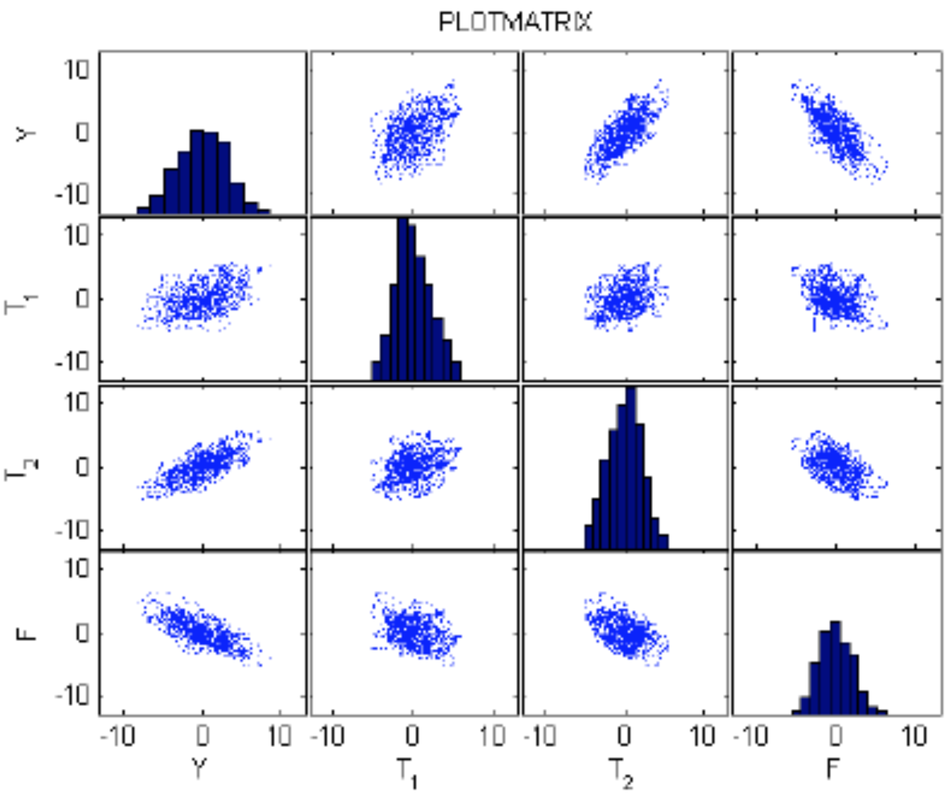
CHE 573- Signal Processing for Process Engineers

Problem-based case study
Use of Correlation Analysis and regression for Softsensor Design

Objectives: To obtain a hands-on experience in designing a simple sofstsensor* using cross-correlation analysis and multiple linear regression.

a. Plot the matrix of data using 'plotmatrix'



The MATLAB code is as follows:

```
clc
clear all
close all
load('Softsense2')
zz=detrend(zz,'constant');
%The handles to the x and y axes of the plot are returned in 'AX'
[H,AX,BigAx,P]=plotmatrix(zz)
ylabel(AX(1,1),'Y');ylabel(AX(2,1),'T_{1}')
ylabel(AX(3,1),'T_{2}');ylabel(AX(4,1),'F')
xlabel(AX(4,1),'Y');xlabel(AX(4,2),'T_{1}')
xlabel(AX(4,3),'T_{2}');xlabel(AX(4,4),'F')
title('PLOTMATRIX')
```

Observations from the plotmatrix:

1-There is significant positive correlation between Y and T2.

2-There is significant negative correlation between Y and F.

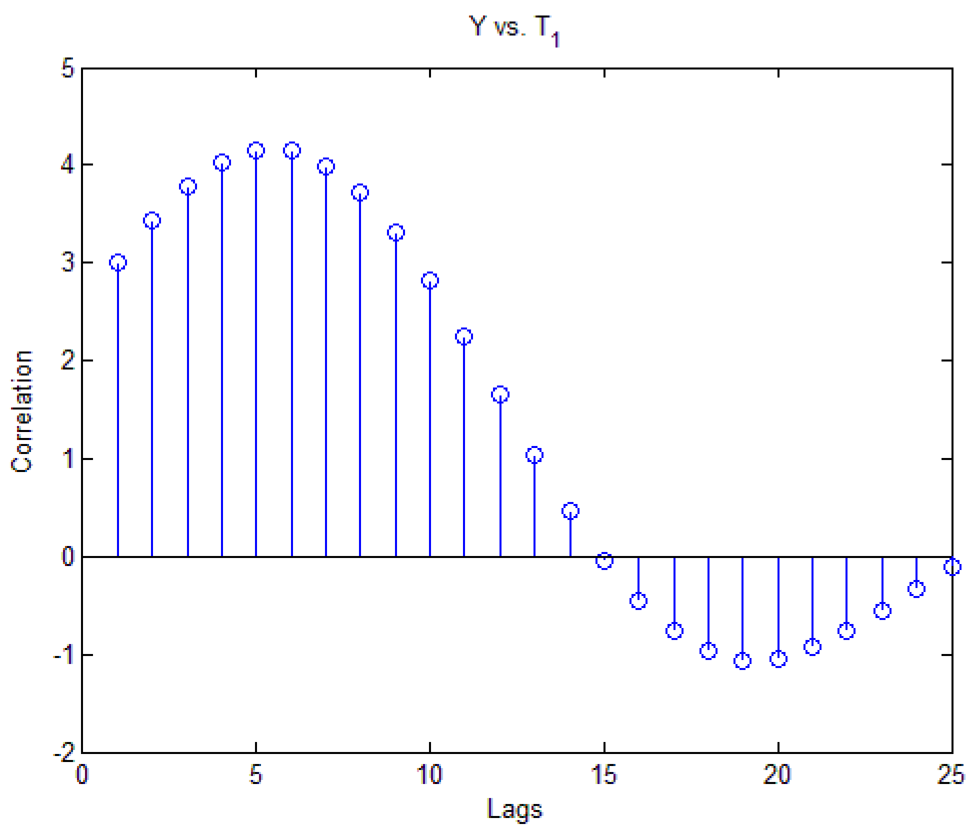
3-The correlation between Y and T1 is smaller in comparison with (Y,T2) and (Y,F) but it cannot be ignored.

In order to investigate the correlation between variables for different lags, the cross-correlation plots should be examined.

b. **<u>Note that the first 750 data points are used to find the appropriate lag.</u>**

b.1. Cross-Correlation between Y and T_1:

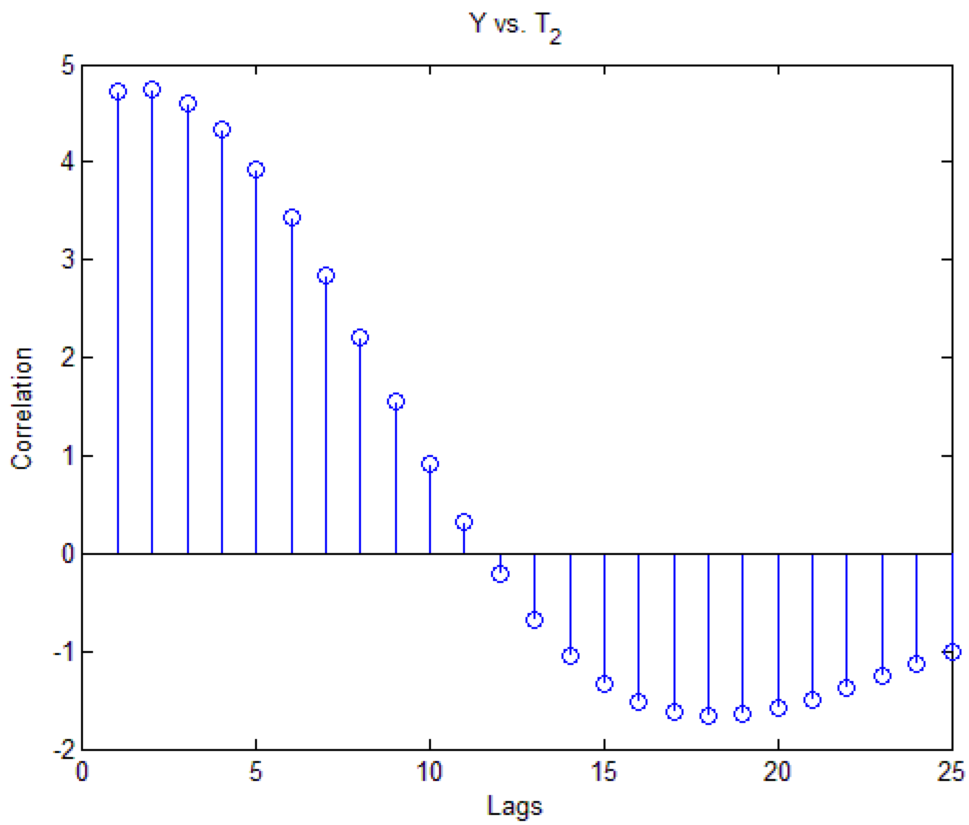
```
[c1,lags1]=xcorr(zz(1:750,1),zz(1:750,2),25,'unbiased');
figure(1)
stem(lags1(26:51),c1(26:51))
title('Y vs. T_{1}')
xlabel('Lags');ylabel('Correlation')
```



As it can be seen the xcorr graph, the maximum correlation happens at lag=5 and hence in design of the softsensor the pair $(Y(k), T_1(k-5))$ should be considered.

b.2. Cross-Correlation between Y and T_2:

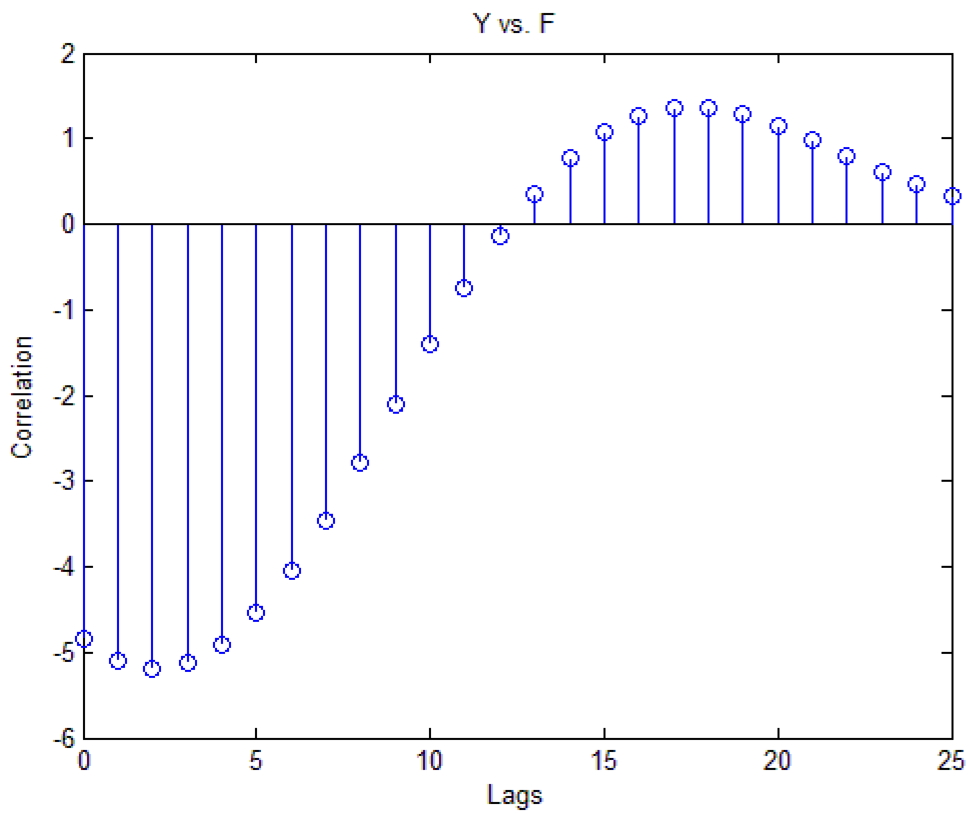
```
[c2,lags2]=xcorr(zz(1:750,1),zz(1:750,3),25,'unbiased');
figure(2)
stem(lags2(26:51),c2(26:51))
title('Y vs. T_{2}')
xlabel('Lags');ylabel('Correlation')
```



The highest correlation occurs at lag=2 and consequently the pair $(Y(k), T_2(k-2))$ will be considered in the design of softsensor.

b.3. Cross-Correlation between Y and F:

```
[c3,lags3]=xcorr(zz(1:750,1),zz(1:750,4),25,'unbiased');
figure(3)
stem(lags3(26:51),c3(26:51))
title('Y vs. F')
xlabel('Lags');ylabel('Correlation')
```



The maximum correlation happens at lag=2 and as a result of this the pair $(Y(k), F(k-2))$ is chosen for the design of softsensor.

Note that negative correlation exits between Y and F.

c. In this part, the data should be split to 2 parts. The first part which includes 2/3 of the total data points will be used for training and the remaining data points will be used for validation.

Using the results from part b, the basic form of the softsensor can be as follows:

$$Y(k) = a + bT_1(k-5) + cT_2(k-2) + dF(k-2)$$

$$\begin{bmatrix} Y(6) \\ Y(7) \\ \vdots \\ Y(n) \end{bmatrix} = \begin{bmatrix} 1 & T_1(1) & T_2(4) & F(4) \\ 1 & T_1(2) & T_2(5) & F(5) \\ \vdots & \vdots & \vdots & \vdots \\ 1 & T_1(n-5) & T_2(n-2) & F(n-2) \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}$$

By means of the linear least squares the parameters can be found. Recall that:

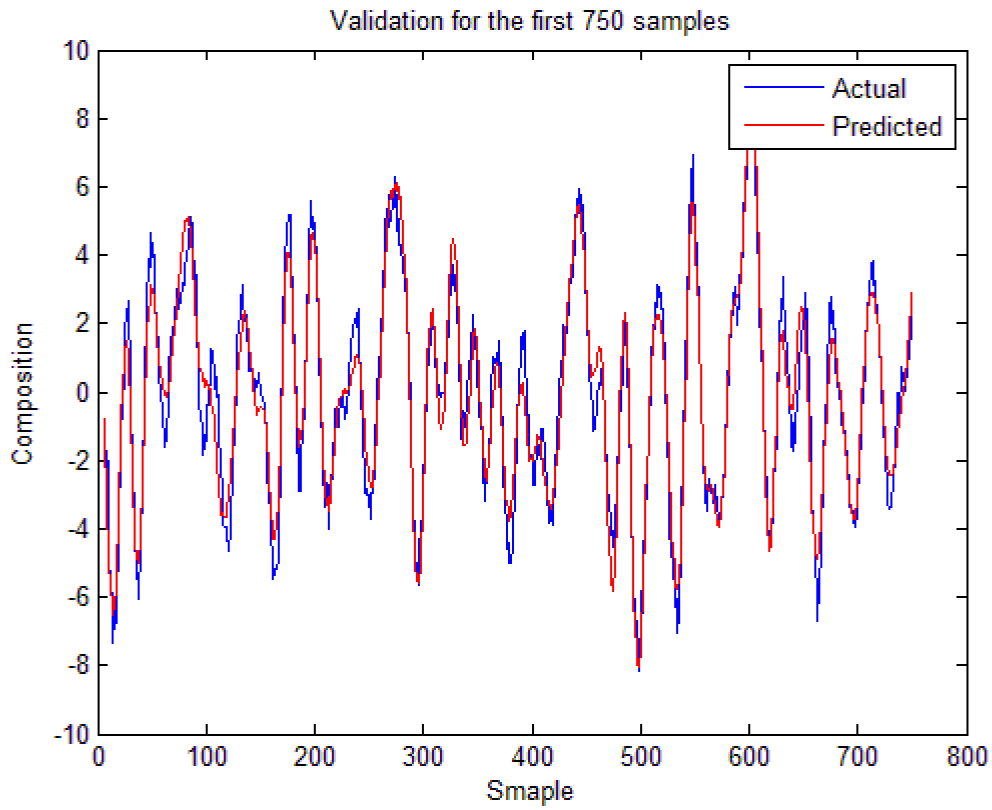
$$Y = X\theta + \epsilon \rightarrow \ \hat{\theta} = (X^T X)^{-1} X^T Y$$

Using this method $\hat{\theta}$ can be found to be:

$$\hat{\theta} = \begin{bmatrix} 0.0113 \\ 0.5726 \\ 0.5485 \\ -0.7052 \end{bmatrix}$$

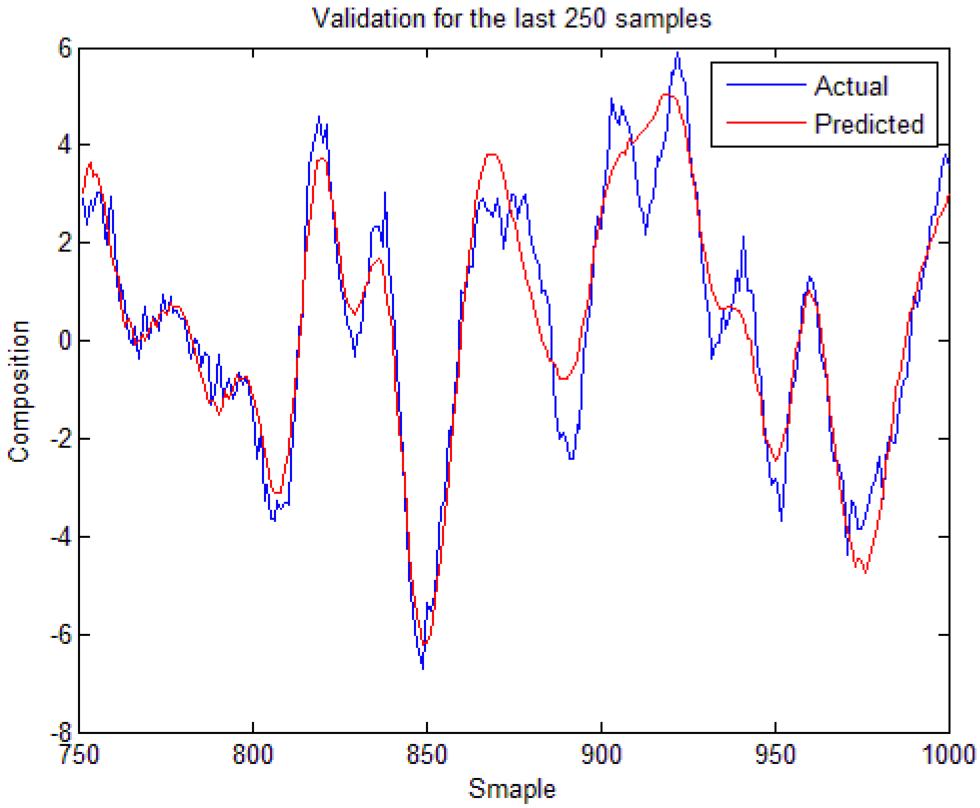
The MATLAB code for designing the softsensor based on the first 750 samples is as follows:

```
%Training data set (first 750 samples)
Y=zz(6:750,1);
X=[ones(size(Y)),zz(1:750-5,2),zz(4:750-2,3),zz(4:750-2,4)];
Theta_hat=inv(X'*X)*X'*Y;
%You can use the alternative:  Theta_hat=(X'*X)\(X'*Y);
Y_hat_training=X*Theta_hat;
plot([6:750],Y,[6:750],Y_hat_training,'r');
legend('Actual','Predicted')
ylabel('Composition')
xlabel('Smaple')
title('Validation for the first 750 samples')
```



The MATLAB code for validating the softsensor using the remaining 250 samples is as follows:

```
%Validation data set
Y_val=zz(751:1000,1);
X_val=[ones(size(Y_val)),zz(751-5:1000-5,2),zz(751-2:1000-2,3),zz(751-2:1000-
2,4)];
Y_hat_validation=X_val*Theta_hat;
plot([751:1000],Y_val,[751:1000],Y_hat_validation,'r');
legend('Actual','Predicted')
ylabel('Composition')
xlabel('Smaple')
title('Validation for the last 250 samples')
```



As it can be seen in the validation graph, the prediction is acceptable.

d.

In order to account for the dynamics of the process so that the softsensor could also work under the dynamic conditions, the adjacent (+ and -ve) lags to the maximum lag as observed in the correlation should be included in the softsensor equation. The following equation can be considered as an example:

$$\begin{aligned} Y(k) = a &+ b_1T_1(k-3) + b_2T_1(k-4) + b_3T_1(k-5) + b_4T_1(k-6) + b_5T_1(k-7) \\ &+ c_1T_2(k-1) + c_2T_2(k-2) + c_3T_2(k-3) + c_4T_2(k-4) + d_1F(k-1) \\ &+ d_2F(k-2) + d_3F(k-3) + d_4F(k-4) \end{aligned}$$