ADAPTIVE CONSENSUS PRINCIPAL COMPONENT ANALYSIS FOR ON-LINE BATCH PROCESS MONITORING

Dae Sung Lee* and Peter A. Vanrolleghem
BIOMATH, Ghent University, Coupure Links 653, B-9000 Gent, Belgium
(author for correspondence, e-mail: DaeSung.Lee@biomath.rug.ac.be)*

Abstract. As the regulations of effluent quality are increasingly stringent and influent loads is, the on-line monitoring of wastewater treatment processes becomes very important to meet ever increasing effluent water quality. Multivariate statistical process control such as principal component analysis (PCA) has found wide application in process fault detection and diagnosis using existing measurement data. In this work we propose a consensus PCA algorithm for adaptive wastewater treatment process monitoring. The method overcomes the problem of changing operation conditions by updating the covariance structure recursively. The algorithm is also based on a consensus algorithm, which does not require any estimation for batch processes monitoring compared to typical multiway PCA models. With this method process disturbances are detected in real time and the responsible measurement are directly identified. The presented methodology is successfully applied to a pilot-scale sequencing batch reactor for wastewater treatment.

Keywords: process monitoring, adaptive, principal component analysis, batch, wastewater treatment

1. Introduction

With the advent of improved on-line sensor technology and automation, wastewater treatment processes (WWTPs) with well-equipped computerized measurement devices produce large amounts of data. Multivariate statistical process control (MSPC) can be very useful for effectively extracting relevant information from the collected measurement data for process monitoring and supervision and has been applied to industrial processes (MacGregor and Kourti, 1995, Wise and Gallagher, 1996). One of the most popular MSPC techniques is principal component analysis (PCA). PCA is a projection method for mapping original high dimensional data onto a lower dimensional space with minimum loss of useful information. In recent years, PCA has also been applied to continuous WWTPs (Rosen and Olsson, 1998; Rosen and Lennox,

2001; Lee *et al*., 2002).

A sequencing batch reactor (SBR) process has a unique cyclic batch operation for biological wastewater treatment (Demuynck, *et al*., 1994; Lee and Park, 1998). Most of the advantages of SBR processes may be attributed to the flexibility capable to meet many different treatment objectives. The SBR process is highly nonlinear, time varying and subject to significant disturbances like hydraulic changes, composition variations and equipment defects. Small changes in concentrations or flows can have a large effect on the kinetics of biological reactions leading to batch-to-batch variability in effluent quality and microorganism growth. Relative to continuous WWTPs, SBR operation data have an added dimension of the batch number in addition to the measured variables and sample times. Therefore, the most basic method of conventional PCA is not directly applicable to SBR processes.

Nomikos and MacGregor (1994) presented the MPCA approach for monitoring batch processes. MPCA is an extension of PCA for three-dimensional batch data (*batch number* × *variables* × *time*). In addition, due to the non steady-state behavior of batch processes, MPCA explain the variation of the measured variables around the average trajectories as calculated from the nominal operation data set. However, MPCA models require all the completed measurement data to on-line monitor the progress of a new batch. This means that it is necessary to estimate the values of all process measurements from the current time to the end of the batch operation as the new batch evolves. Several different estimation methods have been proposed to handle this problem (Nomikos and MacGregor, 1995). MPCA models without need of estimating the uncompleted portion of the batch has also been proposed, based on either a hierarchical approach or a moving window method (Ränner *et al*., 1998; Lennox *et al*., 2001).

In this study adaptive consensus PCA is proposed to develop an on-line monitoring system for SBR processes. The proposed method does not estimate any future deviations of the ongoing batch from the average trajectories. Since the SBR process itself evolves over time as the microorganisms adapt to changing operating conditions like surrounding temperature and varying process loads, the consensus model is recursively updated to overcome this problem. The methodology is applied to a bench-scale SBR that is used to grow sludge that is as stable in properties as possible as this sludge is to be used in a comprehensive study of flocculation (Nopens *et al*., 2002). When the SBR is free from major upsets the sludge is likely to be more stable than when it is subject to significant disturbances. The PCA supervision is used as a tool to monitor the stability of the sludge on the basis of simple on-line data.

2. Materials and Methods

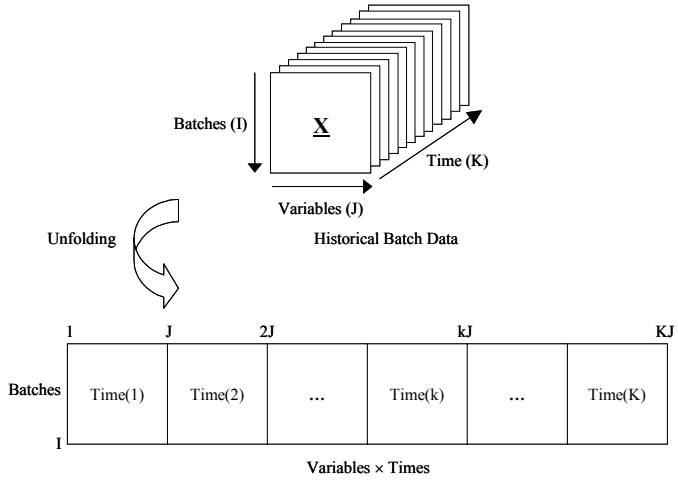


Figure 1. Decomposition of a three-way batch data matrix $\underline{\boldsymbol{X}}$.

2.1. MULTIWAY PRINCIPAL COMPONENT ANALYSIS

In a typical batch run $j = 1,\ldots, J$ variables are measured at each of $k = 1,2,\ldots,K$ time intervals throughout the batch. Similar data will exist on a number of batches $i = 1,2,\ldots,I$. All the data can be summarized in the $\underline{\boldsymbol{X}}(I\times J\times K)$ of a historical database (Fig. 1). Multiway PCA is equivalent to unfolding the three-dimensional data matrix $\underline{\boldsymbol{X}}$ into a large two-dimensional matrix $\boldsymbol{X}$, and then performing a regular PCA (Nomikos and MacGregor, 1994). In case of monitoring batch processes, it is important to determine differences between batches and to project new batches on the model. Therefore, in this work, $\underline{\boldsymbol{X}}$, was unfolded in such a way as to put each of its vertical slices (I) side by side to the right, starting with the one corresponding to the first time interval. The resulting two-dimensional matrix has size ($I\times JK$). This unfolding allows for analyzing the variability among the batches in $\underline{\boldsymbol{X}}$ by summarizing the information in the data with respect to both variables and their time variation.

Before a MPCA is performed on the batch data, the data are normalized using the mean and standard deviation of each variable at each time in the batch cycle over all batches. Then MPCA decomposes the data into a series of principal components consisting of score vectors ($\boldsymbol{t}_r$) and loadings ($\boldsymbol{p}_r$), plus residuals ($\boldsymbol{E}$).

$$\boldsymbol{X} = \sum_{r=1}^{R} \boldsymbol{t}_r \boldsymbol{p}_r + \boldsymbol{E}$$

The loading vectors ($\boldsymbol{p}_r$) define the reduced dimension space (R) and are the directions of maximum variability. Each element of the score vectors ($\boldsymbol{t}_r$) corresponds to a single batch and depicts the overall variability of this batch with respect to the other batches in the data base throughout the whole batch duration (Nomikos and MacGregor, 1995). Usually, a few principal

components can express most of the variability in the data when there is a high degree of correlation among the data (R<<min(I, JK)). R is chosen such that most of the systematic variability of the process data is described by these principal components and that the residual matrix $\boldsymbol{E}$ is as small as possible in a least squares sense. The NIPALS (Nonlinear Iterative Partial Least Squares) algorithm can be used for sequential computation of the dominant principal components (Geladi and Kowalski, 1986).

When monitoring a new batch in progress, the measurements between the current time interval k and the end of the batch are unknown. However, the MPCA model needs the vector to be of full length, $K \times J$, in order to calculate the score vectors for the present batch. Nomikos and MacGregor (1994) presented three possible solutions for estimating in the unknown data in $\boldsymbol{X}_{\text{new}}$. One of these solutions assumes that the future deviations from the mean trajectories will remain for the rest of the batch duration at their current values at time interval k (filling method 2). This estimation method is similar to the one made in model predictive control algorithms and works very well in practice (Tates *et al*., 1999; Lennox *et al*., 2001).

Control charts that are used in monitoring on-line batch processes are generally based on D-statistic and the squared prediction error (SPE) in which control limits are used to determine whether the process is in control or not. The assumption behind these approximate confidence limits is that the underlying process exhibits a multivariate normal distribution with a population mean zero. This is to be expected since any linear combination of random variables, according to the Central Limit Theorem, should tend toward a normal distribution.

The D-statistic measures the degree to which data fit the calibration model:

$$D = \boldsymbol{t}_r^{\mathrm{T}} \boldsymbol{S}^{-1} \boldsymbol{t}_r \sim \frac{I(I-R)}{R(I^2-1)} F_{\mathrm{R,I\text{-}R}}$$

where $\boldsymbol{S}$ is the estimated covariance matrix of the scores. The D-statistic gives a measure of the Mahalanobis distance in the reduced space between the position of a batch and the origin that designates the point with average batch process behavior. The distribution of the D-statistic for all batches can be approximated by a F-distribution, $F_{\mathrm{R,I\text{-}R}}$, and confidence limits for the D-statistic are calculated from this F-distribution.

The SPE_k at time interval k is calculated as follows:

$$\mathrm{SPE}_k = \sum_{c=(k-1)J+1}^{kJ} \boldsymbol{e}^2(\mathrm{c})$$

where $\boldsymbol{e}$ is the residuals for the new batch $\boldsymbol{X}_{\text{new}}$. The distribution of SPE can be well approximated by a weighted chi-squared distribution $g\chi^2_{(h)}$, where g is a weight and h is the degrees of freedom. The values of g and h can easily be estimated by matching the moments of $g\chi^2_{(h)}$ distribution with the moments of the observations of the SPE from the reference

distribution at each time interval k. Then the approximate control limits for the SPE_k with a significance level of α are obtained:

$$SPE_{k,\alpha} = (\frac{v}{2m})\chi^2_{2m_2/v,\alpha}$$

where m and v are the mean and variance of the SPE calculated for each time interval k.

2.2. APAPTIVE MPCA

A SBR process itself evolves over time as the microorganisms adapt to changing operating conditions like surrounding temperature and varying process loads. When a typical MPCA model is used to monitor processes with time-varying behaviors, false alarms often result, which significantly compromise the reliability of the monitoring system (Li *et al*., 2000).

To overcome the problem of changing process conditions, an adaptive MPCA model based on a moving window can be developed (Lee and Vanrolleghem, 2002). A window is a set of data containing a fixed number of batches. When a new batch is available, another window is created by dropping the first batch in the set and by adding the new batch to the window. Hence, the new window overlaps all but one of the batches of the old window and includes new information. In this approach a new covariance structure is identified for each new batch and all batches inside the window frame will have a constant influence on the model until it leaves the window.

2.3. CONSENSUS PCA FOR BATCH MONITORING

Consensus PCA methods have been proposed in the recent literature in order to improve the interpretability of multivariate models (Wold *et al*. 1987, Westerhuis *et al*., 1998). The consensus PCA approach may have significant benefits when monitoring SBR processes since the model can adapt to different phases of the SBR operation in a cycle. In this application, a variant consensus algorithm is proposed to monitor SBR processes. The advantage of this technique is that it does not require any estimation of the future measurement data.

The data matrix $\boldsymbol{X}$ ($I \times KJ$) is divided into K blocks ($\boldsymbol{X}_1$, $\boldsymbol{X}_2$,…,$\boldsymbol{X}_K$) in such a way that the variables from each time instant can be blocked in the same block. Then the consensus PCA algorithm is applied to each time slice at a time rather than all the blocks at once as proposed by Ränner *et al*. (1998). Figure 2 shows the arrow scheme for the consensus PCA algorithm for batch applications. In the lower layer of the model the block $\boldsymbol{X}_k$ at time instant k is considered as a separate source of information and the details of the block is modeled. In the super layer

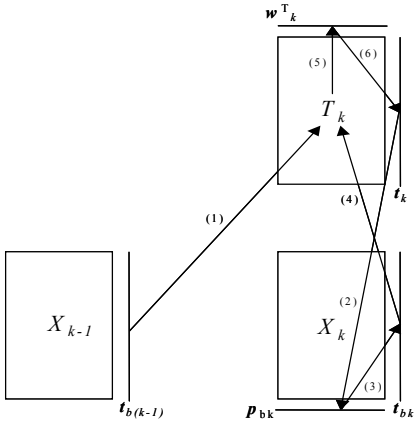


Figure 2. Consensus PCA algorithm for on-line batch process monitoring

information from the previous block, block scores $\boldsymbol{t}_{b(k-1)}$, is combined with the block score vector from the lower layer.

The proposed algorithm derives from a sequence of 'NIPALS steps' and has the following formulation:

1. Arbitrarily choose a column of one of the blocks $\boldsymbol{X}_k$ as $\boldsymbol{t}_k$

   (1) Put the previous block score vector $\boldsymbol{t}_{b(k-1)}$ into a super block $\boldsymbol{T}_k$.

   (2) Block loadings $\boldsymbol{p}_{bk}$ are obtained by regressing $\boldsymbol{t}_k$ on the block $\boldsymbol{X}_k$.

   $$\boldsymbol{p}_{bk} = \boldsymbol{X}^{\mathrm{T}}{}_k \cdot \boldsymbol{t}_k / \boldsymbol{t}^{\mathrm{T}}{}_k \cdot \boldsymbol{t}_k$$

   Then normalize $\boldsymbol{p}_{bk}$ to $\| \boldsymbol{p}_{bk} \| = 1$.

   (3) From $\boldsymbol{p}_{bk}$, block scores $\boldsymbol{t}_{bk}$ for the current block are calculated.

   $$\boldsymbol{t}_{bk} = \boldsymbol{X}_k \cdot \boldsymbol{p}_{bk}$$

   (4) The block scores are combined into the super block $\boldsymbol{T}_k$.

   $$\boldsymbol{T}_k = [\boldsymbol{t}_{b(k-1)}, \boldsymbol{t}_{bk}]$$

   (5) Super weight $\boldsymbol{w}$ is obtained by regressing $\boldsymbol{t}_k$ on the super block.

   $$\boldsymbol{w}_k = \mathbf{T}^{\mathrm{T}}{}_k \cdot \boldsymbol{t}_k / \boldsymbol{t}^{\mathrm{T}}{}_k \cdot \boldsymbol{t}_k$$

   Then normalize $\boldsymbol{w}_k$ to $\| \boldsymbol{w}_k \| = 1$.

   (6) A new $\boldsymbol{t}_k$ is calculated.

   Until the super score $\boldsymbol{t}_k$ converges, a new iteration, (2)-(6) starts.

2. Deflate block data matrix $\boldsymbol{X}_k$

   $$\boldsymbol{p}_{bk} = \boldsymbol{X}^{\mathrm{T}}{}_k \cdot \boldsymbol{t}_k / \boldsymbol{t}^{\mathrm{T}}{}_k \cdot \boldsymbol{t}_k$$

   $$\boldsymbol{X}_k = \boldsymbol{X}_k - \boldsymbol{t}_k \cdot \boldsymbol{p}^{\mathrm{T}}{}_{bk}$$

Go to step 1 to calculate the next principal component.

2.4. SBR PROCESS

The data used in this study were collected from a bench-scale SBR system shown in Fig. 3. A fill-and-draw sequencing batch reactor (SBR) system with a 80-liter working volume is operated in a 6 h cycle mode and each cycle consists of 1 h fill/anaerobic, 2 h 30 min aerobic, 1 h anoxic, 30 min aerobic and 1 h settling/draw phases. The hydraulic retention time (HRT) and solid retention time (SRT) are maintained at 12 hrs and 10 days, respectively. Loading amounts of COD as synthetic municipal-like sewage, NH_4^+-N and PO_4^{3-}-P per cycle in standard conditions are 440, 60 and 9.5 mg/l, respectively.

The controls of the duration/sequence of phases and on/off status of peristaltic pumps, mixer and air supply are automatically achieved by a Labview data acquisition and control (DAC) system. The DAC system consists of computer, interface cards, meters, transmitters and solid state relays (SSR). Electrodes for pH, ORP (oxidation-reduction potential), DO (dissolved oxygen), temperature, weight and conductivity are installed and connected to the individual meters. The status of the reactor is displayed on the computer and the time series of the electrode signals are stored in a data file.

A set of on-line measurements is obtained every one minute (360 time instants) in pH, ORP, DO, conductivity, temperature and weight of the SBR reactor. Thus, no advanced or expensive measurement devices have to be installed in order to make the methods work. All the measurements are simple standard measurements. These measurements were stored for 280 past cycles (=70 days) forming a database of historical information about the process.

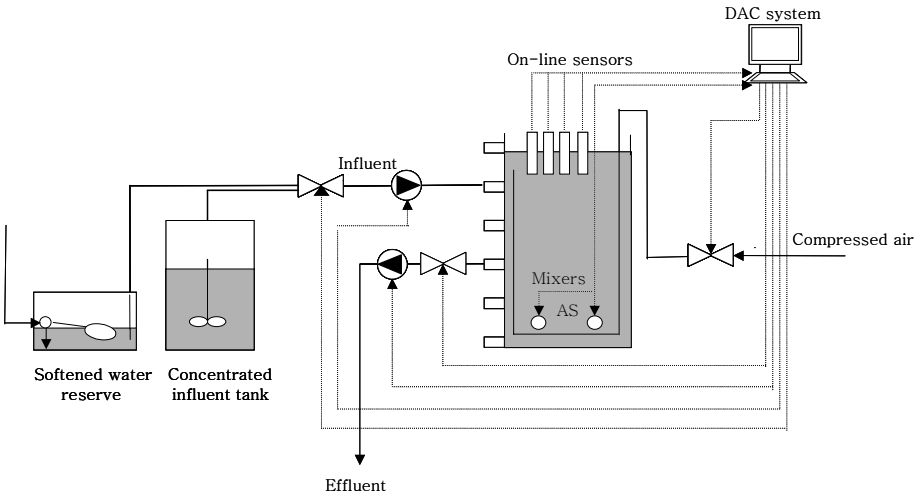


Figure 3. Schematic diagram of SBR process.

3. Results and Discussions

3.1. PRETREATMENT OF SBR DATA SETS

It has been reported before that the on-line sensor values collected in the SBR are related with dynamic characteristics of the nutrient concentrations (NH_4^+, PO_4^{3-} and NO_3^-) in SBRs (Demuynck *et al*., 1994; Chang and Hao, 1996). The derivatives of pH, ORP and DO profiles can accurately detect the ends of phosphate release, ammonia conversion and phosphate uptake, and be a useful information source. Therefore, first and second derivatives of pH, ORP and DO were calculated from the on-line sensor profiles and included into the database. Since the differencing operation magnifies the noise it is necessary to smooth the data beforehand. This can be done by Savitzky-Golay smoothing which is a moving window method fitting a polynome by least squares (Savitzky and Golay, 1964). In addition, only the measurement data from the first 300 sampling time instants were used to develop monitoring models since biological reactions in the settling and drawing phases (corresponding to those of the last 60 time instants) were assumed as negligible. The MPCA algorithms were applied to the three-way data array $\underline{\boldsymbol{X}}$ with dimensions 280×12×300.

3.2. ANALYSIS OF HISTORICAL PROCESS DATA SETS

By examining the process data in the reduced projection spaces defined by a small number of latent variables, it is often possible to extract very useful information to interpret the behavior of the SBR process. A MPCA model was developed from all the historical data sets. Figure 4 shows a score plot of the collected data in the space of the resulting first two principal components. The score plot can give a picture that well represents the process behavior. Figure 4 exhibits two clusters: cluster 1 (batch i=1-162); cluster 2 (batch i=165-280). The separation between cluster 1 and cluster 2 is due to disturbances (batch i=163-164). The scatter character of this plot indicates that the operating data started from the lower part of cluster 1 and gradually moved up; after the process instability occurred, the projected process data returned to the confidence limits inside cluster 2. This implies that static MPCA is not useful since the SBR process itself evolves over time as it is exposed to various disturbances such as influent composition variations, temperature changes, and equipment defects. This problem can be overcome by use of adaptive MPCA. More detailed information about comparison between the performance of typical MPCA and that of adaptive MPCA in the SBR process can be found Lee and Vanrolleghem (2002).

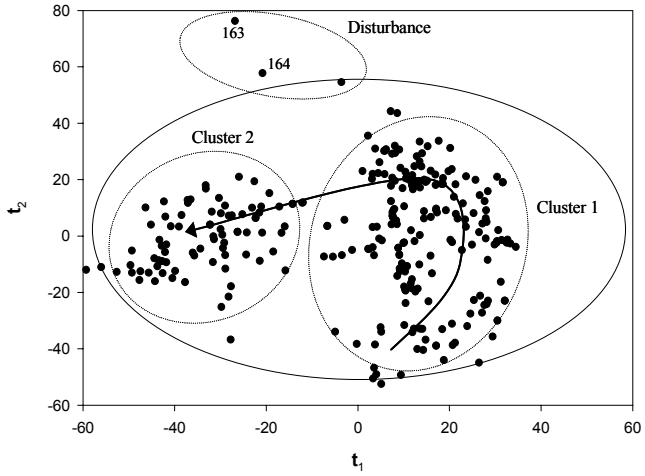


Figure 4. Score plot for all 280 batch data sets. The solid ellipse corresponds to 95% confidence limit.

3.3. APPICATION OF ADAPTIVE MPCA

An adaptive MPCA model based on a moving window strategy was developed from the historical data set of the SBR. In this application, different window lengths ranging from 10 to 50 batches were tested. The time span of the moving window was optimally set to 7.5 days (30 batches). The criterion for the selection of the window size was how fast and correctly the model could detect known disturbances in the validation data sets. Since the number of significant principal components can change over time, it is necessary to determine the number of principal components recursively. The optimal number of significant principal components is calculated recursively using the cumulative percent variance (CPV) method (Li *et al*., 2000).

Figure 5 shows the SPE and *D*-statistic for the first 30 consecutive batches with 99% confidence limits. These plots indicate that the MPCA model at batch number 30 describes adequately the reference batches (no batch is above the 99% confidence limit). The MPCA monitoring model is tested on batch number 31 (Fig. 6(a)). The batch is monitored for every time instant *k* with the SPE and *D*-chart values using filling method 2. It can be seen that batch number 31 stays below the upper control limits for both charts. These results displayed in Fig. 6(a) are fairly representative of the results of normal batches. An abnormal batch (batch number 197) is monitored for every time instant *k* with a MPCA model. Figure 6(b) shows that this abnormal batch has a large deviation in the SPE chart since the 211^{th} time instant of the batch operation. This indicates a special disturbance occurs from the 211^{th} sampling time instant. The disturbance will be discussed below.

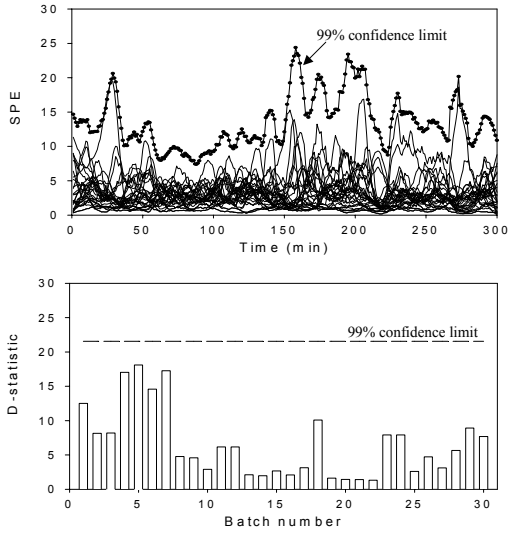
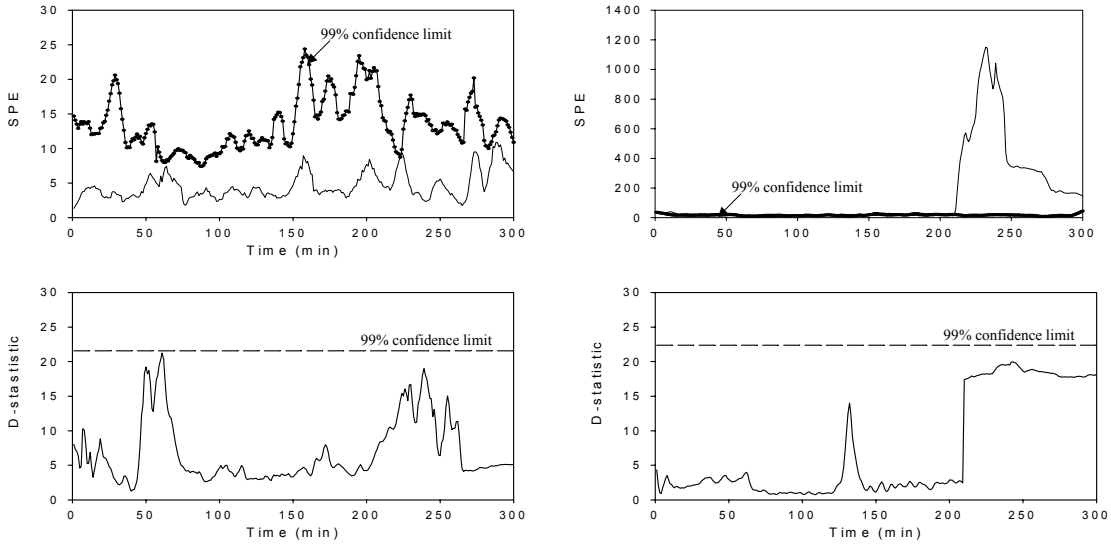


Figure 5. Adaptive multiway PCA. SPE and D-statistic charts with 99% confidence limits.



(a) On-line monitoring charts for batch number 31.

(b) On-line monitoring charts for batch number 197.

Figure 6. Adaptive multiway PCA. SPE and D-statistic charts for on-line monitoring a normal batch (batch number 31) and an abnormal batch (batch number 197).

3.4. APPICATION OF CONSENSUS PCA

A MPCA model with adaptive covariance structure as well as a consensus approach was used to monitor the same historical data set. The data matrix $\underline{\boldsymbol{X}}$ is decomposed into 300 blocks. The window size and method to determine the number of principal components for the adaptive consensus MPCA were equal to those of the adaptive MPCA-both for simplicity and ease of

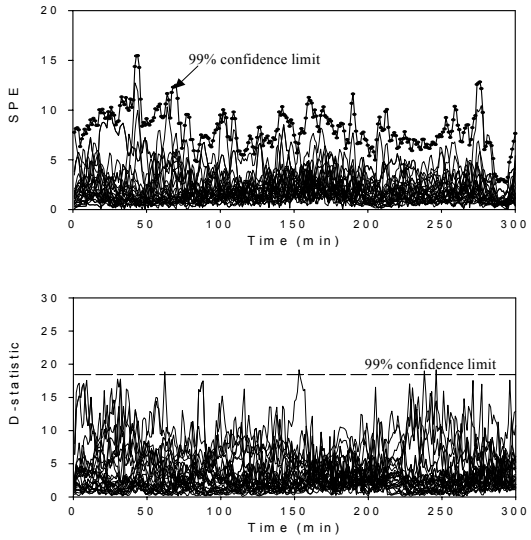
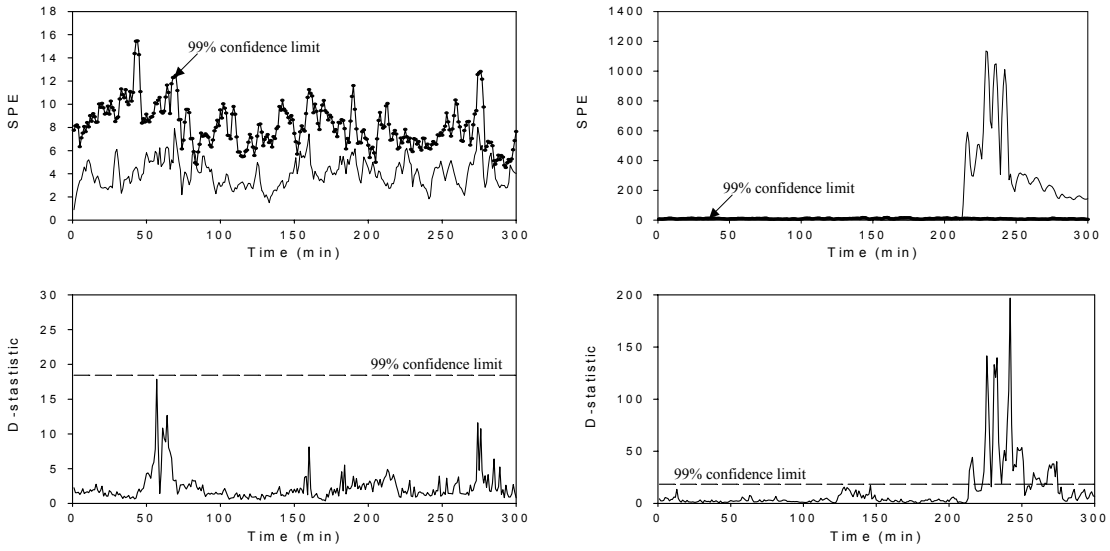


Figure 7. Adaptive Consensus PCA. SPE and D-statistic charts with 99% confidence limits.



(a) On-line monitoring charts for batch number 31.

(b) On-line monitoring charts for batch number 197.

Figure 8. Adaptive consensus PCA. SPE and D-statistic charts for on-line monitoring a normal batch (batch number 31) and an abnormal batch (batch number 197).

comparison. The weighting factor *d* in the proposed consensus algorithm is a tuning parameter that varies depending on how fast the process can change and the value is set to 0.95.

Figure 7 shows the SPE and *D*-statistic charts for the first 30 consecutive batches with 99% confidence limits. These charts show that the consensus PCA model describes adequately the reference database. Figure 8 displays the SPE and D-statistic charts for the normal batch 31 and the abnormal batch 197 using the consensus PCA method. The results are very comparable to

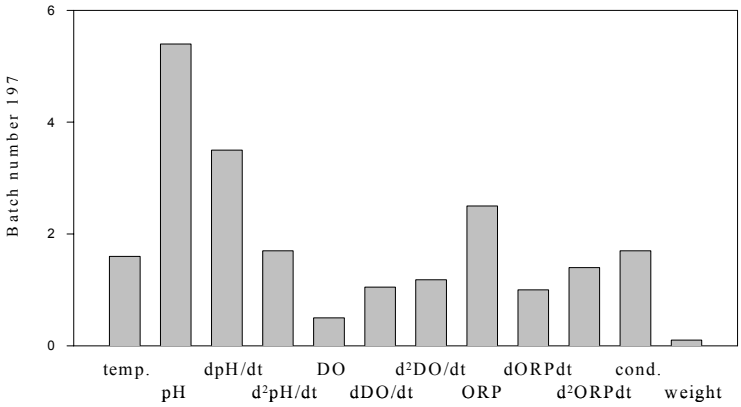


Figure 9. Contribution to the SPE value for all process variables at time k=211 for abnormal batch 197.

those obtained for the typical MPCA model. Further tests using the adaptive consensus approach also showed that it produced similar results to those obtained using the adaptive MPCA. Figure 9 shows the calculated contributions to the SPE value at time point k=211. It is obvious the pH is mainly contributing to the disturbances. This is seen on the original data as an accidental feeding of carbon source to the SBR process at the start of anoxic phase.

4. Conclusions

An adaptive approach based on the consensus algorithm is presented for monitoring the progress of wastewater treatment processes in real-time. The proposed monitoring method is built only from historical measurement data sets of batch processes. Adaptive PCA in terms of updating the covariance structure overcomes the problem of changing operational conditions in WWTPs. The historical operational data sets are split into time slice blocks in such a way that the variables at each sampling instant can be blocked in the same block. Then the proposed consensus algorithm is applied to each time block at a time to develop a corresponding local PCA model. The method can detect faults and isolate disturbance sources with two on-line monitoring charts when it is applied to a pilot-scale sequencing batch reactor. The monitoring performance of the consensus method is comparable to that of the MPCA model proposed by Nomikos and MacGregor (1994). However, the proposed method does not require estimating in the unknown data in $\boldsymbol{X}_{\text{new}}$ between the current time interval k and the end of the batch compared to the typical MPCA algorithm. Furthermore, the methods proposed here can be easily applied in most batch or fed-batch processes for development of easily applicable and effective monitoring models.

Acknowledgements

This work was financially supported by Fund for Scientific Research- Flanders (F.W.O.) and the Ghent University Research Fund.

References

Chang, C.H. and Hao, O.J.: 1996, 'Sequencing batch reactor system for nutrient removal: ORP and pH profiles', *J. Chem. Tech. Biotechnol*. **67**, 27-38.

Demuynck, C., Vanrolleghem, P.A., Mingneau, C., Liessens, J. and Verstraete, W.: 1994, 'NDBEPR process optimization in SBRs: Reduction of external carbon source and oxygen supply', *Water Sci. Tech*. **30**, 169-179.

Geladi, P. and Kowalski, B.R.: 1986, 'Partial least squares regression: A tutorial', *Anal. Chim. Acta*. **185**, 1-17.

Lee, D.S. and Park, J.M.: 1998, 'Neural network modeling for on-line estimation of nutrient dynamics in a sequentially-operated batch reactor', *J. Biotechnol*. **75**, 229-239.

Lee, D.S., Jeon, C.O., Park, J.M. and Chang, K.S.: 2002, 'Hybrid neural network modeling of a full-scale industrial wastewater treatment process', *Biotechnol. Bioeng*. **78**, 670-682.

Lee, D.S. and Vanrolleghem, P.A.: 2002, 'Monitoring of a sequencing batch reactor using adaptive multiblock principal component analysis', *Biotechnol. Bioeng*. (accepted).

Lennox, B., Montague, G.A., Hiden, H.G., Kornfeld, G. and Goulding, P.R.: 2001, 'Process monitoring of an industrial fed-batch fermentation', *Biotechnol. Bioeng*. **74**, 125-135.

Li, W., Yue, H., Valle-Cervants, S. and Qin, S.J.: 2000, 'Recursive PCA for adaptive process monitoring', *J. Proc. Cont*. **10**, 471-486.

MacGregor, J.F. and Kourti, T.: 1995, 'Statistical process control of multivariate processes', *Cont. Eng. Practice* **3**, 403-414.

Nomikos, P. and MacGregor, J.F.: 1994, 'Monitoring batch processes using multiway principal component analysis', *AIChE J*. **40**, 1361-1375.

Nomikos, P. and MacGregor, J.F.: 1995, 'Multivariate SPC charts for monitoring batch processes', *Technometrics* **37**, 41-59.

Nopens, I., Biggs, C.A., Clercq, B.D., Govoreanu, R., Wilen, B.M., Lant, P. and Vanrolleghem, P.A.: 2002, 'Modeling the activated sludge flocculation process combining laser diffraction particle sizing and population balance modeling (PBM)', *Water Sci. Tech*. **45**, 41-49.

Rännar, S., MacGregor, J.F. and Wold, S.: 1998, 'Adaptive batch monitoring using hierarchical PCA', *Chemom. Intell. Lab. Syst*. **41**, 73-81.

Rosen, C. and Olsson, G.: 1998, 'Disturbance detection in wastewater treatment systems', *Water Sci. Tech*. **37**, 197-205.

Rosen, C. and Lennox, J.A.: 2001, 'Multivariate and multiscale monitoring of wastewater treatment operation', *Water Res*. **35**, 3402-3410.

Savitzky, A. and Golay, M.J.E.: 1964, 'Smoothing and differentiation of data by simplified least squares procedures', *Anal. Chem*. **36**, 1627-1639.

Tates, A.A., Louwerse, D.J., Smilde, A.K., Koot, G.L.M. and Berndt, H.: 1999, 'Monitoring a PVC batch process with multivariate statistical process control charts', *Ind. Eng. Chem. Res*. **38**, 4769-4776.

Westerhuis, J.A., Kourti, T. and MacGregor, J.F.: 1998, 'Analysis of multiblock and hierarchical PCA and PLA models', *J. Chemom*. **12**, 301-312.

Wise, B.M. and Gallagher, N.B.: 1996, 'The process chemometrics approach to process monitoring and fault detection,' *J. Proc. Cont*. **6**, 329-348.

Wold, S. Hellberg, S., Lundstedt, T., Sjostrom, M. and Wold, H.: 1987, 'PLS modeling with latent variables in two or more dimensions', *Proceedings of Symposium on PLS Model Building: Theory and Application*, Frankfurt am Main, Germany, September 1987, pp.23-25.