KV_C045
Villez, K., Habermacher, J. (2015). Qualitative trend analysis for fault detection in batch processes. Proceedings of the 9th IWA Symposium on Systems Analysis and Integrated Assessment (Watermatex2015), Gold Coast, Queensland, Australia, June 14-17, 2015, appeared on-line.

Watermatex 2015
9th IWA Symposium on Systems Analysis and Integrated Assessment
11-14 June 2015, Gold Coast, Australia

Qualitative trend analysis for fault detection in batch processes

Villez, K.* and Habermacher, J.*
* Eawag: Swiss Federal Institute of Aquatic Science and Technology, Duebendorf, Switzerland

Keywords: fault detection, qualitative trend analysis, process monitoring, sequencing batch reactor, side stream reactor

Summary of key findings

A new method for fault detection in biological sequencing batch reactors is developed on the basis of shape constrained spline function fitting. The proposed method is based on the optimal fitting of a shape constrained spline function in the presence of discontinuities of its derivatives. The method is favorably compared to a conventional use of principal component analysis (PCA) thanks to *(i)* its intuitive interpretation and *(ii)* an improved fault detection performance.

Background and relevance

Effective fault detection in wastewater treatment systems remains an elusive goal today due to many factors complicating its execution. These include the fact that *(i)* involved processes are poorly understood, *(ii)* biological process are subject to stochastic variations which are hard to account for, and *(iii)* representative data of confirmed fault events are hard to come by. Faced with these challenges, many researchers have chosen statistical models, such as Principal Component Analysis (Rosén & Lennox, 2001) and derivative forms (Aguado *et al.*, 2007; Villez *et al.*, 2008). While often successful, application of such models requires frequent updating to account for process changes. This is difficult in practice because of lack of training in statistical modelling or associated time requirements. In addition, conventional statistical models are hard to interpret. Therefore, there is a need for fault detection methods which deliver high chances of fault detection while remaining straightforward to interpret.

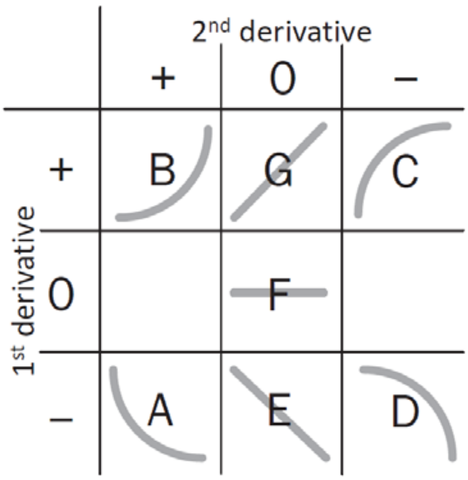

Figure 1. Overview of commonly applied primitives and associated characters.

The proposed method is an extension of the shape constrained splines (SCS) method for qualitative trend analysis developed in Villez *et al.* (2013) for batch fault diagnosis. Basically, a spline function is optimized in the least-squares sense while satisfying a given shape for this function. This shape is defined as a sequence of episodes which are characterized by combinations for the signs of the first and second derivatives (Figure 1). A deterministic optimization algorithm is given in Villez *et al.* (2013). This algorithm exhibits a number of limitations. One of those limitations is that the fitted spline function cannot be discontinuous in the derivatives considered for the definition of shapes. Practically, this means that the permitted shapes are restricted. The SCS method was therefore extended to permit such discontinuities while retaining the global deterministic optimality of the optimization algorithm (Villez, 2014).

Results and discussion

Figure 2 shows a typical time series for the oxidation-reduction potential in the main line sequencing batch reactor (SBR) of an experimental side-stream reactor setup at Eawag. The first 513 data points are shown which correspond to the first 85 minutes of the 6-hour batch cycle. It is relatively easy to see that this time series exhibits a shape which can roughly be characterized by three consecutive episodes with *(i)* a decreasing linear trend (E), *(ii)* a decreasing convex trend (A), and *(iii)* an increasing concave trend (C). These episodes correspond to the following first three operational stages of this batch process' recipe: *(i)* pumping of sludge between the SSR and the SBR (7 min.), *(ii)* addition of fresh wastewater (10 min.) under anoxic conditions, and *(iii)* aerobic conditions in the SBR (total: 285 min.). The EAC shape as recognized by the human eye is considered the default shape corresponding to normal operation of the SBR. Both the EA and AC transitions exhibit a discontinuity in the first and second derivative which is why the new SCS method is necessary.

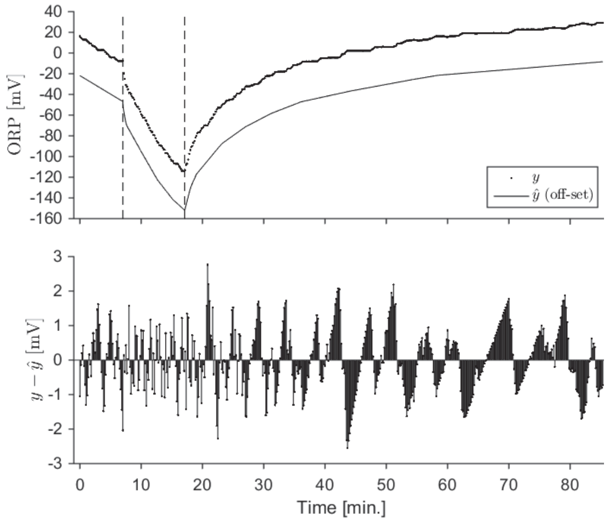

Figure 2: Exemplary time series corresponding to normal operation. Top: raw data and fitted function. Dashed lines indicate changes between the EAC primitives. Bottom: Residuals between the measurements and the fitted function.

The SCS consists of fitting a spline function constrained to exhibit the EAC shape. The locations in time where the shape changes from the E to A, and A to C primitives are considered unknown a priori and are estimated. In Figure 2, the fitted function with EAC shape is shown together with the residuals. It is easy to verify that the shape corresponds well to the imposed EAC shape. The function is an interpolating natural cubic spline function. This means that in the absence of any shape constraints, the function would fit the data perfectly. Any deviation between the measurements and the fitted function is thus due to a mismatch between the imposed shape and the data. This property is exploited for fault detection by computing the sum of squared residuals (SSR_{SCS}), which is the sum of squared deviations between the measurements and the fitted function. The higher it is, the less likely the measured data series matches the imposed shape.

To evaluate the potential of the SCS method for fault detection, all data series corresponding to 410 batch cycles were interpreted by two operators of the experimental plant by answering the question "whether the observed time series can be explained by normal circumstances alone" by means of a simple yes or no. These 410 time series were collected with the same SBR recipe and settings for the bang-bang aeration controller. For 16 cycles out of 410, no consensus could be reached between the two operators. The remaining 394 time series are included for further study. The SCS method is applied to these cycles. In parallel, a PCA model was calibrated with the first 100 time series which were assessed to be normal by both operators. The calibrated PCA model consists of 2 principal components which capture 86.9% and 9.6% of the variance of the calibration data (total: 96.5%). To use the PCA model for fault detection, the sum of squared residuals (SSR_{PCA}) is computed as the sum of squared deviations between the measured data and their reconstructions on the basis of the calibrated PCA model (Jackson & Mudholkar, 1979; Kresta *et al.*, 1991).

The resulting statistics (SSR_{SCS} and SSR_{PCA}) are shown in Figure 3. In this plot, the fault detection limit is set at the highest value for the statistics obtained for a normal cycle. This reveals a number of interesting results. First of all, the SCS method leads to detection of 39 anomalous cycles and the PCA method leads to detection of 30 anomalous cycles, out of 46 abnormal cycles in total. Twenty-nine (29) of the 46 cycles are detected by both methods. The improved result by the SCS method can in part be explained by the fact that the SSR_{SCS} computed for normal cycles is subject to rather small variations over time in comparison to the SSR_{PCA} statistic. The SCS method thus seems more robust to process variations in time compared to the PCA method.

Initial results obtained with an extended method for shape constrained spline show that *(i)* fitting of shape constrained splines with discontinuous behavior is possible and *(ii)* the use of shape constrained spline fitting is promising for the monitoring of biological sequencing batch reactors. Particular advantages of the SCS method include *(i)* an intuitive interpretation of the resulting statistic as a measure of divergence from normalcy and *(ii)* the apparent robustness of the method against parametric process changes. Continued work is aimed at the evaluation whether the same benefits can be maintained when considering *(i)* the use of the methods under slightly different operational schemes (e.g., change of aeration control settings) and *(ii)* in different reactor setups (e.g., differences in process history and microbial culture).

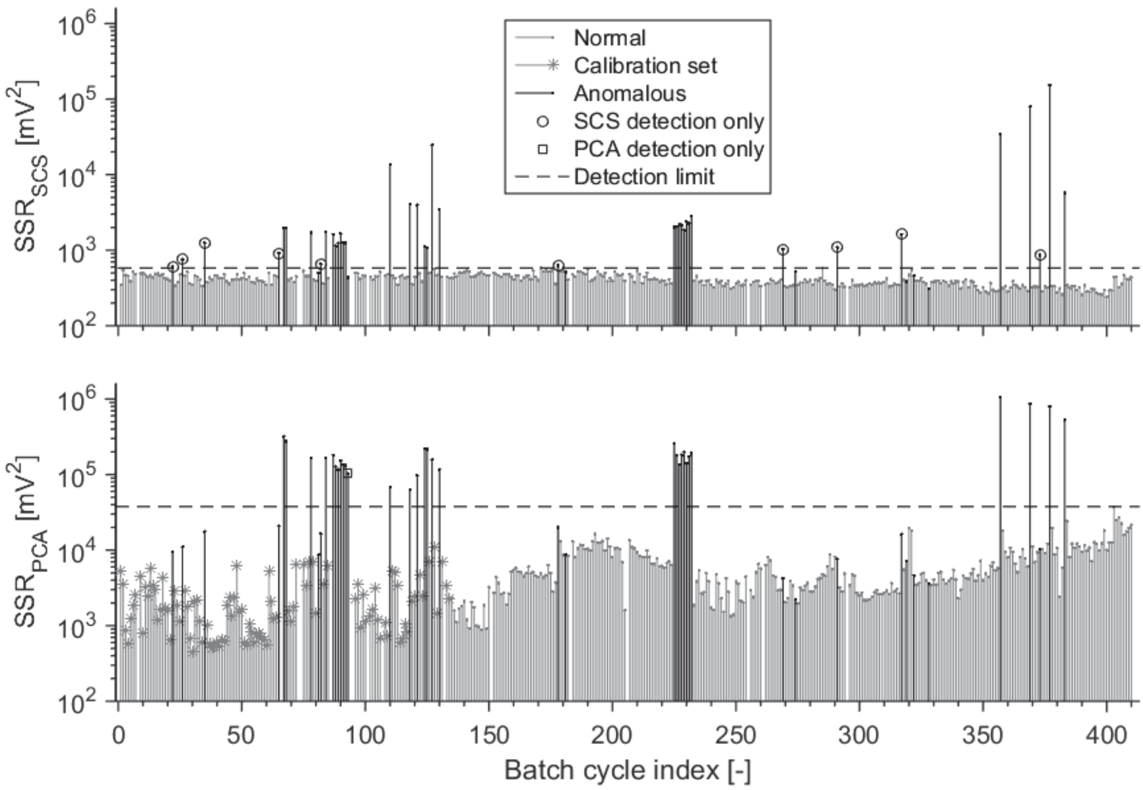

Figure 3: Sum of squared residuals (SSR) obtained with SCS (top) and PCA (bottom).

Acknowledgments

The authors want to thank Nicolas Derlon at Eawag for classifying the studied data series and for creative feedback during this study. All computations were executed with Matlab (The MathWorks Inc., 2012); the FDA toolbox (Ramsay and Silverman, 2002); and MOSEK (MOSEK ApS, 2012).

References

Aguado, D., Ferrer, A., Ferrer, J., & Seco, A. (2007). Multivariate SPC of a sequencing batch reactor for wastewater treatment. Chemometrics and Intelligent Laboratory Systems, 85(1), 82-93.

Jackson, J. E., & Mudholkar, G. S. (1979). Control procedures for residuals associated with principal component analysis. Technometrics, 21(3), 341-349.

Kresta, J. V., MacGregor, J. F., & Marlin, T. E. (1991). Multivariate statistical monitoring of process operating performance. The Canadian Journal of Chemical Engineering, 69(1), 35-47.

Rosén, C., & Lennox, J. A. (2001). Multivariate and multiscale monitoring of wastewater treatment operation. Water research, 35(14), 3402-3410.

Villez, K., Ruiz, M., Sin, G., Colomer, J., Rosen, C., & Vanrolleghem, P. A. (2008). Combining multiway principal component analysis (MPCA) and clustering for efficient data mining of historical data sets of SBR processes. Water Science and Technology, 57(10), 1659-1666.

Villez, K., Venkatasubramanian, V., & Rengaswamy, R. (2013). Generalized shape constrained spline fitting for qualitative analysis of trends. Computers & Chemical Engineering, 58, 116-134.

Villez, K. (2014). Bounding Procedures for Shape Constrained Splines with Discontinuities. Eawag, Dübendorf, Switzerland, Technical Report, TR-001-01-0.