European Symposium on Computer Aided Process Engineering – 15
L. Puigjaner and A. Espuña (Editors)

On-Line Dynamic Monitoring of the SHARON Process for sustainable nitrogen removal from wastewater

Kris Villez[a*], Christian Rosen[b], Stijn Van Hulle[a], ChangKyoo Yoo[a] and Peter A. Vanrolleghem[a]

[a] BIOMATH, Ghent University
Coupure Links 653, B-9000 Gent, Belgium

[b] IEA, Lund University
Box 118, SE-221 00 Lund, Sweden

Abstract

The goal of this work is the development of a suitable monitoring module, which is to be the first module of an integrated fault detection and control system for the SHARON process. To model the process properly, different PCA models are tested. As a first step, PCA is used in an iterative manner to exclude data not considered to represent normal operational conditions and process behaviour from the original data set.

To improve the performance of the identified model, it is decided to account for dynamics in the SHARON process by means of auto-regressive exogenous (ARX) structuring of data before the identification. A fruitful replacement of missing values for this purpose is done by means of a static PCA model. It is shown that the different criteria used in model selection lead to the same DPCA model. In this paper all steps of the monitoring module design are explained and the performance of different models is analyzed.

Keywords: environmental biotechnology, wastewater treatment, statistical monitoring, fault detection, Dynamic PCA

1. Introduction

In the SHARON process (Single Reactor High Activity Ammonia Removal Over Nitrite) the oxidation of ammonia to nitrite is achieved. The conversion of nitrite to nitrate is avoided by applying a high temperature (above 25°C) and an appropriate hydraulic residence time. When half of the ammonia is converted, coupling with an Anammox unit becomes economically interesting, as in the Anammox process equimolar amounts of ammonia and nitrite are removed (van Dongen *et al.*, 2001). The concept's success relies however on the control of the SHARON process since the requirements for equimolar concentrations of ammonia and nitrite and absence of nitrite inhibition in the Anammox reactor have to be fulfilled (van Dongen *et al.*, 2001).

To control highly dynamic processes as the SHARON process, a supervisory control system is developed. Monitoring, fault detection and diagnosis and the actual supervisory control are part of this system. A PCA-based monitoring module is created first. Applications for industrial process monitoring have been established in the past

decade (MacGregor and Kourti, 1995, Kourti and MacGregor 1995, Wise and Gallagher, 1996). Applications to wastewater treatment plants have emerged as well in more recent research (Rosén and Lennox, 2001, Yoo *et al.*, 2003, Lee *et al.*, 2004).
In this study the development of a dynamic PCA (DPCA) model for monitoring is described. Simple static PCA modeling and data selection is discussed firstly. Secondly, the replacement of missing data is explained. Afterwards, different DPCA models are constructed and compared to eachother. Finally, a monitoring model is selected.

2. Materials and methods

For a detailed description of the data set the reader is referred to Van Hulle *et al.* (2003). The data set contains 503 data samples of 10 variables: (1) hydraulic residence time (HRT), (2) ammonia load, (3) the influent ratio of total inorganic carbon to total ammonia nitrogen (TIC:TAN), (4) daily mean dissolved oxygen (DO), (5) daily mean pH, (6) daily mean acid addition, (7) daily mean base addition, (8) effluent ammonia, (9) effluent nitrite and (10) effluent nitrate concentrations.
Ku *et al.* (1995) developed a procedure for dynamic PCA (DPCA) modeling, which is adapted in this work. The procedure uses the calculation of the newly found inner relations in the data set by a step-wise increase in the window length of the ARX structure. As long as the number of new relations is more than zero, the window length is increased. Ultimately, the last model which still reveals new relations is selected.

3. Results and discussion

3.1. Simple static model

In the model for "normal data" selection, two principal components (PC's) were retained as a compromise between guaranteed detection of abnormal data samples and an acceptable number of false alarms. Despite its simplicity, a two-component PCA model allows to discern the following states: (1) low ammonia load and high nitrate values, (2) moderate ammonia load and low nitrate values and (3) high ammonia load and low nitrate values.

3.2. "Normal data" selection

A proper selection of the data used for modelling the normal behaviour is imperative in statistical modeling. Data which caused Hotelling's T^2 or SPE to be violated at 95% confidence levels were investigated in detail. Data that did not represent normal behaviour were omitted. In the next iteration, only the data samples that caused either the 99%-limit for SPE or Hotelling's T^2 to be violated were investigated. As only 10 further data samples were removed in this iteration, no further iteration was performed. The number of samples was cut from 440 to 378 (14% reduction).

3.3. Completion of incomplete samples by PCA projection method

To base the dynamic models on continuous series which extend as long as possible, missing values for off-line measurements were estimated. The missing values were estimated by backward calculation from the scores (inverse PCA), which were estimated by the single component-projection method (Nelson *et al.,* 1996). Negative estimates for concentrations were set to zero and remaining missing variables were estimated again by the same method. Afterwards, samples with estimates which caused Hotelling's T^2 and/or SPE to rise above the 95% level were omitted.

3.4. Construction of models and comparison

In Table 1, the calculated inferences from Ku *et al.* (1995) are listed along with the percentage of captured variance, the false alarm ratio and the undetected failure ratio. The approach of Ku *et al.* (1995) leads to a model with ARX window length equal to two because this model is the last one by which new relations are found.
Typically, the captured variance is important for prediction purposes. In this case, all featured models are considered to satisfy (all captured variances are 70% or higher).

Table 1: Summary of model characteristics

w (window length)	1	2	3	4
m (variables)	10	20	30	40
c (principal components)	4	4	6	6
$r = m - c$ (inner relations found)	6	16	24	34
r_{new} (new inner relations found)	6	4	-2	-
captured variance (%)	74.7	70.0	77.9	75.7
false alarm ratio (%)	2.5	4.8	6.4	8.2
undetected failure ratio (%)	76.7	23.9	28.4	27.8

A third and key criterion is the success in discerning abnormal from normal behaviour. The ratio of false alarms to normal samples (false alarm ratio) and the ratio of undetected abnormal samples to total abnormal samples (undetected failure ratio) relate to this performance. An alarm was induced when either the 95%-level of Hotelling's T^2 or SPE was violated. The false alarm ratio rises with increasing window lengths, but is acceptable for all models (see Table 1). A minimal undetected failure ratio is found for a window of two. The undetected failure ratio is 76.7% in the static PCA model, while it is lower than 30% for all DPCA models. Accounting for the dynamics is thus essential.
Based on the used criteria, the window length of the ARX structure should either be two or three. Therefore, the relations that were observed between the model and the mechanistic knowledge of the real system are summarised only for these two models.

3.4.1 Model with window length = 2 (one time delay) and four PC's

HRT, acid addition and DO values are affecting the 3rd score the most (not shown). Also, the effect of a temporary low value of HRT was noticed very clearly in this score.
pH (variable 5), nitrate (variable 10) and base addition (variable 7) are the dominating variables of the fourth PC (see Figure 4, left). A remarkable distinction of two periods is observed: (1) rather unstable pH values and a rather high base addition (days 99 to 400) and (2) more stable pH values and lower base addition. Discussion with the operator of the labscale SHARON setup reveals that these differences in pH dynamics had not been detected before. Even though the cause could not be determined, the most sensible hypothesis is that the ingrowth of second step nitrifiers resulted in the conversion of nitrite to nitrate, which in turn reduced the toxic effects of nitrite on the first step nitrifiers. Since this was not observed before the PCA modeling, it shows that PCA modeling can be helpful in understanding the characteristics of biological processes.

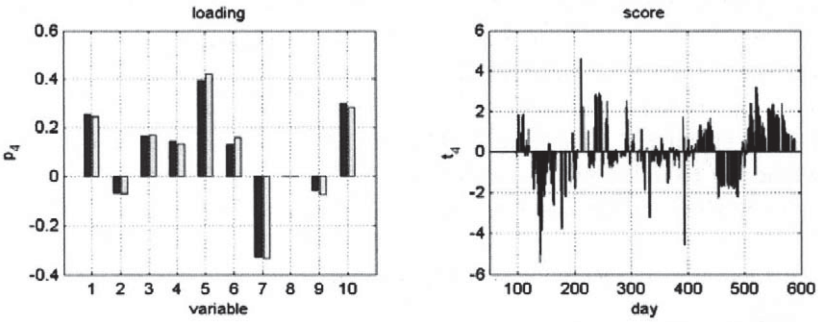

Figure 1: Loading plot and score plot of component 4 in the model with ARX window equal to 2. Red bars show the loadings of the delayed variables (time instant –1) while the yellow bars show the actual variables (time instant 0).

3.4.2 Model with window length = 3 (two time delays) and six components

To a large extent, the PC3 and PC4 captured the same effects as in the former model. Acid effects are not captured in the third but in the fourth PC. Base addition is only influential in PC5. Nitrate build-up is reflected in this score as well. However, the latter effects appear yet in the fourth PC of the previous model. PC6 is dominated by the DO and effluent nitrate. The resulting trends are however observed yet in the first score. As the last two PC's seem only to capture information which is captured in the four PC's of the 1st model, they lead only to a larger captured variance, which is not of prime interest.

3.5. Selection

In the final selection step, the false alarm ratio was omitted as a criterion as the values are comparable for both models. In Table 2, the relevant results are summarized. In three out of four criteria, the model of window length of two is performing the best. Thus, a window length of two is selected, which leads to a four-component DPCA model (see Table 1).

Table 2: Summary of selection procedure: for each of the criteria, the preferenced model is marked by "+"

w	2	3
r_{new}	+	
captured variance		+
undetected failure ratio	+	
model description	+	

3.6. Application of the model

Figure 4 shows the SPE and Hotelling's T^2 values. The following list of errors or abnormal situations that are detected by the model confirms that PCA monitoring is a valid approach in monitoring of biological processes :

- startup phenomena (73-94)
- sensor failures: DO sensor (188-190, 431), pH sensor (121-125, 193-195, 386, 431)

- actuator failures: acid addition (138-140, 241-243), base addition (178, 181, 209-214), oxygen supply (251-256, 466), influent supply (278-281, 292-293), effluent withdrawal (171-174)
- operational changes: change of HRT (449), change of TIC:TAN ratio (541)
- abnormal conversion rates: high (206), low (333, 395)
- addition of biomass (168-170)
- excessive water evaporation due to supply of dry rather then moist air supply (220-222)
- running of fast-dynamic kinetic experiments (308)

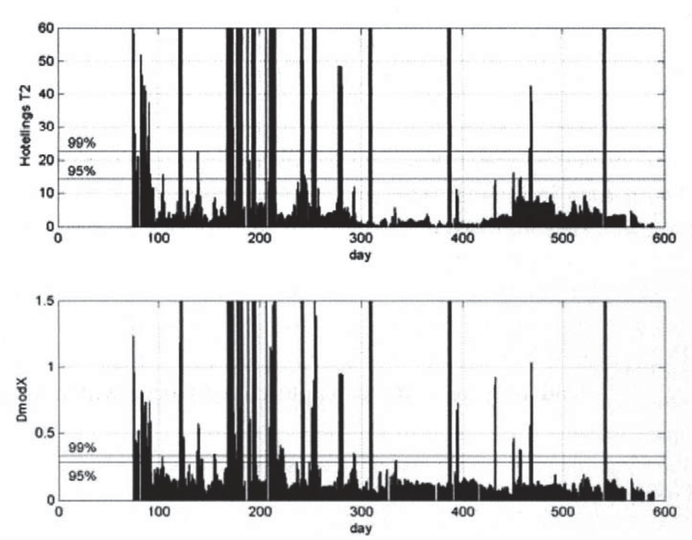

Figure 2: Hotelling's T^2 and DmodX (distance to the model plane) control charts based on the selected model I

3.7. Effect of data estimation

Since considerable effort is put into the estimation of missing data it is interesting to know what effect it has on the model performance. The results for the selected model structure with and without estimates are summarised in Table 3. Captured variances are comparable and the false alarm ratio is equal for both models. The undetected failure ratio seems to be influenced largely by the effect of gaps in the data, as this ratio is almost doubled when no estimates are used.

Table 3: Effect of the use of estimation data on model performance

inference	with estimated data	without estimated data
captured variance (%)	70.0	70.7
false alarm ratio (%)	4.8	4.8
undetected failure ratio (%)	23.9	46.0

4. Conclusions

This paper describes the development of a module for PCA based monitoring of the SHARON process. A dynamic PCA (DPCA) model is selected from several DPCA models, with different window lengths. For monitoring model selection, the criteria include captured variance, added new relations (Ku *et al*., 1995), false alarm ratio, undetected failure ratio and mechanistic relations between the components and the process characteristics.

It is shown that ARX structuring of data clearly improves the monitoring performance compared to a static PCA model. Furthermore, estimation of missing data concerning the nitrogen species improves the performance of a DPCA models.

In the selection, loss of captured variance is traded for an improved undetected failure ratio and a smaller span of the PCA space. The method of Ku *et al.* (1995) for DPCA model selection lead to the same model, which approves their procedure as a relative fast and good-quality method for DPCA model selection.

Acknowledgements

This work was supported by the Institute for Encouragement of Innovation by means of Science and Technology in Flanders (IWT), the Visiting Postdoctoral Fellowship of the Fund for Scientific Research-Flanders (FWO) and the ICON Project No. EVK1-CT2000-054.

References

Kourti, T. and J.F. MacGregor, 1995, Process analysis, monitoring and diagnosis, using multivariate projection methods, Chem. Intell. Lab. Syst. 28, 3.

Ku, W., R.H. Storer and C. Georgakis, 1995, Disturbance detection and isolation by dynamical principal component analysis, Chem. Intell. Lab. Syst. 30, 179.

Lee, D.S. and P.A. Vanrolleghem, 2004, Adaptive consensus principal componenent analysis for on-line batch process monitoring, Environmental Monitoring and Assessment 93(1-3), 119.

MacGregor, J.F. and T. Kourti, 1995, Statistical process contol of multivariate processes, Control Eng. Practive 3(3), 403.

Nelson, P.R.C, P.A. Taylor and J.F. MacGregor, 1996, Missing methods in PCA and PLS: Score calculations with incomplete observations, Chem. Intell. Lab. Syst. 35, 45.

Rosen, C. and J.A. Lennox, 2001, Multivariate and multiscale monitoring of wastewater treatment operation, Wat. Res. 35(14), 3402.

van Dongen, L.G.J.M., M.S.M. Jetten and van Loosdrecht M.C.M., 2001, The Combined Sharon/Anammox Process, A Sustainable Method for N-removal from Sludge Water. IWA Publishing, Water and Wastewater Practicioner Series: Stowa Report, pp 64.

Van Hulle, S., S. Van Den Broeck, J. Maertens, K. Villez, G. Schelstraete, E. Volcke and P.A. Vanrolleghem, 2003, Practical experiences with start-up and operation of a continuously aerated lab-scale SHARON reactor, Proceedings 16th Forum Applied Biotechnology, Comm. Agri. Appl. Biol. Sci. 68/2(a), 77-84.

Wise, B.M. and N.B. Gallagher, 1996, The process chemometrics approach to process monitoring and fault detection, J. Proc. Contr. 6 (6), 329.

Yoo, C.K, P.A. Vanrolleghem and I. Lee, 2003, Nonlinear modeling and adaptive monitoring with fuzzy and multivariate statistical method in biological WWTP, J. of Biotechnology, 105(1-2), 135.