Assignment 3: Linear Regression

In this assignment you will be conducting a multiple linear regression on the Atlantic Multidecadal Oscillation (AMO) index. The AMO is the first mode of variability in the North Atlantic Ocean and exhibits a multidecadal oscillation with a period of 60-80 years. The data for this assignment is similar to the AMO index used in Assignment 1 but this time you will be using unaltered data that has not been detrended. The data can be found in the class data folder `/ulteosrv1/s0/meteo515/data` and is called `AMO_nodetrend.txt`. This data was originally obtained from the ESRL at `https://www.esrl.noaa.gov/psd/data/timeseries/AMO/`.

You may complete the assignment using your programming language of choice. Feel free to use built in functions but make sure you have read the documentation about these functions and are confident they are indeed conducting the calculations you intend. Please submit your assignment, which should be a document that includes your figures and explanations, and the code used to generate any results by uploading the files to Canvas by the assignment due date. Your code should be well commented so that others can easily understand what has been done and marks may be removed from your assignment if this is not the case.

1. In this first question, you will conduct a simple linear regression on the annual average AMO index with time as the predictor (time trend) over the entire time period (omit 2018 for this assignment as the data is incomplete).

- Create a plot of the original data with the regression.
- Include in your report the full regression equation with the estimated coefficients and their p-values.
- Include in your report two measures of goodness of fit of your regression, MSE and R^2.
- Plot the residuals and discuss what they tell you about your regression. Note: If you were to plot the data minus this simple linear regression trend you should obtain the same data that you used in Assignment 1.

2. We will now add an additional level of complexity to our problem by adding a second predictor that will represent the oscillation of the AMO. To do this you will need to choose an appropriate oscillation period (it can range within 60-80 years) and a phase. Once you are happy with your new predictor variable conduct the same analysis as above.

- Create a plot of the original data with the regression.
- Include in your report the full regression equation with the estimated coefficients and their p-values.
- Include in your report two measures of goodness of fit of your regression, MSE and R^2.
- Plot the residuals and discuss what they tell you about your regression and how it compares to the previous regression.

3. In the initial calculation of the simple linear trend, the first predictor was a trend beginning in 1856, however there is evidence that the trend in global mean SST does not begin until around 1900. In this regression your first predictor will be a vector that will allow you to conduct a linear trend beginning in the year 1900 and the second will be the same predictor variable in question 2 that represents the AMO oscillation.

- Create a plot of the original data with the regression.
- Include in your report the full regression equation with the estimated coefficients and their p-values.
- Include in your report two measures of goodness of fit of your regression, MSE and R^2.
- Plot the residuals and discuss what they tell you about your regression and how it compares to the previous two regressions.

4. For the regression in question 3, produce prediction intervals for the predictand y and confidence intervals for the regression.

- Plot the original data, the regression, the prediction intervals, and the confidence intervals.