PHY224 Fitting Exercise 1: Annual Average of daily measurements of atmospheric carbon dioxide since 1960 from the top of Mauna Loa in Hawaii

Rui Geng (student num. 1003048454)

Due Sunday January 26th 2025

Also note: Any constructive feedback or recommendation about formatting or otherwise for future improvement(s), is greatly appreciated.

1 Linear Model fitted over the last 20 years

1.1 Linear Model Specifics:

In this section, we are only looking at the last 20 years of data provided by the National Oceanic and Atmospheric Administration (file named: *co2_annmean_mlo.csv*), in other words, the data from and including 2004 to and including 2023.

Our set up follows the instruction provided by exercise 1 in the fitting documentation. Where we adopted the model of $f(x) = ax + b$ to also include an offset of 2010 years to reduce the prediction error, and x is the year. Our model is then:

$$f_l(x) = A(x - 2010) + B, \forall (A, B) \in \mathbb{R}^2$$

By applying the function $curve\_fit$, we can obtain an estimated values for both parameters A and B rounded to 3 decimal places with uncertainty obtained by taking the square root of the diagonalized covariance variable, i.e. $\sqrt{np.diag(lin\_pcov)}$ then taking the value at index 0 for A's uncertainty and index 1 for B's uncertainty. They are:

$$A = 2.303 \pm 0.005, \text{ and } B = 390.521 \pm 0.031$$

By applying the function $curve\_fit$ we obtain a linear function that best captures the given dataset to be:

$$f_l(x) = 2.303(x - 2010) + 390.521 \text{ in unit of ppm} \tag{1}$$

If we also include the uncertainties associated with the parameters A and B, then our model would be:

$$f_l(x) = (2.303 \pm 0.005)(x - 2010) + (390.521 \pm 0.031) \text{ in unit of ppm} \tag{2}$$

Furthermore, from this data we can predict the value of both 1960 (for verification with the actual measured data) and 2060 (the future). We obtain the values:

Measured Data (1960): 316.91±0.12 ppm

Linear Model Prediction (year 1960): 275.365±0.235 ppm

Linear Model Prediction (year 2060): 505.677±0.235 ppm

Note: Since we are only using a linear model to approximate only with the last 20 years of data, it is to be expected that 1960's data is way off.

1.2 Linear Model Uncertainties and reduced chi-squared value χ_r^2

The calculation of the reduced chi-squared value follows the same function that Prof. Christopher Lee demonstrated in his fitting python tutorial. The result we obtained for this linear model is:

$$\boxed{\chi^2_{r_{lin}} = 35.399}$$

The calculation of the total/absolute uncertainty for this model follows the principle from lecture 1 and 2, and lab 1. Where we first divided the function into 2 separate components, calculate their respective uncertainties, and then obtain the total uncertainty from the model:

1. $g(x) = Ax \rightarrow u(g) = g\sqrt{\left(\frac{u(A)}{A}\right)^2 + \left(\frac{u(x)}{x}\right)^2}$
2. $h(x) = g(x) + B \rightarrow u(h) = \sqrt{u(g)^2 + u(B)^2}$

Note: since the variable x corresponds to the exact number of year, and do not have a situation of say 1960 ± 1 year. Then the uncertainty corresponding to x is 0, namely $u(x) = 0, \forall x \in available\, years$. Then we obtained the value of both A, B's uncertainties via the calling and reading of the variable lin_pcov. This variable is an array of 2 elements since our model only has 2 elements. In this case $u(A) = lin\_pcov[0]$ and $u(B) = lin\_pcov[1]$. Now we can compute the uncertainty of these and obtain:

1. Let $y = (x - 2010)$. Then $g(y) = Ay \rightarrow u(g) = g\sqrt{\left(\frac{u(A)}{A}\right)^2 + \left(\frac{u(y)}{y}\right)^2} = g\sqrt{\left(\frac{u(A)}{A}\right)^2 + \left(\frac{0}{y}\right)^2} = g\sqrt{\left(\frac{u(A)}{A}\right)^2} = g\frac{u(A)}{A} = Ay\frac{u(A)}{A} = y * u(A)$
2. Let $h = g(y) + B$. Then $u(h) = \sqrt{u(g)^2 + u(B)^2}$

Since the variable for years is an array, then our uncertainty would be applied to the individual year and then be populated into a new total uncertainty array corresponding to our linear model. This will be done via a for loop.

1.3 Linear Model Plot

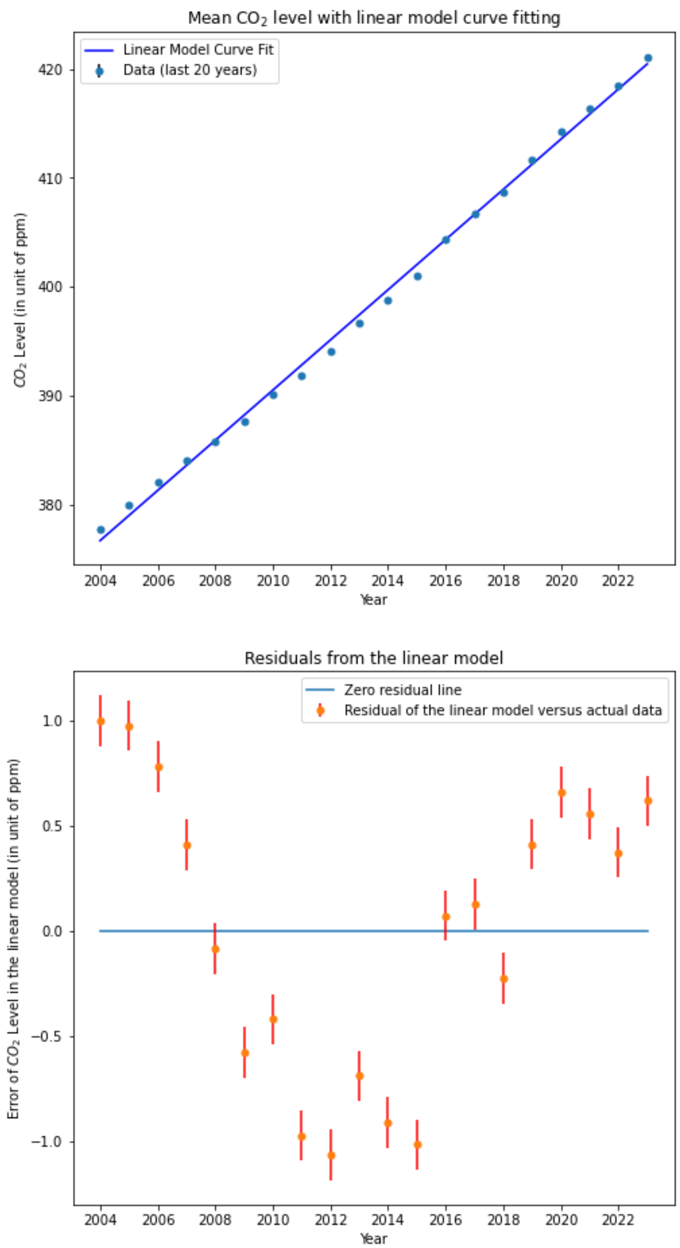

Figure 1: Linear Model Fitting and Residuals: $f_l(x) = (2.303 \pm 0.005)(x - 2010) + (390.521 \pm 0.031)$ in unit of ppm

2 Power Law models fitted over the entire data set

2.1 Quadratic fit

2.1.1 Quadratic Model Specifics:

Following the same structure as the linear model's. Now we will use a quadratic model to perform the curve fitting for the entirety of the dataset, ranging from 1959 to 2023.

The quadratic model is based on the format of: $f(x) = Ax^2 + Bx + C$, where x is the year. More specifically, after considering the year offset to avoid runtime error from the curve fitting function, we have arrived at the formula of:

$$f_q(x) = A(x-1959)^2+B(x-1959)+I, \forall(A,B) \in \mathbb{R}^2, I = \text{initial measured data in 1959}$$

By applying the *curve_fit* function, we obtained the estimated values for parameters A and B rounded to 5 decimal places (to accommodate A's uncertainty) to be:

$$A = 0.01348 \pm 0.00003, \text{ and } B = 0.77451 \pm 0.00161$$

By applying the *curve_fit* function, we obtain a quadratic function that best captures the entire given dataset to be:

$$f_q(x) = 0.01348(x-1959)^2 + 0.77451(x-1959) + 315.98000 \text{ in unit of ppm} \tag{3}$$

If we also include the uncertainties associated with the parameters and the initial condition, our function would then be:

$$\begin{aligned} f_q(x) = (0.01348 \pm 0.00003)(x-1959)^2 + (0.77451 \pm 0.00161)(x-1959) \\ +(315.98000 \pm 0.12000) \text{ in unit of ppm} \end{aligned} \tag{4}$$

Using this quadratic model to predict the values of 1960 and 2060, we obtain:

Measured Data in 1960: 316.91±0.12 ppm

Quadratic Model Prediction (year 1960): 316.77±0.12 ppm

Quadratic Model Prediction (year 2060): 531.69±0.35 ppm

Noticeably, the quadratic model is performing much better than did the linear model.

2.1.2 Quadratic Model Uncertainties and reduced chi-squared value χ_r^2

Similar to that of the linear model, we calculated the reduced chi-squared value to be:

$$\chi_r^2 = 37.850$$

The calculation of the total/absolute uncertainty for this model is similar to that of the linear model, and follows the processes demonstrated in the class. We will require the following:

1. $A(x) = x^2 \to u(A) = A\sqrt{\left(\frac{u(x)}{x}\right)^2 + \left(\frac{u(x)}{x}\right)^2} = A\sqrt{2\left(\frac{u(x)}{x}\right)^2}$
2. $B(x) = aA(x) =\to u(B) = B\sqrt{\left(\frac{u(a)}{a}\right)^2 + \left(\frac{u(A)}{A}\right)^2}$
3. $C(x) = Cx \to u(C) = C\sqrt{\left(\frac{u(C)}{C}\right)^2 + \left(\frac{u(x)}{x}\right)^2}$
4. $D(x) = B(x) + C(x) \to u(D) = \sqrt{u(B)^2 + u(C)^2}$
5. $E(x) = D(x) + I \to u(E) = \sqrt{u(D)^2 + u(I)^2}$

Note: just like with the linear model section, since we are looking at the exact number of year, the uncertainty of the year $u(x)$ is 0. Also, we are given the uncertainty of the measured data, where $u(I) = 0.12$. Then:

1. $$A(x) = x^2 \to u(A) = A\sqrt{\left(\frac{u(x)}{x}\right)^2 + \left(\frac{u(x)}{x}\right)^2} = A\sqrt{2\left(\frac{u(x)}{x}\right)^2} = 0$$
2. $$B(x) = aA(x) =\to u(B) = B\sqrt{\left(\frac{u(a)}{a}\right)^2 + \left(\frac{u(A)}{A}\right)^2} = aA(x)\sqrt{\left(\frac{u(a)}{a}\right)^2 + \left(\frac{0}{A}\right)^2} = A(x)u(a)$$
3. $$C(x) = Cx \to u(C) = C\sqrt{\left(\frac{u(C)}{C}\right)^2 + \left(\frac{u(x)}{x}\right)^2} = C\sqrt{\left(\frac{u(C)}{C}\right)^2 + \left(\frac{0}{x}\right)^2} = u(C)$$
4. $$D(x) = B(x) + C(x) \to u(D) = \sqrt{u(B)^2 + u(C)^2}$$
5. $$E(x) = D(x) + I \to u(E) = \sqrt{u(D)^2 + u(I)^2} = \sqrt{u(D)^2 + (0.12)^2}$$

This is then applied to the python code, and is done via a for loop to account for the year being an array of many years.

2.1.3 Quadratic Model Plot

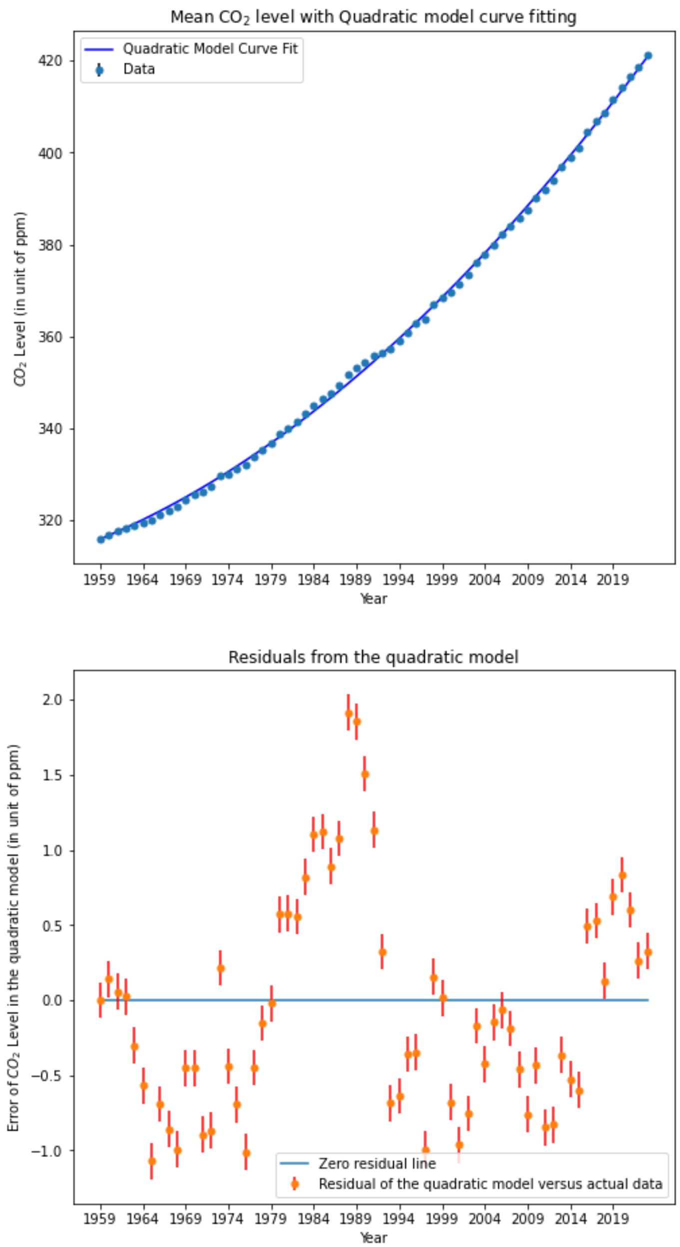

Figure 2: Quadratic Model Fitting and Residuals: $f_q(x) = (0.01348 \pm 0.00003)(x - 1959)^2 + (0.77451 \pm 0.00161)(x - 1959) + (315.98 \pm 0.12)$ in unit of ppm

2.2 Power law fit

2.2.1 Power Model Specifics:

Let's now repeat our process for the quadratic model and apply to a power model. In this section, we will adopt a model in the format of $f(x) = Ax^B + C$, where x is the year. More specifically, considering the same year offset as we did in our quadratic model. Our model becomes:

$$f_p(x) = A(x - 1959)^B + I, \forall (A, B) \in \mathbb{R}^2, I = \text{ initial measured data in 1959}$$

Then using the *curve_fit* function again, we obtained the estimated values for the parameters A and B to be:

$$A = 0.304 \pm 0.001, \text{ and } B = 1.401 \pm 0.001$$

Using these fitted values, our model now can be expressed as:

$$f_p(x) = 0.304(x - 1959)^{1.401} + 315.980 \text{ in unit of ppm} \tag{5}$$

By including the uncertainties, we obtain:

$$\begin{aligned} f_p(x) = (0.304 \pm 0.001)(x - 1959)^{(1.401 \pm 0.001)} \\ +(315.980 \pm 0.120) \text{ in unit of ppm} \end{aligned} \tag{6}$$

Using this power model, we predicted the values of CO_2 to be:

Measured Data in 1960: 316.91±0.12 ppm

Power Model Prediction (year 1960): 316.28±0.12 ppm

Power Model Prediction (year 2060): 511.44±0.12 ppm

Interestingly the power model did not predict as accurate of a result as did the quadratic model. This suggests that in comparison, the quadratic model is actually more accurate than the power model. Below, we can see that indeed, the reduced chi-squared value for this model is significantly higher than that of the quadratic model. Indicating that the power model's expected values are in fact, further away from the measured data than those of the quadratic model.

2.2.2 Power Model Uncertainties and reduced chi-squared value χ_r^2

Just as we did with the linear and quadratic models, the reduced chi-squared value for the power model is:

$$\chi_r^2 = 81.860$$

Similar to the quadratic model, we will use the following equations to calculate the absolute uncertainty via propagation, of the power model.

$$1.\ A(x) = x^n \rightarrow u(A) = A\sqrt{\left(n\frac{u(x)}{x}\right)^2 + (ln(x)u(n))^2}$$
$$2.\ B(x) = bA(x) \rightarrow u(B) = B\sqrt{\left(\frac{u(b)}{b}\right)^2 + \left(\frac{u(A)}{A}\right)^2}$$
$$3.\ C(x) = B(x) + I \rightarrow u(C) = \sqrt{u(B)^2 + u(I)^2}$$

Again, setting $u(x) = 0, u(I) = 0.12$. We can calculate the total uncertainty to be:

$$1.\ A(x) = x^n \rightarrow u(A) = A\sqrt{\left(n\frac{u(x)}{x}\right)^2 + (ln(x)u(n))^2} =$$
$$A\sqrt{\left(n\frac{0}{x}\right)^2 + (ln(x)u(n))^2} = A * ln(x)u(n)$$
$$2.\ B(x) = bA(x) \rightarrow u(B) = B\sqrt{\left(\frac{u(b)}{b}\right)^2 + \left(\frac{u(A)}{A}\right)^2}$$
$$3.\ C(x) = B(x) + I \rightarrow u(C) = \sqrt{u(B)^2 + u(I)^2} = \sqrt{u(B)^2 + (0.12)^2}$$

This is then implemented into the Python code via a for loop as well to account for the array nature of the year variable.

2.2.3 Power Model Plot

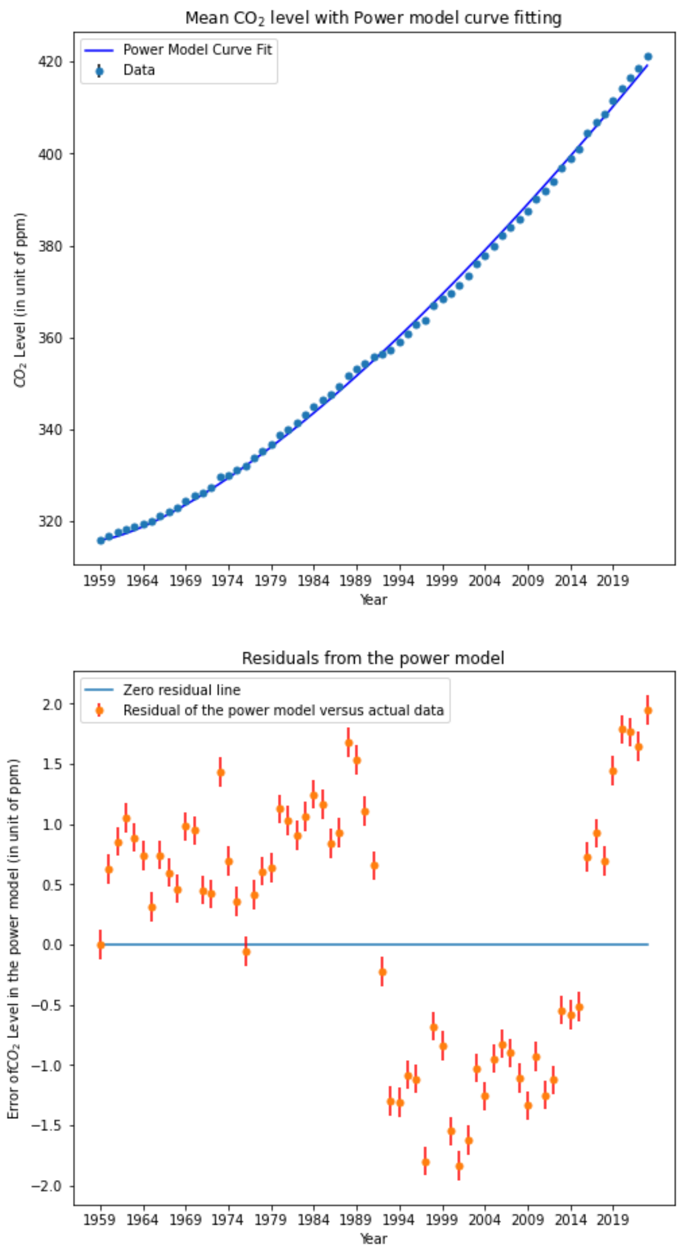

Figure 3: Power Model Fitting and Residuals: $f_p(x) = (0.304 \pm 0.001)(x - 1959)^{(1.401 \pm 0.001)} + (315.980 \pm 0.120)$ in unit of ppm

3 Conclusion

In conclusion, we can observe that the quadratic model has performed the best out of the 3 models we attempted to curve fit to the given dataset. This is to be expected as from the class we had learned that the closer the χ_r^2 value is to 1, the more accurate the model is.

However, while the quadratic model's reduced chi-square value is the closest to 1. It is still significantly large. This could be caused by outliers in the measured data. As can be seen on all of the model plots, the data points do not line up in a perfectly smooth line.

As an exploratory idea, to best fit the given dataset, it might be helpful to combine the idea behind our power and the quadratic models. With the possibility of implementing the summation of a series of polynomials of different degrees.

4 Appendix: Useful Models, Equations and Formulas used in this lab.

4.1 Equations and Formulas

Chi-Squared: $\chi^2 = \sum_{i=1}^{N} \left(\frac{y_i - f(x_i)}{u(x_i)}\right)^2$

Reduced Chi-Squared: $\frac{\chi^2}{N-m}$

Residuals: $\epsilon = y_i - f(x_i)$