

Monitoring of a Sequencing Batch Reactor Using Adaptive Multiblock Principal Component Analysis

Dae Sung Lee, Peter A. Vanrolleghem

Department of Applied Mathematics, Biometrics and Process Control,
Ghent University, Coupure Links 653, B-9000 Gent, Belgium; telephone:
+32-9-264-5935; fax: +32-9-264-6220; e-mail:
daesung.lee@biomath.rug.ac.be

Received 19 August 2002; accepted 21 October 2002

DOI: 10.1002/bit.10589

Abstract: Multiway principal component analysis (MPCA) for the analysis and monitoring of batch processes has recently been proposed. Although MPCA has found wide applications in batch process monitoring, it assumes that future batches behave in the same way as those used for model identification. In this study, a new monitoring algorithm, adaptive multiblock MPCA, is developed. The method overcomes the problem of changing process conditions by updating the covariance structure recursively. A historical set of operational data of a multiphase batch process was divided into local blocks in such a way that the variables from one phase of a batch run could be blocked in the corresponding blocks. This approach has significant benefits because the latent variable structure can change for each phase during the batch operation. The adaptive multiblock model also allows for easier fault detection and isolation by looking at the relationship between blocks and at smaller meaningful block models, and it therefore helps in the diagnosis of the disturbance. The proposed adaptive multiblock monitoring method is successfully applied to a sequencing batch reactor for biological wastewater treatment. © 2003 Wiley Periodicals, Inc. *Biotechnol Bioeng* 82: 489–497, 2003.

Keywords: process monitoring; adaptive multiway principle component analysis (MPCA); multiblock; wastewater; sequencing batch reactor (SBR)

INTRODUCTION

With increasingly stringent regulations for effluent quality, the on-line monitoring of wastewater treatment processes becomes very important for enhancement of process performance by detecting disturbances leading to abnormal process operation at an early stage. The earlier a potential fault can be detected, the less severe its influence will be, and the corresponding corrective action will consequently be more constrained (Lennox et al., 2001). Traditionally, wastewater treatment plants have been monitored by using time-series charts wherein operators can view the different variables as

historical trends and judge the deviation from the norm. However, as the number of variables increases from modern industrial plants with well-equipped computerized measurement devices, it becomes difficult or impossible to interpret all measurement data simultaneously. Therefore, a more systematic way to handle and analyze data is needed to effectively extract relevant information for monitoring and supervision.

In recent years, multivariate statistical process control, such as principal component analysis (PCA) and partial least squares (PLS), has become increasingly popular in many industrial fields (MacGregor and Kourti, 1995; Tates et al., 1999; Wise and Gallagher, 1996). These techniques can be used to extract the state of the system from the enormous volume of stored data via application of statistical methods. Recently, they have also been applied to wastewater treatment process (WWTP) operation (Lee et al., 2002; Rosen and Olsson, 1998; Teppola et al., 1998). However, most of the applications in WWTP have been found for continuous processes.

In the application investigated here, a variant of the PCA technique, multiway principal component analysis (MPCA), is used as a basis to develop a monitoring system for a sequencing batch reactor (SBR) process. The SBR process is highly nonlinear, time-varying, and subject to significant disturbances, such as 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. This behavior distinguishes the SBR process from continuous WWTP. Relative to continuous processes, batch 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 batch 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 \times variables \times time).

Correspondence to: D. S. Lee

Contract grant sponsors: Fund for Scientific Research–Flanders (FWO); Ghent University Research Fund

However, a major limitation of PCA-based monitoring is that PCA models assume that the process conditions do not change significantly, whereas this is rarely the case for WWTP. When a time-invariant PCA model is used to monitor processes with changing process conditions, false alarms often result. Process operators will quickly lose confidence with a monitoring system that puts forth many false alarms. Fortunately, this problem can be solved by making the PCA model adaptive (Li et al., 2000; Rosen and Lennox, 2001).

The SBR process has a cyclic nature, each cycle consisting of several phases — fill, anaerobic, aerobic, anoxic, and draw — depending on the objective of SBR operation. For multiphase batch processes, the basic phenomenon is that there are different covariance structures for the different phases. Based on this knowledge, it is natural to consider using different models for the various phases to achieve better results. Multiblock PCA methods have been proposed in the recent literature to improve the interpretability of multivariate models (Westerhuis et al., 1998). These multiblock methods have been used in cases in which the number of variables is large and additional information is available for blocking the variables into conceptually meaningful blocks. Applications include modeling and monitoring of large chemical processes (MacGregor et al., 1994; Wold et al., 1996). The multiblock PCA approach may have significant benefits when monitoring SBR processes. Each local model is intended to capture specific aspects of the corresponding phases of SBR processes.

In this study, a real-time monitoring system for batch processes is developed using an adaptive multiblock MPCA model to detect abnormal batch behaviors and to identify the major sources of process disturbances. Detection of abnormalities directly after completion of a batch cycle may even prevent detrimental effects on the following batch run. In this application the monitoring frequency is appropriate because SBR process monitoring is especially focused on the microorganism activity changes that have a time constant of a few days. The methodology is applied to a bench-scale SBR used to grow sludge that has the most stable properties possible, as this sludge is to be used in a comprehensive study of flocculation (Nopens et al., 2002). A SBR free from major problems is likely to be more stable than one subject to significant disturbances. The proposed MPCA supervision is used as a tool to monitor the stability of the sludge on the basis of simple on-line measurements.

MONITORING ALGORITHMS

Multiway Principal Component Analysis

In a typical batch run, $j = 1, \dots, J$ variables are measured at each of $k = 1, 2, \dots, K$ time intervals throughout the batch. Similar data will exist on a number of batches, $i = 1, 2, \dots, I$. All data can be summarized in the $\underline{X}(I \times J \times K)$ of a historical database (Fig. 1). Multiway PCA is equivalent to unfolding the three-dimensional data matrix, \underline{X} , into a large two-dimensional matrix, \underline{X} , and then performing a regular

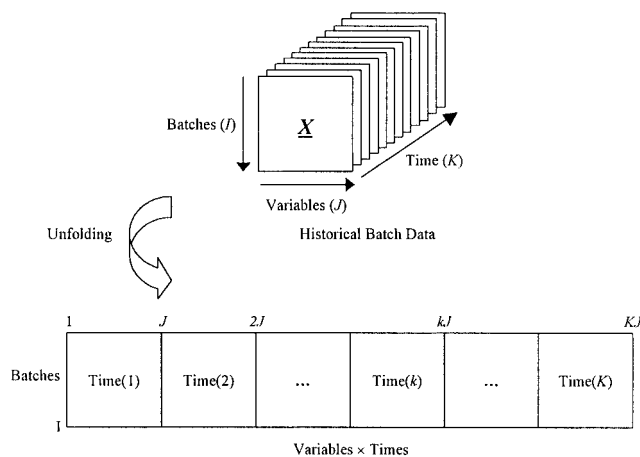


Figure 1. Decomposition of a three-way data matrix, \underline{X} , by MPCA.

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{X} is 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 analysis of the variability among the batches in \underline{X} by summarizing the information in the data with respect to both variables and their time variation.

Before a PCA is performed on the batch data, the data are normalized using the mean and standard deviation of each variable at each time interval in the batch cycle over all batches. Subtracting the average batch trajectory generally eliminates the major nonlinear behavior of the process (MacGregor et al., 1994). MPCA then decomposes the data into a series of principal components consisting of score vectors (\underline{t}_r) and loadings (\underline{p}_r), plus residuals (\underline{E}):

$$\underline{X} = \sum_{r=1}^R \underline{t}_r \underline{p}_r + \underline{E} \quad (1)$$

The loading vectors (\underline{p}_r) define the reduced dimension space (R) and are the directions of maximum variability. Each element of the score vectors (\underline{t}_r) corresponds to a single batch and depicts the overall variability of this batch with respect to the other batches in the database throughout the entire 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 ($R \ll \min(I, JK)$). 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, \underline{E} , is as small as possible in a least-squares sense. The nonlinear iterative partial least squares (NIPALS) algorithm can be used for sequential computation of the dominant principal components (Geladi and Kowalski, 1986).

Abnormal behavior of new batches is identified by projecting the new batches onto the model. Control charts that

are used in monitoring batch processes are generally based on the Q -statistic and D -statistic, 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 of zero. This is to be expected, because any linear combination of random variables, according to the central limit theorem, should tend toward a normal distribution.

The Q -statistic is a measure of the lack of fit with the established model. For batch number i , Q_i is calculated as:

$$Q_i = \sum_{j=1}^J \sum_{k=1}^K (e_{jk})^2 \sim g\chi_{(h)}^2 \quad (2)$$

where e_{jk} are the elements of E . Q_i indicates the distance between the actual values of the batch and the projected values onto the reduced space. The distribution of the calculated Q_i values can be approximated by a chi-square distribution, $g\chi_{(h)}^2$, where g is a constant and h is the effective degrees of freedom of the chi-square distribution.

The D -statistic, or Hotelling T^2 statistic, measures the degree to which data fit the calibration model:

$$D_i = \mathbf{t}_i^T \mathbf{S}^{-1} \mathbf{t}_i \sim \frac{I(I-R)}{R(I^2-1)} F_{R,I-R} \quad (3)$$

where \mathbf{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 an F -distribution, $F_{R,I-R}$, and confidence limits for the D -statistic are calculated from this F -distribution.

Adaptive MPCA

PCA monitoring is based on the assumption that the process conditions are time-invariant, whereas most real WWTPs are time-varying due to changes of influent characteristics, temperature, and microorganism activity. The time-varying characteristics of WWTPs include their mean, variance, and correlation among their measurements (Rosen and Lennox, 2001). When a time-invariant PCA model is used to monitor processes with time-varying behaviors, false alarms often result, significantly compromising 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. A window is a fixed-length batch data set of unfolded data matrix \mathbf{X} with respect to batch number. When a new batch is available, another window is created by omitting 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. The model for batch i is based on the following covariance matrix:

$$\mathbf{X}^T \mathbf{X}(i) = \sum_{n=0}^W (x_{jk}(i-n))^T x_{jk}(i-n) \quad (4)$$

where $\mathbf{X}^T \mathbf{X}(i)$ is the covariance matrix for batch i , W is the length of the window, and x_{jk} are the elements of \mathbf{X} .

Multiblock MPCA

Many industrial batch processes are operated in different phases. Usually, the input variable profiles and reaction conditions vary greatly from one phase to the next. A better alternative is then to divide the batch data sets into several blocks and build local MPCA models for each data block. This approach has significant benefits because the latent variable structure is allowed to change at each phase (Qin et al., 2001; Ranner et al., 1998; Wold et al., 1996). Analyzing the data with a multiblock model also allows for detecting more specific locations of faults in a process (Smilde et al., 2000).

Figure 2 shows the arrow scheme for a multiblock PCA algorithm. The data matrix \mathbf{X} ($I \times KJ$) is divided into B blocks ($\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_B$). In the lower layer of the model, each data block is considered as a separate source of information and the details of the blocks are modeled by corresponding block models. In the super layer, information from all blocks on the lower data level is combined and the relative importance of the different blocks, \mathbf{X}_b , for each dimension is obtained.

The multiblock PCA algorithm derives from a sequence of ‘‘NIPALS steps’’ and has the following formulation (Westerhuis et al., 1998):

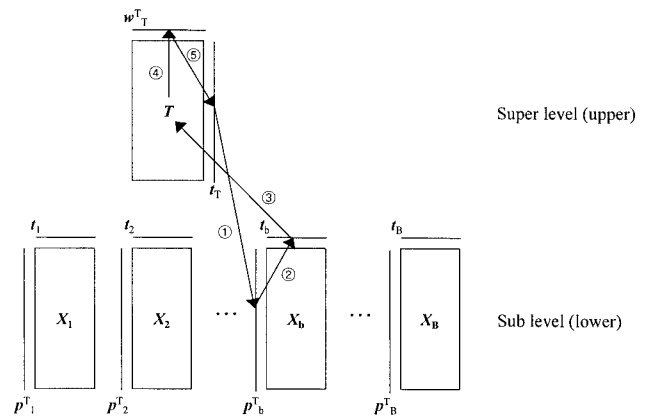


Figure 2. Arrow scheme for multiblock PCA algorithm.

1. Arbitrarily choose a column of one of the blocks, X_b , as t_T :

(1) Block variable loadings, p_b , are obtained by regressing t on all blocks, X_b :

$$p_b = X_b^T \cdot t_T / t_T^T \cdot t_T \quad (5)$$

then normalize p_b to $\|p_b\| = 1$.

(2) From p_b , block scores, t_b , for all blocks are calculated:

$$t_b = X_b \cdot p_b \quad (6)$$

(3) All block scores are combined into a super block, T :

$$T = [t_1, t_2, \dots, t_B] \quad (7)$$

(4) Super weight, w_T , is obtained by regressing t_T on the super block:

$$w_T = T^T \cdot t_T / t_T^T \cdot t_T \quad (8)$$

then normalize w_T to $\|w_T\| = 1$.

(5) A new t_T is calculated.

Until the super score, t_T , converges, a new iteration (1) to (5) starts.

2. Deflate block data matrix, X_b :

$$p_b = X_b^T \cdot t_T / t_T^T \cdot t_T \quad (9)$$

$$X_b = X_b - t_T \cdot p_b^T \quad (10)$$

Go to step 1 to calculate the next principal component.

The block Q_b - and D_b -statistics for batch i can be calculated as follows (Qin et al., 2001):

$$Q_{i,b} = \sum_{j=1}^J \sum_{k=1}^K (e_{b,jk})^2 \quad (11)$$

$$D_{i,b} = t_{i,b}^T S_{ib}^{-1} t_{i,b} \quad (12)$$

where e_{bjk} are the elements of the block residuals and $S_{i,b}$ is the covariance matrix of $t_{i,b}$.

PROCESS DESCRIPTION

Sequencing Batch Reactor

The data used in this study were collected from a bench-scale SBR system shown in Figure 3. A fill-and-draw sequencing batch reactor (SBR) system with an 80-L 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 h 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 a com-

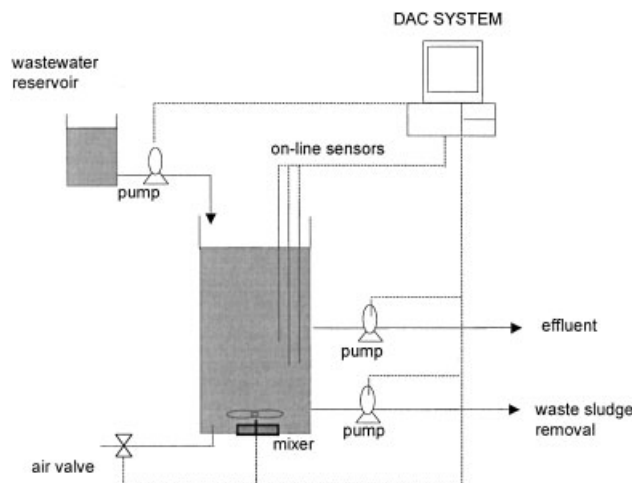


Figure 3. Schematic diagram of sequencing batch reactor system.

puter, interface cards, meters, transmitters, and solid-state relays (SSRs). 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 1 min for pH, ORP, DO, conductivity, temperature, and weight of the SBR reactor. Thus, no advanced or expensive measurement devices need to be installed to make the methods work. All measurements are simple, standard measurements. These measurements were stored for 280 past cycles (= 70 days), forming a database of historical information regarding the process.

RESULTS AND DISCUSSION

Pretreatment of SBR Data Sets

It has been reported that the on-line sensor values were somehow related to the dynamic characteristics of the nutrient concentrations (NH_4^+ , PO_4^{3-} , and NO_3^-) in the SBRs (Chang and Hao, 1996; Demuyne et al., 1994; Lee et al., 2001). The derivatives of the pH, ORP, and DO profiles can accurately detect the ends of phosphate release, ammonia conversion, and phosphate uptake, and can 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. Because 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, because biological reactions in the settling and drawing phases (corresponding to those of the last 60 time instants)

were assumed to be negligible. Moreover, due to the absence of mixing, the sensor signals were unreliable. The MPCA algorithms were applied to the three-way data array, \underline{X} , with dimensions $280 \times 12 \times 300$.

Analysis of Historical Process Data

Initially, the whole data set of 280 batches was analyzed by the multivariate statistical projection method of MPCA (see later). 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 for interpreting the behavior of the SBR process.

Figure 4 shows a score plot of the collected data in the space of the resulting first two principal components. The score plot can provide an adequate representation of the process behavior. Figure 4 presents two clusters: cluster 1 (batch $i = 1$ to 162) and cluster 2 (batch $i = 165$ to 280). The separation between cluster 1 and cluster 2 is due to disturbances (batch $i = 163$ to 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 the SBR operation was changing slowly, but switched to another operating state after disturbances occurred in the process. The disturbances are discussed in what follows.

Multiway PCA

A MPCA model was developed from the historical data set of a consecutive 30 batches. Five principal components, which explained approximately 73% of the total variability, were determined by cross-validation (Krzanowski, 1987). Validation data consisting of the remaining 250 batches were projected onto the model space. Figure 5 shows the Q - and D -statistic charts. It can be seen that the D -statistic barely exceeds its limit, whereas the Q -statistic is far above

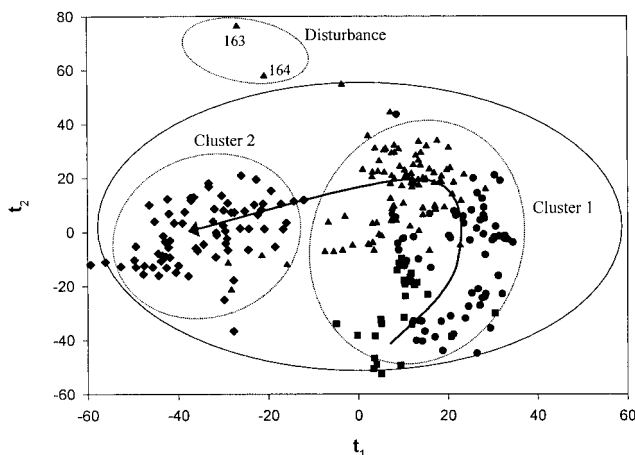


Figure 4. Score plot for all 280 batch data. The solid line corresponds to 95% confidence limit.

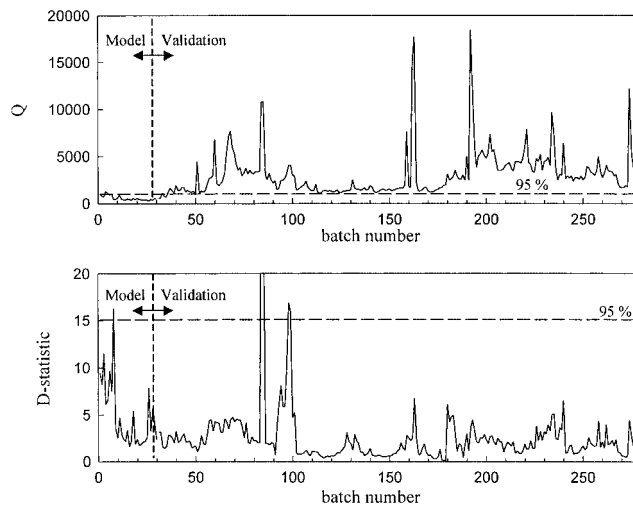


Figure 5. Multiway PCA. Q - and D -statistic charts with 95% confidence limits.

the 95% limit from batch 37 on. This indicates that the process variation in the validation batches cannot be described by the model developed. It is obvious that the static MPCA model is not valid because the fixed reference database is not representative of the SBR process that is time-varying and exposed to various disturbances such as influent composition variations, temperature changes, and equipment defects.

Adaptive MPCA

A monitoring model with adaptive covariance structure was subsequently developed to reduce the problem of changing process conditions. 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) to allow detection of slower disturbances as well as fast ones. 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. When a new block of batch data becomes available, the covariance matrix is updated over the selected window. Because the number of significant principal components can change over time, it is necessary to determine the number of principal components recursively. However, the cross-validation approach that was used for the MPCA model indicated is not suitable, because old data are not representative for the current process. Therefore, the number of significant principal components is calculated recursively using the cumulative percent variance (CPV) method (Li et al., 2000). The CPV is a measure of the percent variance captured by the first R principal components:

$$CPV(R) = \frac{\sum_{l=1}^R \lambda_l}{\text{trace}(V)} \times 100\% \quad (13)$$

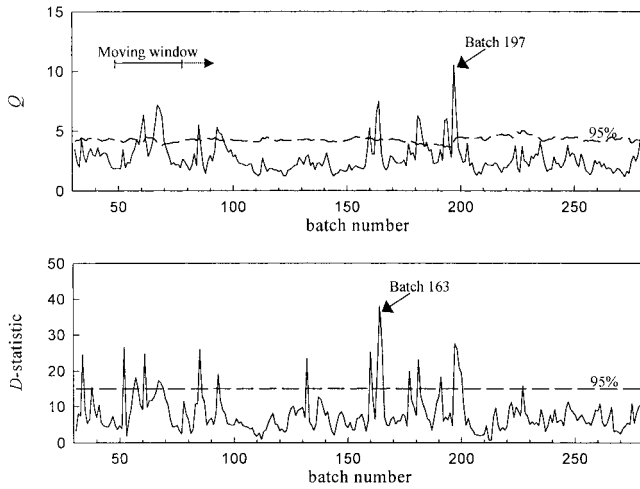


Figure 6. Adaptive MPCA. Q - and D -statistic charts with 95% confidence limits.

where the λ_l represents the eigenvalues of \mathbf{X} and $\mathbf{V} = \mathbf{E}\mathbf{E}^T / (I - 1)$. The number of principal components is chosen when CPV reaches a predetermined limit (75%). A potential adaptation problem is that the model may adapt not only to normal process evolution, but also to disturbances and failures. To prevent this, model updating was skipped when Q_i and D_i indices of new batches were exceeding certain limits.

Figure 6 shows Q - and D -statistics calculated from the adaptive PCA model as well as their 95% confidence level. The Q -statistic here was mostly well inside the confidence limits; this implies that as the covariance structure adapted to the new process conditions, and the updating model effectively captured the variability of the process and significantly reduced false alarms. The Q -statistic for batch number 197 to 198 far exceeded the 95% confidence limit, indicating that an unusual event occurred resulting in a change

of the process mean or covariance structure. The D -statistic violates its limit more distinctively compared with the static model in Figure 5. This implies that most of the variations are now in the model plane (D -statistic). The reason is that some of the variations in the Q -statistic chart have been transferred to the D -statistic chart as the model captured the variations in a correct way (Rosen and Lennox, 2001).

Adaptive Multiblock MPCA

The SBR cycle consists of four distinct phases wherein the latent variable structure can change due to the different environments. It is thus logical to break up the data matrix \mathbf{X} into four different blocks in such a way that the variables from one phase of a cycle can be blocked in the corresponding blocks: anaerobic; first aerobic; anoxic; and second aerobic (Fig. 7). The reference data set, $\mathbf{X}(280 \times 12 \times 300)$, is divided into four parts: $\mathbf{X}_1(280 \times 12 \times 60)$; $\mathbf{X}_2(280 \times 12 \times 150)$; $\mathbf{X}_3(280 \times 12 \times 60)$; and $\mathbf{X}_4(280 \times 12 \times 30)$.

The results presented indicate that adaptive MPCA methods should be used to capture the important variability of the SBR process. Therefore, a model with an adaptive covariance structure as well as a multiblock approach was used to monitor the same historical data set. The window size and method to determine the number of principal components for the adaptive multiblock MPCA were equal to those of the adaptive MPCA — both for simplicity and ease of comparison.

The multiblock monitoring approach can detect and diagnose a fault using the block $Q_{i,b}$ - and block $D_{i,b}$ -statistics in conjunction with the Q_r - and D_r -statistics of the super block. Because the same variable scalings of the adaptive MPCA were used for the adaptive multiblock MPCA model, the super Q_r - and D_r -statistics are identical to the Q_r - and D_r -statistics shown in Figure 6 (Qin et al., 2001). The

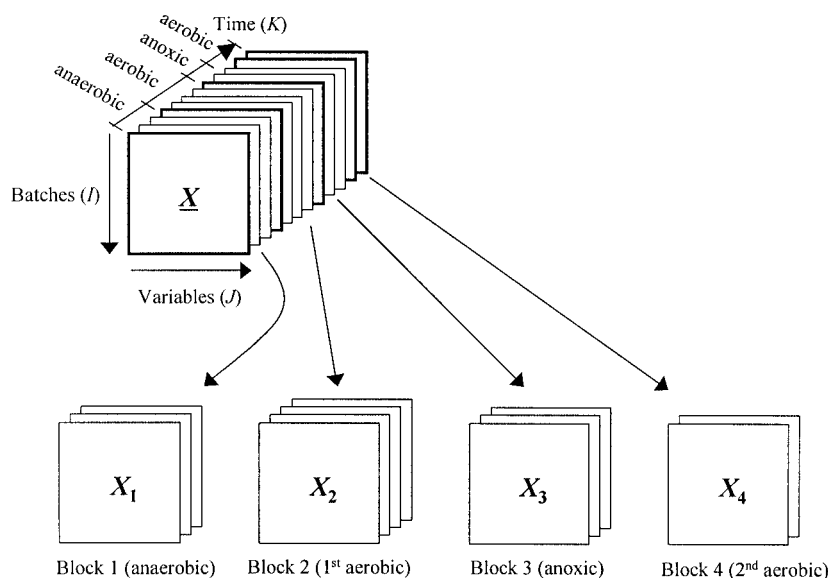


Figure 7. Dividing the three-way data matrix, \mathbf{X} , into four different blocks.

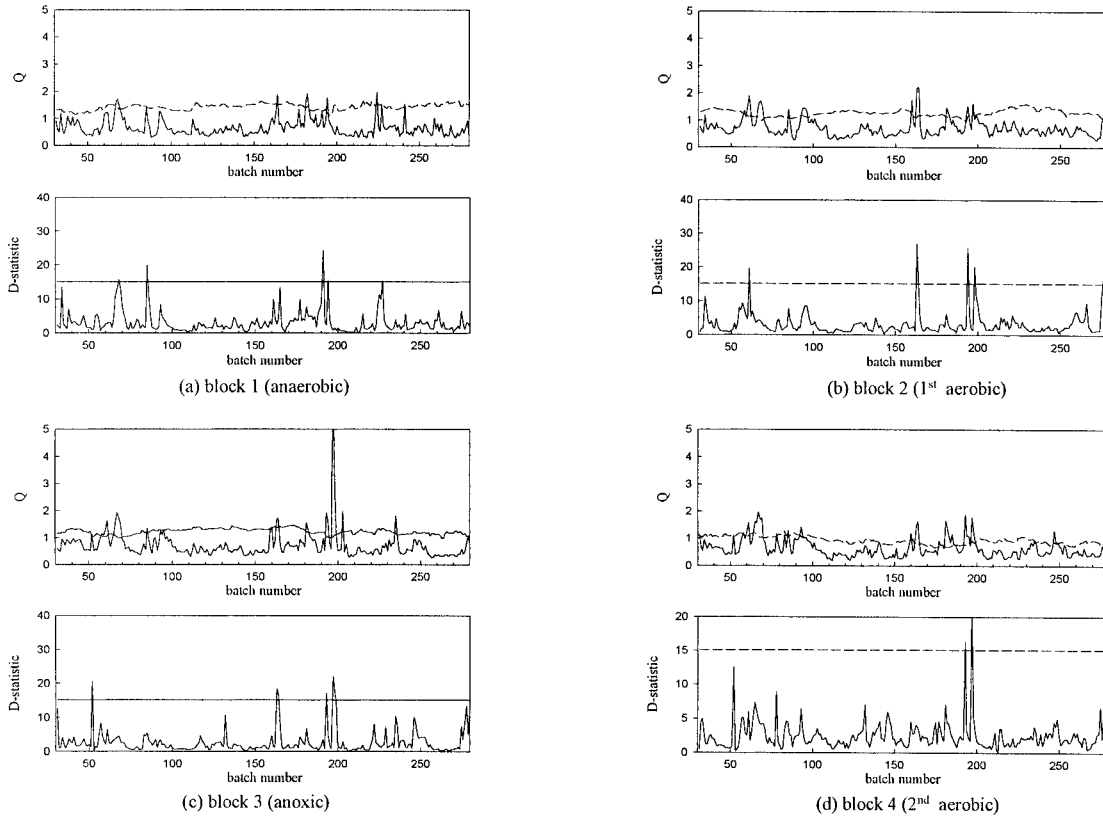


Figure 8. Adaptive multiblock MPCA. Q_i - and D_i -statistic charts for each block with 95% confidence limits.

block $Q_{i,b}$ - and $D_{i,b}$ -statistics were calculated for each block as shown in Figure 8. From Eqs. (2) and (11), the summation of block $Q_{i,b}$ -statistics is simply the Q_i -statistic of the adaptive MPCA in Figure 6. Therefore, adaptive multiblock MPCA monitoring simply groups the contributions to the

Q_i -statistic of adaptive MPCA in terms of blocks (Qin et al., 2001).

Abnormal batches were first monitored with the Q_i - and D_i -statistics in Figure 6. Block $Q_{i,b}$ - and block $D_{i,b}$ -statistics (Fig. 8) were then applied for decentralized monitoring and

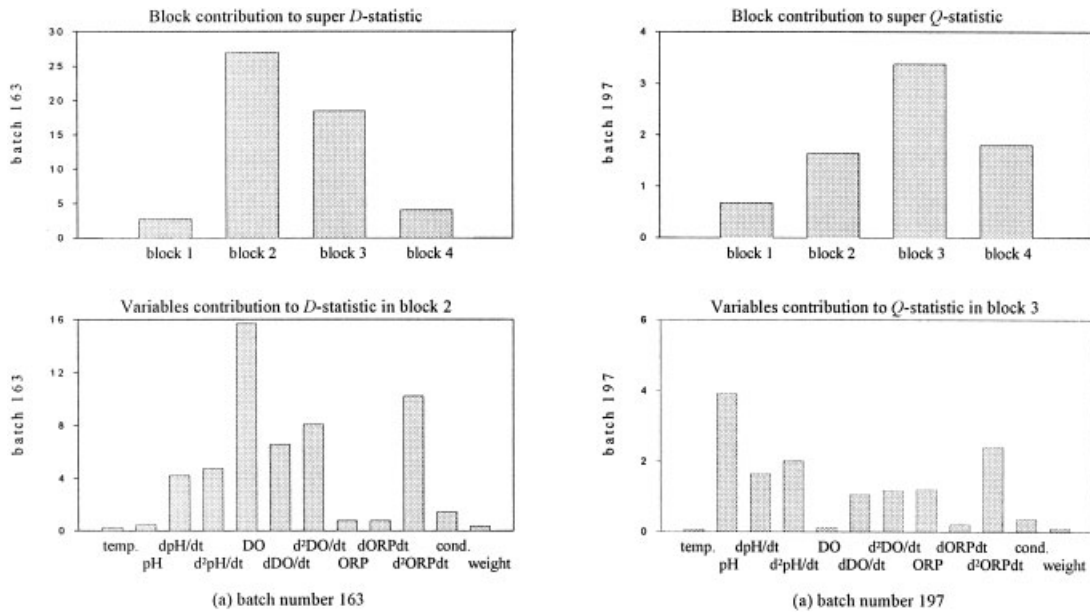


Figure 9. Identification of faulty blocks and variables for batches 163 and 197.

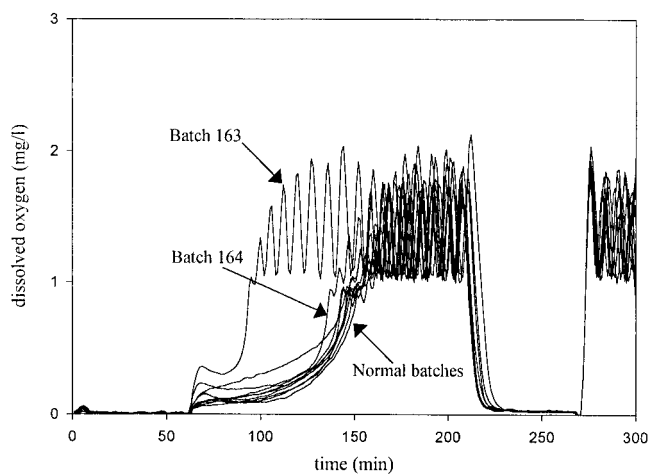


Figure 10. Univariate profile of the DO concentrations for normal batches and batch 163.

fault diagnosis. For batch 163, a disturbance that might have caused the shifting of the operational state (Fig. 4) was clearly alarmed in the D_i -statistic in Figure 6. The $D_{i,b}$ -statistic for each block (Fig. 8) shows that the fault was located mainly in block 2. To identify the disturbance for batch 163 a contribution plot of the block residuals is shown in Figure 9a. Again, it is quantitatively identifiable from Figure 9a that block 2 was the block in which the fault was mainly located. Figure 9a also shows the contribution to the D -statistic for block 2, which clearly identifies that DO contributed to the disturbance. Figure 10 shows the univariate plot of the DO concentrations in the SBR for normal batches and for batch 163. It can be seen that the DO concentrations in batch 163 had highly oscillatory peaks during the aerobic phase. These peaks were caused by an increase of the air flow rate to the SBR.

For batch 197, a disturbance was detected in the super Q -statistic chart of the residuals (Fig. 6). Consequently, block contributions to the Q -statistic should be investigated.

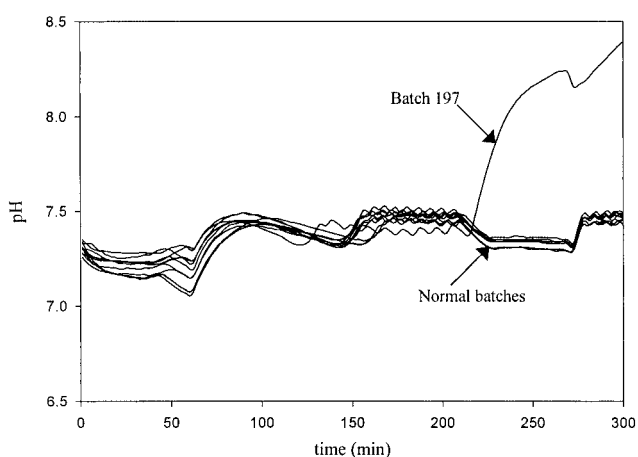


Figure 11. Univariate profile of the pH for normal batches and batch 197.

Figure 9b shows the calculated block contributions to the Q value for batch 197. This simply consisted of grouping the contributions to the Q -statistic for batch 197 in terms of the blocks. It is obvious that the main perturbation was present in block 3. The contributions to the Q -statistic in block 3 (Fig. 9b) show that pH contributed to the disturbances. Figure 11 shows the univariate plot of the pH for normal batches and for batch 197. This higher pH was indeed caused by an accidental feeding of acetic acid solution to the SBR process during the anoxic phase (Sin et al., 2002).

Compared with the adaptive monitoring results, the multiblock model helps to localize the potential cause of the fault or disturbance and gives a much clearer indication of the faulty variables. Therefore, the main advantage of the multiblock approach is to allow for easier interpretation of the data by looking at the relationship between blocks and at smaller meaningful block models.

CONCLUSIONS

A new monitoring algorithm, adaptive multiblock MPCA, has been developed. The method combines the elements of adaptive MPCA and multiblock PCA to monitor batch processes. Adaptive PCA, in terms of updating the covariance structure, overcomes the problem of changing operational conditions. A historical operational data set is split into several blocks based on the multiphase operation strategy of the SBR. Dividing the process data into meaningful blocks based on process knowledge makes it possible to localize the cause of a detected fault and disturbance in a decentralized manner and allows for clear indication of the faulty variables. The application of the proposed algorithm to a SBR process has demonstrated the feasibility and effectiveness of this adaptive process monitoring approach. The methodology is relatively simple, based on simple on-line sensor values, and can easily be included in standard MPCA monitoring. Finally, for further study, multiscale monitoring approaches (Bakshi, 1998; Rosen and Lennox 2001) could be integrated in the adaptive multiblock monitoring approach to adapt to a much wider time scale of changes.

The authors thank I. Nopens, R. Govoreanu, and F. Tola for helping in the operation of the SBR process, and in the pretreatment of on-line sensor values.

NOMENCLATURE

b	index for blocks
B	total number of blocks
CPV	cumulative percent variance
D	Hotelling statistic
E	residual matrix for historical database
$F_{R,I-R}$	F -distribution with R and $I - R$ degrees of freedom
g	constant associated with chi-square distribution
h	degrees of freedom for chi-square distribution
i	index for batches
I	total number of batches
j	index for measurement variables

J	total number of measurement variables
k	index for sampling time instants
K	total number of sampling time instants
n	index for moving window
p	loading vector
Q	squared prediction error
r	index for principal components
R	total number of principal components
S	covariance matrix of t scores
t	score vector
t_T	super score vector
$*V$	residual covariance matrix
W	length of the window
w_T	super weight vector
X	unfolded historical database
\underline{X}	historical database

Greek letters

λ	eigenvalue
χ	chi-square variable

References

- Bakshi BR. 1998. Multiscale PCA with application to multivariate statistical process monitoring. *AIChE J* 44:1596–1610.
- Chang CH, Hao OJ. 1996. Sequencing batch reactor system for nutrient removal: ORP and pH profiles. *J Chem Tech Biotechnol* 67:27–38.
- Demuyneck C, Vanrolleghem PA, Mingneau C, Liessens J, Verstraete W. 1994. NDBEPR process optimization in SBRs: Reduction of external carbon source and oxygen supply. *Water Sci Technol* 304:169–179.
- Geladi P, Kowalski BR. 1986. Partial least squares regression: A tutorial. *Anal Chim Acta* 185:1–17.
- Krzyszowski WJ. 1987. Cross-validation in principal component analysis. *Biometrics* 43:575–584.
- Lee DS, Jeon CO, Park JM. 2001. Biological nutrient removal enhancing anoxic phosphate uptake in a sequencing batch reactor. *Water Res* 35:206–214.
- Lee DS, Jeon CO, Park JM, Chang KS. 2002. Hybrid neural network modeling of a full-scale industrial wastewater treatment process. *Biotechnol Bioeng* 78:670–682.
- Lennox B, Montague GA, Hiden HG, Kornfeld G, Goulding PR. 2001. Process monitoring of an industrial fed-batch fermentation. *Biotechnol Bioeng* 74:125–135.
- Li W, Yue H, Valle-Cervantes S, Qin SJ. 2000. Recursive PCA for adaptive process monitoring. *J Proc Control* 10:471–486.
- MacGregor JF, Jaeckle C, Kiparissides C, Koutoudi M. 1994. Process monitoring and diagnosis by multi-block PLS methods. *AIChE J* 40: 826–838.
- MacGregor JF, Kourti T. 1995. Statistical process control of multivariate processes. *Control Eng Pract* 3:403–414.
- Nomikos P, MacGregor JF. 1994. Monitoring batch processes using multiway principal component analysis. *AIChE J* 40:1361–1375.
- Nomikos P, MacGregor JF. 1995. Multivariate SPC charts for monitoring batch processes. *Technometrics* 37:41–59.
- Nopens I, Biggs CA, Clercq BD, Govoreanu R, Wilen BM, Lant P, Vanrolleghem PA. 2002. Modeling the activated sludge flocculation process combining laser diffraction particle sizing and population balance modeling (PBM). *Water Sci Technol* 45:41–49.
- Qin SJ, Valle S, Piovoso MJ. 2001. On unifying multiblock analysis with application to decentralized process monitoring. *J Chemometrics* 15: 715–742.
- Rännér S, MacGregor JF, Wold S. 1998. Adaptive batch monitoring using hierarchical PCA. *Chemom Intell Lab Syst* 41:73–81.
- Rosen C, Olsson G. 1998. Disturbance detection in wastewater treatment systems. *Water Sci Tech* 37:197–205.
- Rosen C, Lennox JA. 2001. Multivariate and multiscale monitoring of wastewater treatment operation. *Water Res* 35:3402–3410.
- Savitzky A, Golay MJE. 1964. Smoothing and differentiation of data by simplified least squares procedures. *Anal Chem* 36:1627–1639.
- Sin G, Malisse K, Vanrolleghem PA. 2002. An integrated sensor for the monitoring of aerobic and anoxic activated sludge activities in biological nitrogen removal plants. In: *Proceedings of the International IWA Conference on Automation in Water Quality Monitoring*, Vienna, Austria. p 119–126.
- Smilde AK, Westerhuis JA, Boqué R. 2000. Multiway multiblock component and covariates regression models. *J Chemometrics* 14:301–331.
- Tates AA, Louwse DJ, Smilde AK, Koot GLM, Berndt H. 1999. Monitoring a PVC batch process with multivariate statistical process control charts. *Indust Eng Chem Res* 38:4769–4776.
- Teppola P, Mujunen SP, Minkkinen P. 1998. A combined approach of partial least squares and fuzzy c-means clustering for the monitoring of an activated-sludge waste-water treatment plant. *Chemom Intell Lab Syst* 41:95–103.
- Westerhuis JA, Kourti T, MacGregor JF. 1998. Analysis of multiblock and hierarchical PCA and PLA models. *J Chemometrics* 12:301–312.
- Wise BM, Gallagher NB. 1996. The process chemometrics approach to process monitoring and fault detection. *J Proc Control* 6:329–348.
- Wold S, Kettaneh N, Tjessen K. 1996. Hierarchical multiblock PLS and PC models for easier model interpretation and as an alternative to variable selection. *J Chemometrics* 10:463–482.