Statistical Analysis

After cleanup and first data visualization we want to do some statistical analyis. It would be nice to know if there is a correlation with the run number, i.e. the order number when the pressure curve was recorded. If for example gradual clogging is building up, over time the curves should have a bias which changes with the run number.

Therefore, we extract the data in blocks for one experiment, denoise the data, run a PCA and fit the best linear model with interactions up the the first 5 PCA components to predict the run number. The correlation coefficient is recorded. If there is a high correlation coefficient, it means, the linear model is fitting well and there is a change in the pressure curves with run number, even within a single experiment.

Basically, we use the same methods but encapsulate them into some wrapper functions to easily select flowcell the maximum number of principal components and run it for all experiments and then do a boxplot to show the distributions over experiments for all fluidic paths.

Not to clutter the report with the code, we use the option to run the code only but not to show in the report. This would be a report type for a manager...

Boxplot of correlation of individual linear models per experiment
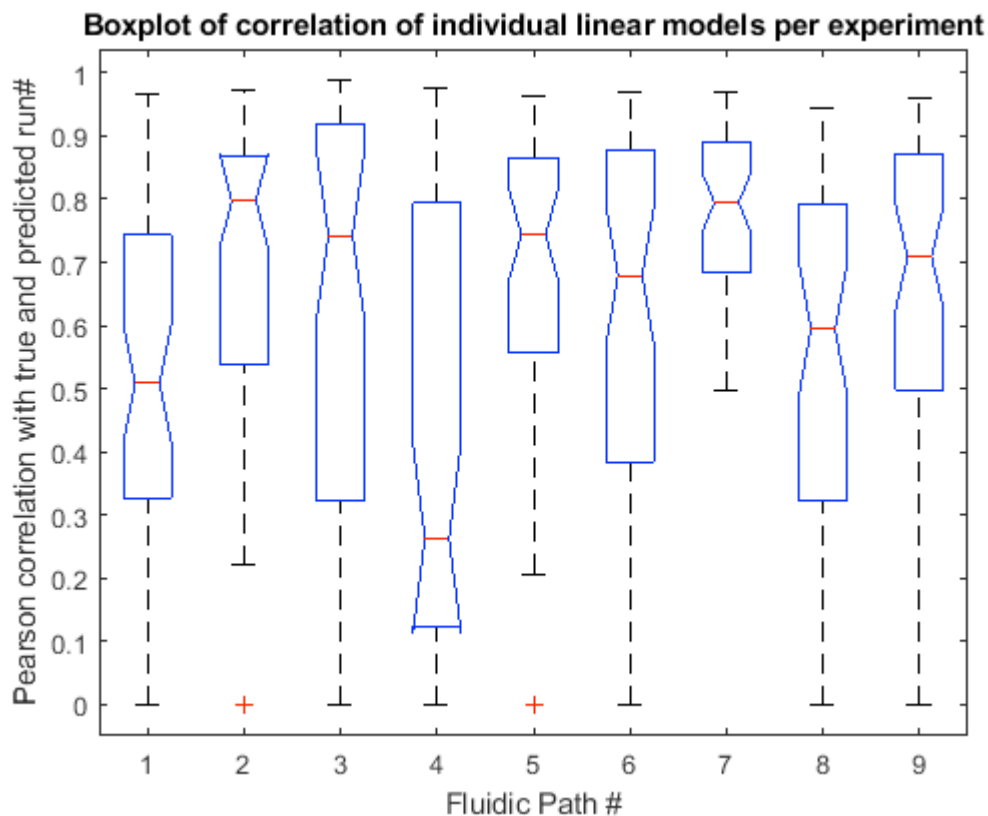
We see on average quite a strong correlation, albeit in some experiments no correlation can be detected. However, in fluidic paths 5 and 7 the median is fairly high and confidence intervals are fairly tight. The chemist has confirmed that

liquids in fluidc paths 5 and 7 have a hgher viscosity than the other fluidc paths. Therefore, we see much more relevant change with run number (time) even during an experiment. It was believed that during an experiment, no clogging effects would occur, but the data analysis clearly shows some effects that subsequent runs give rise in changes in pressure curves.

Many more experiments and with further contextual information may be conducted. This will enable root cause analysis, predictive maintenance algorithms which may trigger cleaing cycles to prevent fluidic paths from being clogged.

Published with MATLAB® R2018a