Multivariate Nonlinear Statistical Process Control of a Sequencing Batch Reactor

Chang Kyoo YOO[1,2], Kris VILLEZ[2],
In-Beum LEE[1] and Peter A. VANROLLEGHEM[2]
[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, San 31 Hyoja Dong, Pohang, 790-784, Korea*

Keywords: Batch Monitoring, Bioprocess, Kernel Principal Component Analysis, Multivariate Statistical Process Control, Nonlinear Multivariate Analysis and Monitoring, Sequencing Batch Reactor (SBR), Wastewater Treatment Process (WWTP)

This research describes the application of a multivariate statistical process control method to a pilot-scale sequencing batch reactor (SBR) using a batchwise nonlinear monitoring technique for a denoising effect. Three-way batch data of normal batches are unfolded batch-wise and then a kernel principal component analysis (KPCA) is applied to capture the nonlinear dynamics within normal batch processes. The developed monitoring method was successfully applied to an 80-*l* sequencing batch reactor (SBR) for biological wastewater treatment, which is characterized by a variety of nonstationary and nonlinear characteristics. In the multivariate analysis and batch-wise monitoring, the developed nonlinear monitoring method can effectively capture the nonlinear relations within the batch process data and clearly showed the power of nonlinear process monitoring and denoising performance in comparison with linear methods.

Introduction

The increase in environmental restrictions in recent times leads to possibilities for advanced process monitoring and control of plant performance. Most of the changes in wastewater treatment plants are slow when the process is recovering from a 'bad' state to a 'normal' state. Early detection and isolation of faults in chemical and biological processes are very effective because they allow corrective action to be taken well before the situation becomes dangerous. Some changes are not so obvious and may gradually grow until they cause any serious operational problem (Kano *et al*., 2001, 2003; 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. The SBR under study has a unique cyclic batch operation, usually with five well-defined phases: fill, react, settle, draw and idle. Most of the advantages of the SBR processes may be attributed to their single-tank designs and the flexibility that allows them to meet many different treatment objectives, which is derived from the possibility of adjusting the duration of the different phases. However, the SBR process is highly nonlinear, time-varying and subject to significant disturbances like hydraulic changes, composition variations and equipment failures. Small changes in concentrations or flows can affect the effluent quality and microorganism growth. Moreover, compared to continuous wastewater treatment processes, SBR operation data have an added dimension of the batch number, in addition to the measured variables and sample times (batches × variables × time), that is, a three-way matrix. Batch processes generally exhibit some batch-to-batch variations in the trajectories of the process variables. However, treatment performance, the key indicator of process performance, is often only examined off-line in a laboratory. Even though operators might be aware that there are some problems in treatment performance, they are often not able to 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, a multivariate analysis and process monitoring of SBRs are crucial to detect faults that can be corrected prior to

Received on May 3, 2005. Correspondence concerning this article should be addressed to C. K. Yoo (E-mail address: ckyoo@postech.edu or ChangKyoo.Yoo@biomath.ugent.be).

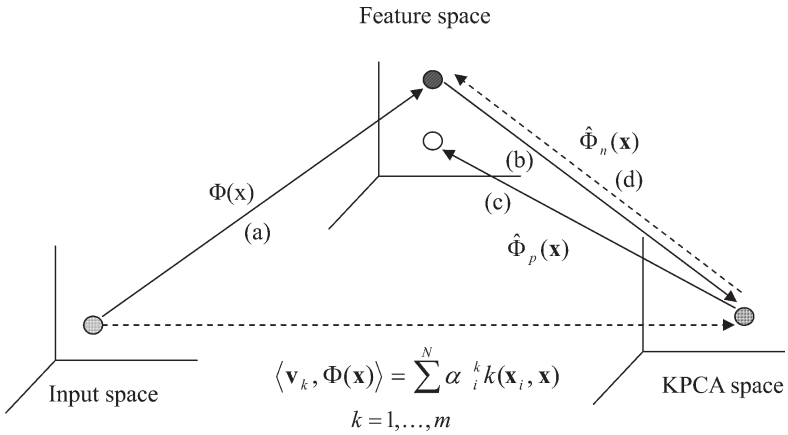

Fig. 1 Conceptual diagram of KPCA

completion of the batch or can be corrected in subsequent batches because the recovery of the biological process from abnormal operation may take several days, weeks or even months (Lee and Vanrolleghem, 2003; Yoo *et al*., 2004).

A multiway principal component analysis (MPCA) and a 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*., 2004). However, for some complicated cases in industrial chemical, biological and environmental processes which especially have nonlinear characteristics, a principal component analysis (PCA) exhibits bad behaviour because of its linearity assumption. They have a shortcoming that the measurement variables of the batch process should have linear correlations. In this work, a multiway kernel principal component analysis (MKPCA) is used to tackle the nonlinear problem and obtain better batch monitoring performance of the pilot-scale SBR. It 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 nonlinear PCA is used to capture the nonlinear characteristics within the batch processes. In this work, it is shown that nonlinear multivariate statistical process control to monitor an 80-*l* SBR for biological wastewater treatment can be used to overcome this drawback and obtain better monitoring performance.

1. Methods

1.1 Kernel principal component analysis (KPCA)

In industrial processes where severe nonlinear correlations exist among process variables, linear statistical techniques are not very effective in reducing the process data dimensions. If a linear PCA is used in these processes, a large number of PCs are required to explain sufficient data variance. For nonlinearly correlated data, the results from linear PCA may be inadequate because minor components can contain important information on nonlinearity. By discarding the minor components, this important information nonlinearity is lost. However, if these minor components are kept, the linear methods may require too much information to be useful. For the process monitoring with nonlinearity, nonlinear statistical techniques are more appropriate (Zhang *et al*., 1997; Yoo *et al*., 2004).

The kernel principal component analysis (KPCA) is an emerging technique to address the nonlinear problems on the basis of PCA. The conceptual framework of the KPCA method is shown schematically in **Figure 1** (Romdhani *et al*., 1999). First, the KPCA performs nonlinear mapping $\Phi(\cdot)$ from an input vector $\mathbf{x}$ to a high-dimensional feature space F (step (a)). Then, a linear PCA is performed in this feature space, which gives score values t_k in a lower p-dimensional KPCA space (step (b)). In order to reconstruct a feature vector) ($\Phi(\mathbf{x})$ from t_k, t_k is projected into the feature space via $\mathbf{v}_k$, giving a reconstructed feature vector $\hat{\Phi}_p(\mathbf{x}) = \sum_{k=1}^{p} t_k \mathbf{v}_k$ (step (c)). Then the squared prediction error (*SPE*) in the feature space is defined as $SPE = \|\Phi(\mathbf{x}) - \hat{\Phi}_p(\mathbf{x})\|^2$. Here, $\Phi(\mathbf{x})$ is identical to $\hat{\Phi}_p(\mathbf{x}) = \sum_{k=1}^{n} t_k \mathbf{v}_k$ if $p = n$, where n is the number of nonzero eigenvalues in the feature space among the total N eigenvalues (step (d)). 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; Lee *et al*., 2004a).

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

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

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

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

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

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

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

$$N\lambda\boldsymbol{\alpha} = \mathbf{K}\boldsymbol{\alpha} \quad (4)$$

where $\boldsymbol{\alpha} = [\alpha_1, ..., \alpha_N]^T$ 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 eigen-problem of Eq. (4). This yields eigenvectors $\alpha_1, ..., \alpha_N$, 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., a score vector, of the k-th observation in the training data, is calculated by projecting $\Phi(\mathbf{x})$ onto eigenvectors $\mathbf{v}_k$ in F where $k = 1, ..., p$.

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

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

Like other nonlinear PCA methods, KPCA has several parameters that must be set in advance. For example, the selection of the kernel function is the most important problem confronting KPCA because the degree to which it captures the nonlinear characteristic of a system depends on the choice of the kernel function. A general question of how to select the ideal kernel for a given process is an open problem. A number of kernel functions have been proposed, the most common being the polynomial kernel, the sigmoid kernel, and the radial basis kernel. The non-linear mapping function would be ideal to be a linearising transform which reflects the process with first exploiting one's knowledge of the physics and chemistry of the process to create meaningful. If one have a nonlinear information (shape) of the process, it could be used to select a kernel functions among polynomial, sigmoid, and radial basis functions in KPCA.

1.2 Multiway principle component analysis (MPCA)

MPCA is used for analysis and monitoring of batch process data. The batch data are reported in terms of batch runs, variables and times. Data are arranged into a three-dimensional matrix $\underline{\mathbf{X}}$ ($I \times J \times K$), where I is the number of batches, J is the number of variables and K is the number of sampling times in a given batch. It can be decomposed using various three-way techniques, including MPCA. Multiway PCA is equivalent to performing an ordinary PCA on a large two-dimensional matrix $\mathbf{X}$ constructed by unfolding the three-way data as shown in approach of **Figure 2** (Nomikos and MacGregor, 1994).

MPCA decomposes the three-way $\underline{\mathbf{X}}$ array into a summation of the product of the score vector t_r and the loading matrix P_r plus the residual array $\underline{\mathbf{E}}$ that is minimized in least squares sense as

$$\underline{\mathbf{X}} = \sum_{r=1}^{R} t_r \otimes P_r + \underline{\mathbf{E}} = \sum_{r=1}^{R} t_r P_r^T + \mathbf{E} = \hat{\mathbf{X}} + \mathbf{E} \quad (6)$$

where $\otimes$ denotes the Kronecker product ($\underline{\mathbf{X}} = \mathbf{t}\otimes\mathbf{P}$ is $\underline{X}(i, j, k) = t(i)P(j, k)$ and R denotes the number of principal components retained. The first equation in Eq. (6) denotes the 3-D decomposition while the second equation displays the more common 2-D decomposition (Nomikos and MacGregor, 1994).

The statistics used for the MPCA are Hotelling's T^2 and a squared prediction error (*SPE*). If a new batch is good and consistent with the normal batches, its scores should fall within the normal range and the *SPE* or Q-statistic should be small. The T^2 and Q-statistics obtained at the end-of-batch for batch i are calculated as

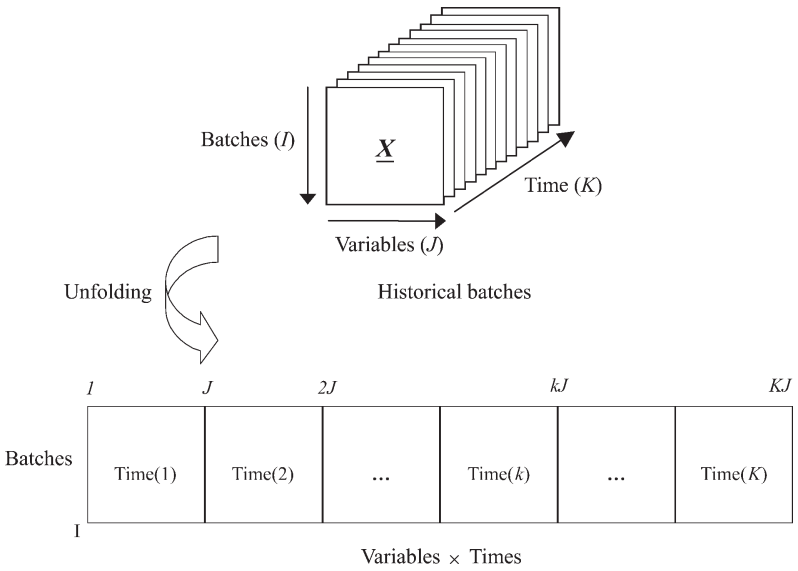

Fig. 2 Unfolding method of MPCA for a three-way batch

$$T_i^2 = \mathbf{t}_{\text{new}}{}^T \mathbf{S}^{-1} \mathbf{t}_{\text{new}} \sim \frac{R\left(I^2-1\right)}{I(I-R)} F_{R,I-R} \tag{7}$$

$$Q_i = \mathbf{e}_i \mathbf{e}_i^T = \sum_{c=1}^{KJ} \mathbf{E}(i,c)^2 \tag{8}$$

where $\mathbf{e}_i$ is the i-th row of $\mathbf{E}$, I is the number of batches in the reference set, $\mathbf{t}_r$ is a vector of R scores, $\mathbf{S}$ is the $(R \times R)$ covariance matrix of the t-scores calculated during the model development (diagonal due to the orthogonality of the t-score values), R is the number of principal components retained in the model, $F_{R,I-R}$ is the F-distribution value with R and $(I-R-1)$ degrees of freedom. Statistical limits on the T^2 and Q-statistics are computed by assuming that the data have a multivariate normal distribution (Nomikos and MacGregor, 1995).

1.3 Nonlinear batch monitoring using MKPCA

In this paper, a nonlinear batch monitoring system of SBR is developed on the basis of the multiway kernel principal component analysis (MKPCA) (Lee *et al.*, 2004b). Batch processes deliver, by nature, a 3-way matrix ($\underline{\mathbf{X}}(I \times J \times K)$) of data. In a typical batch run, $j = 1, 2, ..., J$ variables are measured at $k = 1, 2, ..., K$ time intervals throughout the batch. Similar data exist in several ($i = 1, 2, ..., I$) similar process batch runs. MPCA needs to unfold this matrix in order to obtain a two-way matrix, and then perform PCA. In this paper, we used KPCA instead of PCA to extract the nonlinear structure of the unfolded matrix, $\mathbf{X}(I \times JK)$.

1.3.1 Outline of batch monitoring based on MKPCA

1. Acquire normal operating data $\underline{\mathbf{X}}(I \times J \times K)$ and unfold it batch-wise $\mathbf{X}(I \times JK)$.
2. The data $\mathbf{X}(I \times JK)$ are normalized using the mean and the 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 $\mathbf{x}_k \in R^{JK}$, $k = 1, ..., I$, 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 $\sum_{k=1}^{I} \tilde{\Phi}(\mathbf{x}_k) = 0$.

$$\tilde{\mathbf{K}} = \mathbf{K} - \mathbf{1}_I \mathbf{K} - \mathbf{K}\mathbf{1}_I + \mathbf{1}_I \mathbf{K} \mathbf{1}_I \tag{9}$$

where, $\mathbf{1}_I = \frac{1}{I} \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix} \in R^{I \times I}$.

5. Carry out variance scaling in the feature space for $[1/(I-1)] \sum_{k=1}^{I} \tilde{\Phi}_{\text{scl}}(\mathbf{x}_k)^2 = 1$

$$\tilde{\mathbf{K}}_{\text{scl}} = \frac{\tilde{\mathbf{K}}}{\frac{\text{trace}\left(\tilde{\mathbf{K}}\right)}{I-1}} \tag{10}$$

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

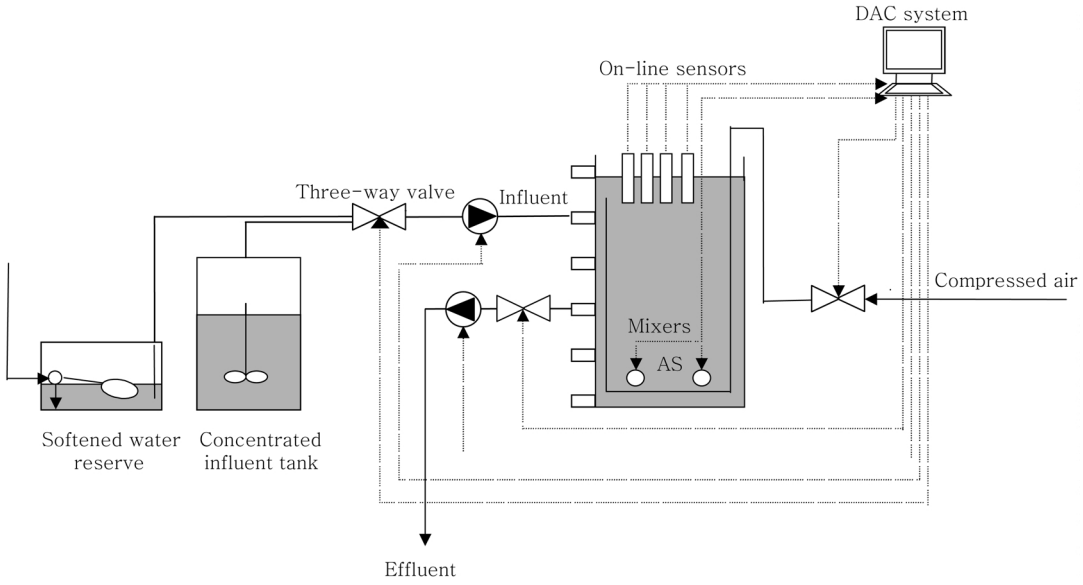

Fig. 3 Schematic diagram of the pilot-scale sequencing batch reactor

$$t_k = \left\langle \mathbf{v}_k, \tilde{\Phi}_{\text{scl}}(\mathbf{x}) \right\rangle = \sum_{i=1}^{I} \alpha_i^k \left\langle \tilde{\Phi}_{\text{scl}}(\mathbf{x}_i), \tilde{\Phi}_{\text{scl}}(\mathbf{x}) \right\rangle$$
$$= \sum_{i=1}^{I} \alpha_i^k \tilde{k}_{\text{scl}}(\mathbf{x}_i, \mathbf{x}) \quad (11)$$

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 the control limits of T^2 and *SPE* charts.

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

$$T^2 = \left[t_1, \ldots, t_p\right] \Lambda^{-1} \left[t_1, \ldots, t_p\right]^T \quad (12)$$

where t_k is obtained from Eq. (5), p is the number of PCs and Λ^{-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.

$$T^2 \sim \frac{p\left(I^2 - 1\right)}{I(I - p)} F_{p, I-p, \alpha} \quad (13)$$

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 a simple calculation of *SPE* in the feature space F suggested by Lee *et al.* (2004b). Then *SPE* in the feature space is defined as

$$SPE = \left\| \Phi(\mathbf{x}) - \hat{\Phi}_p(\mathbf{x}) \right\|^2 \quad (14)$$

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

$$SPE \sim g\chi^2_{h,\alpha}$$
$$g = v/2m, \quad h = 2m^2/v \quad (15)$$

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

2. Results and Discussion

2.1 Pilot-scale 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 an 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 for 12 h and 10 d, respectively. Like synthetic municipal-like sewage, loading amounts of COD, NH_4^+-N and PO_4^{3-}-P per cycle in standard conditions

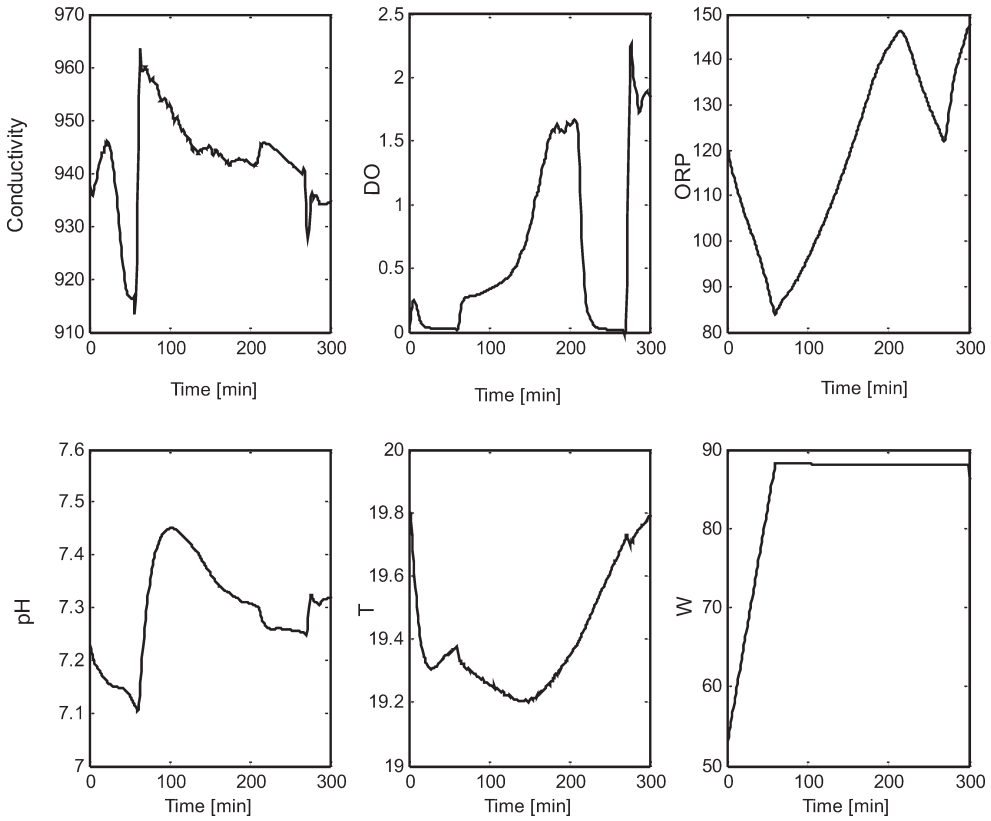

Fig. 4 Typical batch trajectory profiles of an SBR

are 440, 60 and 9.5 mg/*l*, respectively. The control of the duration/sequence of phases and the on/off status of peristaltic pumps, mixer and air supply are automatically achieved by a Labview program. Six electrodes for pH, oxidation-reduction potential (ORP), dissolved oxygen (DO), temperature, conductivity and weight are connected to the individual sensors 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. **Figure 4** shows typical batch profiles of the 6 variables during a batch. (Lee and Vanrolleghem, 2003; Yoo *et al*., 2004).

It has been reported that on-line sensor values collected in the SBR are related with dynamic characteristics of the nutrient concentrations (COD, NH_4^+-N, PO_4^{3-}- and NO_3^-) in SBRs (Chang and Hao, 1996). The derivative of pH, ORP and DO profiles can detect the ends of phosphate release, ammonia conversion, and phosphate uptake, which also are useful information sources. Therefore, the first derivatives of pH, ORP and DO signals were calculated from on-line sensor profiles and included into the database. The second derivatives are too noisy to obtain any valuable information. We considered 150 batches in the historical data set of the SBR for which 9 variables were available at 300 time instants.

2.2 Multivariate batch monitoring of the pilot-scale SBR

Ten principal components (PCs) of the MPCA model for 9 input variables including the first derivatives were retained by the cross-validation method explaining 77.5% of variation of the input space. Twenty PCs of the MKPCA model for 9 input variables including the first derivatives were retained by the broken-stick rule explaining 77.8% of the variation in the feature space (Nomikos and MacGregor, 1994). MKPCA selected the Gaussian kernel, $k(\mathbf{x}, \mathbf{y}) = \exp(-\|\mathbf{x} - \mathbf{y}\|^2/\delta)$ with $\delta = rm\sigma^2$, where r is a constant determined by the process to be monitored (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). According to Schölkopf *et al*. (1999), KPCA has a potential to utilize more PCs to code the structure rather than noise; hence, KPCA outperforms linear PCA in denoising if a sufficiently large number of PCs are used. In this research, a Gaussian kernel was used for the mapping to a high-dimensional feature space since it is found to be appropriate to capture the nonlinearity of the considered system by testing the monitoring performance of a range of kernel functions. Obviously, other kernel functions may be more appropriate for other nonlinear cases.

Figure 5 shows the T^2 and the *SPE* charts of the MPCA and the MKPA models with 9 input variables.

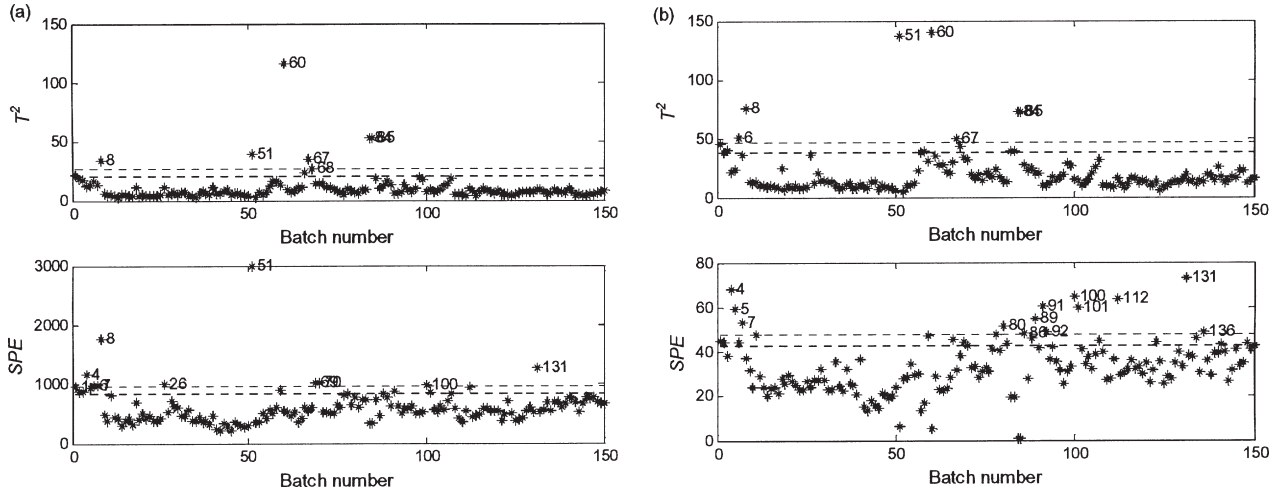

Fig. 5 The Hotelling's T^2 and SPE charts of 150 batches of (a) the MPCA and (b) the MKPCA methods with 9 input variables (six original and three derivative signals). The dotted lines correspond to 95 and 99% confidence limits

There is little difference between the T^2 charts of the MPCA and the MKPCA models. But there is a considerable difference between the *SPE* charts of them. This difference originates from the noise amplification of the derivatives of the three signals. This noise amplification may significantly distort the information of a signal and makes the *SPE* statistic of the MPCA model with 9 input variables less efficient because their *SPE* limits are too high, that is, models do not stay valid for prolonged deviation in the *SPE* limit. This may come from the nonlinear biological kinetics and noise amplification. In the instance of noisy data, MPCA discards the finite variance due to the noise by projection of data onto the main principal components. On the other hand, MKPCA extracts a large number of nonlinear principal components and therefore allows spreading the information regarding the data structure more widely giving a better opportunity to discard some of the eigen directions where the noisy part of the data resides. Thus, MKPCA provides a better capability of feature extraction and denoising. In addition, if the structure to be extracted is nonlinear, then linear PCA has a worse denoising capability than KPCA (Schölkopf *et al*., 1999). This observation is the confirmation of the nonlinear batch monitoring.

In the *SPE* chart, MPCA assigned the eight batches as abnormal since they exceed the 99% control limit, whereas MKPCA assigned several additional batches as abnormal, i.e., batches 89, 91, 92, 101, 112. MKPCA can find nonlinearly abnormal data even though it is regarded as normal according to the linear relation of MPCA. **Figure 6** shows the univariate plots of the original three signals (DO, pH, ORP) and three derivatives (dDO/dt, dpH/dt, dORP/dt) for normal batches and batch 91. In Figure 5(b), it is shown that MKPCA can detect the abnormal behavior of the pH signal and also the derivative of ORP, dOPR/dt in batch 91, whereas MPCA in Figure 5(a) cannot detect this abnormal batch due to the nonlinearity. One classic way of dealing with these nonlinearities in the PCA is to include cross-products of variables as well as nonlinear transformations of variables. A nonlinear mapping function would be ideal as a linearising transformation which reflects one's knowledge of the chemical and biological relations of the process. But this asks to specify a nonlinearity shape in advance. In Figure 6, we cannot define the nonlinear transformation for the first derivatives of pH, ORP and DO signals of SBR.

It is well-known that the first derivatives of pH, ORP and DO signals can give valuable information about the nutrient dynamics, such as the nitrate knee, nitrate apex, ammonia valley, phosphorous release and uptake end points (Chang and Hao, 1996). Although we can use these three additional signals for MPCA to access to this valuable information, derivation always amplifies the noise components in the data set, that is, a high-pass filtering, which makes the SBR data set more nonlinear than the original 6 variables. This requires a bargain between the information-richness and noiseamplification. The impact of nonlinearity was assessed through the use of normal probability plots. **Figure 7** shows the normal probability plot of the *SPE* of the MPCA and the MKPCA model with 9 input variables. From this figure, we can conclude that the MKPCA has better nonlinear feature extraction results than MPCA.

Nonlinear diagnostic charts of an SBR which can capture biological relations can be immediately displayed to the operator as soon as the special biological event is detected. Since operators will use the information on the fault, the detected special event can be corrected prior to completion of the batch or in subsequent batches. When the SBR has no major upsets, the process would be more stable than when it is

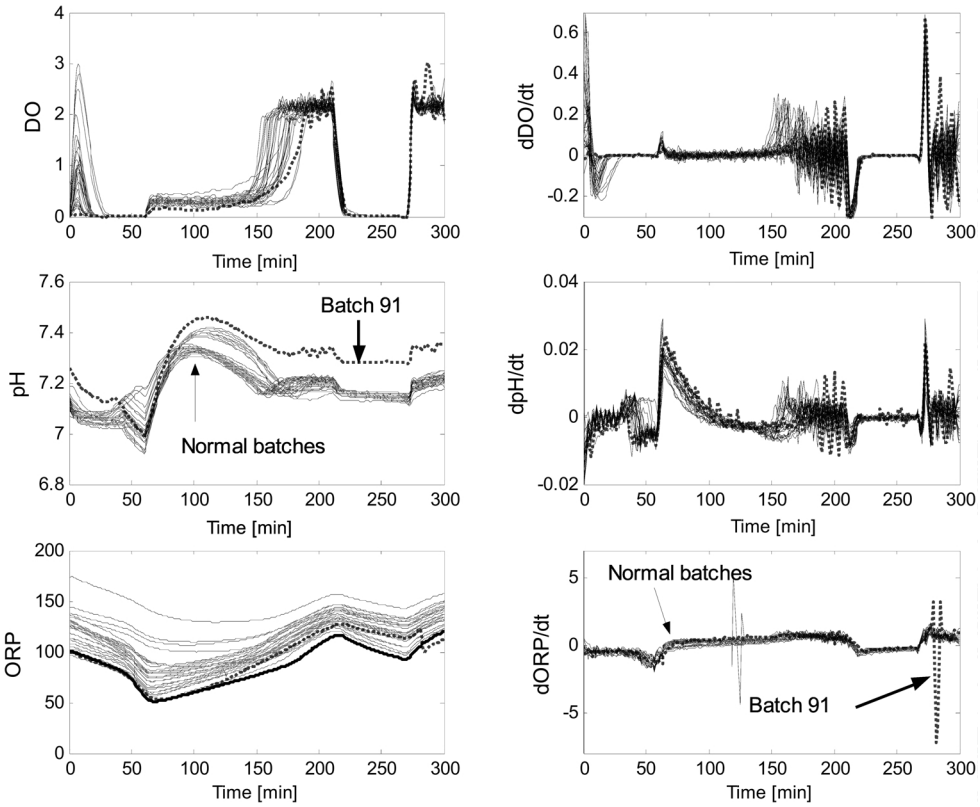

Fig. 6 The univariate plots of DO, pH, ORP, dDO/d*t*, dpH/d*t* and dORP/d*t* for normal batches and abnormal batch 91

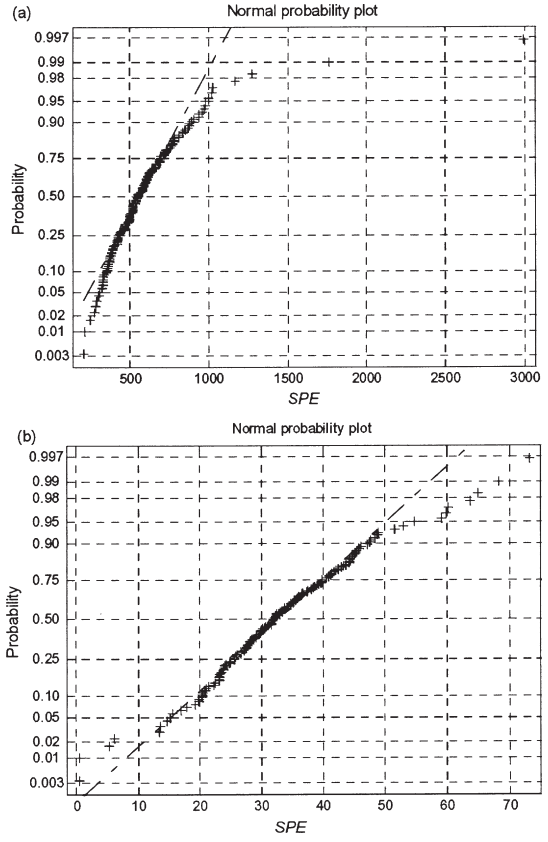

Fig. 7 Normal probability plot of *SPE* of (a) MPCA and (b) MKPCA with 9 input variables

exposed to disturbances.

Conclusion

In this paper, a scheme for nonlinear multivariate statistical process control of a pilot-scale SBR is developed using multiway kernel PCA. Three-way batch data of the normal batch process is unfolded batch-wise and then nonlinear PCA is used to capture the nonlinear relation among batches. This technique has successfully been applied to an 80-*l* SBR for biological wastewater treatment, which is characterized by a variety of non-stationary and nonlinear characteristics. In the multivariate batch monitoring of an SBR, the nonlinear relations among the batch operations of the SBR can effectively be captured and the power and advantages of nonlinear process monitoring in comparison with the linear MPCA method are clearly shown. If nonlinear correlations within the batch exist, MKPCA shows better monitoring and denoising performance than MPCA. MKPCA may be an appropriate tool to monitor process stability and to analyze the nonlinear biological processes using simple on-line measurements and their derivatives.

Acknowledgment

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

Research-Flanders (FWO) and Institute for Encouragement of Innovation by means of Science and Technology in Flanders (IWT). Specially, Lee J. M. is kindly thanked for his valuable discussion.

Nomenclature

$\mathbf{C}^F$	=	covariance matrix in the feature space
$\underline{\mathbf{E}}$	=	residual array
F	=	feature space
$F_{R,I-R}$	=	F-distribution value with R and $(I - R - 1)$ degrees of freedom
I	=	number of batches
J	=	the number of variables
K	=	number of sampling times
$[\mathbf{K}]_{ij}$	=	kernel trick, $=K_{ij} = \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle$
$\mathbf{K} \in R^{N \times N}$	=	gram matrix
$\tilde{\mathbf{K}}$	=	mean-centered kernel matrix
$\tilde{\mathbf{K}}_{scl}$	=	mean-centered and scaled kernel matrix
$k(\mathbf{x}, \mathbf{y})$	=	kernel function, $=\langle \Phi(\mathbf{x}), \Phi(\mathbf{y}) \rangle$
m	=	estimated mean of the *SPE* statistic
N	=	number of data points
P_r	=	loading matrix
$\mathbf{S}$	=	$(R \times R)$ covariance matrix
SBR	=	sequencing batch reactor
SPE	=	squared prediction error
T^2	=	Hotelling's T^2 statistic
t_r	=	principal score vector
v	=	variance of the SPE statistic
$\underline{\mathbf{X}}(I \times J \times K)$	=	three-dimensional matrix
$\mathbf{x}$	=	input vector
$\langle \mathbf{x}, \mathbf{y} \rangle$	=	dot product between $\mathbf{x}$ and $\mathbf{y}$
$\mathbf{v}$	=	eigenvector of the covariance matrix CF
$\Phi(\cdot)$	=	nonlinear mapping function
$\tilde{\Phi}_{scl}(\mathbf{x})$	=	mean centered and variance scaled feature vector of $\Phi(\mathbf{x})$
$\hat{\Phi}_p(\mathbf{x})$	=	reconstructed feature vector with p principal components in the feature space
σ^2	=	variance of the data
λ	=	eigenvalues
$\otimes$	=	Kronecker product

Literature Cited

Chang, C. H. and O. J. Hao; "Sequencing Batch Reactor System for Nutrient Removal: ORP and pH Profiles," *J. Chem. Technol. Biotechnol.*, **67**, 27–38 (1996)

Kano, M., S. Hasabe, I. Hashimoto and H. Ohno; "A New Multivariate Statistical Process Monitoring Method Using Principal Component Analysis," *Comput. Chem. Eng.*, **25**, 1103–1113 (2001)

Kano, M., S. Tanaka, S. Hasabe, I. Hashimoto and H. Ohno; "Monitoring Independent Components for Fault Detection," *AIChE J.*, **49**, 969–976 (2003)

Lee, D. S. and P. A. Vanrolleghem; "Monitoring of a Sequencing Batch Reactor Using Adaptive Multiblock Principal Component Analysis," *Biotechnol. Bioeng.*, **82**, 489–497 (2003)

Lee, J.-M., C. K. Yoo, S. W. Choi, P. A. Vanrolleghem and I.-B. Lee; "Nonlinear Process Monitoring Using Kernel Principal Component Analysis," *Chem. Eng. Sci.*, **59**, 223–234 (2004a)

Lee, J.-M., C. K. Yoo and I.-B. Lee; "Fault Detection of Batch Processes Using Multiway Kernel Principal Component Analysis," *Comput. Chem. Eng.*, **28**, 1837–1847 (2004b)

Mika, S., B. Schölkopf, A. J. Smola, K.-R. Müller, M. Scholz and G. Rätsch; "Kernel PCA and De-noising in Feature Spaces," *Advances Neural Inform. Proc. Syst.*, **11**, 33–42 (1999)

Nomikos, P. and J. F. MacGregor; "Monitoring Batch Processes Using Multiway Principal Component Analysis," *AIChE J.*, **40**, 1361–1375 (1994)

Nomikos, P. and J. F. MacGregor; "Multivariate SPC Charts for Monitoring Batch Processes," *Technometrics*, **37**, 41–53 (1995)

Romdhani, S., S. Gong and A. Psarrou; "A Multi-view Nonlinear Active Shape Model Using Kernel PCA," Proceedings of the British Machine Vision Conference 1999, Nottingham, pp. 483–492, British Machine Vision Association, Malvern, U.K. (1999)

Rosen, C. and J. A. Lennox; "Multivariate and Multiscale Monitoring of Wastewater Treatment Operation," *Water Res.*, **35**, 3402–3410 (2001)

Schölkopf, B., A. Smola and K.-R. Müller; "Nonlinear Component Analysis as a Kernel Eigenvalue Problem," *Neural Comput.*, **10**, 1299–1399 (1998)

Schölkopf, B., S. Mika, C. J. C. Burges, P. Knirsch, K.-R. Müller, G. Rätsch and A. J. Smola; "Input Space versus Feature Space in Kernel-Based Methods," *IEEE T. Neural Network.*, **10**, 1000–1016 (1999)

Yoo, C. K., P. A. Vanrolleghem and I.-B. Lee; "Nonlinear Modeling and Adaptive Monitoring with Fuzzy and Multivariate Statistical Method in Biological Wastewater Treatment Plants," *J. Biotechnol.*, **105**, 135–163 (2003)

Yoo, C. K., D. S. Lee and P. A. Vanrolleghem; "Application of Multiway ICA for On-line Process Monitoring of a Sequencing Batch Reactor," *Water Res.*, **38**, 1715–1732 (2004)

Zhang, J., E. B. Martin and A. J. Morris; "Process Monitoring Using Non-Linear Statistical Techniques," *Chem. Eng. J.*, **67**, 181–189 (1997)