Environmental Monitoring and Assessment (2006) **119:** 349–366
DOI: 10.1007/s10661-005-9030-7

ON-LINE ADAPTIVE AND NONLINEAR PROCESS MONITORING OF A PILOT-SCALE SEQUENCING BATCH REACTOR

CHANG KYOO YOO[1,2,*], IN-BEUM LEE[2] and PETER A. VANROLLEGHEM[1]
[1]*BIOMATH, Department of Applied Mathematics, Biometrics and Process Control, Ghent University, Coupure Links 653, B-9000 Gent, Belgium;* [2]*School of Environmental Science and Engineering, Department of Chemical Engineering, Pohang University of Science and Technology, Pohang, Korea*
*(*author for correspondence, e-mail: ckyoo@postech.edu, ChangKyoo.Yoo@biomath.ugent.be)*

(Received 11 February 2005; accepted 23 August 2005)

Abstract. This article describes the application of on-line nonlinear monitoring of a sequencing batch reactor (SBR). Three-way batch data of SBR are unfolded batch-wisely, and then a adaptive and nonlinear multivariate monitoring method is used to capture the nonlinear characteristics of normal batches. The approach is successfully applied to an 80 L SBR for biological wastewater treatment, where the SBR poses an interesting challenge in view of process monitoring since it is characterized by nonstationary, batchwise, multistage, and nonlinear dynamics. In on-line batch monitoring, the developed adaptive and nonlinear process monitoring method can effectively capture the nonlinear relationship among process variables of a biological process in a SBR. The results of this pilot-scale SBR monitoring system using simple on-line measurements clearly demonstrated that the adaptive and nonlinear monitoring technique showed lower false alarm rate and physically meaningful, that is, robust monitoring results.

Keywords: adaptive batch monitoring, multiway kernel principal component analysis (MKPCA), nonlinear biological process, sequencing batch reactor (SBR)

1. Introduction

The increase in environmental restrictions in recent times has led to an increase in efforts aimed at achieving a better effluent quality of wastewater treatment plants. Achieving this goal requires advanced monitoring of the plant performance. Wastewater treatment plants are slow when they have to recover from a 'bad' state to a 'normal' state. The early detection and isolation of faults in the biological process is therefore very effective since it allows corrective action to be taken well before the situation becomes dangerous. Some process changes are not very obvious and may gradually grow until they become a serious operational problem. Process monitoring and fault detection of the biological processes are very important tasks in process engineering since they aim to ensure the success of the planned operations and to improve the productivity of processes.

In recent industrial process plants, many variables are measured in various operating units and are recorded in abundance. However, such data sets are highly

correlated and are subject to considerable noise. In the absence of an appropriate processing method, only limited information can be extracted, which causes insufficient understanding of the process by the operator and may lead to unstable operation. If properly treated, this data can provide a wealth of information leading to keep the plant operators understand the status of the process and assist them to make appropriate actions to remove abnormalities from the process (Rosen and Lennox, 2001; Yoo *et al.*, 2003).

Sequencing batch reactor (SBR) processes have demonstrated their efficiency and flexibility in the treatment of wastewaters with high concentrations of nutrients (nitrogen, phosphorous), and toxic compounds from domestic and industrial sources. A SBR has a unique cyclic batch operation, usually with five well-defined phases: fill, react, settle, draw and idle. Most of the advantages of SBR processes may be attributed to their single-tank designs and the flexibility that allows them to meet many different treatment objectives, and which is derived from the possibility of adjusting the duration of the different phases. Real-time control of the SBR process can contribute to this. A possible control strategy is based on the identification of the endpoint of a biological reaction. Switching to the next phase shortly after the detection of the reaction endpoint provides an optimum solution for both the process performance and the economics of the plant. In fact, if the duration of a phase is too short, the removal of the pollutants is not complete and the quality of the effluent will not meet the limits imposed by law. On the other hand, cycles which are longer than necessary decrease the capacity of the plant (volume of wastewater treated per day) or increase its operating costs; an aerobic phase which is too long would also mean wasting energy for aeration (Wilderer *et al.*, 2001).

However, the SBR process includes dynamic behaviour that is highly nonlinear, highly complex, carried out by a diverse microbial community, unpredictable and is further compromised by the fact that the effluent concentration is difficult to measure online and may only be available through offline laboratory analysis. Also, the SBR process is subject to significant disturbances like hydraulic changes, variability of influent composition, change in microbiological activity and equipment failures. Small changes in concentrations or flows can affect the kinetics of *nonlinear* biological reactions, which leads to batch-to-batch variability in effluent quality and microorganism growth. Moreover, compared to continuous wastewater treatment processes, SBR operation data have the added dimension of batch number, in addition to the measured variables and sample times (batches $\times$ variables $\times$ time), that is, a three-way matrix. Batch processes generally exhibit some batch-to-batch variation in the trajectories of the process variables. Normal variation is due to typical variations in the operation whereas special variations are due to exceptional phenomena. However, treatment performance, the key indicator of process performance, is often only examined off-line in a laboratory (Lee and Vanrolleghem, 2003; Yoo *et al.*, 2004a). Even though operators are aware that there are some problems in treatment performance, they cannot quickly find out or predict what

the causes are and when the problems will occur because most batch processes are run without any effective form of real-time on-line monitoring. Therefore, multivariate analysis and process monitoring of SBRs are crucial to detect faults that can be corrected prior to completion of the batch or can be corrected in subsequent batches because it may take several days, weeks or even months for the biological process to recover from abnormal operation.

Multiway principal component analysis (MPCA) and multiway independent component analysis (MICA) have been shown to be powerful monitoring tools in many industrial batch processes (Nomikos and MacGregor, 1994; Yoo *et al.*, 2004a,b). However, they have the shortcoming that the measurement variables of the batch process should be linear. Biological wastewater treatment is a complex, nonlinear and multivariate process, where many hydrodynamic and biological reactions occur simultaneously. A new nonlinear batch monitoring technique, called multiway kernel principal component analysis (MKPCA) has been emerging to tackle the nonlinear problem in recent years (Lee *et al.*, 2004). In this work, adaptive and multiway kernel principal component analysis (MKPCA), which extends MKPCA to adaptive biological batch processes, is proposed to overcome this drawback to obtain better batch monitoring performance of the pilot-scale SBR. Kernel PCA can efficiently compute principal components in high dimensional feature spaces by the use of integral operators and nonlinear kernel functions. Three-way batch data of the normal batch process are unfolded batch-wisely, and then the nonlinear multivariate feature extraction method is used to capture the adaptive and nonlinear characteristics within the SBR process.

2. Materials and Methods

2.1. Kernel principal component analysis (KPCA)

Kernel principal component analysis (KPCA) is an emerging technique to address the nonlinear problems not dealt with by PCA. As shown in Figure 1, conceptually, KPCA first performs a nonlinear mapping $\Phi(\cdot)$ from an input vector $\mathbf{x}$ to a high dimensional feature space F and then a linear PCA is performed in this feature space, which extracts the principal components t_k in a lower p dimensional KPCA space. Given any algorithm which can be expressed solely in terms of dot products, i.e. without explicit usage of the variables themselves, this kernel method enables us to construct different nonlinear versions of it. Compared to other nonlinear methods, the main advantage of KPCA is that no nonlinear optimization is involved. Based on these merits, KPCA has shown better performance than linear PCA in feature extraction and classification including nonlinearity (Schölkopf *et al.*, 1998).

To derive KPCA, we first map the data $\mathbf{x}_k \in R^m$, $k = 1, \ldots, N$ into a feature space F where N is the number of samples and compute the covariance

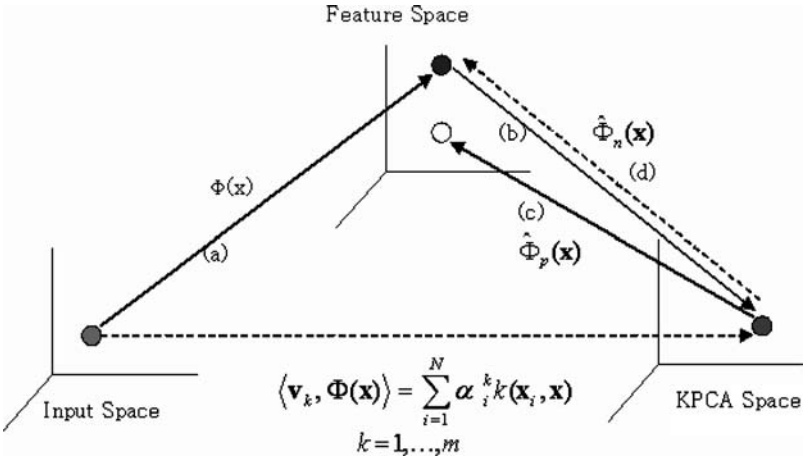

Figure 1. Conceptual diagram of KPCA.

matrix

$$\mathbf{C}^F = \frac{1}{N}\sum_{j=1}^{N} \Phi(\mathbf{x}_j)\,\Phi(\mathbf{x}_j)^T \tag{1}$$

where $\Phi(\cdot)$ is a nonlinear mapping function and it is assumed that $\Phi(\mathbf{x}_k)$ for $k = 1, \ldots, N$ is mean centered and variance scaled. Then, the principal components are computed by solving the eigenvalue problem

$$\lambda \mathbf{v} = \mathbf{C}^F \mathbf{v} = \frac{1}{N}\sum_{j=1}^{N} \langle \Phi(\mathbf{x}_j), \mathbf{v}\rangle\, \Phi(\mathbf{x}_j) \tag{2}$$

where $\lambda \geq 0$ denotes eigenvalues and v denotes the eigenvector of the covariance matrix C^F and $\langle \mathbf{x}, \mathbf{y}\rangle$ means dot product between $\mathbf{x}$ and $\mathbf{y}$. For $\lambda \neq 0$, solution v (eigenvector) can be regarded as a linear combination of $\Phi(\mathbf{x}_1), \ldots, \Phi(\mathbf{x}_N)$, i.e., $\mathbf{v} = \sum_{i=1}^{N} \alpha_i \Phi(\mathbf{x}_i)$. Multiplying $\Phi(\mathbf{x}_k)$ with both sides of Equation (2), we have

$$\lambda \sum_{i=1}^{N} \alpha_i \,\langle \Phi(\mathbf{x}_k), \Phi(\mathbf{x}_i)\rangle = \frac{1}{N}\sum_{i=1}^{N} \alpha_i \left\langle \Phi(\mathbf{x}_k), \sum_{j=1}^{N} \Phi(\mathbf{x}_j)\right\rangle \langle \Phi(\mathbf{x}_j), \Phi(\mathbf{x}_i)\rangle \tag{3}$$

Using the kernel trick, $[\mathbf{K}]_{ij} = K_{ij} = \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j)\rangle$, the eigenvalue problem can be expressed in a simplified form as,

$$N\lambda\boldsymbol{\alpha} = \mathbf{K}\alpha \tag{4}$$

where $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_N]^T$ and $\mathbf{K} \in R^{N \times N}$ is a gram matrix which is composed of K_{ij}. A justification of this procedure is given in Schölkopf *et al.* (1998). Then, performing PCA in the feature space F is equal to resolving the eigenvalue problem of Equation (4). This yields eigenvectors $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_N$ with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$. Dimensionality can be reduced by retaining only the first p eigenvectors. The projection, i.e., score vector, of the kth observation in the training data, is calculated by projecting $\Phi(\mathbf{x})$ onto the eigenvectors $\mathbf{v}_k$ in F where $k = 1, \ldots, p$.

$$t_k = \langle \mathbf{v}_k, \Phi(\mathbf{x}) \rangle = \sum_{i=1}^{N} \alpha_i^k \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}) \rangle \tag{5}$$

To solve the eigenvalue problem of Equation (4) and to project from the input space to the KPCA space using Equation (5), one can avoid the need for performing the nonlinear mappings and computing both the dot products in the feature space through introducing a *kernel function*, that is, $k(\mathbf{x}, \mathbf{y}) = \langle \Phi(\mathbf{x}), \Phi(\mathbf{y}) \rangle$. Representative kernel functions which satisfy Mercer's theorem are the polynomial, sigmoid, and Gaussian kernels.

2.2. Off-line batch monitoring using multiway nonlinear process monitoring

Batch processes are, by nature, leading to a 3-way matrix ($\underline{\mathbf{X}}(I \times J \times K)$) of data. In a typical batch run, $j = 1, 2, \ldots, J$ variables are measured at $k = 1, 2, \ldots, K$ time intervals throughout the batch. There exists similar data on several ($i = 1, 2, \ldots, I$) similar process batch runs. MPCA needs to unfold this matrix in order to obtain a two-way matrix, and then perform PCA. Figure 2 shows the unfolding method for MPCA. By subtracting the mean of each column of the unfolded matrix ($\mathbf{X}(I \times JK)$), the mean trajectory of each variable is removed, so that the major nonlinear behaviour of the process can be eliminated (Nomikos and MacGregor, 1994, 1995). Once the matrix is mean centered and variance scaled and PCA is performed, the results from PCA are the loading vectors and the calculated scores for each batch.

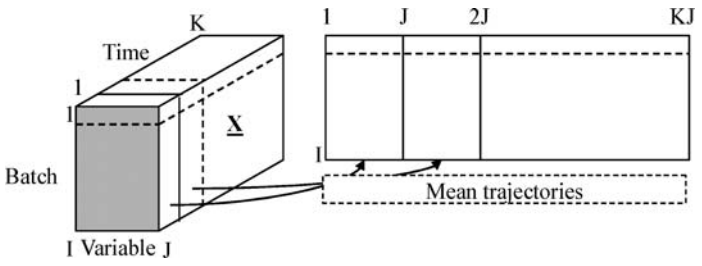

Figure 2. Unfolding method of MPCA for a three-way batch.

The loading vectors have a weight for each variable at each time, representing the history of the process. In this paper, KPCA instead of PCA to extract the nonlinear structure of the unfolded matrix is used (Lee *et al.*, 2003; Yoo *et al.*, 2004c).

2.2.1. *Batch Monitoring Based on MKPCA*
Adaptive batch monitoring scheme based on MKPCA is as follows:

(1) Acquire normal operating data $\underline{\mathbf{X}}(I \times J \times K)$ and unfold it batch-wisely $\mathbf{X}(I \times JK)$.
(2) The data $\mathbf{X}(I \times JK)$ are normalized using the mean and standard deviation of each variable at each time in the batch cycle over all batches.
(3) Given a set of JK-dimensional scaled normal operating data $\mathbf{x}_k \in R^{JK}$, $k = 1, \ldots, I$, compute the kernel matrix $\mathbf{K} \in R^{I \times I}$ by $[\mathbf{K}]_{ij} = K_{ij} = \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle = [k(\mathbf{x}_i, \mathbf{x}_j)]$.
(4) Carry out mean centering in the feature space for $\sum_{k=1}^{I} \tilde{\Phi}(\mathbf{x}_k) = 0$,

$$\tilde{\mathbf{K}} = \mathbf{K} - E\mathbf{K} - \mathbf{K}\mathbf{1}_I + \mathbf{1}_I\mathbf{K}\mathbf{1}_I \tag{6}$$

where each element of $\mathbf{E}$ is equal to $1/I$.
(5) Carry out variance scaling in the feature space for $\frac{1}{I-1} \sum_{k=1}^{I} \tilde{\Phi}_{\text{scl}}(\mathbf{x}_k)^2 = 1$

$$\tilde{\mathbf{K}}_{\text{scl}} = \frac{\tilde{\mathbf{K}}}{\frac{trace(\tilde{\mathbf{K}})}{I-1}} \tag{7}$$

(6) Solve the eigenvalue problem $I\lambda\boldsymbol{\alpha} = \tilde{\mathbf{K}}_{\text{scl}}\boldsymbol{\alpha}$ and normalize $\boldsymbol{\alpha}_k$ such that $\langle \boldsymbol{\alpha}_k, \boldsymbol{\alpha}_k \rangle = \frac{1}{\lambda_k}$.
(7) For normal operating data $\mathbf{x}$ at each batch, extract a nonlinear component via

$$t_k = \langle \mathbf{v}_k, \tilde{\boldsymbol{\Phi}}_{\text{scl}}(\mathbf{x}) \rangle = \sum_{i=1}^{I} \alpha_i^k \langle \tilde{\Phi}_{\text{scl}}(\mathbf{x}_i), \tilde{\Phi}_{\text{scl}}(\mathbf{x}) \rangle = \sum_{i=1}^{I} \alpha_i^k \tilde{k}_{\text{scl}}(\mathbf{x}_i, \mathbf{x}) \tag{8}$$

where $\tilde{\Phi}_{\text{scl}}(\mathbf{x})$ is the mean centered and variance scaled feature vector of $\Phi(\mathbf{x})$.
(8) Calculate the monitoring statistics (T^2 and SPE) at each batch and determine control limits of T^2 and SPE charts.

A measure of the variation summarized within the MKPCA model is given by the Hotelling's T^2 statistic. T^2 is the sum of the normalized squared scores, and is defined as

$$T^2 = [t_1, \ldots, t_p]\Lambda^{-1}[t_1, \ldots, t_p]^T \tag{9}$$

where t_k is obtained from Equation (5), p is the number of PCs and $\boldsymbol{\Lambda}^{-1}$ is the diagonal matrix of the inverse of the variances associated with the retained principal

components. The confidence limit for T^2 is obtained using the F-distribution:

$$T^2 \sim \frac{p(I^2-1)}{I(I-p)} F_{p,I-p,\alpha} \tag{10}$$

where I is the number of batches in the model, p is the number of principal components, and α is the significance level. The measure of goodness of fit of a sample to the PCA model is the squared prediction error (*SPE*), also known as the Q statistic. In this paper, we used the simple calculation of *SPE* in the feature space F suggested by Lee *et al.* (2004):

$$\text{SPE} = \|\Phi(\mathbf{x}) - \hat{\Phi}_p(\mathbf{x})\|^2 \tag{11}$$

where $\hat{\Phi}_p(\mathbf{x}) = \sum_{k=1}^{p} t_k \mathbf{v}_k$ is the reconstructed feature vector with p principal components in the feature space. The confidence limit for the *SPE* can be computed from its approximate χ^2 distribution

$$\text{SPE} \sim g\chi^2_{h,\alpha} \quad g = v/2m, \quad h = 2m^2/v \tag{12}$$

where m and v are the estimated mean and variance respectively of the *SPE* from the reference batches (Nomikos and MacGregor, 1994).

2.3. On-line adaptive batch monitoring using adaptive multiway kernel principal component analysis

When a batch has been monitored, we only know the values from the beginning to the current time. For on-line monitoring, however, test data should be completed until the end of the batch. Several methods for variable trajectory estimation have been proposed to complete the trajectories to the end of the batch. Nomikos and MacGregor (1994, 1995) suggest three different ways of dealing with this problem, i.e. to complete the remaining of the batches: (1) zero deviations, (2) current deviations (3) PCA projection method. Although the selection is dependent on the characteristics of the batch process, the second or the third method suggested by Nomikos and MacGregor (1994, 1995) is mainly used. For on-line monitoring, the distribution of the T_k^2 is approximated by Equation (10) and that of the SPE_k can be approximated by a weighted χ^2 distribution of $\text{SPE}_k \sim (v_k/2m_k)\chi^2_{2m_k^2/v_k}$, where m_k and v_k are the mean and variance of the SPE_k obtained for the data set used for model development at time instant k (Nomikos and MacGregor, 1995). In this work, adaptive and multiway kernel principal component analysis (MKPCA), which extends MKPCA to adaptive biological batch processes, is proposed to overcome this drawback to obtain better batch monitoring performance.

2.3.1. *On-line Adaptive Batch Monitoring*
The on-line adaptive batch monitoring procedure is as follows:

(1) For new batch data until time k, $\mathbf{X}_t(k \times J)$, unfold it to $\mathbf{x}_t^T(1 \times Jk)$. Scale it with the mean and the variance obtained from step 2) of the modeling procedure.
(2) Anticipate the future observations by the filling method which fills in all future measurements with the current deviation from the average batch.
(3) Apply an adaptive mean updating approach with exponential weighted moving average was used to remove non-stationary mean. A previous batch which lies within the 95% confidence limit is used to update the batch mean trajectory.
(4) Given JK-dimensional scaled batch data $\mathbf{x}_t \in R^{JK}$, we compute the kernel vector $\mathbf{k}_t \in R^{1\times I}$ by $[\mathbf{k}_t]_j = [k_t(\mathbf{x}_t, \mathbf{x}_j)]$ where $\mathbf{x}_j$ is the scaled normal operating data of the modeling procedures: $\mathbf{x}_j \in R^{JK}$, $j = 1, \ldots, I$.
(5) The test kernel vector $\mathbf{k}_t$ is adaptively mean centered as follows;

$$\tilde{\mathbf{k}}_t = \mathbf{k}_t - \mathbf{1}_t\mathbf{K} - \mathbf{k}_t\mathbf{1}_I + \mathbf{1}_t\mathbf{K}\mathbf{1}_I \tag{13}$$

where $\mathbf{K}$ and $\mathbf{1}_I$ are obtained from step (4) of the modeling procedure.
(6) The adaptive mean centered kernel vector $\tilde{\mathbf{k}}_t$ is variance scaled

$$\tilde{\mathbf{k}}_{t\,\mathrm{scl}} = \frac{\tilde{k}_t}{\frac{trace\,(\tilde{\mathbf{K}})}{I-1}} \tag{14}$$

(7) For the current batch $\mathbf{x}_t$, extract a nonlinear component via

$$t_k = \langle \mathbf{v}_k, \tilde{\Phi}_{\mathrm{scl}}(\mathbf{x}_t)\rangle = \sum_{i=1}^{I} \alpha_i^k \langle \tilde{\Phi}_{\mathrm{scl}}(\mathbf{x}_i), \tilde{\Phi}_{\mathrm{scl}}(\mathbf{x}_t)\rangle = \sum_{i=1}^{I} \alpha_i^k \tilde{k}_{t\,\mathrm{scl}}(\mathbf{x}_i, \mathbf{x}_t) \tag{15}$$

(8) Calculate the monitoring statistics (T^2 and SPE) of the current batch
(9) Monitor whether T^2 or SPE of the current batch exceeds its confidence limit calculated in the modeling procedure.

The procedures outlined above have various merits for on-line monitoring of adaptive batch processes. First, batch disturbances will be easily detected because the major nonlinear dynamics (mean trajectory) are removed in the preprocessing. Second, MKPCA is used to capture the nonlinearity which occurs in the biological metabolic reaction pathway, which contains the major nonlinear dynamic information across batches. Third, the adaptive mean update through the batches is considered to incorporate the evolving characteristics of the biological batch process, which can contribute to ensuring consistent, long-term performance of the monitoring system.

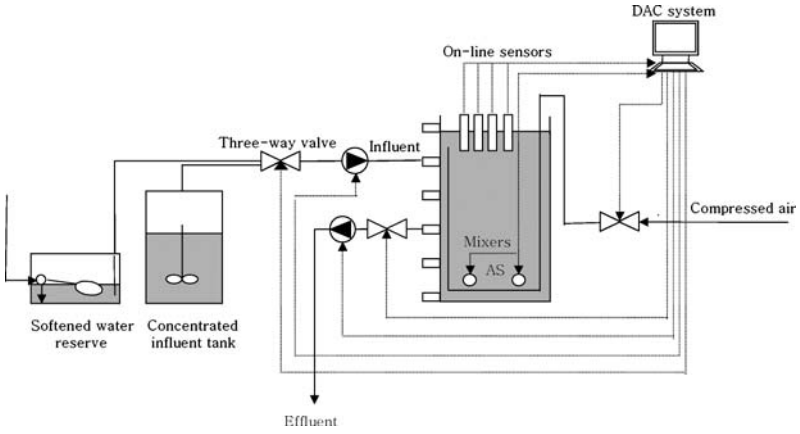

Figure 3. Schematic diagram of the pilot-scale sequencing batch reactor.

2.4. SBR PROCESS

The data used in this research were collected from a pilot-scale SBR system shown in Figure 3. A fill-and-draw sequencing batch reactor (SBR) with a 80-L working volume is operated in a 6 h cycle mode and each cycle consists of fill/anaerobic (1 h), aerobic (2 h 30 min), anoxic (1 h), re-aerobic (30 min) and settling/draw (1 h) phases. The hydraulic retention time (HRT) and the solid retention time (SRT) are maintained at 12 h and 10 days, respectively. The control 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. Six electrodes for pH, oxidation-reduction potential (ORP), dissolved oxygen (DO), temperature, conductivity and weight are connected to the system to check the status of the SBR, where a set of on-line measurements is obtained every one minute. Thus, no advanced nutrient or expensive measurement devices were installed in order to run an on-line monitoring algorithm of the SBR process (Lee and Vanrolleghem, 2003; Yoo *et al.*, 2004a). The status of the SBR reactor is displayed on the computer and the sensor signals are stored. In this research, we considered 150 batches in the historical data set of the SBR. The adaptive batch monitoring algorithms were applied to the three-way data array $\mathbf{X}$ with dimensions $150 \times 6 \times 300$.

3. Results and Discussions

3.1. OFF-LINE BATCH MONITORING OF SBR

Figure 4 shows the Hotelling's T^2 and *SPE* charts of the MPCA and the MKPCA method for all 150 batches with six on-line measurements. To make a fair comparison, we used the same number of principal components for both the MPCA

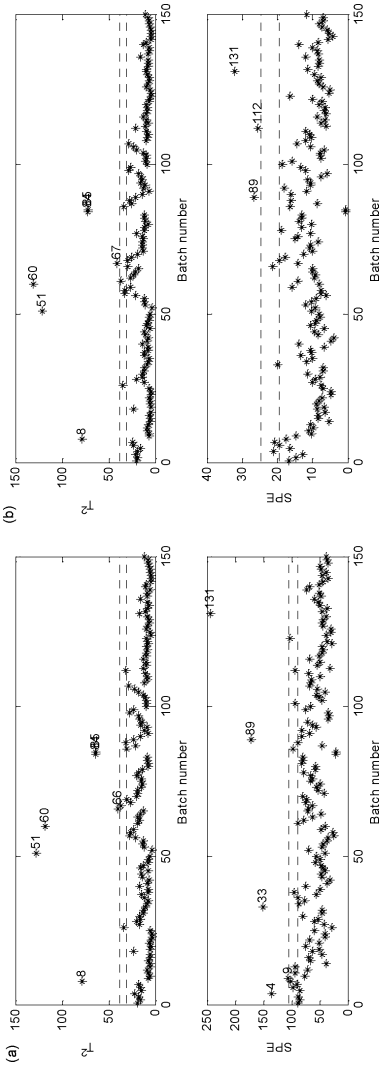

Figure 4. The Hotelling's T^2 and SPE charts of 150 batches of (a) the MPCA and (b) the MKPCA method with the six on-line measurements. The dotted lines correspond to the 99% confidence limits.

and the MKPCA. Sixteen principal components (PCs) in the MPCA model of six on-line measurements were retained, explaining 96.6% of the variation of the input space. MKPCA selected the Gaussian kernel, $k(\mathbf{x}, \mathbf{y}) = \exp(-\|x - y\|^2/\delta)$ with $\delta = rm\sigma^2$, where r is a constant determined by the process to be monitored which is 10 in this research, m is the dimension of the input space, and σ^2 is the variance of the data (Mika *et al*., 1999). Sixteen PCs in the MKPCA model were retained by the broken-stick rule (Nomikos and MacGregor, 1994) with 92.87% of the variation in the feature space explained.

Figure 4 shows the T^2 and the *SPE* charts of the MPCA and the MKPCA model. Compared to MPCA, MKPCA points to a lower number of abnormal batches in the SBR. The T^2 charts of MPCA and the MKPCA model show very similar monitoring results. There are three batches (batches 3, 9, 33) which are different between *SPE* charts of the MPCA and the MKPCA model. In the *SPE* chart, MPCA assigned five batches as abnormal, i.e. being significantly different from the other batches, i.e., batches 4, 9, 33, 89, 131, whereas MKPCA assigned three batches as abnormal, i.e., batches 89, 112, 131. When we checked the historical database, batches 89 and 131 were normal and batches 4, 9, 33, 112 were normal. This effect results from the fact that MKPCA can effectively capture the nonlinear relationship among batches. This result shows that the monitoring result from the *SPE* charts of MPCA may suffer from an oversensitivity to normal batches when MPCA is applied to a nonlinear process such as SBR.

There is a significant difference between the magnitude of *SPE* of the MPCA and the MKPCA model. MPCA is not able to capture the nonlinear dynamics of the SBR, which increases the modeling error of the MPCA. Compared to the MKPCA model, a lot of batches of the MPCA model are in the vicinity of the 95% confidence limit of the *SPE* values. This may come from the *nonlinear* biological kinetics leading to batch-to-batch variability in effluent quality and microorganism growth. On the other hand, MKPCA substantially extracts nonlinear principal components and therefore allows spreading the information regarding the data structure more widely giving a better opportunity to discard some of the eigendirections where the nonlinear part of the data resides. Thus, the MKPCA model has much lower *SPE* values than the MPCA model and MKPCA provides the nice capability of feature extraction and denoising, yielding a robust monitoring system. This observation is the confirmation of the nonlinear batch monitoring.

We can assess the impact of the nonlinearity through the use of QQ plots or normality probability plot. To check the nonlinearity of the SBR due to the periodic influent variations and nonlinear biological reactions, we checked the QQ plot of the principal component (t_2) of the MPCA and the MKPCA model which are composed of only the normal batches. Figure 5 shows the QQ plot of the second score (t_2) of the MPCA and the MKPCA model. From this figure, we can deduce that the SBR process has severe nonlinear dynamic relations and the extracted principal scores of MPCA have a large deviation from linearity. It means that MPCA has lower modeling ability which may lead to higher false alarm rate of a monitoring

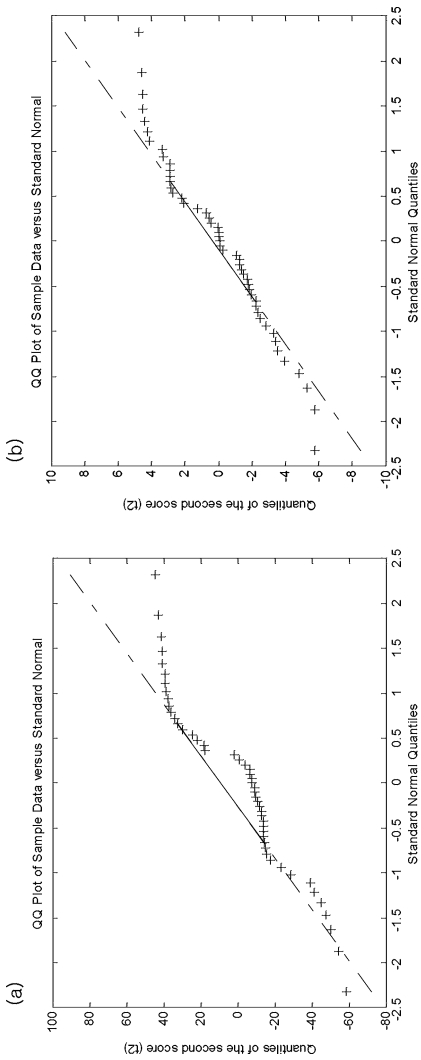

Figure 5. QQ plot for the second score (t_2) of (a) MPCA (b) MKPCA.

result of MPCA. On the other hand, MKPCA can extract the nonlinear principal components, that is, capture the nonlinearity in SBR data.

3.2. On-line adaptive monitoring of SBR

Based on the off-line analysis, 49 batches are modeled for the on-line monitoring of MPCA and MKPCA. For fair comparisons, eight principal components were used for both modeling. To fill in the future values in $\mathbf{X}_{\text{new}}$, we used filling method which fills in all future measurements with the current deviation from the average batch. A SBR process itself evolves over time as the microorganisms adapt to changing operating conditions like surrounding temperature and varying process loads. This often results in false alarms and significantly compromises the reliability of the monitoring system. To overcome the problem of changing process conditions, an adaptive mean updating approach with exponential weighted moving average was used to remove non-stationary mean. This has contributed to ensuring consistent, long-term performance of the monitoring system. In this approach, a previous batch which lies within the 95% confidence limit is used to update the batch mean trajectory. Then the MPCA and MKPCA models are tested against a new batch using a 99% control limit. New on-line data of a batch of SBR are monitored for every time point k with the monitoring charts based on the MPCA and MKPCA model.

Figure 6 shows the on-line adaptive monitoring results of MPCA and MKPCA with the 99% confidence limits in a particular, normal batch (batch number 30). The batch is monitored for every time instant k in terms of its T^2 and *SPE* charts. Both MPCA and MKPCA show that the T^2 charts for this batch are within the control limits for the whole duration of the batch run. However, the *SPE* chart of MPCA in Figure 6(a) exceeds the confidence limit one time, around the 60th sampling time, that is, the MPCA invokes a false alarm. It results from the nonlinear phenomenon occurring during the phase change (from anaerobic to aerobic). On the other hand, the *SPE* chart of MPCA in Figure 6(b) does not show any violation for its confidence limit during the whole duration of the batch run. Therefore, this batch in MKPCA is assigned as being "in control" or "normal". It illustrates that the selected nonlinear principal components in MKPCA can extract the dynamic characteristics of the SBR operation, i.e. the change from the anaerobic to aerobic phase.

Figure 7 represents the on-line adaptive monitoring results of the MPCA and MKPCA model for an *abnormal* batch (batch number 60). Both methods can detect this batch as an abnormal one from off-line monitoring (Figure 5). Both the MPCA and MKPCA methods show similar detection times for this batch. The T^2 and *SPE* charts of MPCA in Figure 7(a) show that this abnormal batch has a large deviation from the 61th time instant until the end of the batch operation, that is, the adaptive monitoring result of MPCA calls that the fault continues until the end of the batch. Indeed, the DO concentration in batch number 60 was increased too early (in the anaerobic phase) but returned to the normal concentration during the aerobic phase,

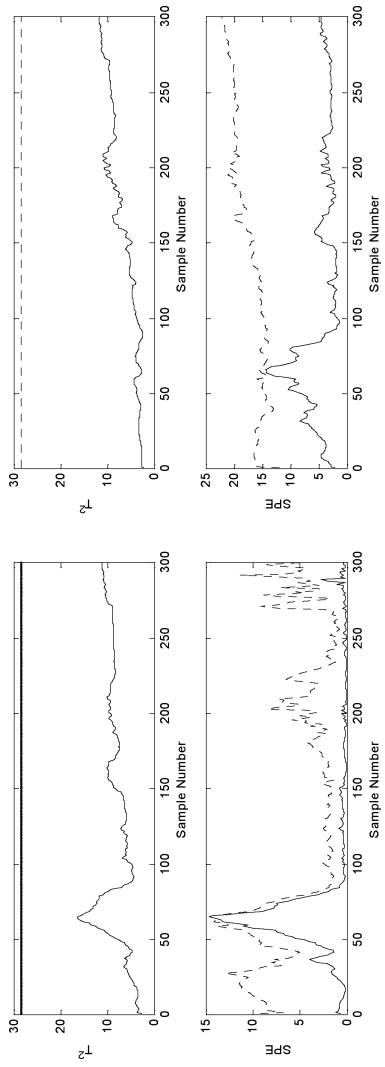

Figure 6. On-line adaptive monitoring charts for (a) MPCA and (b) MKPCA in case of a normal batch (batch number 30). The dotted lines correspond to the 99% confidence limits.

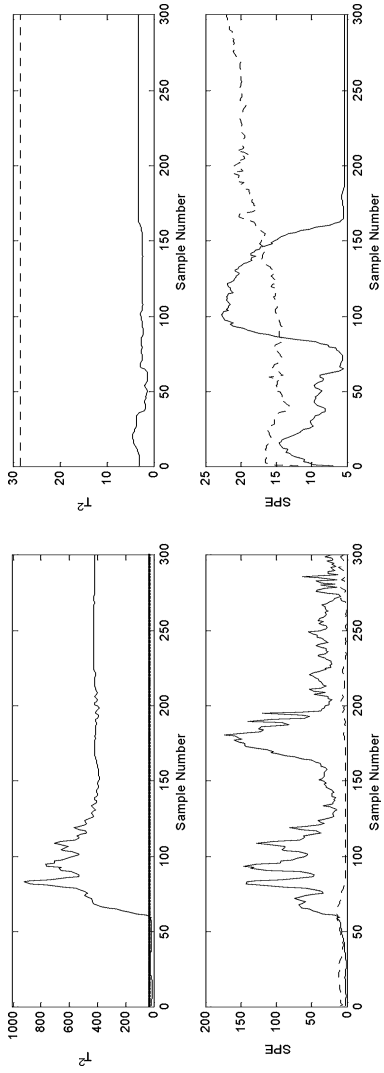

Figure 7. On-line adaptive monitoring charts for (a) MPCA and (b) MKPCA in case of an abnormal batch (batch number 60). The dotted lines correspond to the 99% confidence limits.

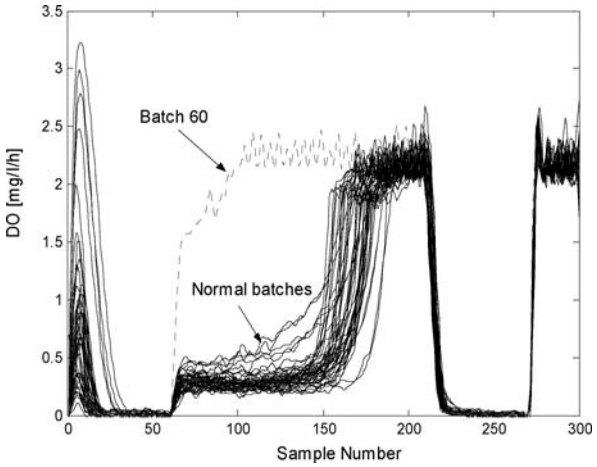

Figure 8. Univariate plot of on-line DO measurement of 49 normal batches and the abnormal batch 60.

see Figure 8. A nonlinearity of the SBR system might cause this lasting false fault detection. On the other hand, the *SPE* charts of AMKPCA in Figure 7(b) detects this abnormal batch behaviour around the 80th time instant but returns below the control limits around the 150th batch instant. It means that AMKPCA can detect this fault during the aerobic phase and return within the control limits after the fault is released.

This false alarm of MPCA comes from a bad modeling result. On the other hand, MKPCA can detect a fault *only* during aeration phase which is a physically meaningful and robust monitoring result. Hence, it results in less false alarms. This result shows that the extracted 8 PCs of MKPCA capture the underlying (nonlinear) factors from the SBR and push an abnormal batch outside the normal operating region and again pull it into the normal operating region. Typically, the false alarm rate of the MPCA is higher as a normal batch might be judged as a non-conforming one. Therefore, this deteriorates the reliability of the monitoring system. The above clearly showed that the nonlinear adaptive monitoring technique can capture the nonlinear dynamics of the SBR better than linear MPCA. Figure 8 shows an univariate plot of the on-line DO measurements of the normal batches and the abnormal batch 60 which was over-aerated at the start of the aeration phase. Thus, when a fault is detected in biological batch process, the batch run will control the current batch and moreover, operators will use the information on the fault to correct the following batch.

In the SBR operation, the influent wastewater is fed into the reactor and mixed with already existing microorganisms. Therefore, the performance of the current batch highly depends on microorganism activity in the previous batches. In addition, the SBR process is subject to significant disturbances like hydraulic changes and

composition variations. Small changes in concentrations or flows can have a large effect on the kinetics of biological rations leading to batch-to-batch variability in effluent quality and microorganism growth. Compared to the previous ones, the proposed method provides more meaningful information on the nonlinear evolving biological process and captures the nonlinear relation among batches, which results in more robust monitoring performance.

4. Conclusions

Biological wastewater treatment such as performed in SBRs is a complex, nonlinear and multivariate process, where many hydrodynamic and biological reactions occur simultaneously. Linear MPCA has the shortcoming that the measurement variables of the batch process should be linear. The developed adaptive and nonlinear monitoring method was successfully applied to an 80 L SBR. In off-line and on-line batch monitoring, it can effectively capture the nonlinear relationship among process variables. The results of the pilot-scale SBR monitoring study clearly showed that the adaptive and nonlinear monitoring technique generates less false alarms, physically meaningful and robust monitoring results in comparison to linear MPCA.

Acknowledgments

This work was supported by Korea Research Foundation Grant (KRF-2004-041-D00151) and a Visiting Postdoctoral Fellowship of the Fund for Scientific Research-Flanders (FWO).

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.

Lee, D. S. and Vanrolleghem, P. A.: 2003, 'Monitoring of a sequencing batch reactor using adaptive multiblock principal component analysis', *Biotechnol. Bioeng.* **82**, 489–497.

Lee, J., Yoo, C. K. and Lee, I.-B.: 2003, 'Fault detection of batch processes using multiway kernel principal component analysis', *Comp. & Chem. Eng.* 2004 (in press).

Mika, S., Schölkopf, B., Smola, A. J., Müller, K.-R., Scholz, M. and Rätsch, G.: 1999, 'Kernel PCA and de-noising in feature spaces', *Adv. Neural Information', Proc. Syst.* **11**, 33–37.

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.

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

Schölkopf, B., Smola, A. and Müller, K.: 1998, 'Nonlinear component analysis as a kernel eigenvalue problem', *Neural Comput.* **10**(5), 1299–1399.

Wilderer, P., Irvine, R. L. and Goronszy, M. C.: 2001, 'Sequencing Batch Reactor Technology', *IWA Scientific and Technical Report*, IWA publishing, London, UK.

Yoo, C. K., Vanrolleghem, P. A. and Lee, I.: 2003, 'Nonlinear modeling and adaptive monitoring with fuzzy and multivariate statistical method in biological wastewater treatment plant', *J. Biotech.* **105**(1/2), 135–163.

Yoo, C. K., Lee, D. S. and Vanrolleghem, P. A.: 2004a, 'Application of multiway ICA for on-line monitoring of a sequencing batch reactor', *Wat. Res.* **38**(7), 1715–1732.

Yoo, C. K., Lee, J., Vanrolleghem, P. A. and Lee, I.: 2004b, 'On-line monitoring of batch process with multiway independent component analysis', *Chemom. and Intel. Lab. Sys.* **71**(2), 151–163.

Yoo, C. K., Villez, K., Lee, J., Lee, I. and Vanrolleghem, P. A.: 2004c, 'Nonlinear multivariate statistical monitoring of a sequencing batch reactor (SBR)', Submitted to IWA symposium on Symposium Systems Analysis and Computing in Water Quality Management, Beijing, China, 5–7 November, 2004.

Zhang, J., Martin, E. B. and Morris, A. J.: 1997, 'Process monitoring using non-linear statistical techniques', *Chem. Eng. J.* **67**, 181–189.