**Lab 3: Using MATLAB® to Import Data and Curve Fit**

Lab 3 is an introduction to importing data from Microsoft Excel, using MATLAB**®** to plot data, and generating a curve of best fit. ***Insert your answers to any questions, your MATLAB*® *code for generating plots, and the plots you create into this document as your lab report.***

This lab is based off of data provided from a United Nations Database website (<http://data.un.org/Explorer.aspx>). The data covers a wide variety of topics including water usage, population data, energy production, etc. Four Excel Files are included in the recitation folder which has the necessary data for this lab (PopulationData.xlsx, USdata\_SolarProduction.xlsx, USdata\_WindProduction, and USdata\_TotalElectricity). Please upload this data from blackboard into your MATLAB**®** folder where you plan to generate the script file.

**Part 1) Plotting the Electricity Production in the United States**

Look at the Excel data file: UNdata\_TotalElectricity.xlsx in Microsoft Excel just to see what columns are in the file.

***Step 1. Import Raw Data:*** Import the Year and Quantity column of data for the United States into MATLAB**®** using the *xlsread* command at the beginning of a script file **(see slide 20 in the LEC3\_CurveFitting\_VideoSlides which is located in the Lecture 3: Pre-Work folder on the MetaSite)**. Give each vector a name (Year and Quantity for example). Confirm that the data was imported by looking in the workspace after you run the xlsread commands.

Note: The data is listed by Year meaning there is a vector from 1991:1:2012. However, it is easier to look at 1991 as year = 0, 1992 as year = 1. This can be done by including the following line in your code: Year = Year-1991. This will better condition your fitted curves and must be done to properly answer the questions at the end of Part 1.

***Step 2. Plot Raw Data:*** Plot Year on the abscissa and Quantity on the ordinate using red \* and make sure to include axis labels. For the ordinate, include the units for the Quantity plotted (the units are shown in the Excel File).

***Step 3. Curve Fitting:*** Add code before your plot command to fit the appropriate 1st order polynomial equation (trend line) to the Quantity data vs Year using polyfit and polyval MATLAB**®** commands. **(See slides 12 & 13 in the LEC3\_CurveFitting\_VideoSlides file)**.

***Step 4. Plot Trend Line:*** Update your plot commands to include the curve fitted data on the same plot as the raw data. The curve fit should be a solid line. Your plot should look similar to the plot on slide 6 of LEC3\_CurveFitting\_VideoSlides except you will only have the linear fit (1st order polynomial) and the equations for the curve fit will not be included.

***Step 5. Residuals:***

1. Add code after your curve fitting commands to calculate the residuals for all data points, which is the difference between the original data and curve fitted values.

2. Update your plot command to use the subplot command to put your original graph from step 3 on the top and a bar chart of the year vs residuals on the bottom in the same figure (see **LEC3\_CurveFitting** slide 6 to see sample). Make sure you are using MATLAB**®** commands to generate the subplot, do not click buttons in the Figure window. You can check your answers by clicking the Basic Fitting option under the Tools menu, but ultimately you want to generate a script file which does all the work for you.

****

%Total Electricity

clear; clc; clf;

%Input

t = xlsread('UNdata\_TotalElectricity','C2:C23');

t = t-1991;

q = xlsread('UNdata\_TotalElectricity','E24:E45');

%Analysis

poly1 = polyfit(t,q,1);

poly1\_vals = polyval(poly1,t);

r = q - poly1\_vals;

%Output

figure(1)

subplot(2,1,1)

plot(t,q,'r\*',t,poly1\_vals,'k-');

xlabel('\bfYears since 1991','FontSize',14);

ylabel('\bfQuantity (kWh million)','FontSize',14);

title('\bfElectricity Production in the US','FontSize',20);

subplot(2,1,2)

bar(r);

**Include the following information below.**

**Curve fit Equation (refer to the variable poly1st on slide 13 to understand how the vectors in poly1st are related to the curve-fit equation):** y= 54510.7250141164x + 3376832.93280632

**Question 1: What are the units for the slope in your 1st order fitted polynomial?**

Units of slope: kWh/yr

**Part 2a) Plotting and curve fitting the Wind Production for the United States**

Look at the Excel data file: USdata\_WindProduction.xlsx in Microsoft Excel just to see what columns are in the file.

Repeat ***Steps 1 to 5*** from Part 1 but do this for the Wind Production data. The one difference is that now you should fit a 3rd order polynomial to this data. (**see** **LEC3\_CurveFitting, slide 13 & 14** for guidance on fitting a higher-order polynomial to data)

**Part 2b) Plotting and curve fitting the Solar Production for the United States**

Use the Excel data file: USdata\_SolarProduction.xlsx to generate a new plot which shows the Wind Production data from Part 2a and the Solar Production data for the United States on the same plot. You should fit a 3rd Order Polynomial to both sets of data. The following line of code will allow you to plot the residuals for Wind and Solar Production on the same plot, where Residuals\_wind and Residuals\_solar should be replaced with the variable you have used to store the residual calculations:

bar(Year,[Residuals\_wind Residuals\_solar]);

****

%Wind and Solar Production

clear;

%Input

t = xlsread('USdata\_WindProduction','C2:C23');

t = t-1991;

q\_wind = xlsread('USdata\_WindProduction','E2:E23');

q\_solar = xlsread('USdata\_SolarProduction','E2:E23');

%Analysis

poly3\_wind = polyfit(t,q\_wind,3);

poly3\_wind\_vals = polyval(poly3\_wind,t);

residuals\_wind = q\_wind - poly3\_wind\_vals;

poly3\_solar = polyfit(t,q\_solar,3);

poly3\_solar\_vals = polyval(poly3\_solar,t);

residuals\_solar = q\_solar - poly3\_solar\_vals;

%Output

figure(2)

subplot(2,1,1)

plot(t,q\_wind,'r\*',t,q\_solar,'b\*'); hold on;

xlabel('\bfYears since 1991','FontSize',14);

ylabel('\bfQuantity (kWh million)','FontSize',14);

title('\bfWind and Solar Production in the US','FontSize',20);

legend('Yearly wind measurements','Yearly solar measurements');

plot(t,poly3\_wind\_vals,'k-',t,poly3\_solar\_vals,'g-');

subplot(2,1,2)

bar(t,[residuals\_wind residuals\_solar]);

**Include the following information below.**

**Curve fit Equation for Wind Production:**

y = 44.7866444768048(x^3) - 800.514761211334(x^2) + 3952.28576777919x - 301.716996047575

**Curve fit Equation for Solar Production:**

y = 4.37757090354346(x^3) - 102.237145234629(x^2) + 611.839303297776x + 181.443083003936

**Part 3a) Plotting and comparing data between multiple countries**

Next, use the UNdata\_TotalElectricityProduction.xlsx and repeat ***Steps 1 to 4*** from Part 1 but do this for the United States and China. Thus, you should have the data for both countries on the same plot. You should also fit a 3rd order polynomial to each dataset. It is not necessary to plot the residuals for this case.

**Paste your script file and plot below (You do not need to include the residual plot):**

 %Total %Total Electricity for US and China

clear;

%Input

t = xlsread('UNdata\_TotalElectricity','C2:C23');

t = t-1991;

q\_china = xlsread('UNdata\_TotalElectricity','E2:E23');

q\_us = xlsread('UNdata\_TotalElectricity','E24:E45');

%Analysis

poly3\_china = polyfit(t,q\_china,3);

poly3\_china\_vals = polyval(poly3\_china,t);

poly3\_us = polyfit(t,q\_us,3);

poly3\_us\_vals = polyval(poly3\_us,t);

%Output

figure(3)

plot(t,q\_china,'r\*',t,q\_us,'b\*'); hold on;

legend('China yearly measurements','US yearly measurements');

plot(t,poly3\_china\_vals,'k-',t,poly3\_us\_vals,'g-');

xlabel('\bfYears since 1991','FontSize',14);

ylabel('\bfQuantity (kWh million)','FontSize',14);

title('\bfElectricity Production in China and the US','FontSize',20);

**Include the following information below.**

**Curve fit Equation for China:**

y = 193.847784688993(x^3) + 5281.85491781338(x^2) + 9936.07041663877x + 773156.793276681

**Curve fit Equation for US:**

y = -118.662284168924(x^3) + 1319.62070343856(x^2) + 74619.5907619829x + 3254902.29683795

**Part 3b) Taking population into account when plotting data (per capita)**

One challenge when comparing data between countries is the fact that they all have different populations. Thus the production data could be divided by the total population in each country to determine if individuals are using more electricity. First, import the population data for both countries from the PopulationData.xlsx file. Take the Quantity vector for each country from Part 3a and dot-divide it by its population vector. Then plot the data a fit a 3rd order polynomial to each dataset. This will now provide you information about production per capita.

**Paste your script file and plot below (You do not need to keep the residual plot). Watch your units!!!**

****

%Total Electricity for US and China (per capita)

clear;

%Input

t = xlsread('UNdata\_TotalElectricity','C2:C23');

t = t-1991;

q\_china = xlsread('UNdata\_TotalElectricity','E2:E23');

q\_us = xlsread('UNdata\_TotalElectricity','E24:E45');

p\_china = xlsread('PopulationData','D2:D23');

p\_us = xlsread('PopulationData','D24:D45');

%Analysis

q\_china\_percap = q\_china./p\_china;

q\_us\_percap = q\_us./p\_us;

poly3\_china\_percap = polyfit(t,q\_china\_percap,3);

poly3\_china\_percap\_vals = polyval(poly3\_china\_percap,t);

poly3\_us\_percap = polyfit(t,q\_us\_percap,3);

poly3\_us\_percap\_vals = polyval(poly3\_us\_percap,t);

%Output

figure(4)

plot(t,q\_china\_percap,'r\*',t,q\_us\_percap,'b\*'); hold on;

legend('China yearly measurements (per capita)','US yearly measurements (per capita)');

plot(t,poly3\_china\_percap\_vals,'k-',t,poly3\_us\_percap\_vals,'g-');

xlabel('\bfYears since 1991','FontSize',14);

ylabel('\bfQuantity (kWh)','FontSize',14);

title('\bfElectricity Production in China and the US','FontSize',20);

**Include the following information below.**

**Curve fit Equation for China:**

y = 0.000116452105934808(x^3) + 0.00452509413148634(x^2) + 0.00182642621133508x + 0.668514797539332

**Curve fit Equation for US:**

y = 0.668514797539332(x^3) + 0.00344753342633886(x^2) + 0.125323361016681x + 12.8586569866980