

## Multivariate monitoring for time-derivative non-Gaussian batch process

Min Han Kim and Chang Kyoo Yoo<sup>†</sup>

College of Environment and Applied Chemistry/Green Energy Center, Kyung Hee University,  
Seocheon-dong 1, Giheung-gu, Yongin-si, Gyeonggi-do 446-701, Korea

(Received 9 May 2007 • accepted 22 January 2008)

**Abstract**—This research is an application of process monitoring on a pilot-scale sequencing batch reactor (SBR) using a batchwise multiway independent component analysis method (MICA) for denoising effect, which can extract meaningful hidden information from non-Gaussian data. Three-way batch data of SBR are unfolded batch wise, and then a multivariate monitoring method is used to capture the non-Gaussian and nonlinear characteristics of normal batches. It is successfully applied to an 80 L SBR for biological wastewater treatment, which is characterized by a variety of error sources with non-Gaussian characteristics. In the monitoring result, multiway principal component analysis (MPCA) can detect the abnormal batches with a false alarm rate of 47.5%, whereas MICA charts show less false alarm rate of 4.5%. The results of this pilot-scale SBR monitoring system using simple on-line measurements clearly demonstrated that the MICA monitoring technique showed lower false alarm rate and physically meaningful robust monitoring results.

**Key words:** Batch Monitoring, Denoising, Multiway Independent Component Analysis (MICA), Non-Gaussianity, Sequencing Batch Reactor (SBR), Time-derivative

### INTRODUCTION

Due to increasing environmental constraints and the necessity of reliable wastewater treatment, efficient modeling and monitoring methods are becoming more and more important. Reliable modeling and monitoring techniques of biological wastewater treatment plants (WWTP) are necessary to maintain the system performance as close as possible to optimal conditions. An adequate model enhances the understanding of the biological processes and it can be a basis for better process design, control, and operation. Also, process monitoring and early fault detection methods are efficient to execute corrective actions well before a dangerous situation occurs in biological processes [1-5].

Sequencing batch reactor (SBR) processes have been interested in the point of their flexibility with a very simple physical structure, when treating wastewaters with high concentrations of nutrient, nitrogen, phosphorous, and toxic compounds from domestic and industrial sources. An SBR has a unique cyclic batch operation, usually with five well-defined phases: fill (static-fill, mixed-fill, aerated-fill), reaction (mixed-reaction, aerated reaction), settle, draw and idle (Fig. 1). This flexibility of operation 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. But the SBR process is highly non-linear and non-Gaussian from the nonlinear biological reaction kinetics, time varying and subject to significant disturbances like hydraulic changes, composition variations and equipment failures. Small changes in concentrations or flows can affect effluent quality and microorganism growth [6-9].

Even though operators are aware that there are some problems in treatment performance, they cannot quickly find or predict what the causes are and when the problems will occur because most batch

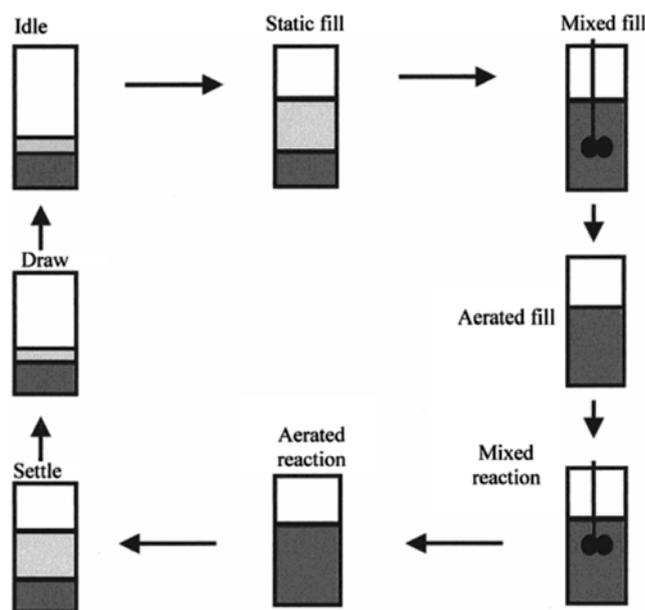


Fig. 1. Schematic of SBR operation during one cycle [8].

processes are run without any effective form of real-time on-line monitoring. The treatment performance of SBR is often only examined off-line in a laboratory, the key indicator of process performance. Therefore, multivariate statistical monitoring of SBRs is crucial to detecting 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 [10,11]. It has been reported that on-line sensor values collected in SBR are related with dynamic characteristics of the nutrient concentrations (COD,  $\text{NH}_4\text{-N}$ ,  $\text{PO}_4^{3-}$  and  $\text{NO}_3^-$ ) in SBRs [6]. 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.

<sup>†</sup>To whom correspondence should be addressed.

E-mail: ckyoo@khu.ac.kr

But these derivatives of the signals may make it a nonlinear behavior. Especially, for the modeling of nonlinear dynamic systems that have a large or even an infinite phase space, principal component analysis (PCA) is able to come up with a lower dimensional subspace and becomes a more non-parsimonious model than a linear dynamic system [12-14].

Multivariate principal component analysis (MPCA), developed by Nomikos and MacGregor [15], has been shown to be a powerful monitoring tool in many industrial batch processes. However, it has the shortcoming that the measurement variables of the batch process should be normally distributed and linear. In this work, it is shown that multivariate independent component analysis (MICA) suggested by Yoo et al. [9] can be used to overcome this drawback for better monitoring performance of a biological batch process. MICA is able to tackle the non-Gaussianity of a biological process and the nonlinearity of a time-derivative of pH, ORP and DO signals. The proposed method is applied to an 80L SBR for biological wastewater and compared with other conventional monitoring method.

## THEORY

### 1. Independent Component Analysis (ICA)

To illustrate the superiority of ICA over PCA, we applied the two types of analysis to a simple example system, similar to that used by Hyvärinen et al. [16] and Lee [17] except that a modified mixing matrix was used in the present work. Let's consider two source variables that have the uniform distributions shown in Fig. 2(a). The source variables are linearly independent, i.e., the values of one source variable do not convey any information about the other source variable. These sources are linearly mixed as follows:

$$\mathbf{x} = \mathbf{A}\mathbf{s} \quad (1)$$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix}$$

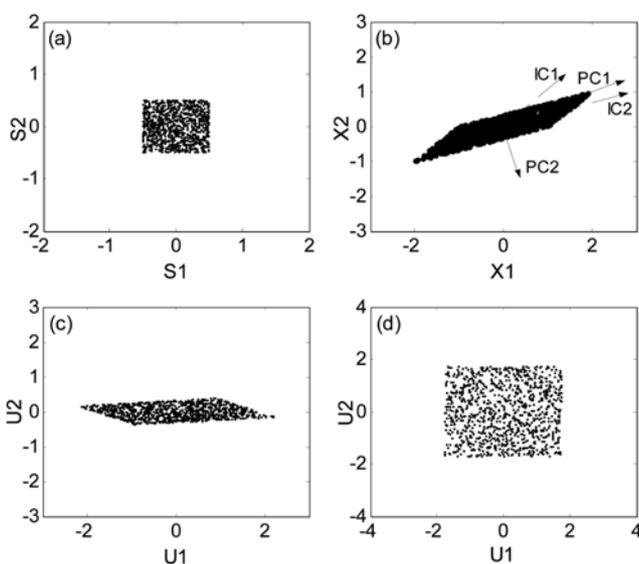


Fig. 2. (a) Scatter plot of the original source data; (b) The mixtures and axes of PCA and ICA; (c) The recovered source data using PCA; (d) The recovered source data using ICA.

where  $\mathbf{x}$  is measured signals,  $\mathbf{A}$  is the unknown mixing matrix,  $\mathbf{s}$  is the source signals. Fig. 2(b) shows the scatter-plot of the mixtures. Note that the random variables  $x_1$  and  $x_2$  are not independent because it is possible to predict the value of one of them from the value of the other. When PCA is applied to these mixed variables, it gives two principal components. The axes of the first and second PCs (PC1, PC2) are shown in Fig. 2(b). The first PC is the axis capturing the highest variance in the data and the second PC is the axis orthogonal to the first PC. Fig. 2(c) shows the PCA solution, which differs from the original because the two principal axes are still dependent. However, the ICA solution shown in Fig. 2(d) can recover original sources since ICA not only decorrelates the data but also rotates it such that the axes of  $u_1$  and  $u_2$  are parallel to the axes of  $s_1$  and  $s_2$  [17]. The axes of the first and second independent components (IC1, IC2) are shown in Fig. 2(b).

In the ICA algorithm, it is assumed that  $d$  measured variables,  $x_1, x_2, \dots, x_d$ , can be expressed as linear combinations of  $m$  ( $\leq d$ ) unknown independent components,  $s_1, s_2, \dots, s_m$ . The relationship between them is given by

$$\mathbf{X} = \mathbf{A}\mathbf{S} + \mathbf{E} \quad (2)$$

where  $\mathbf{X} = [\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(n)] \in \mathbb{R}^{d \times n}$  is the data matrix (ICA uses the transposed data matrix, in contrast to PCA),  $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_m] \in \mathbb{R}^{d \times m}$  is the unknown mixing matrix,  $\mathbf{S} = [\mathbf{s}(1), \mathbf{s}(2), \dots, \mathbf{s}(n)] \in \mathbb{R}^{m \times n}$  is the independent component matrix,  $\mathbf{E} \in \mathbb{R}^{d \times n}$  is the residual matrix, and  $n$  is the number of samples. Here, it is assumed that  $d \geq m$  (when  $d = m$ , the residual matrix,  $\mathbf{E}$ , becomes the zero matrix). The basic objective of ICA is to estimate both the mixing matrix,  $\mathbf{A}$ , and the independent component,  $\mathbf{S}$ , from only the observed data,  $\mathbf{X}$ . Alternatively, one could define the objective of ICA as follows: to find a demixing matrix,  $\mathbf{W}$ , whose form is such that the rows of the reconstructed matrix  $\hat{\mathbf{S}}$ , defined as

$$\hat{\mathbf{S}} = \mathbf{W}\mathbf{X} \quad (3)$$

become as independent of each other as possible [16].

### 2. Monitoring of MICA

The monitoring method based on MICA is similar to that based on MPCA. MICA is equivalent to performing ICA on a large two-dimensional matrix  $\mathbf{X}$  constructed by batchwise unfolding the three-way data matrix  $\underline{\mathbf{X}}$ . MICA decomposes the three-way array  $\underline{\mathbf{X}}$  into a summation of the product of independent vectors  $\mathbf{s}_r$  and loading matrices  $\mathbf{A}_r$ , plus a residual array  $\underline{\mathbf{E}}$  so that the ICs  $\mathbf{s}$  become as independent of each other as possible:

$$\underline{\mathbf{X}} = \sum_{r=1}^R \mathbf{s}_r \otimes \mathbf{A}_r + \underline{\mathbf{E}} = \sum_{r=1}^R \mathbf{s}_r \mathbf{a}_r^T + \underline{\mathbf{E}} = \hat{\underline{\mathbf{X}}} + \underline{\mathbf{E}} \quad (4)$$

where  $\otimes$  denotes the Kronecker product ( $\underline{\mathbf{X}} = \mathbf{s} \otimes \mathbf{A}$  is  $\underline{\mathbf{X}}(i, j, k) = \mathbf{s}(i) \mathbf{A}(j, k)$ ) and  $R$  denotes the number of ICs retained. The  $\mathbf{S}$  and  $\mathbf{A}$  matrices in Eq. (4) can be equivalent to the loading matrix and score matrices by analogy with MPCA, i.e.,  $\mathbf{S}$  can be regarded as the score matrix  $\mathbf{T}$ , and  $\mathbf{A}$  can be treated as the loading matrix  $\mathbf{P}$ . The  $i$ th elements of the independent vector  $\mathbf{s}$  correspond to the  $i$ th batch and summarize the overall variations in this batch with respect to the other batches over the entire history of the batch. The mixing matrix,  $\mathbf{A}$ , summarizes the time variations of the measured variables about their average trajectories. The elements of this matrix are the weights, which give the independent vectors  $\mathbf{s}$  for a batch when applied to

each variable at each time interval within that batch [9].

Batch processes are, by nature, leading to a 3-way matrix ( $\mathbf{X}(I \times J \times K)$ ) of data. In a typical batch run,  $j=1, 2, \dots, J$  variables are measured at  $k=1, 2, \dots, K$  time intervals throughout the batch. There exists similar data on several ( $i=1, 2, \dots, I$ ) similar process batch runs. MPCA needs to unfold this matrix in order to obtain a two-way matrix, and then perform PCA. Fig. 3(a) shows the unfolding

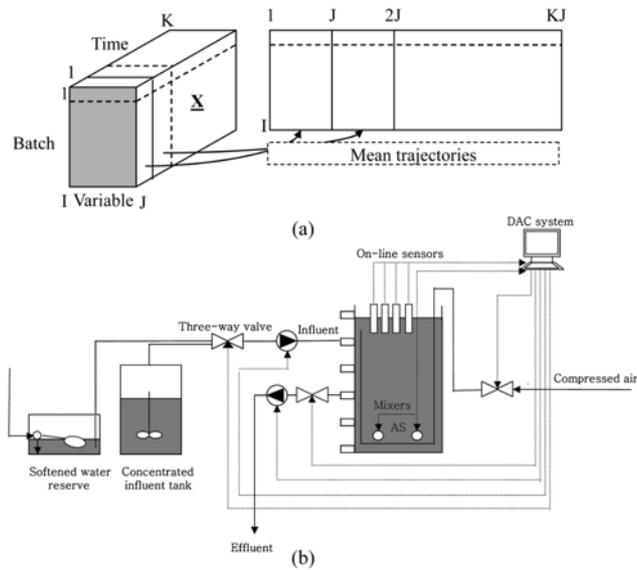


Fig. 3. (a) Batchwise unfolding method of MICA for a three-way batch, (b) schematic diagram of the pilot-scale sequencing batch reactor.

method for MICA. 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 behavior of the process can be eliminated [15] but still remains a non-Gaussian behavior. Once the matrix is mean centered and variance scaled and ICA is performed to tackle the remaining non-Gaussianity, the results from ICA are the loading vectors and the calculated scores for each batch. The loading vectors have a weight for each variable at each time, representing the history of the process. In this paper, MICA instead of MPCA is used to extract the nonlinear structure of the unfolded matrix [9].

In MICA, two statistics are deduced from the process model in normal operation: the D-statistic for the systematic part of the process variation and the Q-statistic for the residual part of the process variation. The D-statistic for a batch  $k$ , also known as the  $\Gamma^2$  statistic, is the sum of the squared independent scores and is defined as follows:

$$\Gamma^2(k) = \mathbf{s}_{newd}(k)^T \mathbf{s}_{newd}(k) \quad (5)$$

The Q-statistic for a batch  $k$ , also known as the SPE statistic, is defined as follows:

$$\text{SPE}(k) = \mathbf{e}(k)^T \mathbf{e}(k) = (\mathbf{x}(k) - \hat{\mathbf{x}}(k))^T (\mathbf{x}(k) - \hat{\mathbf{x}}(k)) \quad (6)$$

where  $\hat{\mathbf{x}}$  can be calculated as follows:

$$\hat{\mathbf{x}} = \mathbf{Q}^{-1} \mathbf{B}_d \hat{\mathbf{s}} = \mathbf{Q}^{-1} \mathbf{B}_d \mathbf{W}_d \mathbf{x} \quad (7)$$

The confidence limits of the  $\Gamma^2$  and SPE statistics in MICA can be obtained by kernel density estimation [18]. Here, the  $\Gamma^2$  value is used to detect faults associated with abnormal variations within an MICA model subspace, whereas the SPE value is used to detect

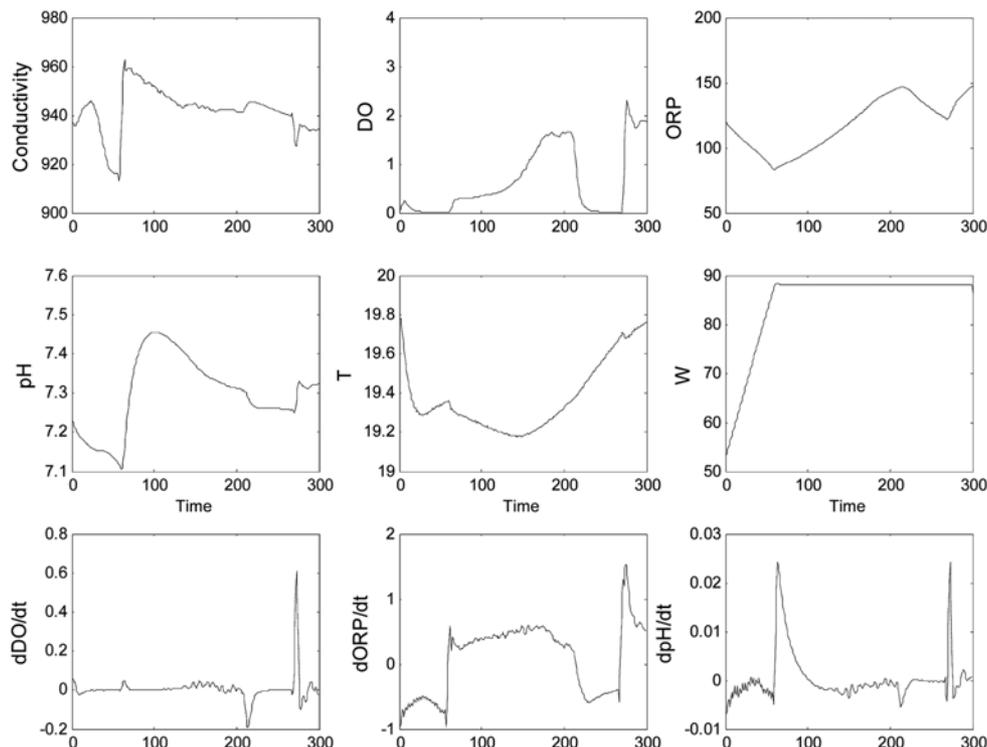


Fig. 4. Batch trajectories of 9 variables including the first derivatives of pH, ORP, DO signals in a SBR.

new events that are not taken into account in an MICA model subspace. For the detailed theory of the MICA method, refer to the Yoo et al.'s [9] paper.

## RESULTS AND DISCUSSION

### 1. Process Description of the Pilot-scale SBR System

The data used in this research were collected from a pilot-scale SBR system shown in Fig. 3(b). A fill-and-draw sequencing batch reactor (SBR) with an 80-liter 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 hrs and 10 days, respectively. Six electrodes for conductivity, dissolved oxygen (DO), oxidation-reduction potential (ORP), pH, temperature (T), weight (W), the first derivative of DO (dDO/dt), the first derivative of ORP (dORP/dt), and the first derivative of pH (dpH/dt), 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. Fig. 4 shows the typical batch trajectories of nine variables in an SBR including the first derivatives of pH, DO, ORP signals. Nine variables in Fig. 4 show the significant signs of the nutrient dynamics, such as the nitrate knee, nitrate apex, ammonia valley, phosphorous release and uptake end points. The historical data set of the SBR consisted of 280 batches (70 days) for which nine variables including the first derivatives of pH, DO, ORP signals were measured at 300 time instants.

### 2. Analysis of Historical SBR Data Set using MPCA and MICA

Fig. 5 shows the monitoring results of all 280 batches of the SBR using the MPCA and MICA methods, where the dotted lines correspond to the 95 and 99% confidence limits. Five components of the MPCA model were selected by the cross-validation method. To ensure comparison of equivalent models, five ICs were selected for the MICA model. From this figure, we notice that the MICA plot shows characteristics dissimilar from the MPCA one. Compared to MPCA, MICA points to a lower number of abnormal batches in SBR. It is well known that changes in the relationships between variables, such as sensor faults, tend to be detected on the SPE chart,

while changes in operating conditions, for example, a grade change and an evolution of new operating condition, are typically identified on the  $T^2$  chart. However, the MPCA model may not be valid because the NOC model is not representative of the SBR process because it does not follow the Gaussian distribution due to the time-varying characteristics of SBR process. This difference can be explained by the density estimation of the SBR data.

The density estimates of the first score ( $t_1$ ) in MPCA do not follow the Gaussian distribution (not shown in the paper) but the 'super-gaussian distribution' in which random variables take relatively more often values that are very close zero or very large. Thus, the  $T^2$  and SPE charts of MPCA that are based on the assumption that the data are Gaussian distributed may cause a false result when used for monitoring. As pointed by Yoo et al. [11], note that the  $T^2$  and SPE statistics do not depend on a certain distribution, only their confidence limits. Because non-Gaussian data may inflate the variance, it tends to reduce the  $T^2$  statistics. Typically, this increases the false alarm rate of the MPCA, which a normal batch might be judged as a non-conforming one. Obviously, this deteriorates the reliability of the multivariate monitoring system and makes it subject to unfavorable criticism. On-line monitoring on multivariate methods is often an issue of minimizing the number of false alarms while true deviations are retained and detected. Robust limits can be determined empirically by using kernel density estimation as used in this paper. Moreover, if the batch length of a new batch is different from the length of old batches, the on-line monitoring method based on MPCA cannot be applied. Actually, many real-world data sets have a supergaussian distribution in which the probability density of the data is peaked in the middle and has heavy tails (large values far from zero), when compared to a Gaussian density with the same variance [4]. This observation is the motivation of the MICA model. MICA is sensitive to modes whose influences on the measured variables follow a supergaussian distribution with large tails and a pronounced peak in the middle.

### 3. Monitoring of SBR using MPCA and MICA

#### 3-1. Off-line Monitoring of SBR

The MPCA and MICA models for the SBR monitoring were developed after an analysis of the historical SBR data set. Fig. 5 shows a multivariate analysis result of all 280 batches with MPCA

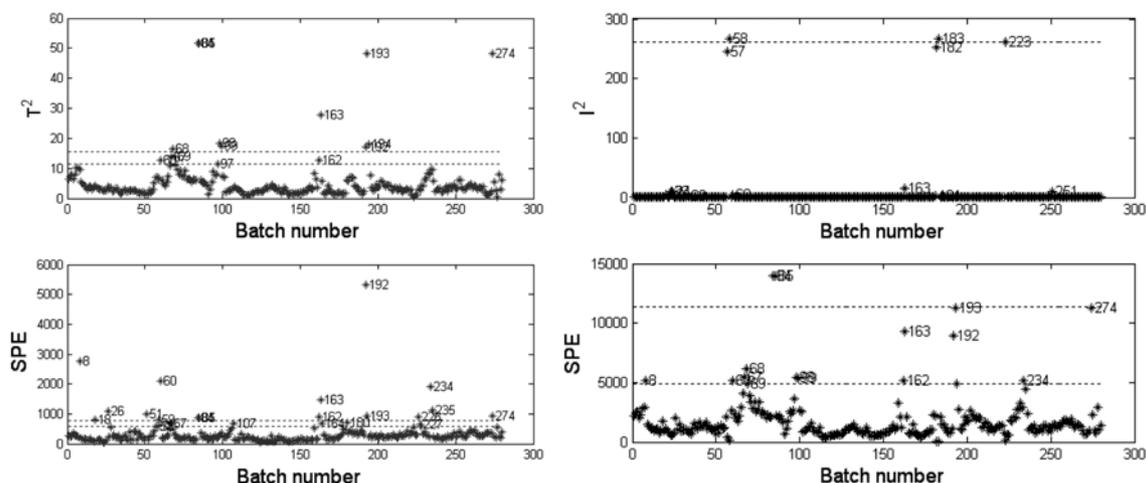


Fig. 5. Multivariate analysis of all 280 batches, (a) MPCA, (b) MICA.

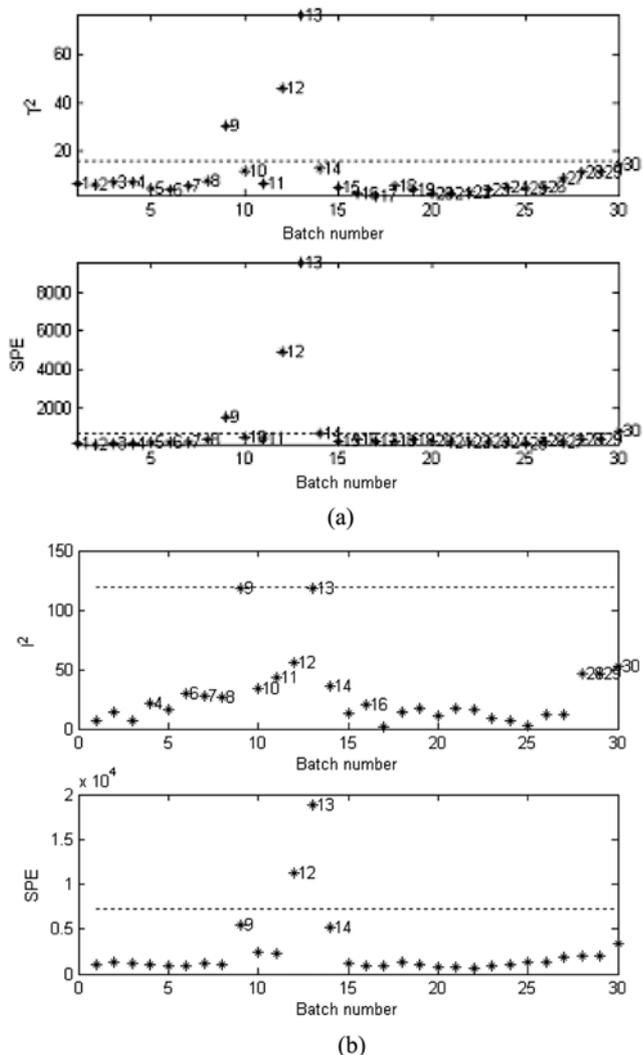


Fig. 6. Monitoring result of 30 test batches. (a) MPCA and (b) MICA. The dotted lines correspond to the 99% confidence limit.

and MICA. The MPCA model selected 143 batches to create a rather broad scope of normal batches, where seven abnormal batches (batch number: 8, 18, 26, 51, 60, 84, 85) were excluded for the normal operating condition (NOC) model. The MICA model selected 146 batches, where four abnormal batches (batch number: 57, 58, 84, 85) were excluded for the normal NOC model. The test data set that consisted of the following 30 batches was projected onto the reduced MPCA and MICA model spaces.

Fig. 6 shows the batch monitoring result of 30 test batches by MPCA and MICA. While both of them could detect two abnormal batches (batch 12, 13), MPCA detected batch 9 as an abnormal batch while MICA left batch 9 as a normal batch. Actually, batch 9 is a normal batch. When MPCA is applied to non-Gaussian data, the  $T^2$  chart of MPCA may suffer oversensitivity for normal batches, e.g., batch 9. As a data set deviates from a Gaussian distribution, the variance tends to increase and hence the  $T^2$  statistic tends to decrease. Typically, this increases the false alarm rate of the MPCA in which a normal batch might be judged as a non-conforming one. Obviously, this deteriorates the reliability of the monitoring system.

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 with non-Gaussian characteristic in effluent quality and microorganism growth. Compared to the previous ones, the MICA method provides more meaningful information on the evolving biological process and captures the biological relation among batches, which results in more robust monitoring performance.

### 3-2. On-line Monitoring of SBR using MPCA and MICA

To estimate the future values in  $X_{new}$ , we used the filling method, which fills in all future measurements with the current deviation from the average batch, since online monitoring of SBR needs a filling method until the batch end [15]. Then the MPCA and MICA models are tested against a new batch of 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 MICA model.

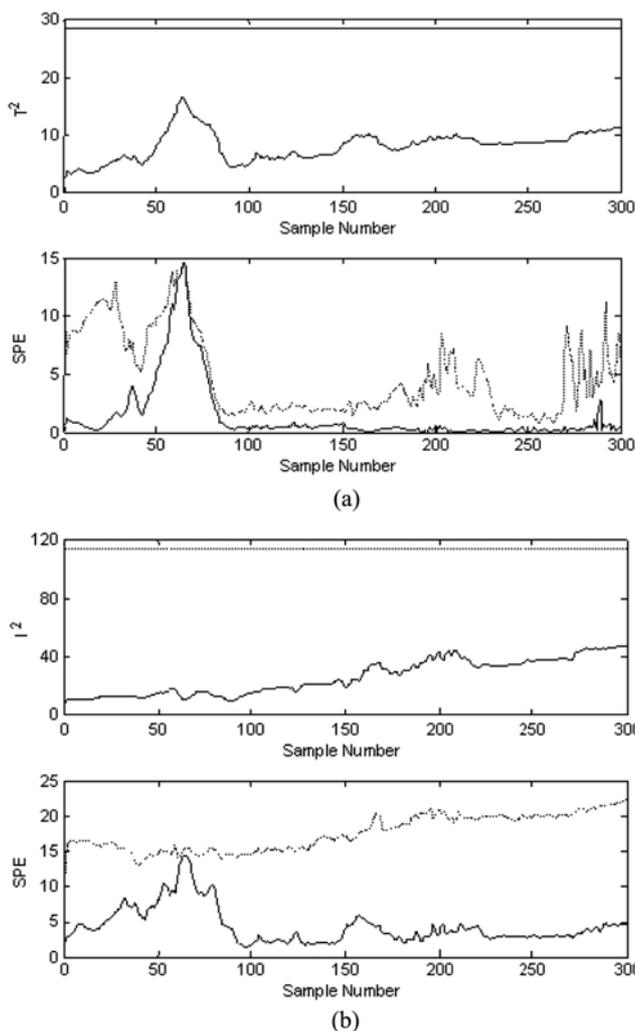


Fig. 7. On-line monitoring charts for (a) MPCA and (b) MICA in case of a normal batch (batch number 1).

Fig. 7 shows the on-line monitoring results of the  $T^2$  and SPE charts of MPCA and the  $I^2$  and SPE charts of MICA with the 99% confidence limits in a particular, normal batch (batch number 1). The batch is monitored for every time instant  $k$  in terms of their control limits of MPCA and MICA. Both MPCA and MICA show that the  $T^2$  charts for this batch are within the control limits for the whole duration of the batch run [4]. However, the SPE chart of MPCA in Fig. 7(a) exceeds the confidence limit one time, around the 60<sup>th</sup> sampling time, that is, the MPCA invokes a false alarm. It results from the non-Gaussian data distribution by a nonlinear biological phenomenon of the change of phosphorous uptake rate, which occurs during the phase change (from anaerobic to aerobic). On the other hand, the  $I^2$  and SPE charts of MICA in Fig. 7(b) do not show any violation for its confidence limit during the whole duration of the batch run. Therefore, this batch in MICA is assigned as being “in control” or “normal”. It illustrates that the selected independent components in MICA can extract the non-Gaussian characteristics of the SBR operation, i.e., the change of the change of phosphorous uptake rate by the phase change from the anaerobic to aerobic phase.

Fig. 8 represents the on-line monitoring results of the  $T^2$  and SPE charts of the MPCA and the  $I^2$  and SPE charts of MICA model for

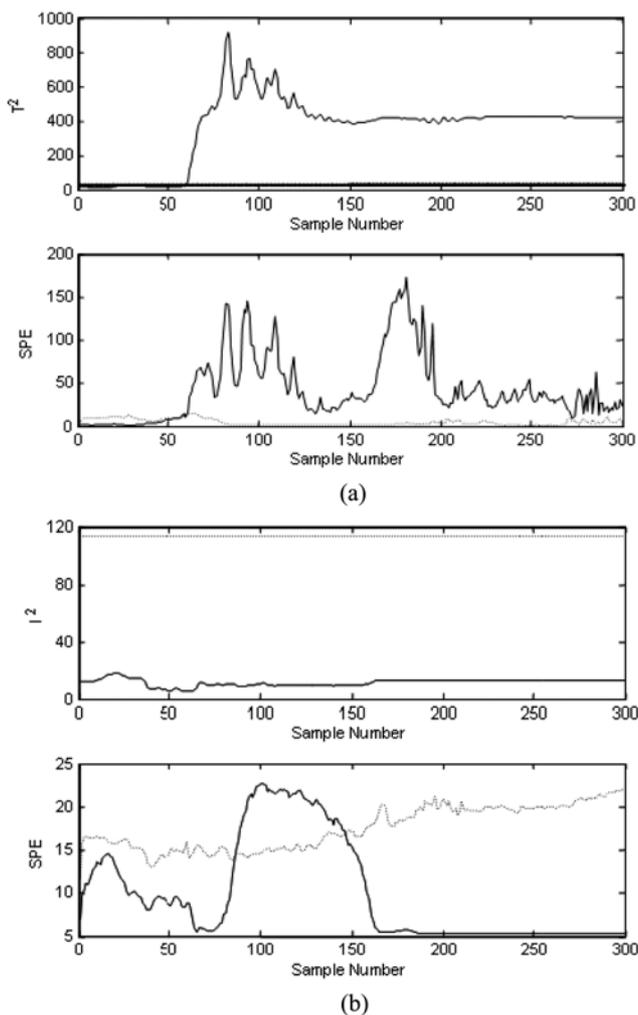


Fig. 8. On-line monitoring charts for (a) MPCA and (b) MICA in case of an abnormal batch (batch number 12).

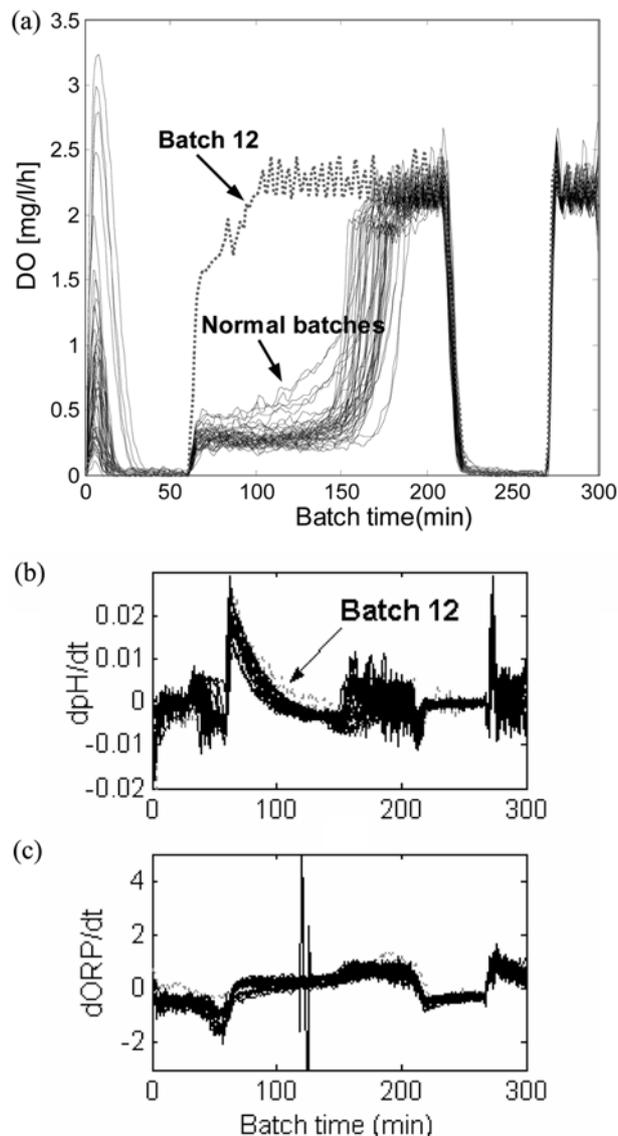


Fig. 9. Univariate plot of 49 normal batches and the abnormal batch 12, (a) DO, (b) dpH/dt, (c) dORP/dt.

an *abnormal* batch (batch number 12). Both methods can detect this batch as an abnormal one from off-line monitoring of Fig. 6. Both the MPCA and MICA methods show similar detection times for this batch. The  $T^2$  and SPE charts of MPCA in Fig. 8(a) show that this abnormal batch has a large deviation from the 61<sup>th</sup> time instant until the end of the batch operation, that is, the monitoring result of MPCA indicates that the fault continues until the end of the batch.

Fig. 9 shows the univariate plots of the original DO and two derivatives of pH and ORP signals (dpH/dt, dORP/dt) for 49 normal batches and batch 12, where DO was over-aerated at the start of the aeration phase in batch 12. Here, the DO concentration in Fig. 9(a) was increased too early (in the anaerobic phase) but returned to the normal concentration during the aerobic phase, and the derivatives of pH and ORP signals in Fig. 9(b) show abnormal behaviors in only the aerobic phase. The non-Gaussianity of the SBR system might cause this lasting false fault detection. On the other hand,

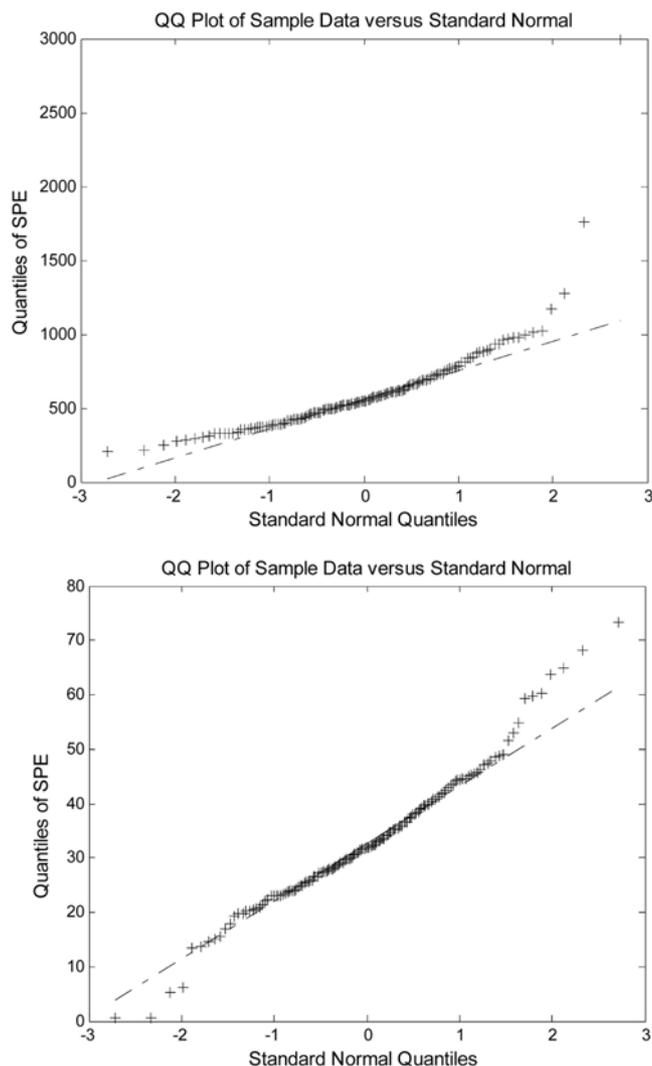
**Table 1. Detection results of offline and online monitoring in the false and missed alarms**

	False alarm rate (Type I error)		Missed alarm rate (Type II error)	
	MPCA	MICA	MPCA	MICA
Offline monitoring	3.3%	0.0%	0%	0%
Online monitoring	47.5%	4.5%	13.2%	15.4%

the  $I^2$  and SPE charts of MICA in Fig. 8(b) detects this abnormal batch behavior around the 80th time instant but returns below the control limits around the 140<sup>th</sup> batch instant.

Table 1 shows the detection results of offline and online monitoring of the MPCA and MICA, where type I and II errors correspond to false and missed alarms, respectively. In offline monitoring, MPCA shows a false alarm rate of 3.3%, while MICA shows a false alarm rate of 0%. In the online monitoring result, the MPCA control charts detect the abnormality of batch 12 too fast with a false alarm rate of 47.5%, whereas the MICA charts successfully show a deviation in their patterns only during aerobic phase with a false alarm rate of 4.5%. Because MICA can detect this fault during the aerobic phase and return within the control limits after the fault is released, MICA can detect a fault *only* during the aeration phase, which results in fewer false alarms and more robust monitoring result. On the other hand, both MPCA and MICA show the missed alarm rate of 0% in the offline monitoring result, that is, no missed alarm. In online monitoring, MICA shows larger missed alarm rate than MPCA. It represents that MICA is a more conservative monitoring method than MPCA. If one focuses on the reduction of false alarms (type I error), MPCA should be preferred, but if one emphasizes minimizing missing alarms (type II error), MICA is preferred in the monitoring system. In general, the operator in the plant prefers a robust and physically meaningful monitoring system with a lower false alarm rate.

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 [6]. Although these three additional signals for MPCA can be used to access to this valuable information, derivation always transforms the noise components in the data set, which makes the SBR data set more nonlinear. And then this nonlinearity from noise transformation makes the SBR data set into more non-Gaussian than the original variables [16]. Fig. 10 shows the QQ plot of the second score of the MPCA and the MICA model. The impact adding the first derivatives to the original variables was assessed through the use of QQ plots. From this figure, it can be deduced that the SBR process has severe non-Gaussian and non-Gaussian relations and the extracted principal scores of MPCA have a large deviation from linearity in Fig. 10(a). It means that MPCA has lower modeling ability, which may lead to a higher false alarm rate of a monitoring result of MPCA. On the other hand, MICA can extract the non-Gaussian independent components, that is, capture the non-Gaussianity from nonlinearity including the first derivatives of the original variables in SBR data in Fig. 10(b). This figure shows that the MICA has better non-Gaussian feature extraction results than MPCA.

**Fig. 10. QQ plot for the first score vector of (a) MPCA (b) MICA.**

## CONCLUSION

This paper describes the application of a pilot-scale SBR monitoring using batchwise MICA. While MPCA has the shortcoming that the measurement variables of the batch process should be Gaussian, MICA can extract meaningful hidden information from non-Gaussian data sets. The results of the pilot-scale SBR monitoring system using simple on-line measurements clearly demonstrated that the MICA monitoring technique showed lower false alarm rate and physically meaningful, that is, robust, monitoring results. It can be easily applied to most batch or fed-batch processes which have non-Gaussian distributed data, for example, chemical and biochemical processes.

## ACKNOWLEDGMENTS

This work was supported by the Korea Research Foundation Grant funded by the Korean Government (MOEHRD) (KRF-2007-331-D00089) and funded by Seoul R&BD Program (CS070160). Mr. Min Han Kim is supported by the excellent research fellowship of graduate school of Kyung Hee University (2008). The author is

grateful for use of the data set from a lab scale SBR reactor in Ghent University in Belgium.

### REFERENCES

1. G. Olsson and B. Newell, *Wastewater treatment systems: Modeling, diagnosis and control*, IWA publishing, UK (1999).
2. J. Chen and J. H. Yen, *Korean J. Chem. Eng.*, **20**, 1000 (2003).
3. S. Lee, S. Yeom and K. S. Lee, *Korean J. Chem. Eng.*, **21**, 575 (2004).
4. C. K. Yoo, D. S. Lee and P. A. Vanrolleghem, *Wat. Res.*, **38**(7), 1715 (2004).
5. K. S. Kim and J. W. Ko, *Korean J. Chem. Eng.*, **22**, 26 (2005).
6. C. H. Chang and O. J. Hao, *J. Chem. Tech. Biotechnol.*, **67**, 27 (1996).
7. P. A. Wilderer, R. L. Irvine and M. C. Goronszy, *Sequencing batch reactor technology*, IWA publishing, UK (2001).
8. Mace and Mata-Alvarez, *I&EC*, **41**, 5539 (2002).
9. C. K. Yoo, J. Lee, P. A. Vanrolleghem and I. B. Lee, *Chemom. and Intel. Lab. Sys.*, **71**(2), 151 (2004).
10. C. K. Yoo, B. H. Bang, I. B. Lee, P. A. Vanrolleghem and C. Rosén, *Korean J. Chem. Eng.*, **21**, 1087 (2004).
11. C. K. Yoo, I.-B. Lee and P. A. Vanrolleghem, *Environmental Monitoring and Assessment*, **119**(1-3), 1573 (2006).
12. J. M. Lee, S. J. Qin and I. B. Lee, *AIChE J.*, **52**(10), 3501 (2006).
13. O. Stanimirovic, H. C. J. Hoefsloot, P. K. de Bokx and A. K. Smilde, *J. Chemometrics*, **20**, 43 (2006).
14. Z. Ge and Z. Song, *Ind. Eng. Chem. Res.*, **46**, 2054 (2007).
15. P. Nomikos and J. F. MacGregor, *AIChE J.*, **40**(8), 1361 (1994).
16. A. Hyvärinen, J. Karhunen and E. Oja, *Independent component analysis*, John Wiley & Sons, INC., USA (2001).
17. T. Lee, *Independent component analysis: Theory and applications*, Kluwer Academic Publishers, Boston, USA (1998).
18. Q. Chen, R. J. Wynne, P. Goulding and D. Sandoz, *Cont. Eng. Prac.*, **8**, 531 (2000).