W271 Group Lab 2

The Keeling Curve

Adam Kreitzman, Hailee Schuele, Lee Perkins, Paul Cooper

Contents

1 Introduction

Beginning in 1958, regular measurements of atmospheric CO2 levels have been taken at the Mauna Loa Observatory in Hawaii, under the guidance of scientist Charles David Keeling. Many interesting conclusions have been drawn about the Co2 data, including noting daily and annual peaks and valleys, coinciding with the respiration cycles of vegetation. Most alarmingly, the readings from Mauna Loa showed a consistent upward trend, starting around 315 parts per million (ppm) in the '50s and rising to around 370 ppm by the current year, 1997.

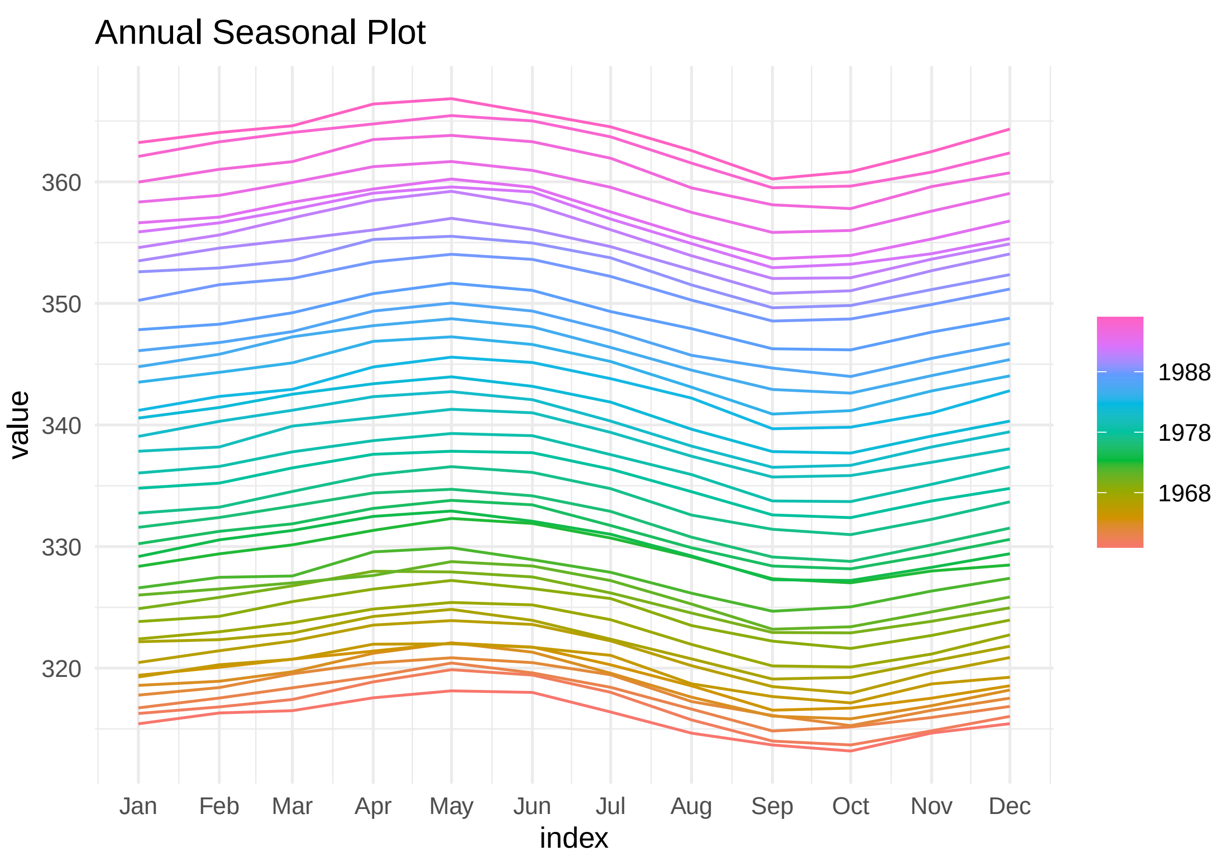
The graph displaying both the cyclical nature and upward trajectory of atmospheric Co2 concentrations has come to be known as the Keeling Curve, and it is one of the hallmark findings in understanding man-made impacts on climate change. In this report, we seek to better understand how successful historical models of the Keeling Curve have been, and generate up-to-date projections of future Co2 concentrations, so that we can better understand any past modeling shortcomings, and better evaluate the impacts of ongoing carbon-emission reduction efforts going forward.

2 Data and Measurement

Data is captured at Mauna Loa for a multitude of reasons. Perhaps most obviously, it has the longest history of CO2 measurement of any currently operating site. Additionally, Mauna Loa has the advantage of being very isolated from most of the world's vegetation, and at a high altitude, meaning measurements mostly eliminate the effect of local plant and human activities. Data from Mauna Loa has been compared to data from many other locations, including Antarctica, Washington DC, and Switzerland, and while locality plays an undeniable role in short term measurements, the long-term patterns and trends are corroborated by other measurements.

Co2 data shows several readily apparent patterns. First, daily levels peak at night and are lowest in the afternoons. More important for our purposes, there is clear seasonal behavior coinciding with the rise and fall of plant life on a macroscopic scale, with atmospheric averages peaking in May and bottoming out in September, as plants work to remove Co2 via photosynthesis.

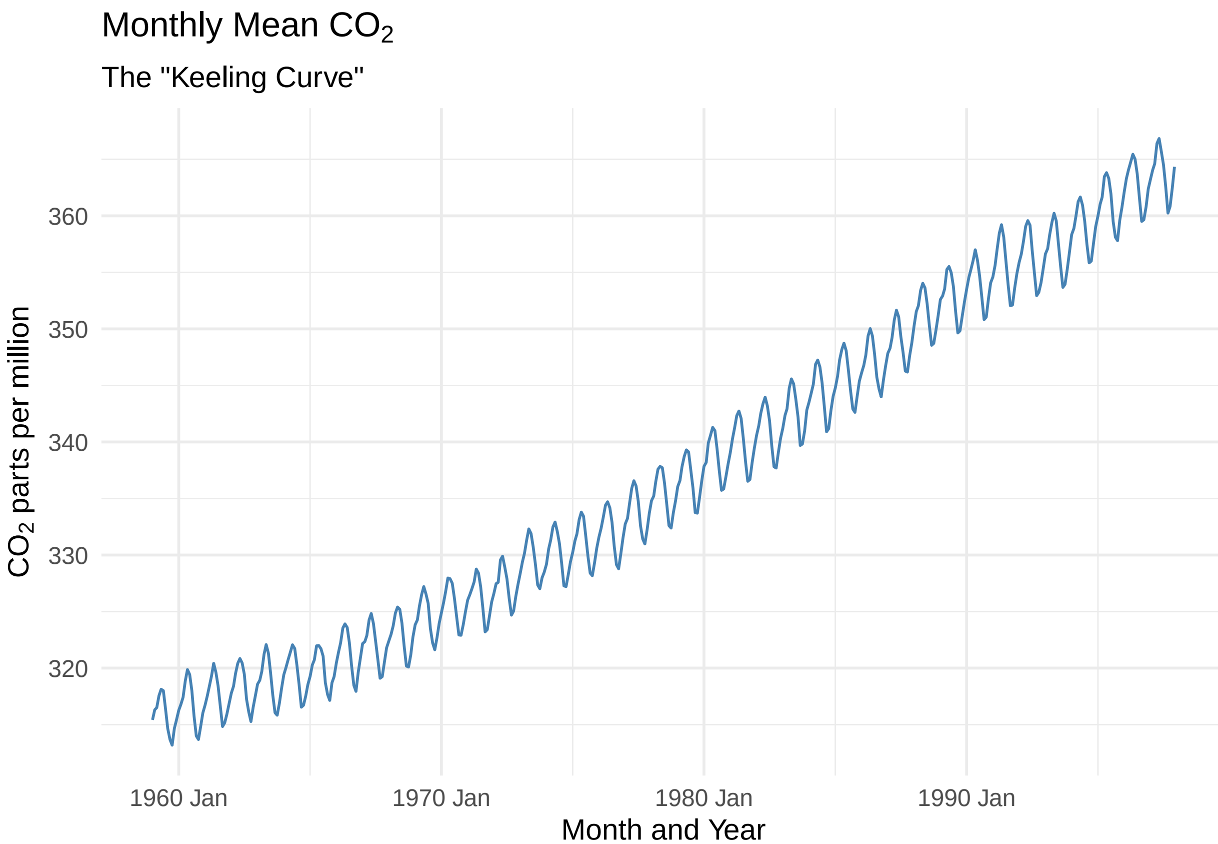
```
co2_ts %>%
  tsibble::fill_gaps() %>%
  gg_season(value, period = "year") +
  labs(title = "Annual Seasonal Plot")
```



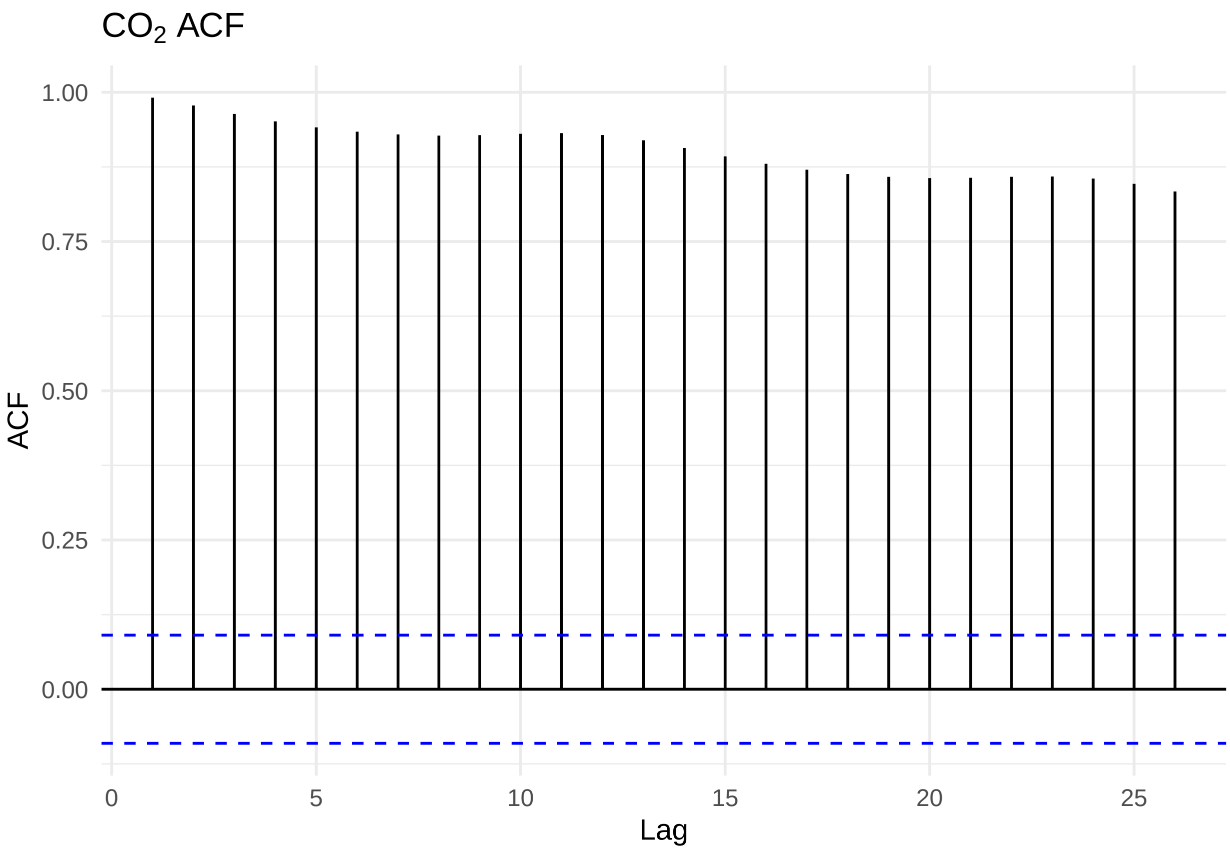
Most notably, though, is the strong increasing trend on the long-term averages, with a rough average increase of 2.5 ppm per year. Additionally, the ACF and PCF plots show some characteristics of an AR(1) model:

```
tsibble::as_tsibble(co2) %>%
    ggplot() +
```

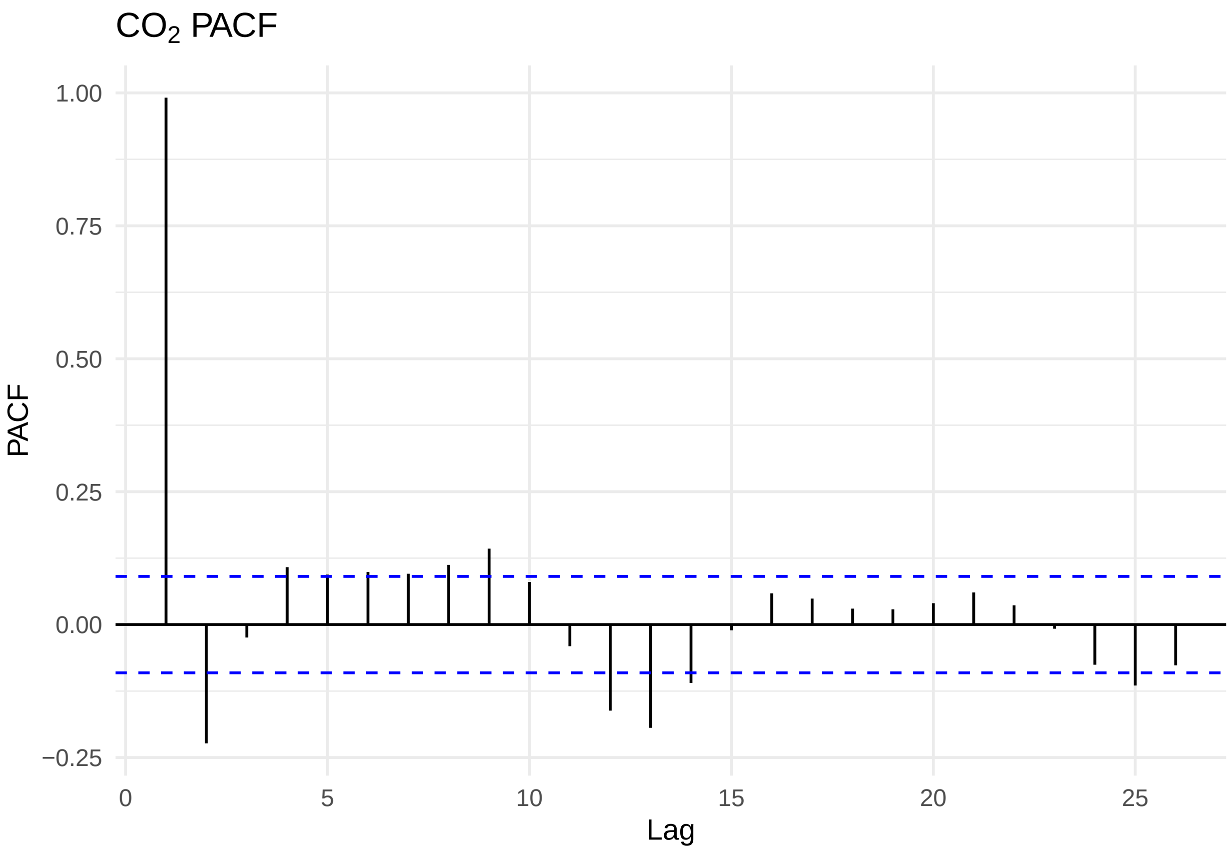
```r
    aes(x=index, y=value) +
    geom_line(color = 'steelblue') +
    labs(
        title = TeX(r'(Monthly Mean $CO_2$)'),
        subtitle = 'The "Keeling Curve"',
        x = 'Month and Year',
        y = TeX(r'($CO_2$ parts per million)')
    )
```




```r
tsibble::as_tsibble(co2) %>%
    pull(value) %>%
    ggAcf() +
        labs(title = TeX(r'($CO_2$ ACF)'))
```



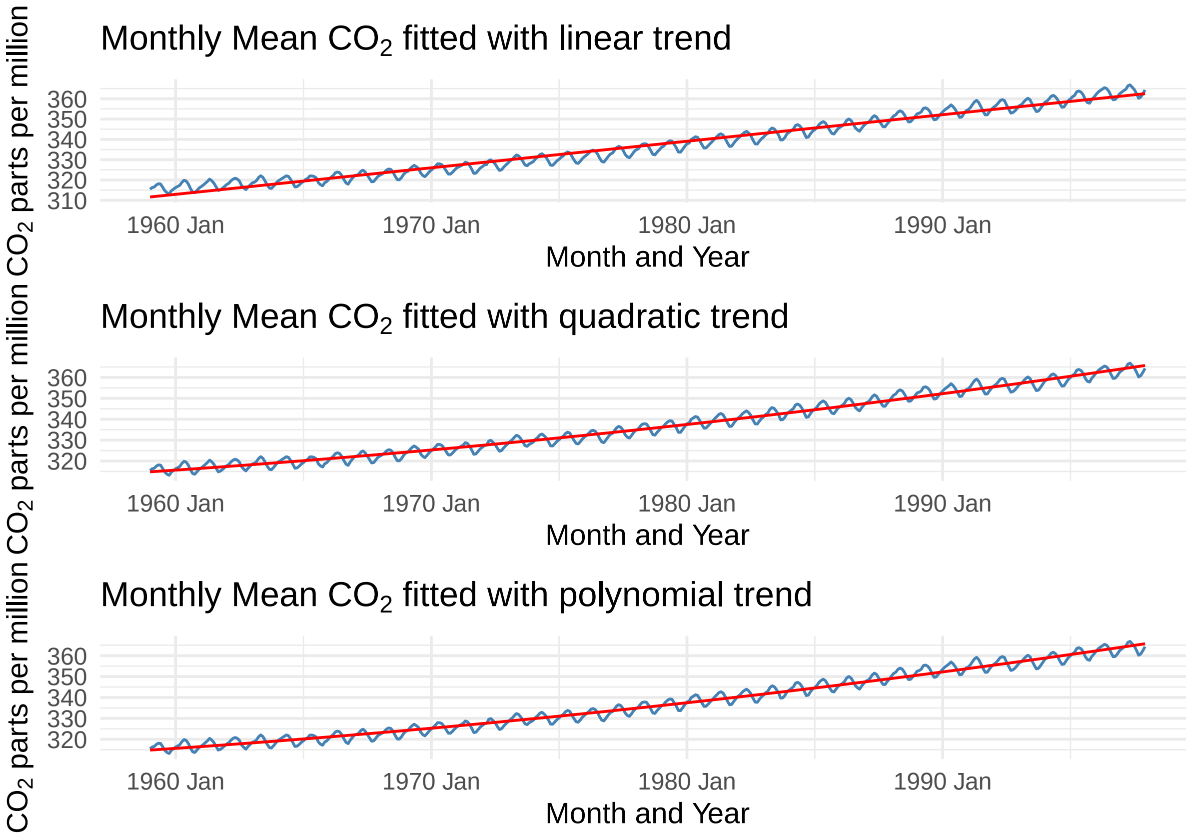

```r
tsibble::as_tsibble(co2) %>%
    pull(value) %>%
    ggPacf() +
        labs(title = TeX(r'($CO_2$ PACF)'))
```



3 1997 Models

With the Mauna Loa data at hand, we attempt a series of modeling techniques, starting with the simplest and building to the most robust. Before beginning modeling, we considered taking a log transform of the data, but that is not appropriate given the data; we do not visually observe accelerating growth in either the trend or the cyclic patterns. Models using the log transform of the Co2 values were not improvements, and will not be used in the remainder of the report.

Fitting linear and quadratic models to the overall dataset does a reasonable job of picking up the long-term trends of the data, but doesn't capture the seasonal aspects appropriately. A stronger model is obtained by using a polynomial model with dummy variables for month included, to capture the annual seasonal effects:



Finally, we fit an ARIMA model to the data, following the Box-Jenkins methodology. The Augmented Dickey-Fuller test failed to reject the null hypothesis, indicating stationarity. We proceeded to select our ARIMA model by minimizing BIC over reasonable parameters, and fit the following model:

```
# <Either nicely formatted text of the ARIMA(1,0,1)(4,1,0)[12], or the output of the model %>%
co2_model <- co2_ts %>% model(ARIMA(value ~ 1 + pdq(0:15,0:2,0:15) + PDQ(0:10,0:2,0:10), ic="b
co2_model %>% report()
```

```
## Series: value
## Model: ARIMA(1,0,1)(4,1,0)[12] w/ drift
##
## Coefficients:
##          ar1      ma1     sar1     sar2     sar3     sar4  constant
##       0.9819  -0.3751  -0.7144  -0.5526  -0.3335  -0.1504    0.0630
## s.e.  0.0098   0.0495   0.0488   0.0581   0.0580   0.0497    0.0081
##
## sigma^2 estimated as 0.09338:  log likelihood=-101.22
## AIC=218.45   AICc=218.77   BIC=251.43
```

This matches our intuition from the EDA; the PCF has one significant lag term, indicating an AR(1) pattern, and the seasonal pattern is well modeled with the 12 month lag.

4 1997 Forecasts

We produced forecast ranges for January 2022 each of the four models derived:

```
# <forecast table for each model, with lower 95, mean, upper 95>
co2_2022_forecast <- co2_model %>% forecast(h="24 years")
```

Taking the ARIMA model, we also provide some estimates for when CO2 will cross several high thresholds. We've provided a lower bound, indicating the year and month when our forecast confidence interval first contains the threshold, an upper bound, indicating the last time our forecast interval contains the threshold, as well as our mean estimate:

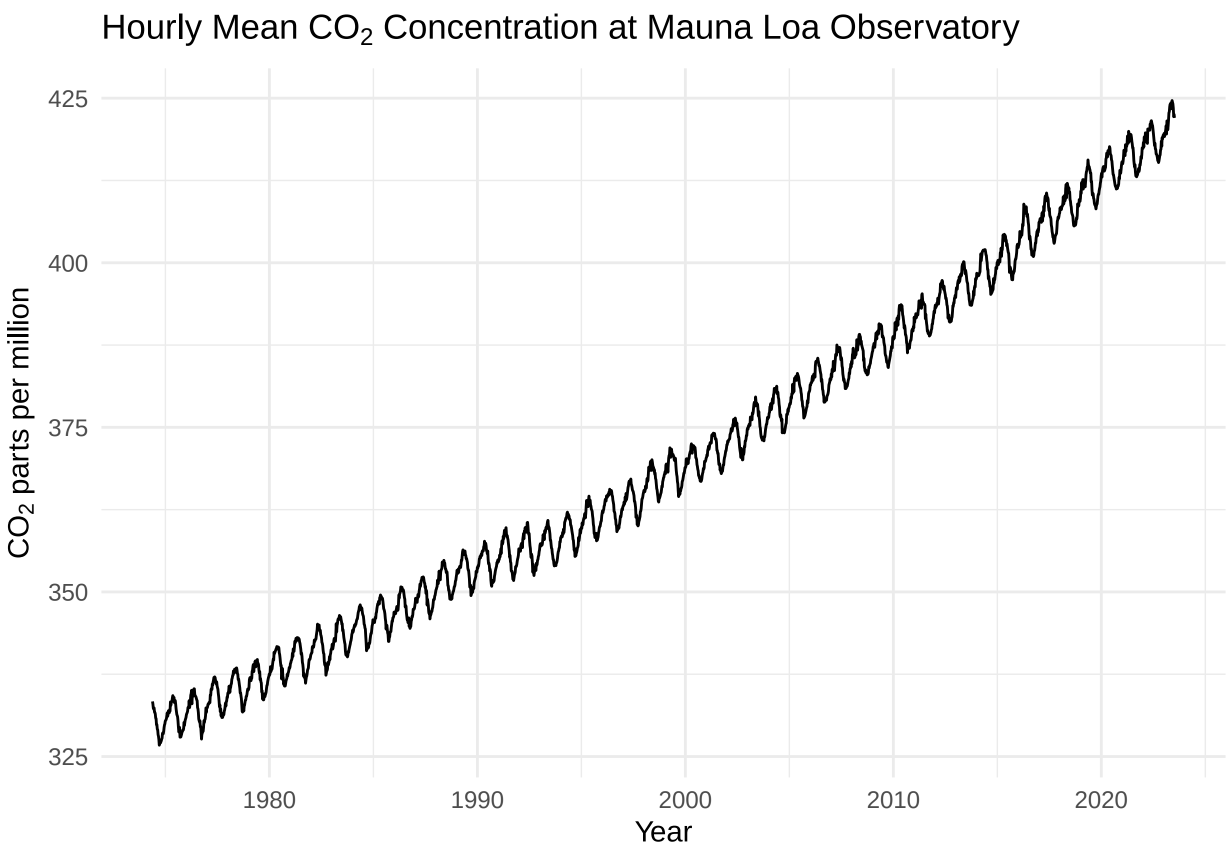
```
# <forecasts for 420/500 ppm>
h_calc <- (2100 - 1998) * 12 + 1
# Forecast atmospheric CO2 growth
co2_forecast <- co2_model %>% forecast(h=h_calc)

fc_hilo <- co2_forecast %>%
    hilo() %>%
    unpack_hilo('95%')

fc_hilo <- fc_hilo %>%
    rename(upper_95 = `95%_upper`) %>%
    rename(lower_95 = `95%_lower`)

first_420_upper <- fc_hilo %>% filter(upper_95 > 420)
# first_420_upper$index[1]

last_420_lower <- fc_hilo %>% filter(lower_95 > 420)
# last_420_lower$index[1]

# 500 ppm
first_500_upper <- fc_hilo %>% filter(upper_95 > 500)
# first_500_upper$index[1]

last_500_lower <- fc_hilo %>% filter(lower_95 > 500)
# last_500_lower$index[1]
```

5 2023 - Revisiting and Revising

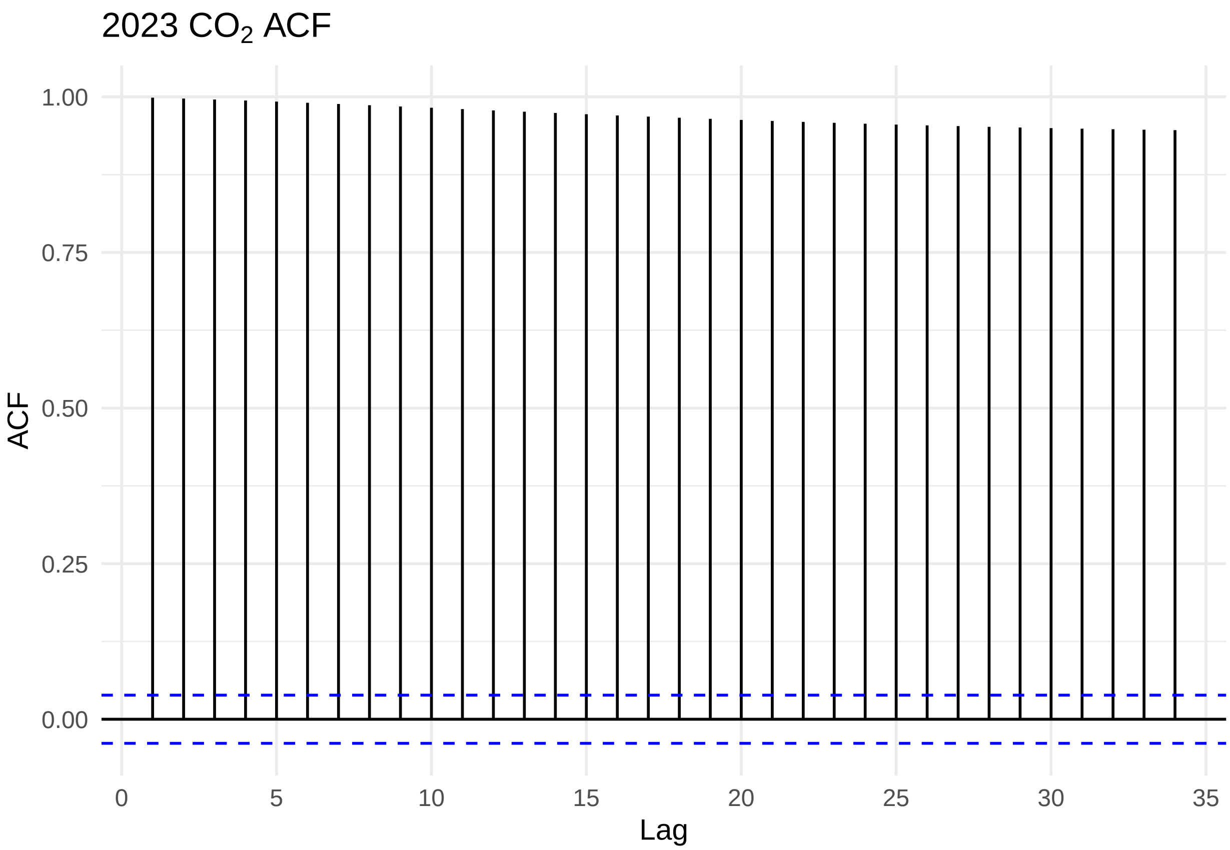
Back in 1997, we made modeling decisions and predictions about atmospheric CO2 levels, based on data collected from the Muana Loa observatory. 26 years later, we revisit those projections and compare them to the actual data collected in the interim. Our goal is to evaluate our modeling techniques from before, and provide updated estimates based on the data now available. Data for this report is sourced from the United States' Nation Oceanic and Atmospheric Administration (NOAA).

The Keeling Curve has many similar aspects to our 1997 models, but there is a notable acceleration in the trend that was not fully anticipated. Still, we see strong indications from the ACF and PCF plots of an AR(1) model, with annual seasonal effects.

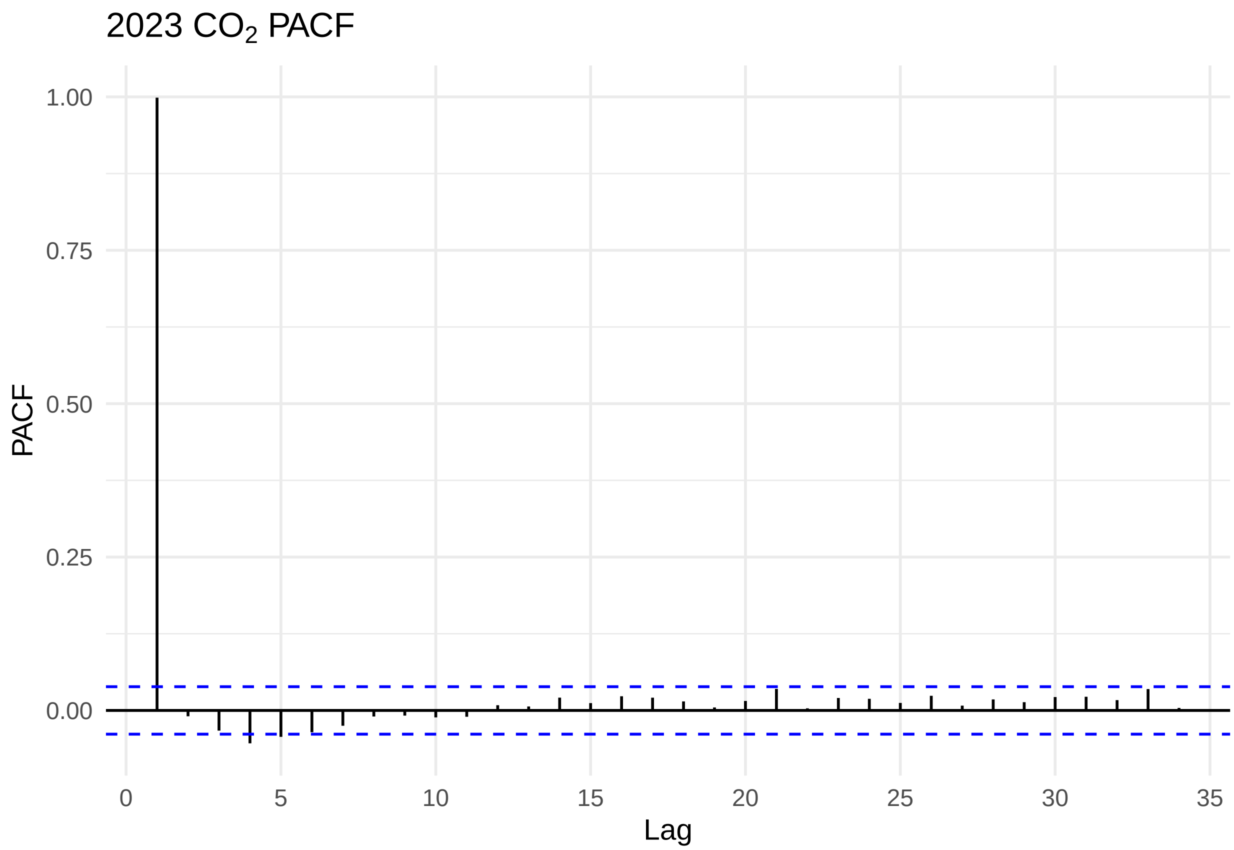
```r
co2_present_ts %>%
    ggplot(aes(x = time_index, y = average)) +
    geom_line() +
    labs(
        title = TeX(r'(Hourly Mean $CO_2$ Concentration at Mauna Loa Observatory)'),
        x = 'Year',
        y = TeX(r'($CO_2$ parts per million)')
    )
```




```r
co2_present_ts %>%
    pull(average) %>%
    ggAcf() +
        labs(title = TeX(r'(2023 $CO_2$ ACF)'))
```



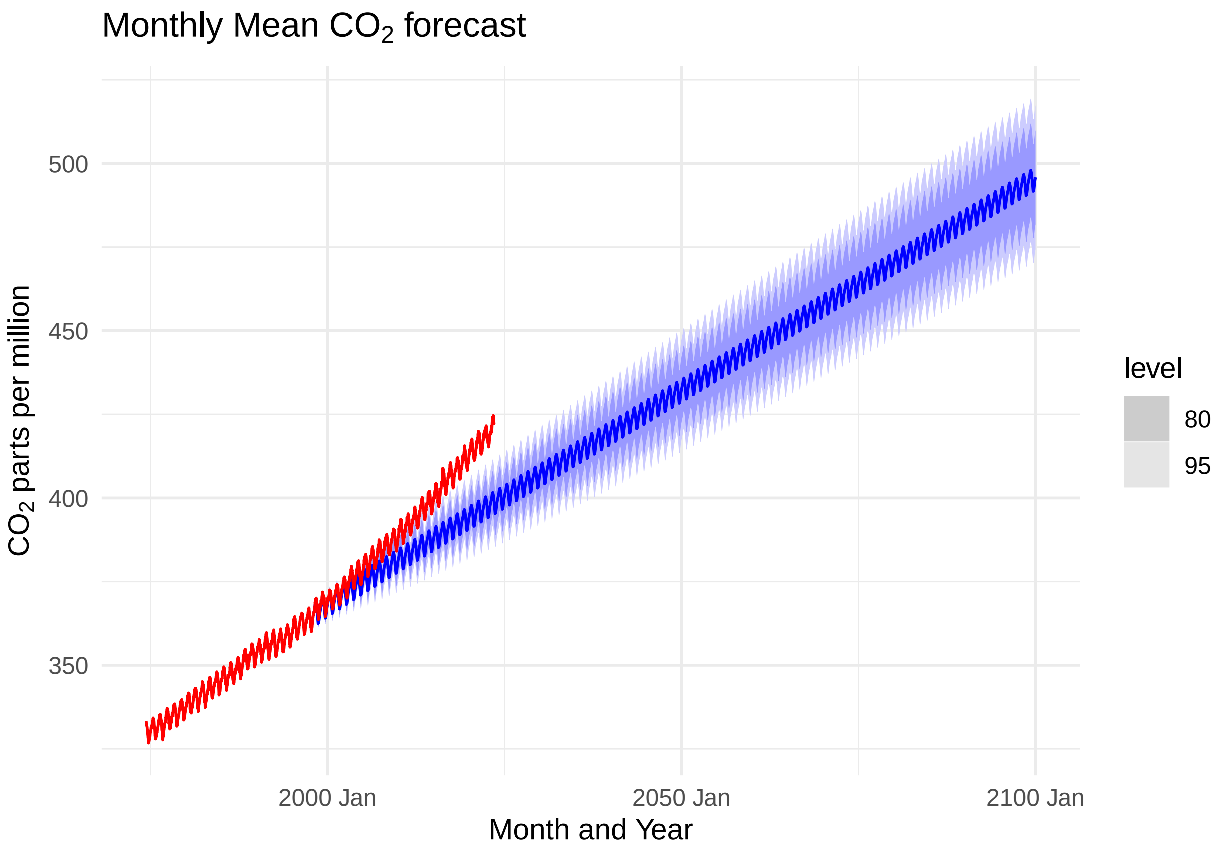

```r
co2_present_ts %>%
    pull(average) %>%
    ggPacf() +
        labs(title = TeX(r'(2023 $CO_2$ PACF)'))
```



6 Evaluation 1997 Forecasts

In 1997, we made predictions about what the CO2 levels would be in January 2020. Below is a summary of those predictions, compared with the actuals measured at Mauna Loa:

```
# <That awesome plot of actuals vs our ARIMA prediction, with a subtitle>
# [0:480,]
co2_forecast %>%
    autoplot() +
    geom_line(aes(y = co2_present_ts$average, x = co2_present_ts$time_index), color = "red") +
    labs(
        title = TeX(r'(Monthly Mean $CO_2$ forecast)'),
        x = 'Month and Year',
        y = TeX(r'($CO_2$ parts per million)')
    )
```




```
# <Table of forecasts and their errors for all 1997 models. Should be copy paste with a few ex
co2_monthly_ts <- co2_present_ts %>%
    data.frame() %>%
    filter(time_index >= "1997-01-01") %>%
    mutate(ym = yearmonth(time_index)) %>%
    select(-time_index) %>%
    group_by(ym) %>%
    summarise(average = mean(average)) %>%
    mutate(
        t = ym - ym[1],
        index = ym,
        value = average
    ) %>%
    as_tsibble(index = ym)

lm_model <- co2_ts %>% model(lm = TSLM(value ~ index))
co2_lm_forecast <- forecast(lm_model, h = h_calc)

lm_poly_model <- lm(value ~ poly(t, degree = 2, raw = TRUE), data = co2_ts)
co2_lm_poly_forecast <- predict(lm_poly_model, newdata = data.frame(t = (nrow(co2_ts)+1):(nrow

rbind(
    accuracy(co2_lm_forecast, co2_monthly_ts)[4:5],
    data.frame(
```

```
        RMSE = RMSE(co2_lm_poly_forecast, co2_monthly_ts$value),
        MAE = MAE(co2_lm_poly_forecast, co2_monthly_ts$value)
    ),
    accuracy(co2_2022_forecast, co2_monthly_ts)[4:5]
)
```

```
## Warning: The future dataset is incomplete, incomplete out-of-sample data will be treated as
## 918 observations are missing between 2023 Aug and 2100 Jan
```

```
## # A tibble: 3 x 2
##    RMSE   MAE
##   <dbl> <dbl>
## 1  57.1  46.7
## 2 663.  635.
## 3  21.5  17.4
```

As shown, our forecasts fell short of actuals. The Keeling Curve shows an acceleration in the trend that our models failed to anticipate. It appears the long-term acceleration slowed during the 1990s, which unfortunately lowered our extrapolated estimates considerably.

Additionally, we made forecasts about when atmospheric CO2 levels would cross 420 ppm, which we now know first occurred in . The bottom of our confidence interval occurred in April 2029, meaning we hit this high level of CO2 about a decade ahead of schedule.

7 Updating the models

8 Conclusion

Using our NSA forecast, we generated predictions 7800 weeks out (150 years) so that 2122 would be included in the prediction. We saw that for the 420ppm prediction, it crossed for the first time at

[fill this in]