MONT 104N – Modeling the Environment
Chapter 4 Project
October 15, 2018

Background Information

The Mauna Loa Observatory (located at an elevation of about 3400 meters on the Mauna Loa volcanic mountain on the "big island" of Hawaii) is a research station maintained and staffed by the National Oceanic and Atmospheric Administration (NOAA), the major agency of the U.S. government that studies weather and climate phenomena. NOAA maintains a web site, for instance, with the National Weather Service day-to-day forecasts and severe weather warnings that form the basis for much local weather reporting in the media. Among the data collected regularly at Mauna Loa are measurements of atmospheric concentrations of a host of trace gases, including carbon dioxide, or CO_2. The units are ppm (CO_2 molecules per million air molecules, after water vapor has been removed). The data set of those measurements goes back to 1958 and is one of the most complete records of the recent evolution of this aspect of the Earth's atmosphere. In this project we will apply the modeling techniques introduced in this chapter to try to understand what this data set is saying about changes in atmospheric CO_2 over time.

Important Note

This is a well-known data set and you can find all sorts of discussions of various aspects of it on the web, if you look. *Please do not look at these* until after you have worked through at least questions A through D below. The idea is for you to approach this as "fresh" as possible, make your own observations and analysis, and draw your own conclusions.

Getting Started

The data we will need comes from a text file that you will download from the course homepage. Look under the Examples, Class Notes, Etc. heading for the *Data for Chapter 4 Project*. This is a large text file with columns of data set up for reading into GoogleSheets. Download it to your computer, then use the File/Import with the Upload option to read it in. Make sure GoogleSheets is using the Separator Type/Detect Automatically option.

Questions

The first thing you will notice if you look at the CO_2 levels from month to month is that there is *a lot* of up-and-down variation.

(A) Is it completely random, though? And is there an underlying trend? To start to answer this question:

(1) Create scatter charts of the CO_2 monthly averages for the calendar years 1965, 1975, 1985 (individually), versus the month from Column `B` or the decimal year from Column `C`. This

will require picking out the correct range of rows in Columns B or C, and E for each of these years.

(2) Looking at these scatter plots, what do you notice about the way CO_2 levels vary over these years? Describe what happens over the course of a typical year, and hypothesize a reason why the annual pattern works this way. Note: Mauna Loa is in the Northern Hemisphere and typical mixing patterns in the atmosphere mean that most of the air that passes over this location has come from other areas in the Northern Hemisphere. What happens through the months of May, June, July, August in the Northern Hemisphere, and how might that affect atmospheric CO_2 levels?

(3) How might you model the yearly variation of the CO_2 levels? Suggest mathematical function(s) that might be useful and how you might apply them.

(B) Condensing the Data to a More Manageable Form. Our goal is to model how atmospheric CO_2 levels have been changing over this period (but on the year-to-year level, not on the much more variable month-to-month level). This will be much more manageable if we identify some way to compute a "summary value" for each year to use as the representative CO_2 level for that year.

(1) Identify (at least) three different ways that might be used to produce that sort of "summary value" and describe why they would be suitable.

(2) Choose one of your proposed ways to do this and give a reason for why you think that will be the most reasonable way to "condense" the data for each year.

(3) Create new columns in your spreadsheet giving the number of years since 1959, and your summary CO_2 value for the year. Since we don't have complete values for 1958, start with 1959. Also include only the most recent year for which you have data for the full year, namely 2017.

(C) "Let the modeling begin!"

(1) Fit a linear model to your "condensed" data set and record your results. Give the equation of the regression line as a function of the years since 1959 (that is, $x = 0$ correspond to 1959, $x = 1$ corresponds to 1960, and so forth). Also give the R^2 value as a measure of goodness of the fit. Note: The way we discussed of adding the equation of the regression line to the scatter plot only gives a small number of decimal places in the computed slope and intercept. To see more precise values, you can use the SLOPE and INTERCEPT functions in Google Sheets.

(2) The *residuals* for a linear model are the differences between the data values and the predicted values from the model. That is, if (x_i, y_i) are the data values, and the computed regression line is

$$y = \widehat{m}x + \widehat{b},$$

then the residual for the ith data point is

$$y_i - (\widehat{m}x_i + \widehat{b}).$$

Compute those and generate a scatter plot showing the residual as a function of the year. Is there a consistent pattern there? What does this indicate about the suitability of a linear model for this data? (You will probably want to consult the discussion of the residuals in the text for this.)

(3) What does your model predict concerning the CO_2 level in 2020? (This is slightly outside the interval 1959 to 2017 of course, but not too far outside. So extrapolation from the linear model is at least a possibility.)

(D) Atmospheric CO_2 levels are of concern, of course, because of the "greenhouse gas" properties of this compound—the way atmospheric CO_2 can trap energy from reflected solar radiation and increase temperatures near the surface of the Earth. The greenhouse effect is necessary for life on Earth as we know it, of course (life as we know it could not exist at the temperatures that would prevail with no greenhouse effect at all because all water would be frozen as ice). But have there been times in the past when CO_2 concentrations were significantly higher than they are now? What were the Earth's climate and sea levels like then? What do we know about carbon dioxide levels over the last 1000 years or so and how do scientists measure that? (This may require some research. Be sure to give the sources you used to compile your information.)

Assignment

Submit your edited spreadsheet with the data and write up answers to the questions above in a separate document.