

Global CO_2 Emissions 1995 and Present

What Keeling missed all these years

Yuna Kim, Jinsoo Chung, Priya Reddy, Jocelyn Thai

Contents

1	Background	2
2	Models and Forecasts	4
2.1	Linear Models	5
2.2	Model Forecast	6
2.3	ARIMA Models	7
2.4	Forecasting Atmospheric CO_2 Growth	8
3	Report from the Point of View of the Present	9
3.1	Introduction	9
3.2	Compare linear and Arima model forecasts against realized CO_2	11
3.3	(3 points) Task 4b: Evaluate the performance of 1997 linear and ARIMA models	11
3.4	(4 points) Task 5b: Train best models on present data	12
3.5	(3 points) Task Part 6b: How bad could it get?	14
4	Conclusions	14

Understanding a changing climate, and what it means for the earth's inhabitants is of growing interest to the scientific and policy community. Although, at this point in 1997 it is not entirely clear what the consequences of this growing awareness will be, in this report we present likely outcomes under "business-as-usual" scenarios. In doing so, our hope, is to establish a series of possible futures, and, as evidence, technology, and policy develop over the coming decades, our goal is that we can then weigh the impacts that carbon-emission reduction efforts might take. In this report we seek to understand if atmospheric CO_2 levels have increased since 1997 and if we anticipate that this trend will continue. To do this we seek to answer 2 main questions: > What is the best model we can use to model CO_2 levels over time? > Using this model, what do we forecast atmospheric CO_2 levels will be? And how confident are we in these predictions?

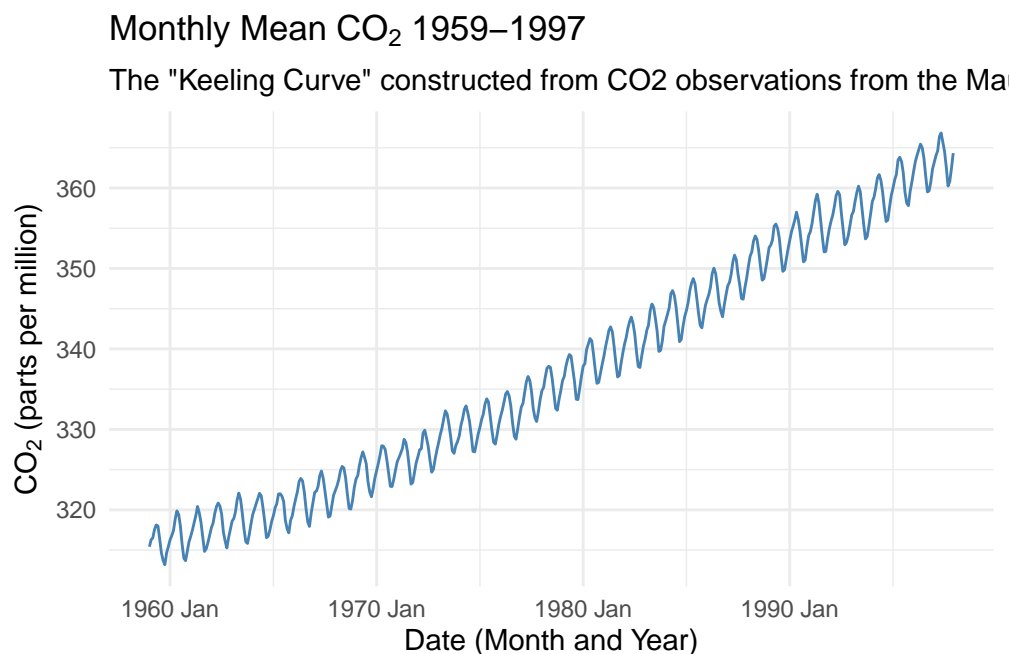
1 Background

1.0.1 Carbon Emissions

High CO₂ levels have been shown to negatively play into climate change, and the increase in atmospheric CO₂ has cascading effects from global warming to ocean acidification. Due to these effects, we may notice food supply chain disruptions, natural disasters of increasing intensity, and habitat disruption of all ecosystems and the animal species living there. In an economical sense, understanding these patterns is important so we can prepare for any shifts and uncertainties in global supply chains that might arise for these new climate patterns. In total understanding the patterns in atmospheric CO₂ concentrations is important both so that we can mitigate the existing effects of this increase and prevent continuing upwards trends in CO₂. In addition if our report findings are significant, the findings could persuade other governmental organizations to increase their CO₂ monitoring efforts and global preparedness efforts.

1.0.2 Historical Trends in Atmospheric Carbon

In 1958 Charles Keeling began continuous monitoring of atmospheric carbon dioxide concentrations from the Mauna Loa Observatory in Hawaii and soon observed a trend increase carbon dioxide levels in addition to the seasonal cycle. He was able to attribute this trend increase to growth in global rates of fossil fuel combustion. This trend has continued to the present, and is known as the “Keeling Curve,” and is plotted below. At first glance we can see in this curve both an overall upwards trend in CO₂ levels as the mean seems to be increasing over time. We can also see some sort of cyclic or seasonal pattern of atmospheric CO₂ fluctuations each year.



Looking at the data there are three areas of interest that we may want to look into further – the upwards trend, the seasonality, and the irregularities we observe in the data.

To begin with we can look at the trend in our data. To better visualize the trend we can plot the yearly average CO₂ level (thereby eliminating the seasonality we observed in our initial plot) over time.

We can see in Figure 1, that when we aggregate by year, there is a clearly year-over-year upwards trend in the CO₂ levels in the data. The lack of seasonality in this graph also indicates that our previous assumption that the seasonality was yearly is correct.

This seasonality can be better visualized by plotting CO₂ levels for each year that we have observations in our dataset.

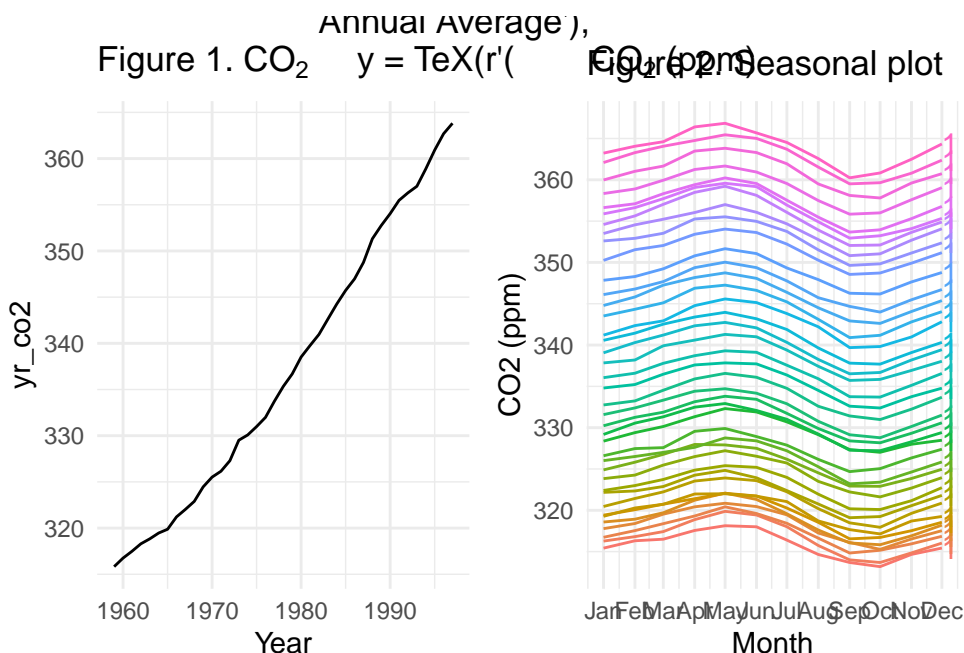


Figure 2 shows a clear seasonal pattern in our data. It appears as though CO₂ levels peak at around May and then dip to a minimum in September and October. To further explore our initial observations on seasonality we can look at our ACF, PACF, and Lag plots.

Fig 3A. ACF of Monthly CO2

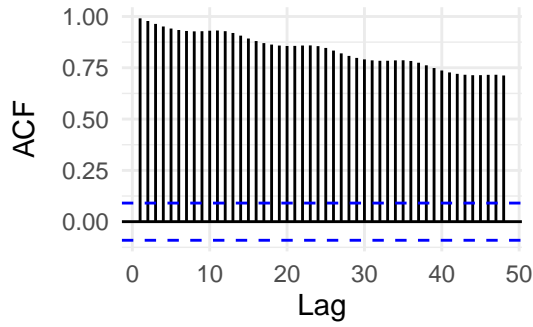


Fig 3B. PACF of Monthly CO2

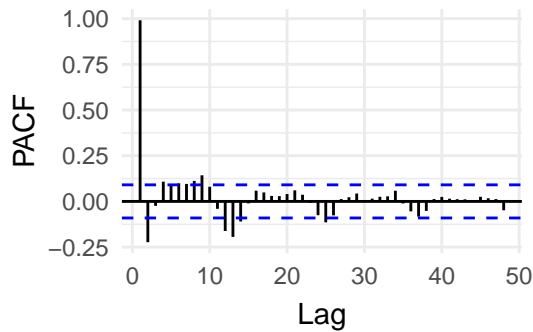
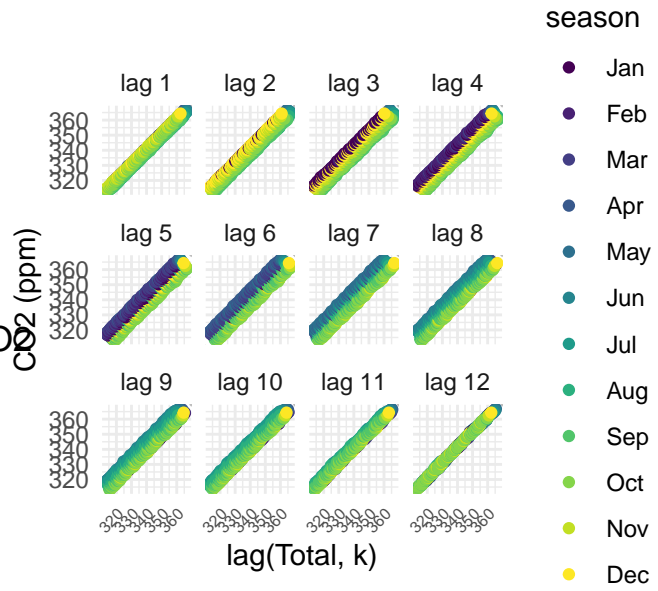


Fig 3C. CO2 Lagged scatterplots



In Figure 3C, the lag plot, there is a very strong positive correlation across all lags, reflecting the seasonality of our data. This is also reflected in the ACF plot, Figure 3A, which decays gradually indicating a trend in our data, and also has a subtle rise in correlation at around every 12th lag. This behavior in the ACF plot indicates a seasonal or cyclic pattern in our data. The Figure 3B PACF cuts off to zero after lag 2 and stays that way. However, at lags 12 and 13, the values become significant again. This suggests that although the model exhibits AR model-like behavior, there's some deviances in the data.

In looking at the patterns in our data we are also interested in assessing whether the upwards trend we observed is accelerating – that is we want to look at the trend of the growth rate of CO2 levels. At a high level, knowing this could help researchers assess the efficacy of global efforts to curb CO2 emissions.

2 Models and Forecasts

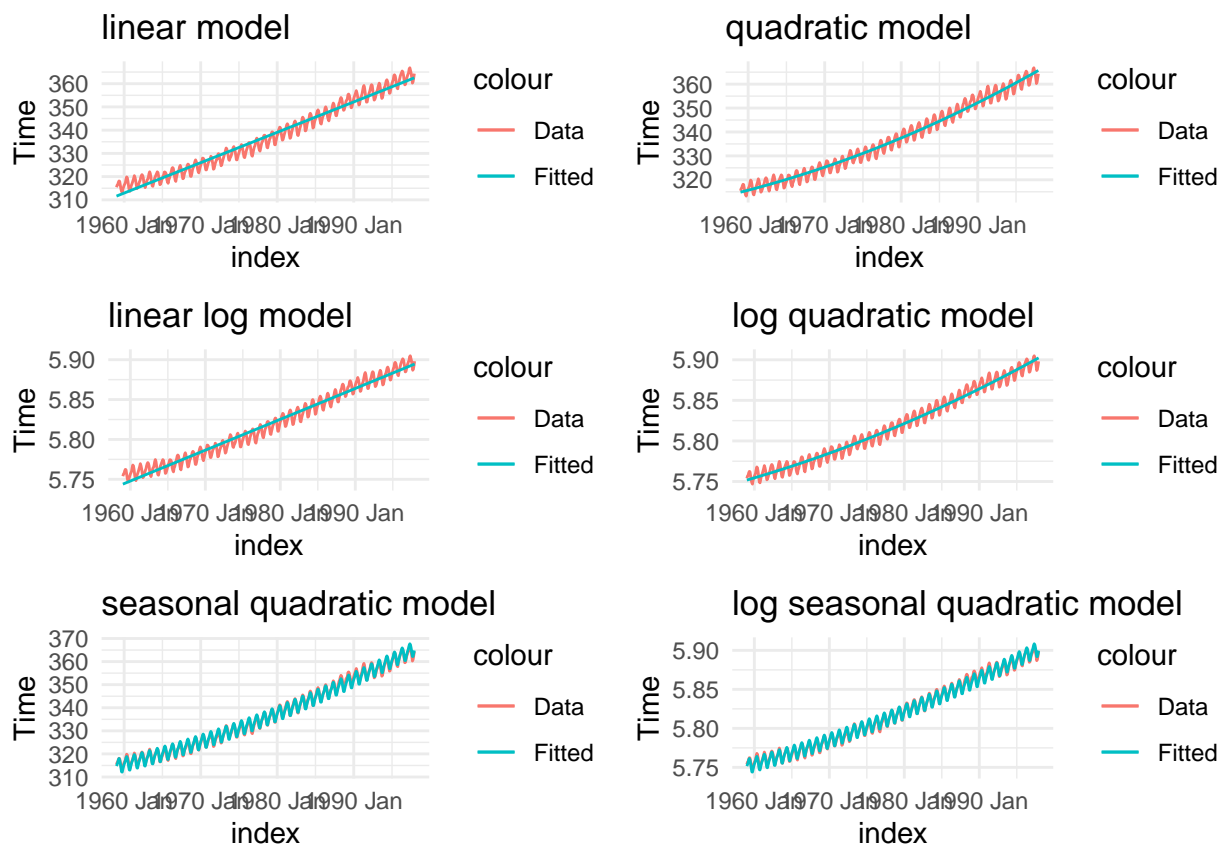
In this section, we evaluate two classes of models for answering our questions – a linear time model and an ARIMA model to assess which time series model is most appropriate to use to model CO2 levels over time.

2.1 Linear Models

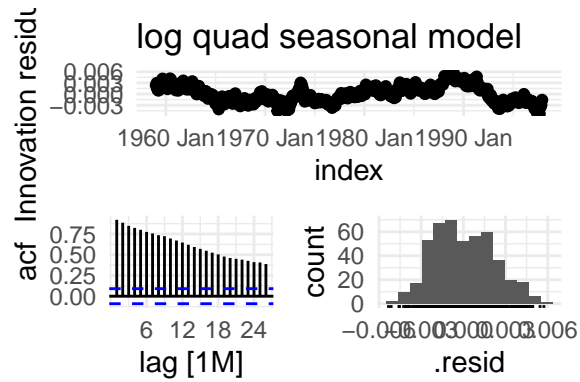
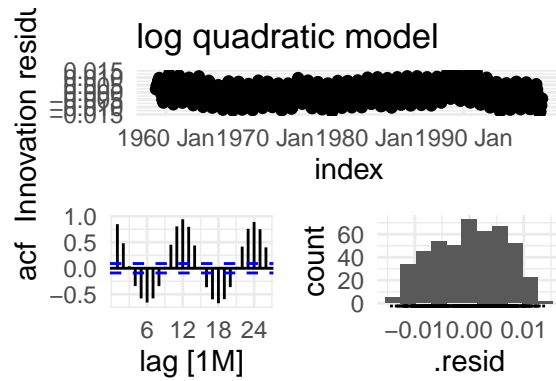
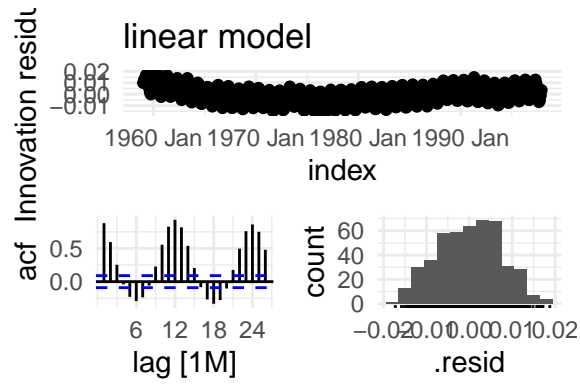
To fit linear time trend model to the `co2` series we will compare a regular time trend linear model to a quadratic time trend model. We will also fit a polynomial time trend model that incorporates seasonal dummy variables, and use this model to generate forecasts to the year 2020. Note that we will be evaluating using both the unscaled CO2 and the $\log(\text{CO}_2)$ value for constructing our models.

To begin our analysis, we fit a linear model of the form to our data:

$$\text{CO}_2 = \phi_0 + \phi_1 + \epsilon_{eit} \quad (1)$$

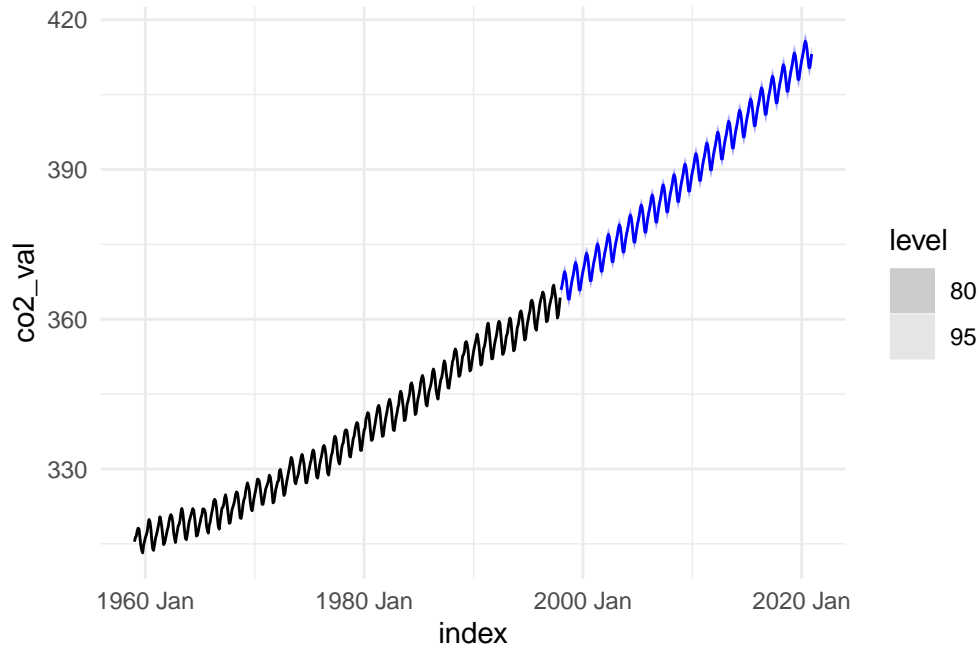


We can see that the log models appear to fit the data slightly better (some overshoot in the linear model around 1980s, whereas less in log linear model). Same with the quadratic models, appears like it's slightly overestimating in the quadratic model and is overestimating less (ie. in the middle of the seasonal variation) in the log quad model. We can see this in the seasonal model too. Moving forward with working on the $\log(\text{co}_2)$ data with this data transformation we have a more stationarity of the residuals as well.



2.2 Model Forecast

We chose the quadratic seasonal model to make prediction forecast of CO₂ up to year 2020.



2.3 ARIMA Models

We will also generate a few ARIMA models and compare them to our linear model. To select our ARIMA model we use the AIC, the AIC selects against models with too many parameters. We know that there is a seasonal component to our data, so we will also select for seasonal AR and MA values and check if seasonal differencing will make the data stationary. I am choosing to use the `ARIMA()` function to select our data. This method will determine the optimal p, d, q, P, D, Q values as well as the optimal levels of differencing for us to have stationary data and will then select a model based on the metric we have chosen as our criteria (AIC). We can interpret the resulting ARIMA model in terms of the number of (seasonal) AR terms, the level of differencing to achieve stationary data. We are including seasonality in our model because we can see in the decomposition as well as the time series plots that there is a strong possibility of a seasonal component in the CO2 levels. If adding a seasonal component improves the AIC score then such a model will be chosen.

```
## <lst_mdl[1]>
## [1] <ARIMA(1,0,1)(4,1,0)[12] w/ drift>

## # A tibble: 7 x 6
##   .model term      estimate std.error statistic  p.value
##   <chr>  <chr>      <dbl>    <dbl>    <dbl>    <dbl>
## 1 ts.model ar1         0.982    0.00984    99.8 5.55e-312
## 2 ts.model ma1        -0.375    0.0495    -7.57 2.05e-13
## 3 ts.model sar1       -0.714    0.0488   -14.6 4.69e-40
## 4 ts.model sar2       -0.553    0.0581    -9.52 1.04e-19
## 5 ts.model sar3       -0.333    0.0580    -5.75 1.66e-8
## 6 ts.model sar4       -0.150    0.0497    -3.03 2.60e-3
## 7 ts.model constant    0.0630    0.00809     7.79 4.41e-14
```

With the AIC as our selection criteria we have estimated the model to be an ARIMA(1,0,1)(4,1,0)[12] with drift. This model has an AIC of 218, since this was the model that was chosen, this must be the smallest AIC value in the models that we have compared. This ARIMA(1,0,1)(4,1,0)[12] model can be interpreted by saying that there is 1 AR term, 1 MA term, 4 seasonal AR terms and the data had to be seasonally differenced once to be made stationary. the [12] at the end of the model also indicates

To be thorough we will also repeat the above process but instead using BIC and AICC as our selection metrics. In general the BIC is stricter than the AIC in penalizing additional parameters in our model, so it is possible that this selection process will result in a different model. When we repeated the above process using the bic or the aicc as our criteria, we found that we were still selecting the exact same ARIMA model, so we will move forwards with this model.

2.4 Forecasting Atmospheric CO2 Growth

Now that we have a few different models that do a decent job at modeling our CO2 data, we want to use these models to generate predictions for us.

2.4.1 When will we reach 420 and 500 ppm CO2?

First we want to see when these models forecast that the atmospheric CO2 levels will reach 420ppm and 500ppm. Since we know there is a overall upwards trend as well as a seasonal component to the CO2 levels, when we generate these predictions we will actually want to look at both the first and the last times that the model predicts the atmospheric CO2 levels to be at these values. We know from our previous predictions for the linear model that even by 2020, the predicted CO2 ppm is not 420 ppm. Of course, there may be some variance in the predictions between the ARIMA and linear model but this does serve as a good guideline.

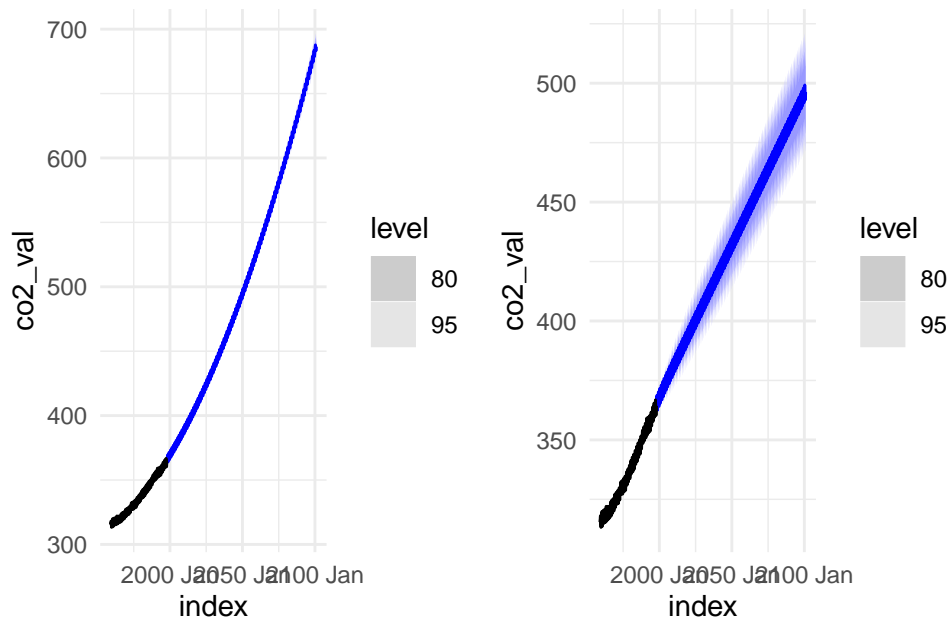
When we do this we can then filter our predictions to find the times when the CO2 levels are predicted to be 420 and 500ppm. We should note that these are approximations as the true predictions from our model are not integers and instead are the mean of a prediction distribution. As a result we had to round these prediction values to find where the predicted CO2 will reach these levels.

Our ARIMA model predicts that the first time that the atmospheric CO2 levels will to be 420ppm is April 2038, and the last time is October 2042. To be more accurate, we found that the predicted CO2 values will reach 420ppm for the first time sometime in between March 2038 and April 2038 and for the last time sometime in between September 2042 and October 2042. We can also see that the first time that the atmospheric CO2 levels are predicted to be 500ppm is March 2101, and the last time is October 2105. Again, to be more accurate, we found that the predicted CO2 values will reach 500ppm for the first time sometime in between March 2101 and April 2101 and the for the last time sometime in between October 2105 and November 2105.

2.4.2 What do we predict CO2 levels will be in 2100 (103 years from now)?

We also want to look at the predictions for our models for the more distant future. Here we will look at what the model predicts will be the CO2 levels in the year 2100. Plot on the left is model prediction of linear quadratic seasonal and on the right is the ARIMA model. The linear quadratic

seasonal predicts for a steeper increase of the CO2 level up to year 2100 compared to the ARIMA model.



3 Report from the Point of View of the Present

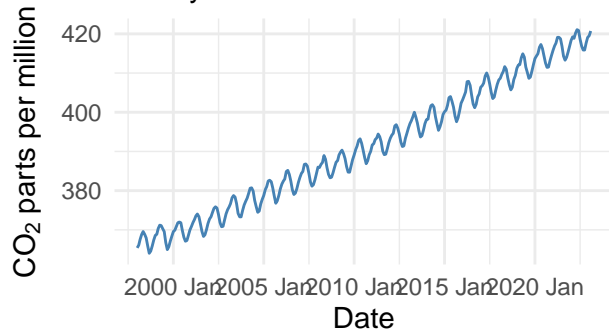
One of the very interesting features of Keeling and colleagues' research is that they were able to evaluate, and re-evaluate the data as new series of measurements were released.

3.1 Introduction

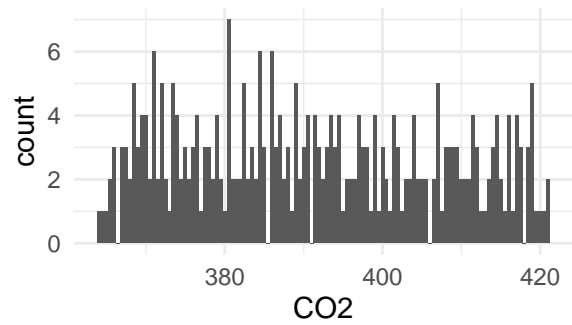
Now there are more dataset that we can collect and up to present on the modern CO2 emission data. We will evaluate our previous data prediction model to the actual measure on newly collect data from United States' National Oceanic and Atmospheric Administration.

```
## # A tsibble: 6 x 2 [1M]
##   time_index average
##   <month>   <dbl>
## 1  2022 Sep    416.
## 2  2022 Oct    416.
## 3  2022 Nov    418.
## 4  2022 Dec    419.
## 5  2023 Jan    420.
## 6  2023 Feb    421.
```

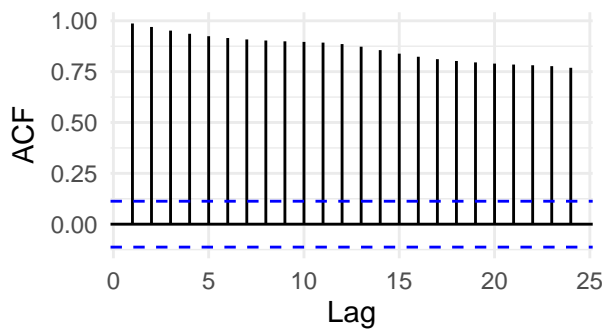
Mean CO₂ from 1997 to Present
weekly data



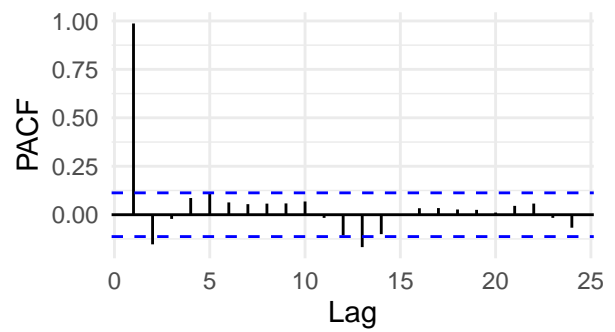
Histogram of CO₂ levels



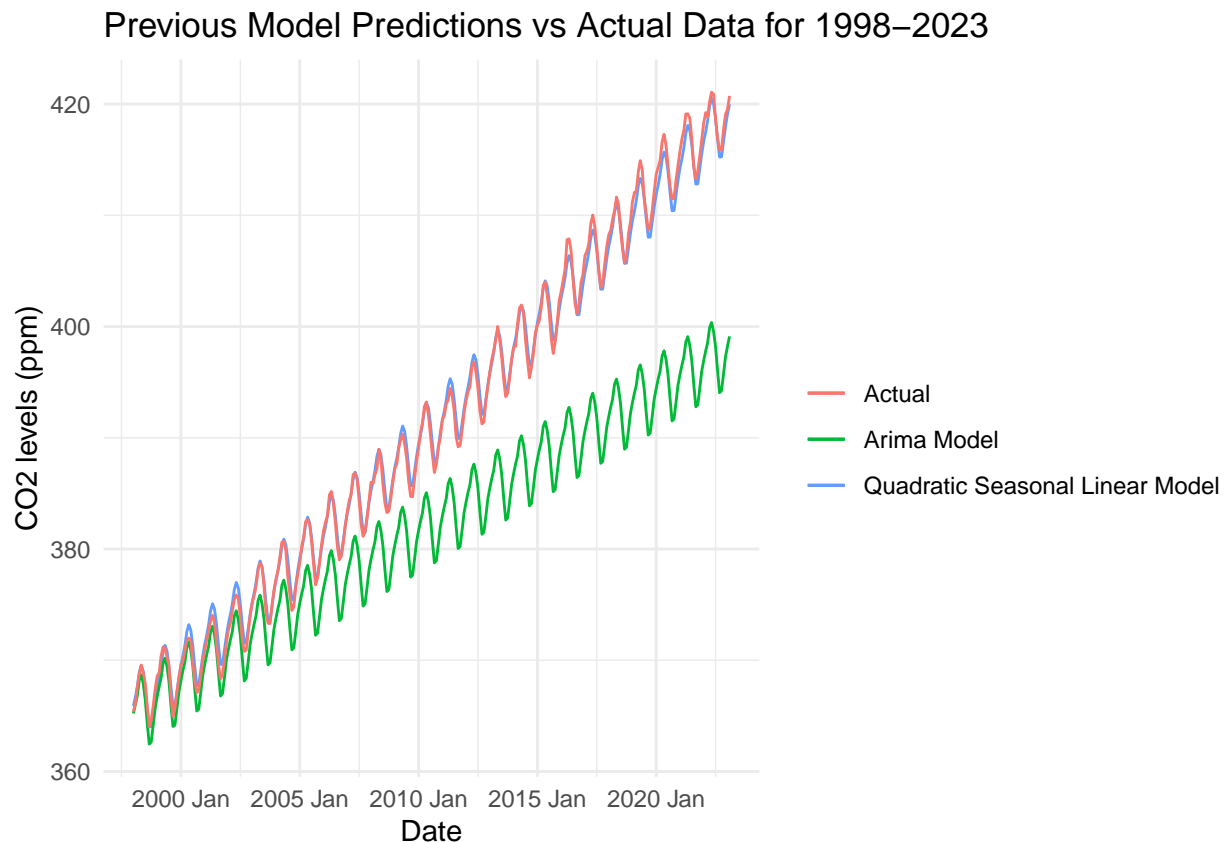
ACF of CO₂ Levels



PACF of CO₂ Levels



3.2 Compare linear and Arima model forecasts against realized CO2



The linear model is a fairly close match to the actual values, it does seem like from 2000-2005 it has a tendency to slightly overestimate the CO2 levels, and then from 2015 to 2020 it appears to slightly underestimate the CO2 levels.

The ARIMA model deviates from the actual values. The model values are similar until 2000, though the model does tend to underestimate the CO2 levels even during this time period. The model really deviates from the actual values at around 2003, as it starts to severely underestimate the CO2 levels. It appears that the Keeling curve has shown some acceleration in the increase of the mean over time, and this might explain the increasing gap between the ARIMA model performance and the true CO2 levels as we move away from 1997.

3.3 (3 points) Task 4b: Evaluate the 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?

In 1997 we predicted that the CO2 levels would cross 420 ppm at around April 2038, but from the actual data we can see that the CO2 levels cross 420 ppm by February 2022, which is an about a 15 year difference. This is in-line with what we observed in our ARIMA model predictions as the ARIMA model was underestimating the CO2 levels in comparison to the actual CO2 levels.

We can also assess the overall performance of our model using a test metric such as the RMSE. The RMSE is the root mean squared error and is a metric that captures on average how far our predictions are from the actual values. It is calculated with the formula $\sqrt{\sum_{i=1}^N \frac{(y_i - \hat{y}_i)^2}{N}}$ where N is the number of observations, y_i is the ith actual CO2 value and \hat{y}_i is the ith CO2 prediction – in essence we are squaring the sum of the differences between the actual and predicted values, and then dividing by the number of observations before taking the square root. When we calculate the RMSE for our two models of interest we can see that the ARIMA model performs worse than the quadratic seasonal model by an order of roughly 10. The Quadratic Seasonal model's RMSE is 0.79235, meaning that on average the Quadratic Seasonal model prediction is off by about 0.79235 ppm. By contrast the ARIMA model RMSE is 8.6259 meaning that on average the ARIMA model is off by about 8.6259 ppm.

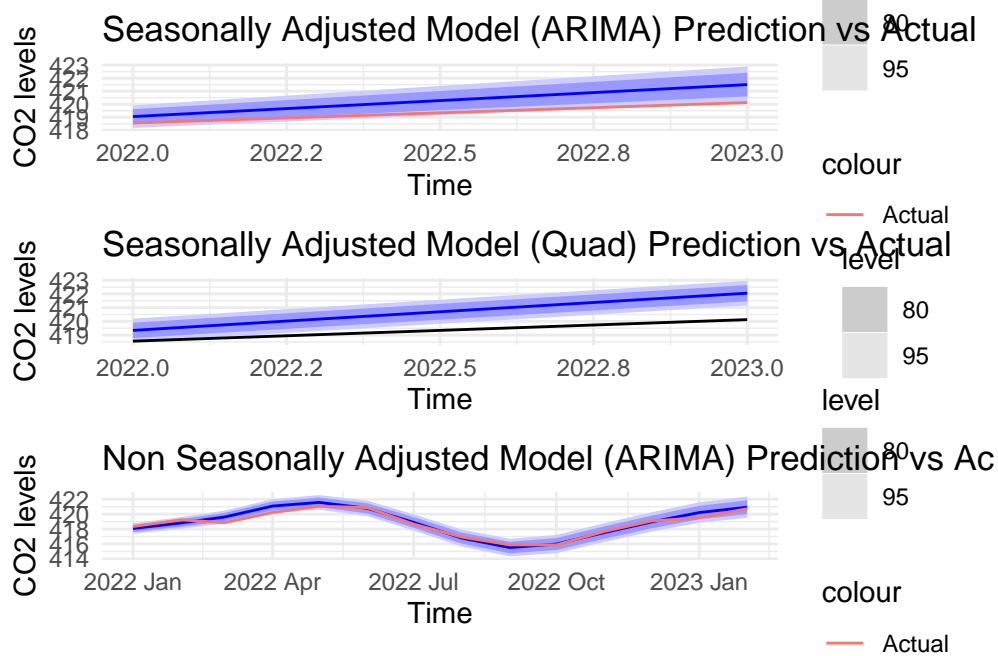
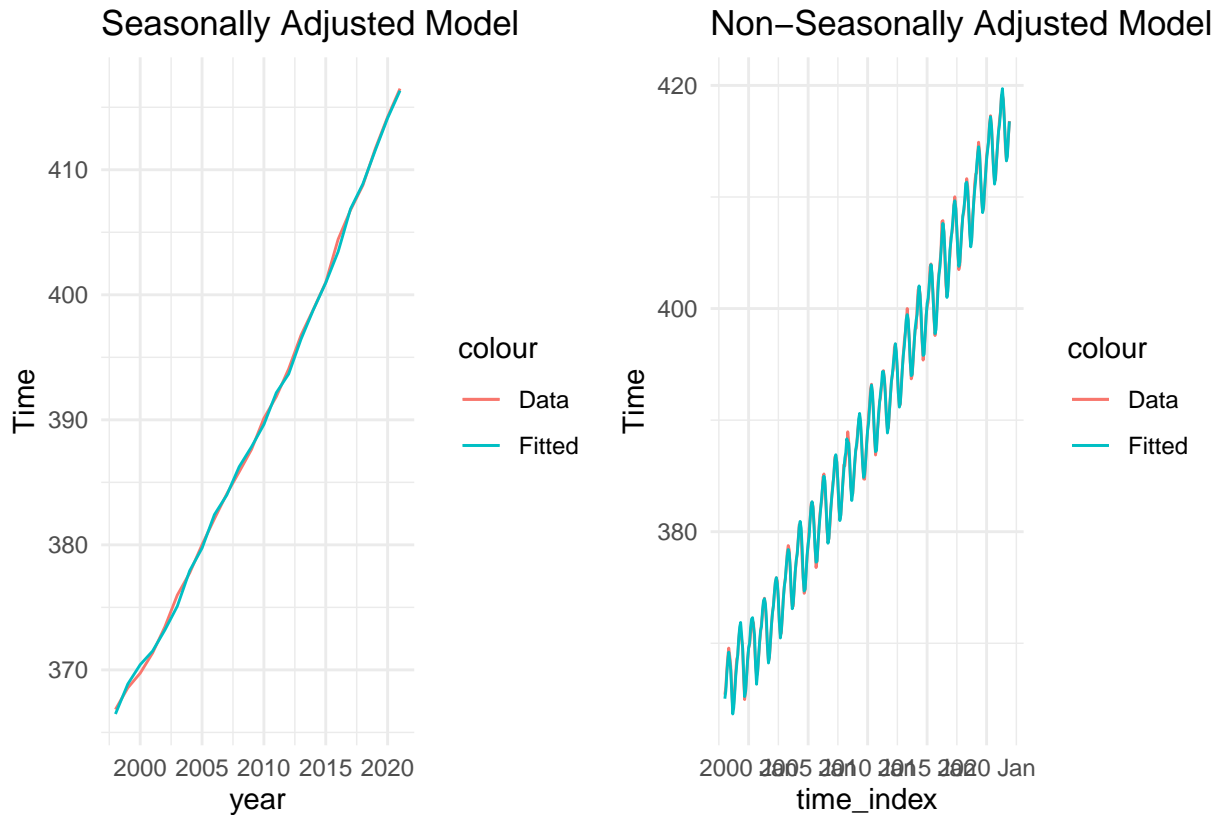
```
## Quadratic Seasonal Model RMSE = 0.79235NULL
```

```
## ARIMA model RMSE = 11.301NULL
```

3.4 (4 points) Task 5b: Train 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.

```
## Warning in sqrt(diag(best$var.coef)): NaNs produced
```

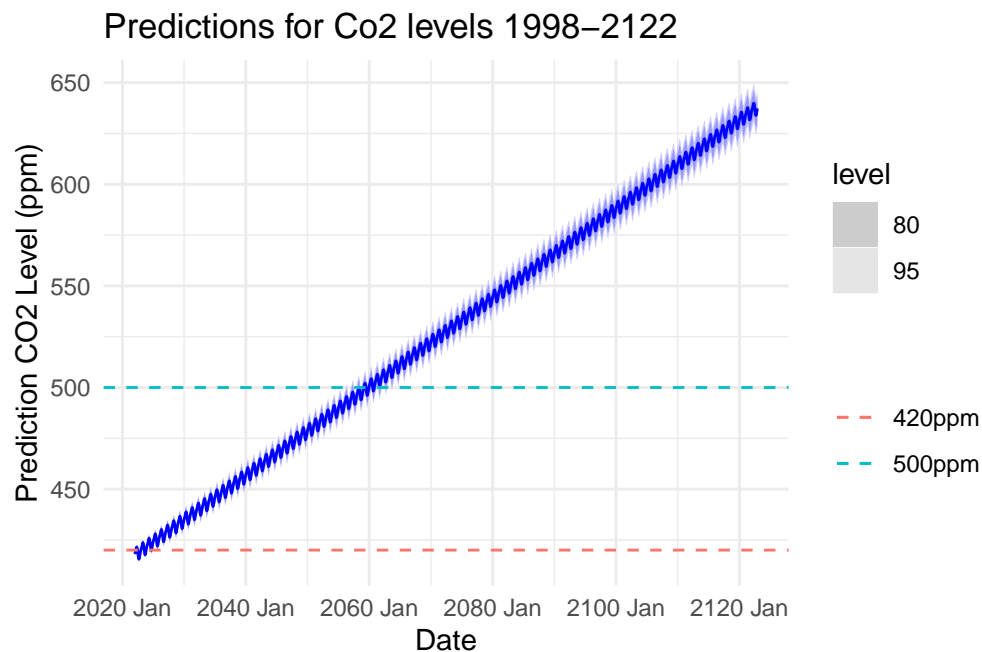


Both seasonally adjusted and non-seasonally adjust models perform well. The data fits perfectly within the training dataset. with forecasts, the seasonally adjust model falls short, deviating from the actual model. The non-seasonally adjusted ARIMA model predicts close to what the actual

values were. The quadratic model performs the worst compared to the ARIMA model when using Seasonally Adjusted models.

3.5 (3 points) Task Part 6b: 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?



```
## [1] 636.94
```

It's predicted that the average co₂ level in 2122 would be 637.42 ppm. There's a possibility that the actual values are worse than predicted judging from the predictions in the Keeling curve. If there's no mitigation in the co₂ levels in the future, the co₂ levels won't likely decline.

4 Conclusions

We have conducted a CO₂ prediction modeling through two sets of data set. Initially, we have generated the prediction from Keeling's data set to predict CO₂ emission. Data transition was explored examining log and seasonal transition. We have also conducted linear, quadratic linear and ARIMA models to compare the prediction behavior of the models. Ultimately evaluating the model with the up to present data of CO₂ emission data we were able to confirm that the non-seasonally adjusted ARIMA model had the model of best prediction.