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.

The file 'Softsense' contains simulated data from a simple process. This file can be downloaded from the 'eClass' web-page for CHE 573. The first column of this matrix contains overhead composition data, Y, in a distillation column as available from an analyzer. The second and third columns contain tray temperatures, T_1 and T_2 , for the same column and the last column contains the feed rate. All variables are in normalized units. The analyzer is expensive to maintain as it requires significant maintenance and often malfunctions in which case no composition data is available. You have to do the following:

a) Plot the data and see if there appears to be any correlation between the response variable, Y, and the tray temperatures and the feed flow rate to the column. You can use the 'plotmatrix' command for this.

b) It is to be expected that there will be some time lag between the tray temperature readings and the composition and between the feed flow rate and the composition. You are to investigate this time lag between Y and $T_{1,}$ Y and T_2 and between Y and F. Hint: use 'xcorr' to find out these time lags. Be aware of causality constraints.

c) Using multiple linear regression (MLR) find the relationship between Y and the lagged variables T_1 ,T_2 and F. (For this case consider a simple model, with 'lag-adjusted' variables, such as: $Y = a + bT_1 + cT_2 + dF$, and estimate the coefficients a, b, c and d). Validate this model of the softsensor by plotting the measured composition with respect to time and superimposing the predicted value of the composition using the model that you have developed.

d) Please note that this exercise will serve as a very simplistic illustration about how to develop soft-sensor models. In practice many of the explanatory variables may be correlated in which case one will have to use methods such as 'PLS' or partial least squares. The model that we are considering is strictly a steady-state model and will suffice as an approximate sensor. What modifications would you consider so that it can also serve as a dynamic softsensor?