Analysis of the effect of CO_2 Concentration on temperature and weather patterns

By: Frason Francis

Introduction:

Over the past few decades, the records for global mean temperature have steadily increased. With every additional increment in temperature, there's an exponential increase in a controlled heat within the atmosphere to cause changes in local and global weather patterns. As a consequence, during this era, we witnessed a gentle an increase in weather-related disasters like prolonged droughts, heavier precipitation and flooding, severe tropical cyclones, and more recently with the sudden spike in wildfires across California. The long-term impacts of temperature change include shifting weather patterns, coastal erosion, and sea-level rise. Carbonic acid gas (CO_2) is a crucial heat-trapping (greenhouse) gas, which is released through human activities and burning fossil fuels. In line with the world organization Intergovernmental Panel on temperature change (IPCC) report, since the start of the commercial revolution, the CO_2 concentrations have risen from 280 ppm (parts-per-million) to 417 ppm as of January 2021. During the identical period, the world means temperature increased approximately 1ºC above the pre-industrial baseline.

Data Collection

Importing the Data information from the NOAA FTP Site into R studio.

We use read.table to load the info from ESRL into a data frame and use head to look at the first few lines of the data.

	V1	V2	V3	V4	V5	V6	V7	V8	V9
1	1974	5	19	1974.380	333.37	5	-999.99	-999.99	50.40
2	1974	5	26	1974.399	332.95	6	-999.99	-999.99	50.06
3	1974	6	2	1974.418	332.35	5	-999.99	-999.99	49.60
4	1974	6	9	1974.437	332.20	7	-999.99	-999.99	49.65
5	1974	6	16	1974.456	332.37	7	-999.99	-999.99	50.06
6	1974	6	23	1974.475	331.73	5	-999.99	-999.99	49.72

Preparing the Data

We can now filter out the decimal years and historical comparisons from the table and keep the year, month, day and the carbon concentration observed. We will only need the first, second, third, and fifth columns from the table. After filtering out the data, we will name the columns as per year, month, day, co2ppm.

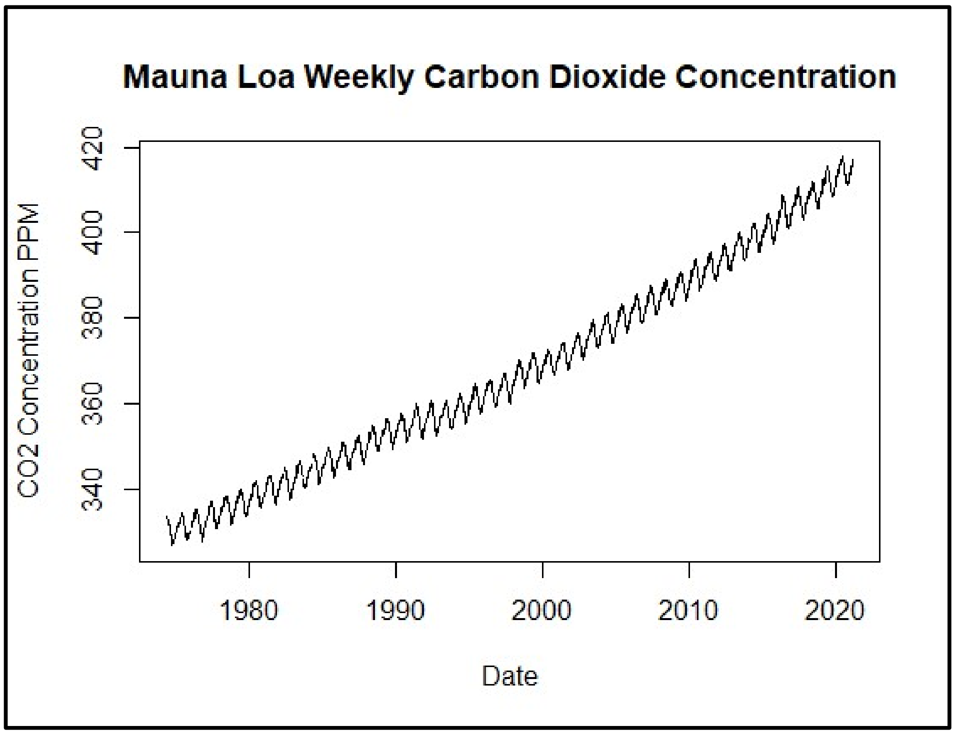
	year	month	day	co2ppm
1	1974	5	19	333.37
2	1974	5	26	332.95
3	1974	6	2	332.35
4	1974	6	9	332.20
5	1974	6	16	332.37
6	1974	6	23	331.73

date	co2ppm
Min. :1974-05-19	Min. :-1000.0
1st Qu.:1986-01-20	1st Qu.: 347.0
Median :1997-09-24	Median : 364.8
Mean :1997-09-24	Mean : 357.8
3rd Qu.:2009-05-29	3rd Qu.: 387.5
Max. :2021-01-31	Max. : 417.7

In order to analyse this data over time, I have convert the year, month, and day columns into a data type as dates. By looking at the carbon concentration data there is a bias error with the minimum 1000 CO2ppm. Noticing that the original file says (-999.99 = no data). Therefore, the value -999.99 is being used to mean that data that isn't available. Therefor we don't want that value in our calculations so, we use a special value called NA instead. And then we examine our data for further analysis.

Examining the Data

By our observations lowest observed CO_2 concentration in the Mauna Loa record is 326.7 ppm, and the highest is 408.7 ppm. The ppm unit stands for [parts per million](). It's a way of expressing very small ratios, equivalent to 0.0001%. We can plot carbon dioxide concentration over time using the date column as the x-axis to look at the data visually. The plot function in R is used to examine the relationship between variables.



Quantifying the Trend

By observation it is very clearly that there is a linear trend from the 326 ppm minimum in 1974 to the 417 ppm maximum in 2021. We then quantify that trend with a linear regression using lm with CO2 concentration as the dependant variable and date as the independent variable.

```
Call:
lm(formula = mauna_loa_weekly$co2ppm ~ mauna_loa_weekly$date)

Residuals:
    Min      1Q  Median      3Q     Max 
-7.7411 -2.3038 -0.0811  2.2305  9.2272 

Coefficients:
                       Estimate Std. Error t value Pr(>|t|)    
(Intercept)           3.180e+02  1.495e-01  2127.3   <2e-16 ***
mauna_loa_weekly$date 4.911e-03  1.325e-05   370.8   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.204 on 2418 degrees of freedom
  (18 observations deleted due to missingness)
Multiple R-squared:  0.9827,	Adjusted R-squared:  0.9827 
F-statistic: 1.375e+05 on 1 and 2418 DF,  p-value: < 2.2e-16
```

We can notice that the p-value <2e-16 *** for date. This means that the chance of this pattern occurring randomly is essentially zero. Also notice the R-squared value of 0.9824. This means that there is a 98% of the change of the CO2 concentration.

Examining Seasonality

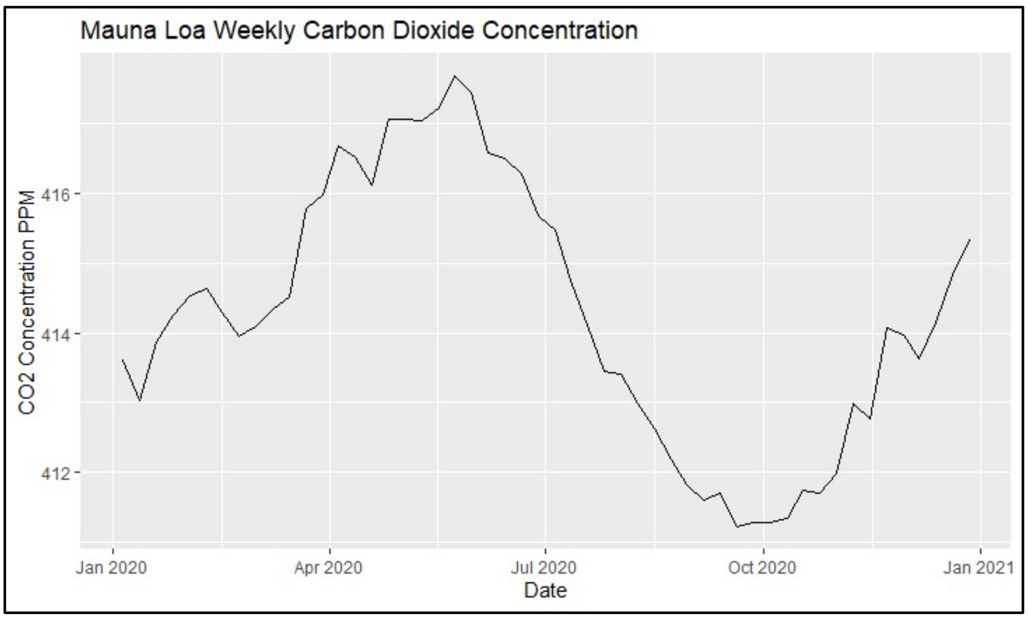
So we can see the long term trend in CO2 concentration is going up in almost a straight line over time, and dipped just below this trend in the 90s.

It looks like it cycles each year, so it should be seasonal variation. Let’s take a closer look at a couple years.

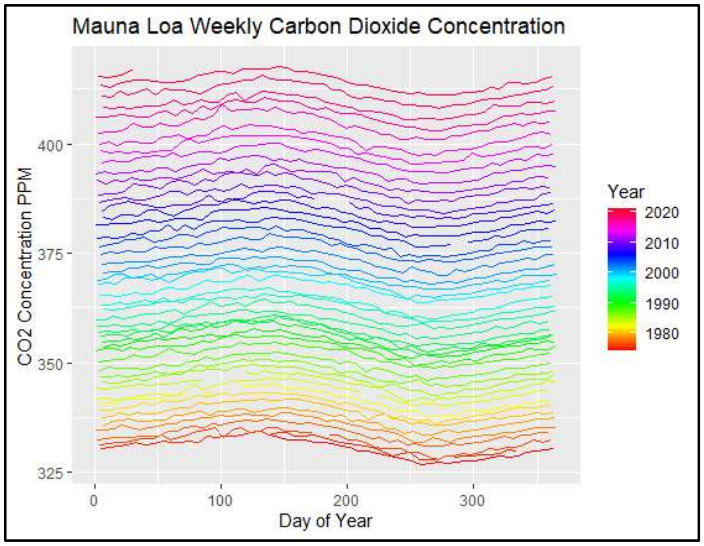
```
> mauna_loa_weekly %>% subset(year(date) == 2021) %>% head()
           date co2ppm
2434 2021-01-03 415.38
2435 2021-01-10 415.07
2436 2021-01-17 415.42
2437 2021-01-24 416.01
2438 2021-01-31 417.12
> mauna_loa_weekly %>% subset(year(date) == 2020) %>% tail()
           date co2ppm
2428 2020-11-22 414.07
2429 2020-11-29 413.97
2430 2020-12-06 413.63
2431 2020-12-13 414.15
2432 2020-12-20 414.84
2433 2020-12-27 415.33
```

```
      date                co2ppm
 Min.   :2020-01-05   Min.   :411.2
 1st Qu.:2020-04-03   1st Qu.:412.9
 Median :2020-07-01   Median :414.1
 Mean   :2020-07-01   Mean   :414.3
 3rd Qu.:2020-09-28   3rd Qu.:415.8
 Max.   :2020-12-27   Max.   :417.7
```

The year 2020 closed a little at a value of 415.33 ppm and the year 2021 started of with a higher value of 417.12 ppm. We should also make note of the range of the values by looking at the summary of 2020.The highest concentration observed in 2020 was 417.7 ppm, and the lowest observed was 411.2 ppm. The average or mean of all observations that year was 414.1 ppm.



It looks like the starting point of 413.60 ppm at the beginning of the year 2020 was close to the mean value of 414.3, and from there, it went up to the maximum of 417.7 ppm around late spring early summer in the northern hemisphere, then steeply dropped in the fall. The maximum value was observed on May 24th and the minimum value was observed on September 20th. The pattern looks very similar. The max value of CO2ppm in 2021 so far is 417.12 ppm. We can look at the seasonal trend in all years observed by looking at the following below plot.



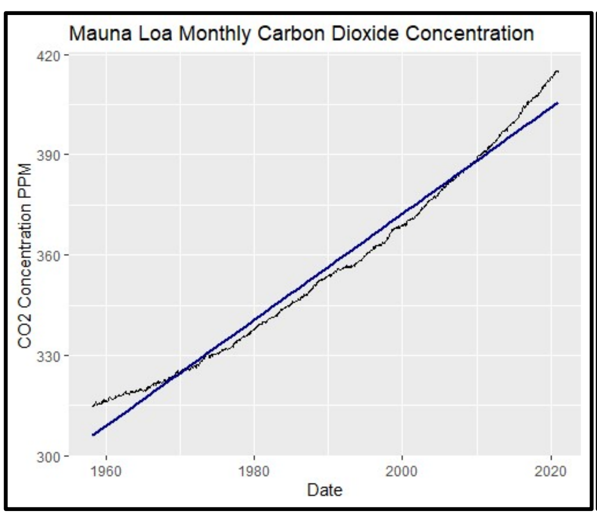
This seasonal cycle repeats each year, but is overcome by the long-term increasing trend.

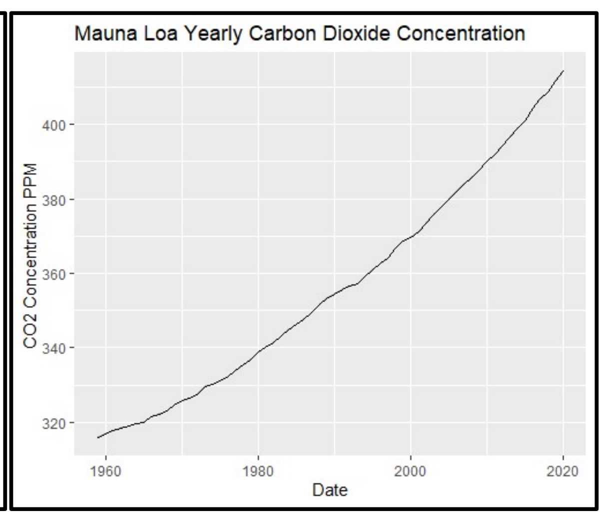
Monthly vs Early Data Analysis

This cycle and trend might be easier to see on a monthly vs yearly basis. Returning to the source data, ESRL publishes monthly as well as yearly Mauna Loa observations.

Monthly: ftp://aftp.cmdl.noaa.gov/products/trends/co2/co2_mm_mlo.txt

Yearly: ftp://aftp.cmdl.noaa.gov/products/trends/co2/co2_annmean_mlo.txt





```
> head(mauna_loa_monthly)
  year month co2ppm       date
1 1958     3 314.43 1958-03-01
2 1958     4 315.16 1958-04-01
3 1958     5 314.71 1958-05-01
4 1958     6 315.14 1958-06-01
5 1958     7 315.18 1958-07-01
6 1958     8 316.18 1958-08-01
```

```
> head(mauna_loa_yearly)
  year co2ppm uncertainty
1 1959 315.98        0.12
2 1960 316.91        0.12
3 1961 317.64        0.12
4 1962 318.45        0.12
5 1963 318.99        0.12
6 1964 319.62        0.12
```

The monthly record at Mauna Loa goes back to 1958. Raw data averages are provided, but as our convenience, an interpolated column is provided that fills in gaps in data.

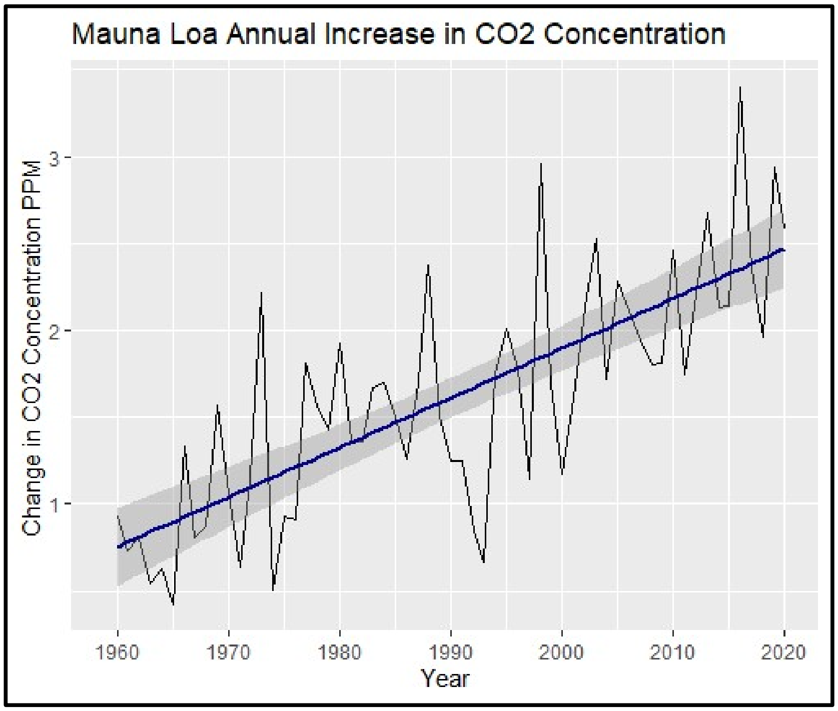
Looking back to 1958 at the longer monthly record, the straight linear progression we found from 1974 to 2020 is starting to deviate. In the past half century, the increase of CO2 in the atmosphere is curving up and growing at a faster rate of time. The yearly data includes an estimation of uncertainty, which is how far off you think your estimation of a variable is. For this dataset, the uncertainty is constant and very small, at 0.12 ppm.

This annual trend isn't perfectly linear. It looks like it's curving up over time.

In all years observed, the change of yearly average CO2 concentration was a positive increase, they varied from a minimum change of 0.420 ppm up to a maximum change of 3.400. The average rate of change of 1.611 ppm per year.

```
   year co2ppm uncertainty co2ppm.inc
7  1965 320.04        0.12       0.42
58 2016 404.41        0.12       3.40
>
```

The year of the lowest increase was 1965, and year of the highest increase was 2016. The graph below shows it's increases in general trend over time.

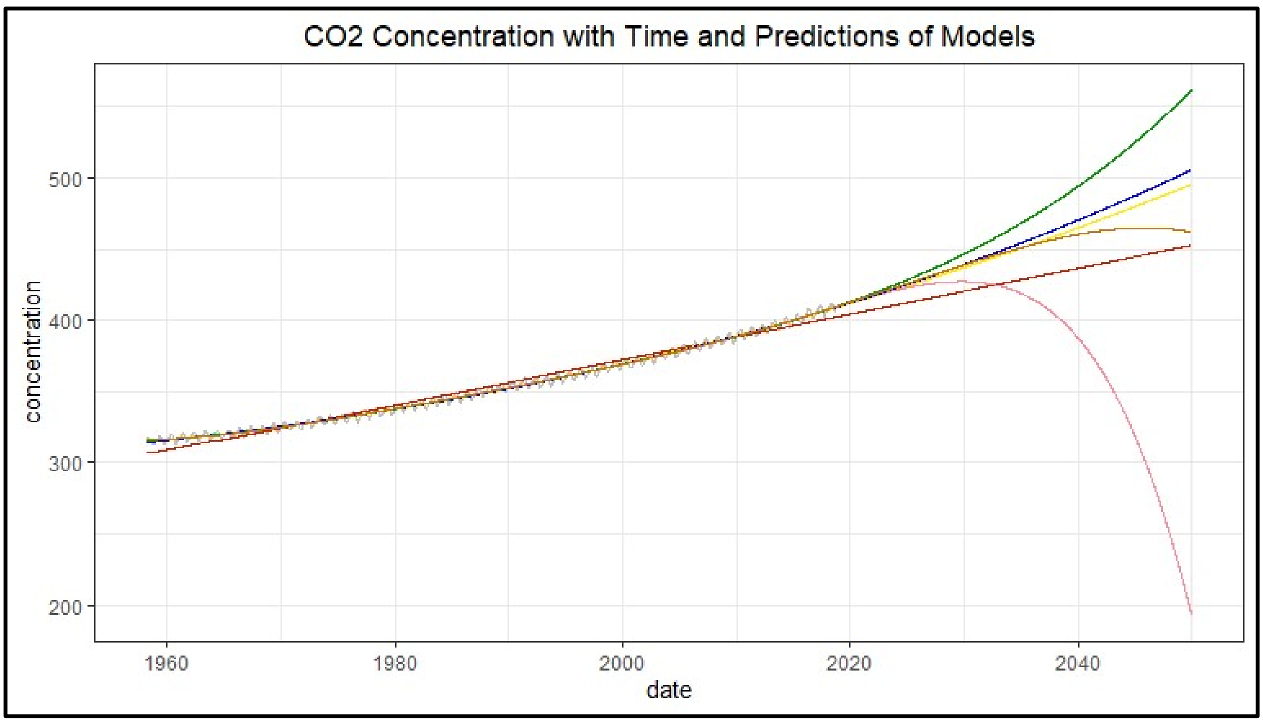


We can now see the variation in this data. But the overall trend within this variation is the rate at which CO2 is increasing from year to year is itself increasing. So now we've seen that not only has CO2 been accumulating in the atmosphere the past half century, but it's been accumulating faster over time.

Model selection to predict CO2 concentrations

I have used the ANOVA function of R to compare the performance of models and explain the variance. I have created 7 models. Where *m1* is the linear model and m2 to m7 are the higher order polynomial models. With the help of ANOVA function, I saw that adding 7 terms did not contribute to explaining variance in the data so I have chosen model 6 as my final model to predict the future.

Later, I observed that model 6 (m6) is making a strange extrapolation which is not expected. The higher order polynomial model explains the variance in the sample data better than lower order polynomial models but it is not making good future predictions. I observed this fact by experiments.



The Above graph shows the extrapolations of different models with colorful lines. After performing various experiments, I decided to use model *m4* model which is a polynomial with degree 4 to extract the CO2 concentration to 2050 (green line on the graph).

Conclusion:

Clearly, countries around the world aren't doing enough to achieve global climate change goals, and are operating business-as-usual. With increasing temperatures, we will expect to work out the frequency and severity of maximum weather-related disasters to extend, melting glaciers and coastal inundation causing mass migrations, droughts and fire destroying rainforest and global food supply chains. More must be done to combat temperature change and build a low-carbon future. there's a general perception that low-carbon future means losing out on conventional jobs and a stronger economy. Contrary to the final belief, companies like Tesla who are at the forefront of the fight against carbon emissions, and showing the way to getting it done.

REFRENCES:

https://www.esrl.noaa.gov/gmd/ccgg/trends/

https://www.usgs.gov/volcanoes/mauna-loa/geology-and-history

https://www.esrl.noaa.gov/gmd/ccgg/about/co2_measurements.html

https://towardsdatascience.com/interpreting-climate-change-through-data-science-321de6161baf

https://scrippsco2.ucsd.edu/data/atmospheric_co2/mlo.html

https://research.noaa.gov/article/ArtMID/587/ArticleID/2636/Rise-of-carbon-dioxide-unabated