

A new algorithm for selecting the unfolding method and the number of sub-models in batch process modelling with PCA

There is a wide range of approaches for modelling batch process data with Principal Component Analysis (PCA). All these approaches are based on following a number of fixed steps. To mention some methods: Nomikos and MacGregor (1994) propose to unfold the data in the batch direction and calibrate a PCA model; Wold *et al.* (1997) propose to unfold variable-wise the data, calibrate a PLS model to auto-align the data, rearrange the scores batch-wise and then model with PCA; Ramaker *et al.* (2005) propose to calibrate as many PCA models as number of sampling times in the process.

One of the most critical decisions is how to arrange the three-way data in two dimensions: from batch-wise unfolding to variable-wise unfolding and from a single data matrix –and thus a single PCA model- to multiple data matrices –multiple PCA models. It is well known that different modelling approaches are able to capture different features of a process. For instance, if there are several different phases in a process, calibrating multiple models may be convenient. Therefore, rather than using the same modelling approach always, evaluating which arrangement of the data is appropriate for a specific process may be more advantageous. No general mechanism to evaluate and compare PCA models calibrated from different arrangements of the data has been defined yet. This is the aim of this poster.

The approach presented here is based on a cross-validation algorithm recently proposed by Camacho *et al.* (2007). The sum of squares of prediction error (PRESS) associated to a model is computed in such a way that models fitted after different unfolding procedures (batch-wise, variable-wise, batch dynamic), with different number of sub-models (single model and multi-models) and number of PCs can be directly compared. Thus, just like it is done to select the number of significant PCs in two-way, the PRESS curve associated to the data of a batch process can be used to select the appropriate unfolding method, number of sub-models and number of PCs of all the submodels. Data from several batch processes of very different nature (a polymerization process, a fermentation process and a waste-water treatment process) will be used to show the performance of the new algorithm. Also, some guidelines for improving its efficiency are pointed out. Since the PRESS curves associated to the three processes present different shape, the conclusion is that the best arrangement of the data to fit the PCA model/s is very dependant on the current process. Thus, this arrangement can be understood as a parameter which has to be calibrated together with the PCA model. This, in turn, makes clear the need of an evaluation algorithm such as the one proposed here.

1. P. Nomikos, J. MacGregor, Monitoring batch processes using multiway principal components analysis, *AIChE Journal* 40 (1994) 1361-1375.
2. S. Wold, N. Kettaneh, H. Friden, A. Holmberg, Modelling and diagnostics of batch processes and analogous kinetic experiments, *Chemometrics and Intelligent Laboratory Systems* 44 (1998) 331-340.
3. H. Ramaker, E.N.M. van Sprang, J.A. Westerhuis and A.K. Smilde, Fault Detection properties of global, local and time evolving models for batch process monitoring. *Journal of Process Control* 15 (2005) 799-805.
4. Camacho, J., Picó, J. and Ferrer A. (2007), “Cross-validatory Identification of the Number of Components in Principal Components Analysis”, submitted to *Technometrics*.