

HW4 (due on May/1 12:30PM)

Question 1: Use example1.m and load ding-data1.txt (time, var1, var2, var3 &var4) (30 points)

1. Plot all variables together (5 points)
2. Use fitlm (see example.m) to regress var4 on var1, var2 and var3 ($\text{var4}_{\text{pred}} = a + b \cdot \text{var1} + c \cdot \text{var2} + d \cdot \text{var3}$) and plot the regressed $\text{var4}_{\text{pred}}$ with var4. Based on the output of fitlm, discuss which variable is the most important in the regression model. (10 points)
3. Calculate the correlation between var4 and the regressed $\text{var4}_{\text{pred}}$ obtained above (5 points)
4. Use the data kernel method to redo the regression and compare it with the results derived from fitlm (10 points)

Question2. Load ding-data2.txt (year, X, Y) (50 points)

1. Plot X and Y (vs time), build a linear regression model ($Y = a + bX$) and calculate the linear trends of X and Y, respectively (in unit */decade) (15 points)
2. Remove the linear trends from X and Y, respectively and plot the residual parts, and calculate the correlation between the two residual time series. Build a linear regression model again for the residual X and Y ($Y_{\text{residual}} = a + bX_{\text{residual}}$). Compare this new regression with the one obtained from 1, and explain how the coefficients (a and b) vary. (15 points)
3. Repeat 1 and 2 but using the regressed quadratic trend. Do you think which one (the linear or quadratic trend) is more reasonable to capture the long term change in X and Y. (20 points)

Question 3: Use example2.m (20 points)

1. Use the example2.m to plot the sum of first 10 Sine terms to reconstruct the square wave. (10 points)
2. Modify example2.m to plot the sum of first 100 Sine terms illustrating 5 full periods of the oscillation of the square wave. (the total time points are still equal to 101). (10 points)