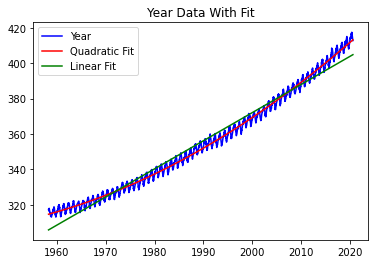
# Time Series Lab Report - Sam Freed

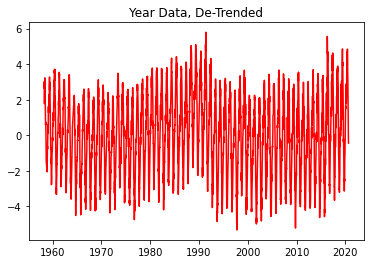
In this lab, we were introduced to PyPlot, NumPy, and SciPy, learned how to import data from various file formats, and then modified and reformatted the data to truly understand its impact.

### Section 1: Year Data and Fitting



The figure above showcases the given data, CO2 levels in situ across more than 50 years. The yearly cycle and a general increase in the values can be seen, but this data alone would not help us to predict the general trend. In order to do this, we can fit the data to polynomials of various orders. First-order (Ax + B) and second-order (Ax2 + Bx + C) fits are shown above; the second-order fit, shown in red, will be used for the remainder of this lab.

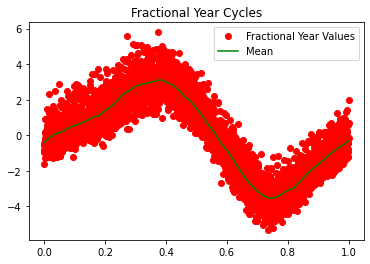
If we subtract the values of this fit from the initial data, we can then see how intense the yearly cycles are relative to the general trend.



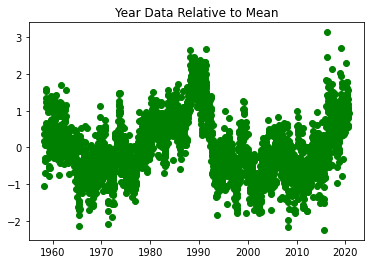
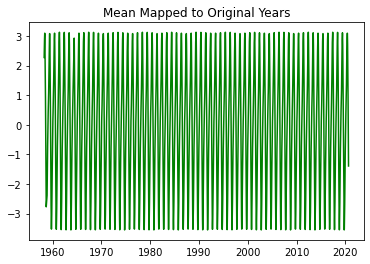
This new visualization reveals that, despite a relatively similar appearance in the yearly data above, the yearly cycles are by no means uniform.

### Section 2: Isolating the Yearly Values

In this section, the yearly data is broken down into a year-long frame so that the yearly cycle can be seen and utilized.



The figure above was achieved by removing the year from its individual week data points, represented by the red dots. Then, each value was binned, the mean of each bin was calculated, and then the means were connected as a line. This allows for an idea of a “standard” year’s fluctuations to be determined based on the dataset.



This mean can then be re-mapped to the original year values instead of the fractional year using interpolation, and then the true year data can be compared to the expected mean as shown above.

### Section 3: Other Plots and Statistics

First, using the Augmented Dickey-Fuller test, the probability of the existence of a unit root and stationary data can be calculated. This test produced the following values for this dataset:

ADF Statistic: 1.348346

p-value: 0.996866

Critical Values:

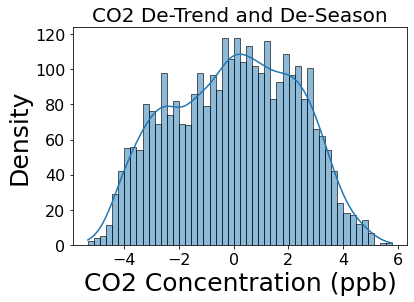
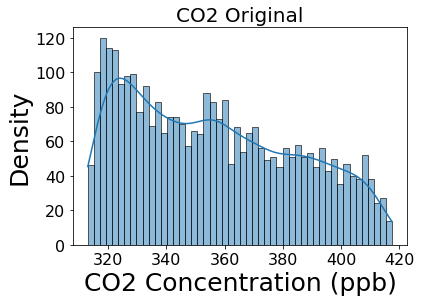
1%: -3.432

5%: -2.862

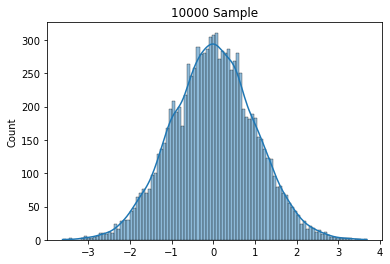
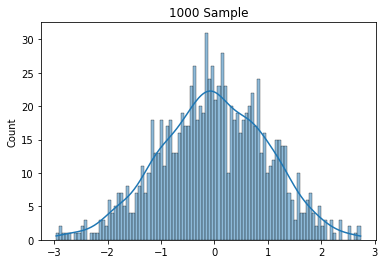
10%: -2.567

The most significant value of these is the p-value, which in this case determines whether or not a unit root exists in the data. With a p-value of 0.9968, the data is 99.68% likely to be non-stationary and we cannot reject that there is a unit root.

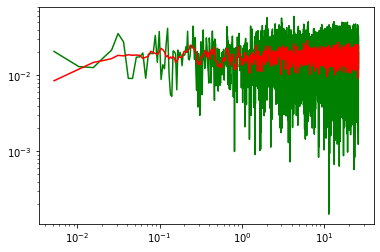
Second, the data is represented in another way: using histograms. These allow us to view the density of each data point and also are very easy to compare distributions with known, standard distributions, such as Gaussian distributions.

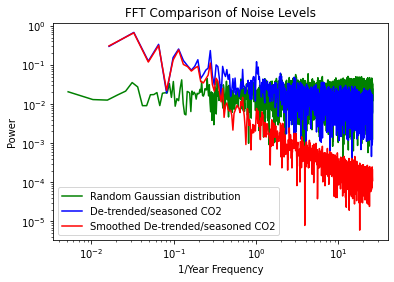
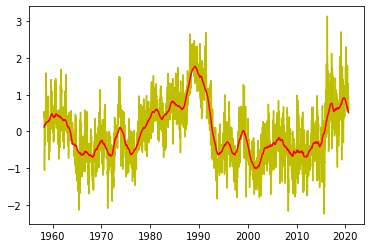


The histogram on the left shows the density of the raw data values, which showcases the larger distribution of the lower concentrations due to the lower end of the exponential curve in the 1960s and the 1970s. The histogram on the right then shows the de-trended, de-seasoned data, which is very easy to compare to the Gaussian distributions shown below.



Finally, the importance of smoothing can be shown using the Fast Fourier Transform on the Gaussian and de-trended/de-seasoned data. In smoothing, one hopes to remove noise from the data and flatten out the Fourier transform as much as possible, which can be seen below. The green data is not smoothed, and there are a lot more frequencies involved in its Fourier breakdown as a result. The red data is smoothed, and the magnitude of the Fourier frequencies are therefore not nearly as large.





The red line on the left is the smoothed CO2 data, which is then put through the FFT. Its resulting Fourier signal is much flatter and the power levels are much lower, showing the success of our smoothing.