The objectives for this assignment are:

1. to gain experience in modelling time series data using linear dynamic models,

2. to build some experience using MATLAB's System Identification Toolbox.

You may work together in groups to complete the assignment, but you must hand in your own assignment solution. If you work with a group, please identify the people that you worked with on your solution. Computer printout may be included with your solution as an appendix, but please do not provide these as your entire solution report.

1. On the course web page you will find three data files named 'che572_assn5_q1a_2011.mat', 'che572_assn5_q1b_2011.mat', and 'che572_assn5_q1c_2011.mat'. These files contain time-series data collected for a single process variable y. These data files were produced with the same stochastic model, but contain 51, 201, and 1001 data points, respectively. The model used to produce the data was:

$$y_t = \frac{(1 - 0.9z^{-1})}{(1 - 0.85z^{-1})(1 - z^{-1})} \epsilon_t \ , \qquad \epsilon_t \in N(0, 1)$$

In this question you will investigate the effect of the amount of data and of differencing nonstationary data sets on the quality of the resulting model.

a) For each data set, without differencing the data,
   i. determine the noise model (including estimates of the model parameters),
   ii. determine the accuracy of your model and your parameter estimates.

b) Repeat part (a), but use differenced data this time.

c) Comment on the effect of differencing the data on the predictive accuracy of the model, the estimated accuracy of the parameter estimates (i.e., the covariance of the estimates) and the how close the parameter estimates are to the actual parameter values (i.e., the bias in the parameter estimates). *Be quantitative.*

d) Comment on the effect of the amount of available data on the predictive accuracy of the model, the estimated accuracy of the parameter estimates (i.e., the covariance of the estimates) and the how close the parameter estimates are to the actual parameter values (i.e., the bias in the parameter estimates). *Be quantitative.*

2. On the course web page you will find two data files named 'che572_assn5_ q2a_2011.mat' and 'che572_assn5_q2b_2011.mat'. These files contain data collected from testing done on two different plants. (The second data file was produced by testing one of the stirred tank heaters in CME 274). The input variable is u and the output variable is y. For each data set:

    i. Determine both the process and noise models, and estimate the model parameters,

    ii. Determine the accuracy of your model and your parameter estimates.

    iii. Comment on whether you believe the sampling and switching frequencies were appropriate for the plant tests. *Explain your answer.*