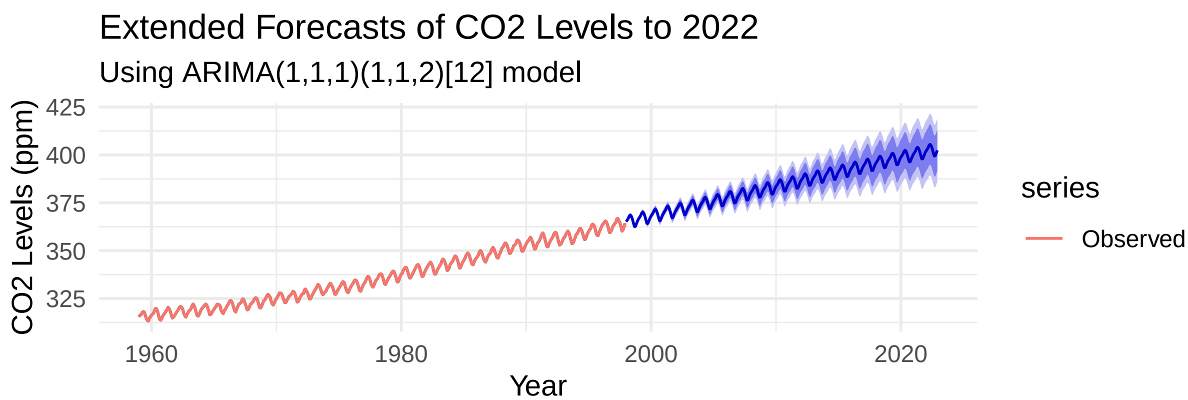


Table 1—ARIMA Forecast CO_2 ppm Threshold Crossings

	Point Forecast	Early PI	Late PI
420 ppm Year	2032.417	2023.583	Never
500 ppm Year	2084.417	2055	Never

Note: Where different points of our confidence interval cross the 420 ppm and 500ppm thresholds. The Early PI is the upper bound of the 95 percentile CI and Late PI is the lower bound of the 95 percentile CI.

Using the ARIMA model we see certain projection thresholds of interest in Table 1. We see that the ARIMA model at the earliest predicts crossing the 420 ppm threshold in 2023 with a point projection of crossing this 420 ppm threshold in 2032. If all goes well however, we could also never cross this threshold, as indicated by the Late PI column.

The models we generated including the linear, quadratic, quadratic with seasonal terms, and ARIMA all agree that the co2 ppm are increasing. We want to interpret the coefficients for one of our best performing models, the quadratic model with seasonal terms. We see that there is a direct trend between increase in time and Co2 levels, with a small positive weight on the quadratic term. Looking at the seasonal terms, we can see that co2 levels increase from January till May, then slowly start to fall back down by July. August through December see the steepest decreases in co2 ppm levels relative to January, with the lowest being -3.24 co_2 ppm in October, until it returns to January levels.

VI. Conclusions

In general, all of our models agreed that the co2 levels are continuing to increase and that indicates that the issue of co2 levels rising is even more pressing now than it was in the past. In particular, our ARIMA model was our best fit for the data even with it's ability to be less preturbed by seasonal trends, it still indicated a clear upwards trend in the atmospheric CO_2 levels over time. This is very concering for the future as we are still unsure of the potential effect of CO_2 levels increasing this quickly.

VII. References

[1] Callendar, G.S. (1938), The artificial production of carbon dioxide and its influence on temperature. Q.J.R. Meteorol. Soc., 64: 223-240. https://doi.org/10.1002/qj.49706427503

[2] Keeling, C.D. (1960), The Concentration and Isotopic Abundances of Carbon Dioxide in the Atmosphere. Tellus, 12: 200-203. https://doi.org/10.1111/j.2153-3490.1960.tb01300.x

VIII. Appendix

Table 2—Performance of Models

	AIC	AICc	BIC	Log Like.	RMSE
Linear	904.83	904.89	917.28	−1113.48	2.612
Quadratic	735.41	735.5	752	−1027.77	2.175
Quadratic w/ Season	−286.49	−285.42	−224.26	−505.82	0.713
ARIMA	180.78	180.97	205.5	−84.39	0.287

Global CO_2 Emissions from 1997 to 2022 and Beyond

By Abbas Siddiqui
Matthew Mollerus
Eric Jung*

In 1997 our team wrote a report detailing the Keeling Curve's evolution up until 1997. We had created linear, quadratic, and ARIMA models to model CO_2 parts per million (ppm) in the atmosphere as measured by the Mauna Loa station in Hawaii. The 1997 report showed a steadily increasing trend of CO_2 ppm which at the time was mildly concerning. At this time, we have more evidence that high CO_2 levels in the atmosphere has profound impact on both climate change and organisms. We now reevaluate the models developed in 1997 and update them with new data to determine if the situation surrounding CO_2 levels has changed. What we find is grim. *Keywords: Replication, Modern Science*

I. Introduction

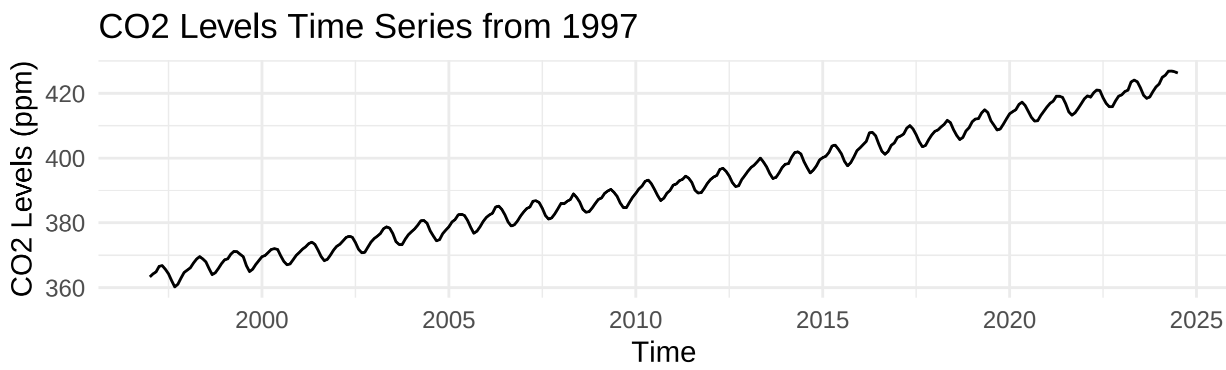
Since our last report there has been more evidence of rising CO_2 levels negatively impacting climate and living organisms [1][2]. We update our report from 1997 by pulling the CO_2 data from the National Oceanic and Atmospheric Administration's Global Monitoring Laboratory up to July 2024. The data is also from Mauna Loa, Hawaii to be consistent with the 1997 report. We aim to check if the 1997 models remain well fit models, and to assess if the CO_2 emissions trends have changed over time.

II. Measurement and Data

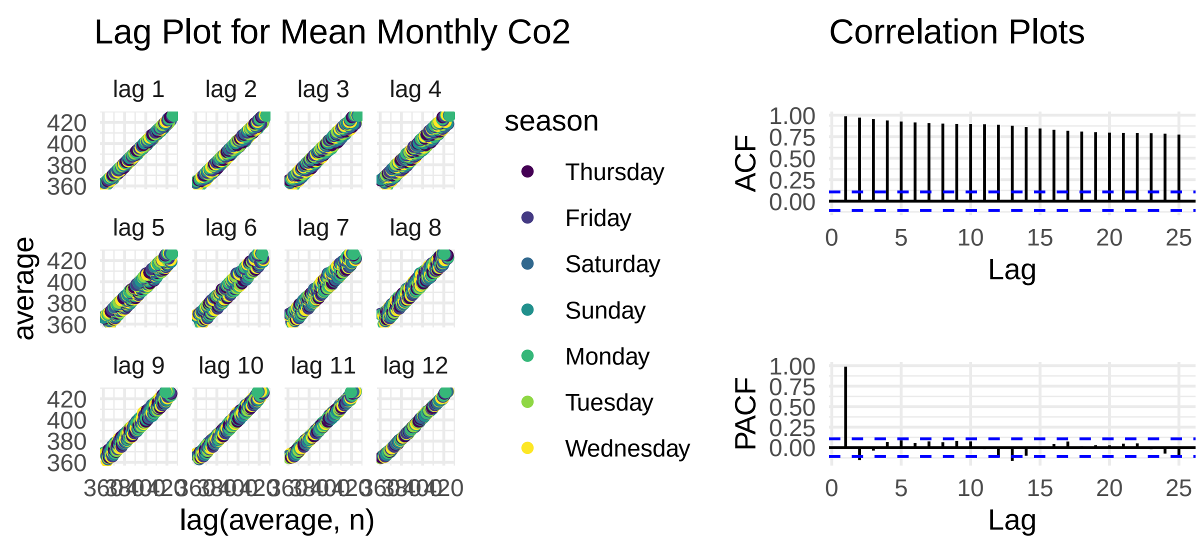
A. Trends in Atmospheric Carbon

We begin by analyzing the trends and patterns in the updated CO_2 time series.

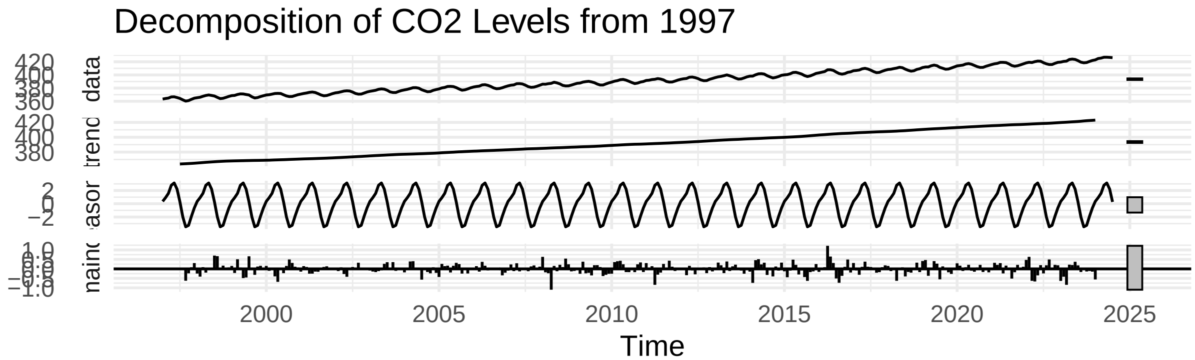
* Siddiqui: UC Berkeley, School of Information, asiddiqui6@ischool.berkeley.edu. Mollerus: , mjmollerus@ischool.berkeley.edu. Jung: , erijung@ischool.berkeley.edu. The authors would like to thank their instructors from MIDS 271.



We can see that the trend of the CO_2 ppm remains fairly similar to the trends we saw preceding 1997. The seasonal trends also look very similar. We now examine the ACF, PACF and lag plots.



The ACF shows a slow decline in value indicating a trend. Additionally, there is some waviness present in the graph suggesting a seasonal component. Looking at the PACF graphs we can see there is some oscillation, suggestive of a MA process. However, taking into account the earlier plot of the data, it is also clear that the time series is non-stationary, thus we believe a ARIMA model would be most appropriate.



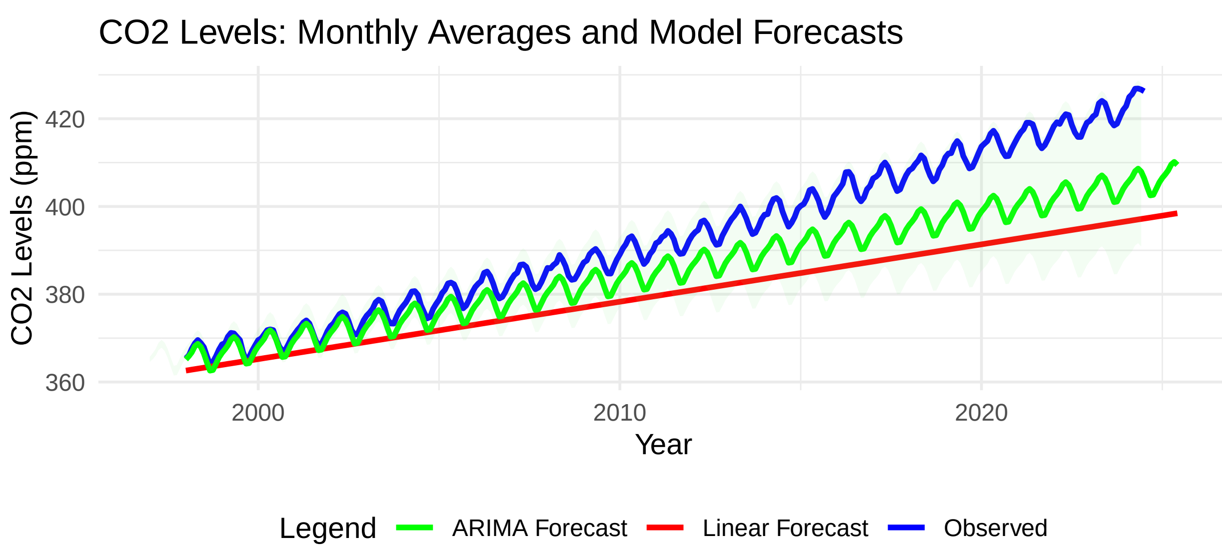
We see in the decomposition that there is a clear trend and seasonal components. Visually, these look similar to what we would expect from the time series plot.

III. Modeling

A. Previous Model Forecasts

We now use the models we built in 1997 to forecast up to the present day.

model	RMSE
Linear Model	14.98
ARIMA	9.25



The model is actually under predicting the co2 levels compared to the true data. From the residuals, upon visual inspection, we can also tell that the predictions seem to fall increasingly short as time increases. The ARIMA prediction looks to have learned the seasonality a lot better. It still under predicted the true co2 levels but was closer than the linear predictions. We conduct a Ljung-Box test to confirm that the ARIMA model's residuals are no longer independently distributed.

```
##
##  Box-Ljung test
##
## data:  arima_resid
## X-squared = 1503, df = 0, p-value <2e-16
```

We reject the null hypothesis of the Ljung-Box test, confirming that our model is no longer fitting well given the updated data. We now recall the forecasts that our model made previously.

The observed co2 crossed 420 ppm between 2021 - 2022, whereas both the ARIMA and linear models' forecasts were a lot more delayed. The ARIMA model predicted 2023 at the earliest and the linear model forecasted around 2050. The observed crossing happened extremely close to the upper bound of the 95 percent confidence interval of our old ARIMA model. This implies that our current course of CO_2 emissions is projected to land us near the upper bound of our earlier predictions, signaling that we are approaching our worst case prediction from our old model.

B. *Updating Models*

Seeing that our models are now significantly under predicting atmospheric CO_2 levels, we aim to update the models using the updated data. We split our data into a train set consisting of CO_2 readings up until 2 years ago, and a test set containing the CO_2 readings from the past 2 years. Using this training set, we construct three models:

- A SARIMA model on non-seasonally adjusted training data
- An ARIMA model on seasonally adjusted training data
- A quadratic trend model on seasonally adjusted training data

We construct the SARIMA on the not seasonally adjusted data using the following declaration:

```
arima_model_nsa <- auto.arima(train_nsa_ts, seasonal = TRUE,
  max.p = 5, max.q = 5, max.d = 2,
  max.P = 5, max.Q = 2, max.D = 5,
  max.order = 5, stepwise=TRUE)
```

We construct the ARIMA on the seasonally adjusted data using the following declaration:

```
arima_model_sa <- auto.arima(train_sa_ts, seasonal = FALSE,
  max.p = 5, max.q = 5, max.d = 2,
  max.P = 10, max.Q = 10, max.D = 10,
  max.order = 5, stepwise=TRUE)
```

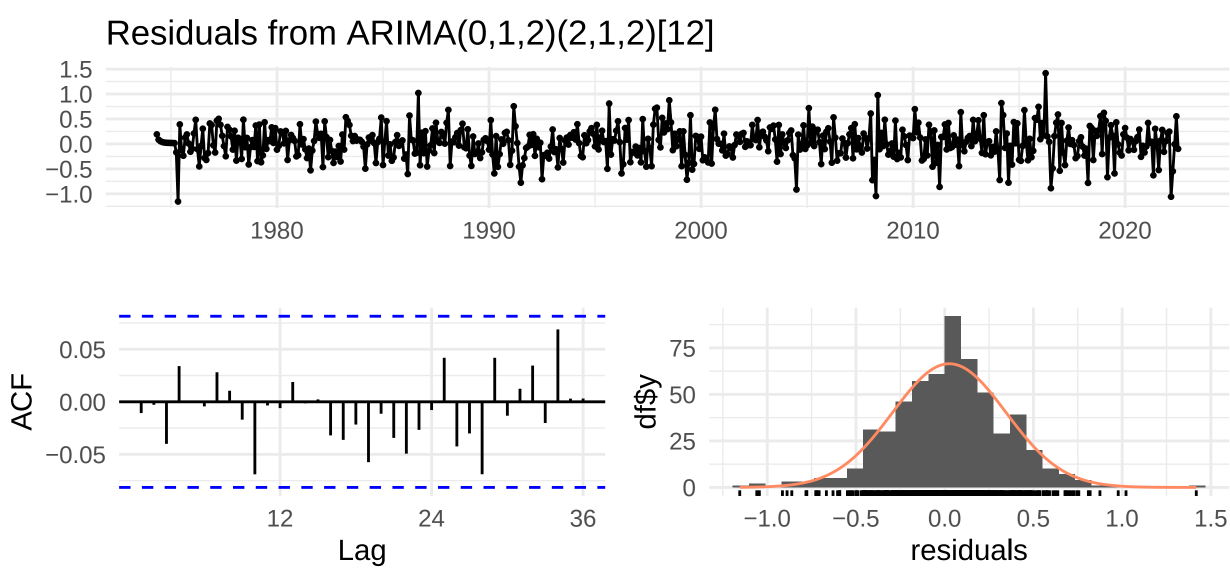
We construct the quadratic trend model on the seasonally adjusted data using the following declaration:

```
poly_trend_model <- train_sa %>%
  model(
    quadratic_trend = TSLM(sa_value ~ trend() + I(trend()^2))
  )
```

C. Updated Model Evaluation

We now use the updated models to forecast for our test set from the last two years and calculate metrics, starting with in-sample metrics.

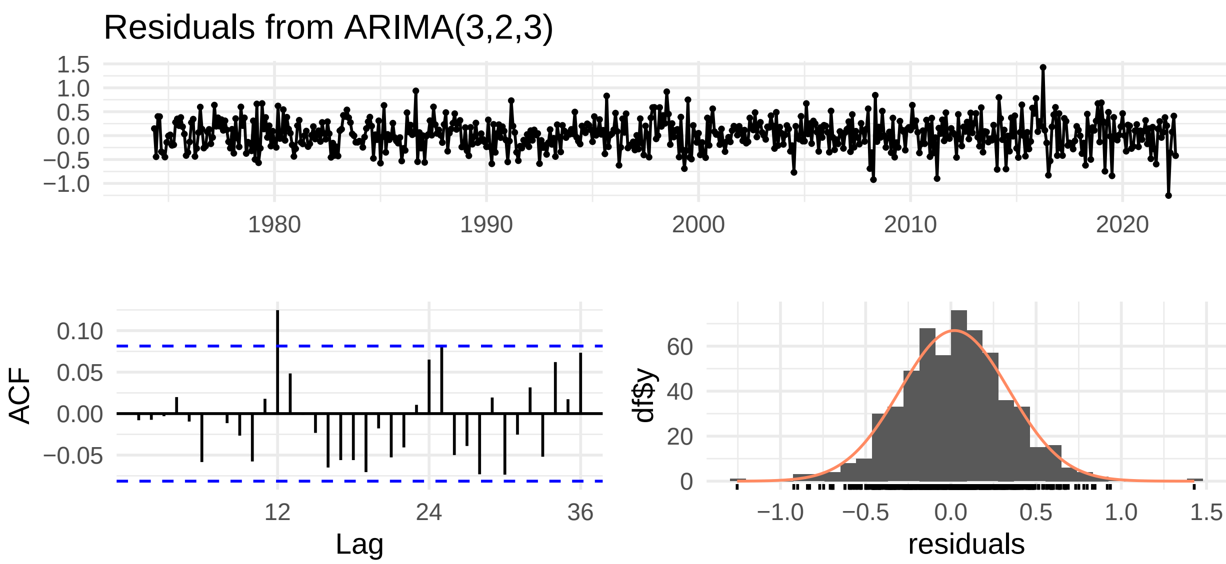
We check the in-sample residuals of the ARIMA models




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(0,1,2)(2,1,2)[12]
```

```
## Q* = 12, df = 18, p-value = 0.9
##
## Model df: 6.   Total lags used: 24
```

We fail to reject the null hypothesis of the Ljung Box test for the SARIMA model. It appears to be well-fit.



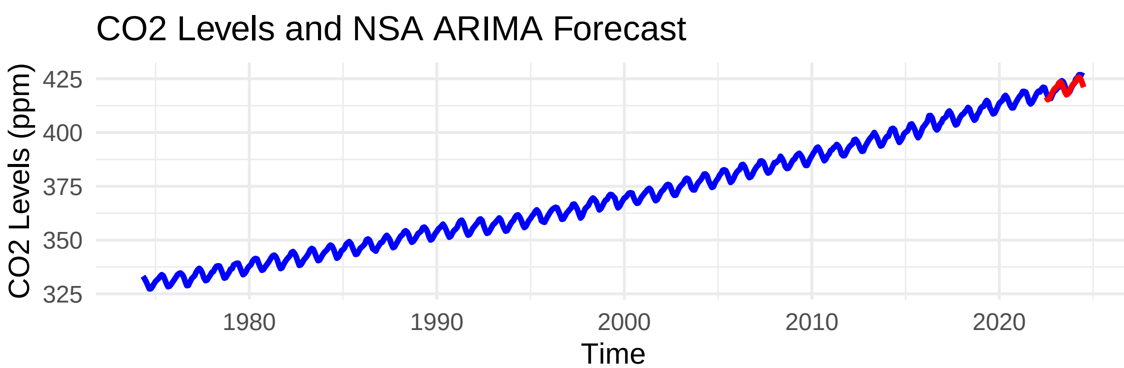

```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(3,2,3)
## Q* = 31, df = 18, p-value = 0.03
##
## Model df: 6.   Total lags used: 24
```

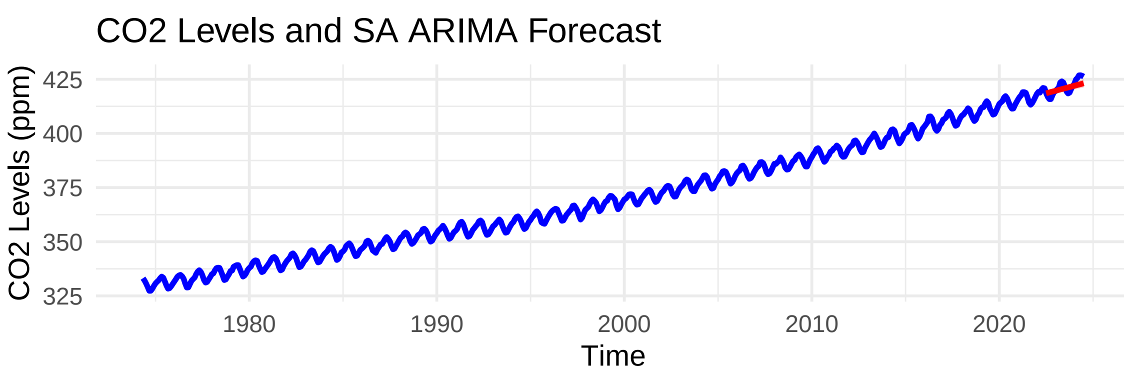
We see that there are is a significant lag in the ARIMA model fit to seasonally adjusted data. We also fail to reject the null hypothesis of the Ljung-Box test. Perhaps a slight adjustment could be done to tune this model. We now use the constructed models to measure RMSE on the test set with the appropriate adjustments made for each model's assumptions.

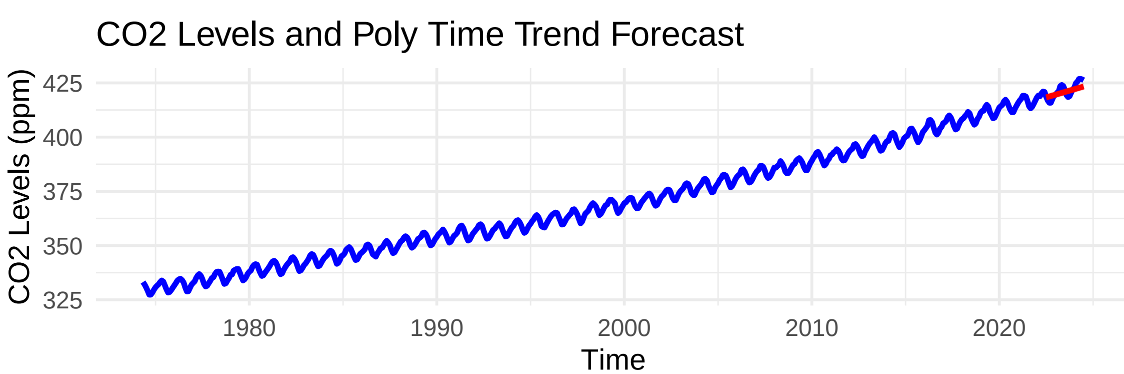
model	RMSE
ARIMA (NSA)	1.837
ARIMA (SA)	0.976
Polynomial Trend (SA)	0.916

From the above results we can see that our quadratic model fitted on seasonally adjusted data had the lowest pseudo out-of-sample RMSE, while the ARIMA model on seasonally adjusted was close behind. As a result, we know that removing the seasonal component of the data helped.

We now forecast on the non seasonally adjusted test set.







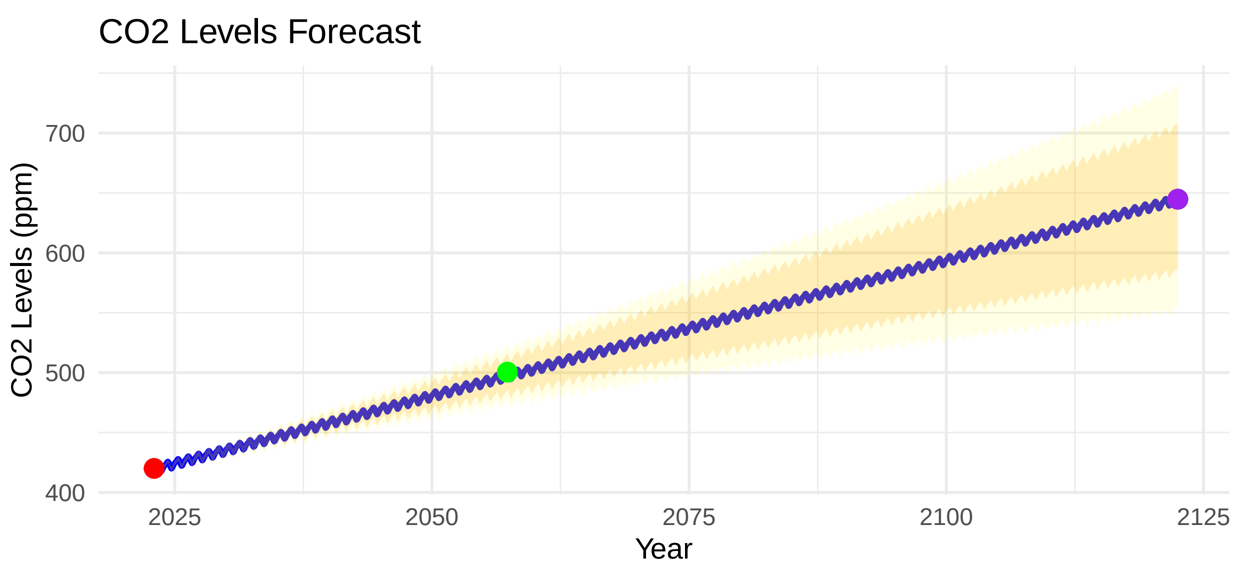
We see that when we return using the nonseasonally adjusted data to forecast, the quadratic and seasonally adjusted ARIMA fail to capture the seasonal variation that the SARIMA model does. This is to be expected given the construction of the models. We now recalculate the RMSEs on the non-seasonally adjusted data for all models.

model	RMSE
ARIMA (NSA)	1.84
ARIMA (SA)	2.50
Polynomial Trend (SA)	2.43

We now see that the SARIMA outperforms the seasonally adjusted model for the out-of-sample test set. We will use the SARIMA model for future forecasts as we wish to capture the seasonal patterns in our projection.

IV. Forecasting

We now use the SARIMA trained on data through July 2022 to forecast until 2122.



Using our ARIMA model on non-seasonally adjusted data we forecast that the first time we will cross 420 co2 ppm will be in 2023 (indicated in red). We anticipate crossing 500 co2 ppm in 2057 (indicated in green) and by 2122 we predict that the co2 ppm will be at 644.77 (indicated in purple).

Table 4—Date Forecasts

	Date	Forecast	Low 80	High 80	Low 95	High 95
First 420	$Jan2023$	420.06	419.419	420.7	419.08	421.04
First 500	$May2057$	500.431	485.401	515.461	477.444	523.418
Final 420	$Jul2122$	644.769	583.985	705.554	551.808	737.731
Final 500	$Jul2122$	644.769	583.985	705.554	551.808	737.731
Forecast 2122	$Jul2122$	644.769	583.985	705.554	551.808	737.731

V. Conclusion

We have reexamined the 1997 report using updated CO_2 data from the same collection location. The models in the past severely under predict the growing trends of atmospheric CO_2 levels with overestimating the time to key thresholds like 420 ppm by over 10 years. We have revamped these models with the updated data. Finally we have selected a SARIMA model to use for forecasting until 2122. The results are quite grim. Our SARIMA model predicts reaching 500 CO_2 ppm in 2057 and by 2122 predicts reaching over 640 ppm.We are not very confident in our predictions as the changes in co2 ppm will be heavily dependent on things our models cannot account for such as public policies, technological advancements, and unforeseen events. These will inevitably lead to potentially severe data and model drift. Given the acceleration of increasing atmospheric CO_2 levels, unless substantial effort is made to combat growing emissions, we worry our model's predictions could still be under predicting, similar to the models constructed in 1997. We urge for more research to be done in this field to help combat growing CO_2 emissions.

VI. References

[1] Stott, P., Stone, D. & Allen, M. Human contribution to the European heatwave of 2003. Nature 432, 610–614 (2004). https://doi.org/10.1038/nature03089

[2] Heuer, Rachael M., and Martin Grosell. "Physiological impacts of elevated carbon dioxide and ocean acidification on fish." American Journal of Physiology-Regulatory, Integrative and Comparative Physiology 307, no. 9 (2014): R1061-R1084.