Automated modelling and updating of models to adjust for changes in raw materials and/or product specification

Professor Olav Kvalheim and colleagues, University of Bergen

Professor Kvalheim summarised research in this field in a talk at a recent CPACT research day entitled “Optimal partial least squares models and predictions of properties using Monte Carlo resampling strategies”. This topic is important in process control especially because for some industries there can be variation in the process feed induced by either raw material variation in continuous processes or by gradual change in feed in batch processes. In these circumstances, it is helpful if automated modelling and updating of models can be implemented to adjust for changes in raw materials and/or product specification.

A recent project involved the production of omega-3 concentrates from fish oil where there can be large variation in the raw materials and large variation in the product specifications in what is a continuous manufacturing process. The aim of the project was to develop a method and prototype that can automatically select the right samples to be used in modelling and determine model dimension, and update the PLS model. The focus of the research was in devising a significant test for model selection. The test developed for automatic and reliable model selection in PLS progression involved a single loop Monte Carlo resampling algorithm to provide a nonparametric test.

Further developments have led to a comparison of single and double loop Monte Carlo resampling algorithms. The outcome of this study concluded that single loop resampling can provide the same prediction of error as the double loop procedure. The resampling methodology was applied to NMR spectra from a database of more than 1000 samples. From the results for 8 PLS models differing in the number of samples and the y-variable, it was shown that prediction estimates from independent sampling tools were statistically similar in 75% of cases. A small, but statistically significant increase in prediction error was observed for 25% of the cases.

Two nonparametric significance tests have been successfully developed from single loop Monte Carlo resampling to allow automated model selection. An important feature is that the procedure provides estimates with confidence intervals. These estimates can be used to assess the reliability and stability of the results and the resulting predictive properties can be used to predict optimal process settings, which means the process system can then be operated with minimal or no operator interaction.