Model the need for cooling in a H2O2 process

Task

This is a non-linear regression task. The process that generated the data is EKA chemicals Hydrogen Peroxide production. Data is available from 1997-06-01 thru midnight 1998-12-31, a total of 19 months'. The interval between each measurement is 10 minutes and some periods of the data is missing e.g., where service have been conducted and some very fast transition. Lastly the data has been decimated to 30 min measurements by averaging over 3 samples.
The output is the valve opening of a valve connected to one (out of two) heat exchangers. This opening is part of a feed-back control loop based on the temperature of the fluid that passes thru the heat exchanger. The goal with the control is to keep the temperature constant and the fluid in liquid format. In cases where a large amount of cooling is required (large opening of the valve) its indicating that the fluid is in transition to gas form and this information is of particular significance.
The goal of the modeling is to construct a model with all or part of the variables available to model the valve opening.

Data

The data supplied in a MATLAB file, ChemTrainNew.mat, consist of three matrices: XtrainDS (4466x65), YtrainDS (4466x1) and XtestDS (2971x65). The input matrix (XtrainDS) contains all variables to the process (the first column is time) and the output matrix (YtrainDS) contain the vale opening. The third matrix (XtestDS) is test data consisting of all inputs from a time period that follow the 19 month training data.

Steps and sub goals

1. Get acquainted with the data, do scatter plots and explore the relations between inputs and output.
2. Estimate the covariance between inputs and output; rank the importance of the variables based on this covariance.
3. Construct a linear model. Select the most important variables by e.g. forward selection, backward elimination or other method. Use cross-validation to estimate the models ability to generalize.
4. Try PCA (principle components analysis) and possibly PLS (partial least squares) to solve the problem.
5. Construct different multi layer perceptions (MLP) for this problem:
 (a) With variables {3, 6, 16, 49, 50, 51, 52, 53, 65} (columns in XtrainDS).
 (b) With variables {4, 10, 19, 27, 28, 38, 42, 50}.
 (c) Use forward selection and backward elimination and try find a model better than the two proposed.
6. Use the models in (a), (b) and (c) above. Make predictions on test data (XtestDS) based on the three models. Evaluate the performance of a committee based on the three models.