320b Multivariate Approaches for the Diagnosis of a Batch Chemical Process
Manuel Zarzo and Alberto Ferrer

In batch processes the quality of the final product is difficult to control. Quite often the critical points of the process that require a more accurate control to avoid batches out of specifications remain unknown. This problem occurs in an industrial process that elaborates the polymer polypropylene oxide. This batch process consists of 4 stages and several substages, and 52 process variables (temperatures, pressures, flows, pH, etc.) are recorded on line every minute. One of the quality parameters analyzed in laboratory in the final polymer is the hydroxyl index (I_{OH}). The problem is that this parameter is out of specifications in about 15% of the produced batches. Data from 69 historical batches of this process have been analyzed in order to diagnose the problems.

As the duration of the different stages is not constant from batch to batch, several alignment methods have been used to synchronize the trajectories of the 52 process variables. This procedure produces a tree-way data structure (batches x process variables x time) that has been unfolded according to the methodology proposed by Nomikos and MacGregor (1995), resulting a large matrix with 9100 variables comprised by 52 aligned trajectories (block of variables formed by the evolution of the corresponding process variable in a scale of pseudo-time for the set of 69 batches). In order to get rid of the main non-linearities, data have been mean centered. To give "a priori" equal weight to all the process variables, data have been scaled to unit variance.

For the diagnosis it is necessary to identify those variables that in a certain part of the process are correlated with the hydroxyl index. Considering it as the response variable, a PLS (Partial Least Squares Regression) can be directly applied to the unfolded matrix, resulting that the first component is statistically significant, with a goodness of prediction by crossvalidation (Q^2) about 33%. In this case the loadings are proportional to the linear correlation coefficient. So, high loadings in absolute value correspond to process variables that in certain instants of time are significantly correlated with the hydroxyl index. But the diagnosis becomes difficult from the analysis of the loading plot, since loadings are scattered and there are no outstanding trajectories with specially high values.

Another approach used is based on the application of PLS for the different blocks of variables, following the idea of Zarzo and Ferrer (2004), and has been called "blockwise PLS". With every trajectory, a PLS has been carried out considering the I_{OH} as response variable; for the first component, the associated latent variable and its goodness of prediction by cross-validation (Q^2) has been calculated. These latent variables contain the projections (scores) of the batches in the directions determined by the PLS components. This method produces a new matrix of 52 latent variables that contains the main information of the original matrix in order to predict the final quality. A further analysis of this matrix will reveal useful information: stages more correlated with the final quality, outliers, shifts in the process, etc. But in this case just the analysis of the 52 Q^2 values highlights the most important information. If the 52 values of Q^2 are charted on a normal probability plot, a linear trend is observed, but the highest 7 values are slightly separated from the straight line, highlighting those trajectories most important from a statistical point of view. This result simplifies the diagnosis. The pressure during the second stage (2PR), the temperature during the first stage (1Tª), the derivative trajectory of this temperature (1Tªd) and the derivative trajectory of pressure (1PRd) are the latent variables with highest Q^2 values.

Regarding the latent variable of pressure (2PR), the highest loadings correspond to the beginning of the second stage. During a period of about 20 minutes, the pressure of the batches with highest hydroxyl index (out of specifications) describes a trajectory than tends to be lower than the mean trajectory. And the opposite occurs for batches with lowest I_{OH}. Thus, during the beginning of the second stage, the

pressure has a negative correlation with the final quality. If a further PLS is conducted with the variables of temperature during this period, the goodness of prediction (Q^2) becomes about 0.6 and this multivariate model could be used for on-line monitoring. But furthermore, these results reveal information regarding the diagnosis: as the correlation appears since the first minutes of the second stage, it seems that the first stage is critical.

Regarding the temperature during the first stage, the highest loadings of this latent variable correspond to the final period of the addition of polyalcohol and during the addition of alkaline solution. Actually, if the original trajectories are observed, during these periods the batches with highest hydroxyl index tend to take a temperature lower than average. But should the trajectories of temperature be more accurately controlled to minimize the deviation from the target trajectory, in order to achieve a reduction in the final quality values? This will be true if this observed correlation between temperature and I_{OH} is due to a cause-effect relationship, but there might be other causes.

One approach to try to identify causal correlation is to check for consistent correlation among different sets of batches. This analysis has been denominated "consistency study". The set of 69 batches can be split in two parts: 38 batches produced in one period and the remaining 31 produced in another one. Two PLS models have been fitted, one for each period, using in both cases variables from the first stage. The variable loading from both models, corresponding to the first PLS component, have been plotted in the same chart to compare the values. The chart reveals that the temperature during the first stage is consistent, since the loadings take similar high values. But it has also been identified that for the second set of batches, the level of the tank during the first stage is highly correlated with the hydroxyl index, and this correlation appears from the first minutes of the addition of polyalcohol, just at the beginning of the stage.

So, probably the problem is the flowmeter that controls the addition of polyalcohol. It is a Coriolis type flowmeter supposed to perform with high accuracy, and it is known that the main error in the measurement is produced when the flow is too low. This happens at the end of the addition, when the total mass added is about to reach the setpoint: the valve that controls the addition is partially closed, and the flow is drastically reduced for some minutes. In order to determine if this is the problem, two variables have been calculated: the flow and duration of this final filling period. Both have certain correlation with the hydroxyl index. In order to obtain further information, a CUSUM chart has been obtained with this flow, another chart for this duration, and a third one for the hydroxyl index. The comparison of these charts reveal a shift that is detected in the 3 charts at about the same batch.

Although from a statistical point of view it is not possible to identify causal correlation from observational predictive models and a design of experiments should be conducted to finally validate the hypothesis, different approaches seem to give enough evidence that an excessive variability in the mass of polyalcohol used as reagent, due to an error in the flowmeter, is the main cause of the variability of the hydroxyl index. This hypothesis is also supported with stoichiometric calculations that highlight that a small variation of the mass used as reagent has a high impact in the final hydroxyl index. So, the advice is to improve the control of this flowmeter in order to achieve a more accurate measure of the mass added to the tank.

REFERENCES

Nomikos, P.; MacGregor, J.F. (1995). Multivariate SPC charts for monitoring batch processes. Technometrics 37: 41-59.

Zarzo, M.; Ferrer, A. (2004). Batch process diagnosis: PLS with variable selection versus block-wise PCR. Chemometrics and Intelligent Laboratory Systems 73: 15-27.