Projecting Atmospheric Carbon Dioxide concentrations

Objectives: The purpose of this assignment is to plot real data; try linear and non-linear least-squares fitting to develop a simple model of the observed behavior; and make testable predictions.

Logistics: All homework assignments will consist of writing programs, one for each part of a problem, and providing a printout of the problem along with the output obtained when the program is run. Organize your explanations and code in order in a **Jupyter notebook**, and then execute the code and save the entire as an html file. Both the 'ipynb' file and the html output should be uploaded under the Assignment on Moodle, two files per homework assignment. **Please write your name and the date at the top.**

The different parts of the problem set will be graded as indicated on the assignment sheet. This includes the following (for 5 points): every program you turn in should start with a **comment or docstring** that lists **what the problem is**. Each part of your program should also have short comments describing how the program works.

Introduction: The National Oceanic and Atmospheric Administration (NOAA) Global Monitoring Division makes public its measurements of atmospheric CO_2 concentrations on their website: https://www.esrl.noaa.gov/gmd/ccgg/trends/

Included in the Moodle assignment are two data files, downloaded from this site, cataloging the measurements of carbon dioxide concentrations at a site in Mauna Loa, Hawaii:

co2_mm_mlo.txt is the full text file, including a header describing the allowed use of the data and the meaning of the various columns in the dataset.

co2_data.txt is a truncated file with the headers stripped and containing just the columns of data.

Part 1: Loading and Plotting the Data

Skim through the original text data file's header and find the description for each data column. The first two columns are the year and month that the data was collected, and then these are combined to make a single "decimal time" in the third column. The fourth column contains the raw CO_2 concentration measurements, but occasionally a measurement is missing, leading to some entries of -99.99 – these are corrected for (using a linear interpolation) in the fifth column. You'll want to make note of the units for the various quantities in each column.

A. (10 pts) Read in the data in co2_data.txt. Store the incoming data as an array and then reference the data columns you will actually use when making your plot. You'll want to use the third data column for the time and the fifth data column for the CO_2 concentration.

B. (10 pts) Plot the data, including axis labels, with units, and a title.

C. (10 pts) You will see that the data has a pronounced oscillation of about 1 year due to seasonal variations in CO_2 levels. Estimate the amplitude of this oscillation – note that this is one measure of the uncertainty in the concentration. Your answer to this part should either be a clearly-labeled comment in your program, or as text displayed to the command window when your program is run.

Part 2. Linear and Quadratic fits

To make our fits a little easier, it will be better to work with a time axis that starts at 0. This means that our new time axis is "time since 1958", the first time point in the data set. Create a new list/array that is equal to the old one minus the initial time value. Check that the new time array has its first entry equal to 0 and its last entry equal to 60.4170. The function you will use to do the fits is *scipy.optimize.curve_fit.* Be sure to include the following in your program: *from scipy.optimize import curve_fit*

A. (20 pts) Perform a linear fit to the data, and find the best values of the intercept and slope. Plot the data and the best fit line together; plot the residuals from the fit in a separate plot. Print the best fit values +/- σ (uncertainties) to the screen (incluce units), as well as the chi-squared value for the fit.

B. (5 pts) Assess your fit: how well do you think the data is fit by a straight line? Be sure to justify your assessment with evidence! Your answer to this part should be included as a notebook cell.

C. (20 pts) Perform a quadratic fit to the data, and extract the best values of the fit parameters. Plot the data and the best fit line together and the residuals from the fit in a separate plot. Print the best fit values +/- σ (uncertainties) to the screen, as well as the chi-squared value for the fit.

D. (5 pts) Assess your fit: how well do you think the data is fit by a quadratic function? Be sure to justify your assessment with evidence! Place your answer to this part in a notebook cell.

Part 3. Projecting future CO_2 levels

Use your linear and quadratic fits to project the CO_2 levels at Mauna Loa in 2058, 100 years since the beginning of the data set.

A. (10 pts) First generate a plot with the data and your two best fits (one linear, one quadratic) with the time axis extended to 100 years.

B. (10 pts) Then, find the value of the CO_2 concentration at t=100 years in your linear fit and your quadratic fit, and print these to the screen, with units.

What to hand in:

Remember to turn in your complete code for all three parts, with output and plots generated in each, as an ipynb file and a html file.

Due date: 11:59 pm on Friday October 12, 2018