ENVIRONMETRICS
Environmetrics 2008; **19**: 602–617
Published online 18 December 2007 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/env.900

Enhanced process monitoring for wastewater treatment systems

Chang Kyoo Yoo[1*,†], Kris Villez[2], Stijn W.H. Van Hulle[3] and Peter A. Vanrolleghem[2,4]

[1] *College of Environment and Applied Chemistry/Center for Environmental Studies, Kyung Hee University, Seocheon-dong 1, Giheung-gu, Yongin-Si, Gyeonggi-Do, 446-701, Korea*
[2] *BIOMATH, Department of Applied Mathematics, Biometrics and Process Control, Ghent University, Coupure Links 653, B-9000 Gent, Belgium*
[3] *PIH, Department of Industrial Engineering and Technology, Hogeschool West-Vlaanderen, Graaf Karel de Goedelaan 5, B-8500 Kortrijk, Belgium*
[4] *modelEAU, Département de Génie Civil, Pavillon Pouliot, Université Laval, Québec G1K 7P4, Canada*

SUMMARY

Wastewater treatment plants (WWTPs) remain notorious for poor data quality and sensor reliability problems due to the hostile environment, missing data problems and more. Many sensors in WWTP are prone to malfunctions in harsh environments. If a WWTP contains any redundancy between sensors, monitoring methods with sensor reconstruction such as the proposed one can yield a better monitoring efficiency than without a reconstruction scheme. An enhanced robust process monitoring method combined with a sensor reconstruction scheme to tackle the sensor failure problems is proposed for biological wastewater treatment systems. The proposed method is applied to a single reactor for high activity ammonia removal over nitrite (SHARON) process. It shows robust monitoring performance in the presence of sensor faults and produces few false alarms. Moreover, it enables us to keep the monitoring system running in the case of sensor failures. This guaranteed continuity of the monitoring scheme is a necessary development in view of real-time applications in full-scale WWTPs. Copyright © 2007 John Wiley & Sons, Ltd.

KEY WORDS: environmental monitoring; faulty sensor; reliability; sensor failure; SHARON

1. INTRODUCTION

The number of sensor and other data in wastewater treatment and water transport systems has increased almost exponentially over the last few decades (Seyfried *et al*., 2001; Jeppsson *et al*., 2002; Olsson and Newel, 1999; Rieger *et al*., 2003, 2004, 2005; Vanrolleghem and Lee, 2003; Yoo *et al*., 2004b, 2006). This does not necessarily mean that the information has increased as much. With increasing sensor availability and stricter effluent quality requirements the operator will need ever improving support from the control systems. This leads to higher demands on reliable fault analysis, data screening, information extraction and condensation, and operator guidance. Next to that, a strong need for guarantee of data quality is expressed (Strotmann *et al*., 1995; Jetten *et al*., 1997; Cox *et al*., 2000; Rosen and Lennox, 2001; Rosen *et al*., 2003).

*Correspondence to: C. K. Yoo, College of Environment and Applied Chemistry/Center for Environmental Studies, Kyung Hee University, Seocheon-dong 1, Giheung-gu, Yongin-Si, Gyeonggi-Do, 446-701, Korea.
†E-mails: ckyoo@khu.ac.kr; changkyoo.yoo@biomath.ugent.be

Received 9 January 2007
Accepted 30 October 2007

Process operators obtain information on the current process conditions from a range of sensor types. Hence, the accuracy of sensors is crucial to successful process control and monitoring and the ability to detect sensor faults is very useful, especially when processes are monitored and controlled based on process information from many sensors. Sensors may exhibit partial failures such as bias, drift or precision degradation as displayed in Figure 1. It causes the accuracy and reliability of the measurement to decrease, which may result in an erroneous control action and false perception of the performance of the monitored system. Faulty sensors that are either completely or partially failing (hard fault or soft fault) provide incorrect information for monitoring and control. This can be detrimental to various data-driven decision schemes. Moreover, data may not be available due to sensor malfunction or communication problems within the data collection system. These data problems make it difficult to extract and interpret information from data. Monitoring or control using the measurements is then problematic.

Conventional engineering methods to find and to correct for sensor faults make use of procedures that check and recalibrate the sensors periodically. Often, this does not satisfy the requirements of the hostile environment in environmental processes, such as water, waste and air pollution. Therefore, prompt detection of the occurrence and correct identification of the location of sensor faults and reliable reconstruction (or recovery) of faulty sensors is of primary importance for efficient operation. In contrast to fault detection and isolation, sensor fault detection and validation is quite a new research area, which is required for use in wastewater treatment, but has few application results (Qin and Li, 1998; Qin, 2003; Volcke *et al*., 2005).

This paper concentrates on formulating a process monitoring system to the problem of ''faulty sensor'' characteristics in wastewater treatment plant (WWTP). The method developed here uses a

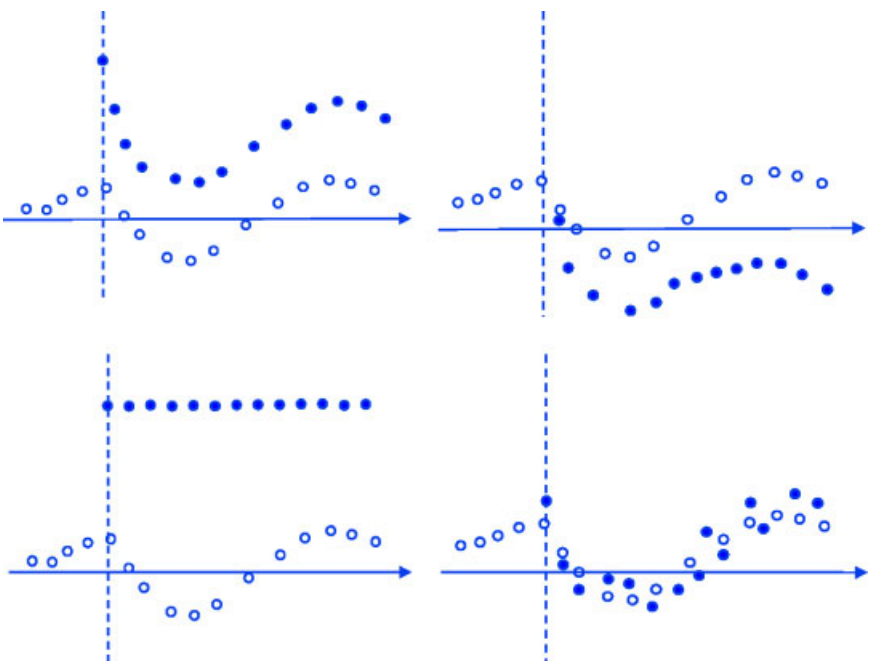

Figure 1. Four types of sensor faults, (a) bias, (b) drifting, (c) complete failure, (d) precision degradation

DOI: 10.1002/env

principal component analysis (PCA) model for process monitoring and a sensor validation model to generate a reconstructed sensor value which can be used to develop a more reliable monitoring model. This approach is organized by combining the PCA model and the faulty sensor reconstruction method. The important tasks involved in the proposed monitoring system are to check the biological process state continuously, and maintaining its capabilities by applying the sensor reconstruction technique under the sensor failure case.

This paper is organized as follows. Section 2 shows the single reactor high activity ammonia removal over nitrite (SHARON). Section 3 introduces the statistical process monitoring, the sensor fault identification and reconstruction methodology and proposes an enhanced process monitoring approach to tackle faulty sensors. The monitoring performances of the proposed method are illustrated through a lab-scale reactor in section 4. Finally, conclusions are drawn in section 5.

2. MATERIALS

2.1. SHARON process

The SHARON process has demonstrated its efficiency and flexibility in the treatment of sludge digestion wastewater which is characterized by high concentration of ammonia nitrogen. In the SHARON process, partial nitrification i.e. biological oxidation of ammonium to nitrite is established by working at high temperature (about 35°C) and maintaining an appropriate sludge retention time (SRT) of 1–1.5 days, so that ammonium oxidizers are maintained in the reactor, while nitrite oxidizers are washed out and further oxidation of nitrite to nitrate is prevented. In this way, significant aeration cost savings are realized in comparison with conventional complete nitrification. When only half of the ammonia is converted, a combination with an Anammox unit becomes economically interesting, as in the subsequent Anammox treatment step equimolar amounts of ammonia and nitrite are converted to nitrogen gas (Hellinga *et al*., 1998; van Dongen *et al*., 2001; Seyfried *et al*., 2001). In comparison with conventional *N*-removal, the coupled SHARON and Anammox processes in theory result in a 40% reduction of the stoichiometrically required oxygen while no carbon source needs to be added and a negligible amount of sludge is produced. The success of this concept is, however, highly dependent on the control of the SHARON process since a stable operation of the Anammox process requires equimolar concentrations of ammonia and nitrite in the SHARON effluent and absence of nitrite inhibition in the Anammox reactor (van Dongen *et al*., 2001; Van Hulle *et al*., 2005; Volcke *et al*., 2006).

2.2. A lab-scale SHARON reactor

A lab-scale SHARON reactor was constructed and operated in the BIOMATH laboratory (Figure 2). The reactor is a 2 L continuously stirred tank reactor (CSTR) without biomass retention. The synthetic influent is pumped with a peristaltic pump from the 5 L influent vessel to the reactor. The pump flow rate of this influent pump determines both the hydraulic residence time (HRT) and the sludge retention time (SRT), since both residence times are equal and defined as the ratio of the volume to the flow rate. The reactor is aerated through a pumice stone using air from a compressor (1 bar) and the normal operational temperature is 35°C. In the reactor the dissolved oxygen (DO) and pH are measured. The pH is controlled through Labview software (National Instruments, www.ni.com) by the addition of acid (HCl) and base ($NaHCO_3$). The data used in the research were collected in the steady-state operation period and consist of 10 variables: (1) HRT, (2) influent ammonium, (3) bicarbonate:ammonium ratio,

DOI: 10.1002/env

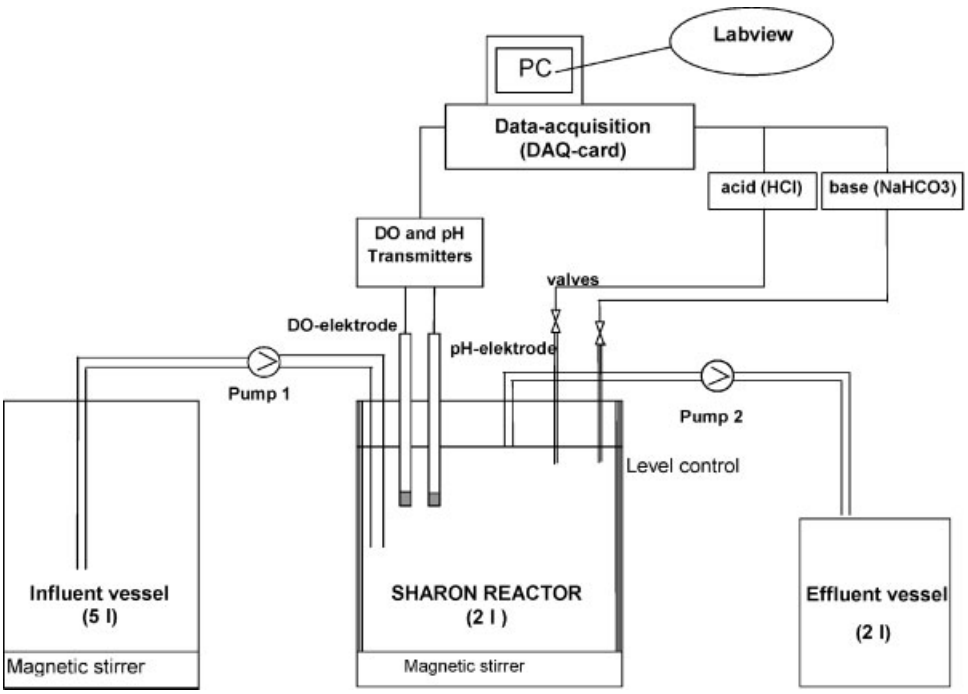

Figure 2. Set-up of a lab-scale SHARON reactor

(4) DO, (5) pH, (6–7) dosage rates of base and acid, and (8–10) daily measurements of ammonia, nitrite, and nitrate. Their typical variable trajectories show three times abnormal behaviors of acid and base addition at samples 65, 67, and 77, at which abnormally high amounts of base and acid are added to the reactor due to actuator failures. Except for these three samples, the SHARON reactor operated normally (Van Hulle *et al*., 2005).

3. METHODS

3.1. Statistical monitoring

The fault detection method, which monitors the process is based on the extraction of information that is latent in the multidimensional data matrix and unobservable by means of univariate methods. A key point is thus how to extract the hidden information from the multidimensional data set. Provided with historical data during normal process operation, most of the common-cause variation in the process can be expressed as a function of a set of variables, which are lower in number compared to the set of original variables. PCA models as used here are able to account for correlation between variables in a monitoring context. PCA decomposes the data matrix ($\mathbf{X}$) which covers m sensors and N samples into a score matrix $\mathbf{T}$ and a loading matrix $\mathbf{P}$ by singular value decomposition (SVD)

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{\mathrm{T}} + \tilde{\mathbf{T}}\tilde{\mathbf{P}}^{\mathrm{T}} = \hat{\mathbf{X}} + \tilde{\mathbf{X}} = [\mathbf{T}\,\tilde{\mathbf{T}}][\mathbf{P}\,\tilde{\mathbf{P}}]^{\mathrm{T}} = \overline{\mathbf{T}}\,\overline{\mathbf{P}}^{\mathrm{T}} \tag{1}$$

where $\hat{\mathbf{X}} = \mathbf{T}\mathbf{P}^{\mathrm{T}}$ is the model matrix and $\tilde{\mathbf{X}} = \tilde{\mathbf{T}}\tilde{\mathbf{P}}^{\mathrm{T}}$ is the residual matrix. The principal component subspace (PCS) is $\mathrm{S_p}$=span$\{\mathbf{P}\}$ and the residual subspace (RS) is $\mathrm{S_r} = \mathrm{span}\{\tilde{\mathbf{P}}\}$. A sample vector $\mathbf{x}$ can

DOI: 10.1002/env

be projected on the PCS and RS respectively:

$$\hat{\mathbf{x}} = \mathbf{P}\mathbf{P}^{\mathrm{T}}\mathbf{x} \equiv \mathbf{C}\mathbf{x} \in \mathrm{S_p} \tag{2}$$

$$\tilde{\mathbf{x}} = \tilde{\mathbf{P}}\tilde{\mathbf{P}}^{\mathrm{T}}\mathbf{x} \equiv (\mathbf{I} - \mathbf{C})\mathbf{x} = \tilde{\mathbf{C}}\mathbf{x} \in \mathrm{S_r} \tag{3}$$

$$\mathbf{x} = \hat{\mathbf{x}} + \tilde{\mathbf{x}} \tag{4}$$

PCA uses the squared prediction error (SPE) as a fault detection index

$$\mathrm{SPE} = \|\tilde{\mathbf{x}}\|^2 = \left\|\tilde{\mathbf{C}}\mathbf{x}\right\|^2 = \mathbf{x}^{\mathrm{T}}(\mathbf{I} - \mathbf{C})\mathbf{x} \leq \delta_\alpha \tag{5}$$

The detection limit for SPE can be determined with Q-statistics. If a sensor fails, which breaks the normal correlation, the residual will increase above detection threshold (Qin, 2003).

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

$$T^2(k) = \mathbf{t}(k)\Lambda^{-1}\mathbf{t}(k)^{\mathrm{T}} \tag{6}$$

where Λ^{-1} is the diagonal matrix of the inverse of the eigenvalues associated with the retained principal components (PCs). The confidence limit for T^2 is obtained using the F-statistic:

$$T^2_{m,n,\alpha} = \frac{m(n-1)}{n-m} F_{m,n-m,\alpha} \tag{7}$$

where n is the number of samples in the model, m is the number of PCs, and α is an appropriate level of significance for performing the test which typically takes the value of 0.05 or 0.01 for the warning and action limits respectively.

The confidence limit for the SPE can be computed from its approximate distribution

$$\mathrm{SPE}_\alpha = \Theta_1 \left[\frac{c_\alpha \sqrt{2\Theta_2 h_0^2}}{\Theta_1} + 1 + \frac{\Theta_2 h_0 (h_0 - 1)}{\Theta_1^2} \right]^{1/h_0} \tag{8}$$

where c_α is the standard normal deviate corresponding to the upper $(1-\alpha)$ percentile and λ_j is the eigenvalue associated with the jth loading vector, $\Theta_i = \sum_{j=m+1}^{d} \lambda_j^i$ for $i = 1, 2, 3$ and $h_0 = 1 - \frac{2\Theta_1\Theta_3}{3\Theta_2^2}$.

For a new on-line sample $\mathbf{x}_{\mathrm{new}}$, if $T^2_{\mathrm{new}} < T^2_{\mathrm{lim}}$ and $Q_{\mathrm{new}} < Q^2_{\mathrm{lim}}$, we consider the process to be in control with $100(1-\alpha)\%$ confidence. Otherwise, the process may be out of control. Here, the T^2 value which indicates a distance is used to detect faults associated with abnormal variations within a model subspace, whereas the Q value which indicates a distance from the residual space is used to detect new events that are not taken into account in the model subspace (Yoo *et al*., 2004a).

3.2. Sensor fault identification and reconstruction

If the normal process model is built as a PCA model, the model residuals are used to detect sensor faults. In the presence of a sensor fault, the sensor measurement ($\mathbf{x}$) will contain the normal values of the process variables and the fault, that is

$$\mathbf{x} = \mathbf{x}^* + \Xi_i \mathbf{f}_i(t) \tag{9}$$

DOI: 10.1002/env

where $\mathbf{x}^*$ is a vector of normal sensor values, $\mathbf{f}_i(t) \in \mathcal{R}^{li}$ is a vector of the fault magnitude, $\boldsymbol{\Xi}_i \in \mathcal{R}^{nxli}$ is a matrix of fault directions, and $\mathbf{f}_i$ is the magnitude of the fault. While $\Xi_i = [0\,0 \cdots 1 \cdots 0]^{\mathrm{T}}$ represents a single sensor fault in the ith sensor, $\boldsymbol{\Xi}_i$ contains the corresponding columns of the identity matrix to represent multiple sensor faults. Using Equation (6), the model residual, $\mathbf{e}(t)$, can be written as follows:

$$\mathbf{e}(t) = \mathbf{B}\mathbf{x}(t) = \mathbf{B}\mathbf{x}^*(t) + \mathbf{B}\boldsymbol{\Xi}_i\mathbf{f}_i(t) = \mathbf{e}^*(t) + \mathbf{B}\boldsymbol{\Xi}_i\mathbf{f}_i(t) \tag{10}$$

where $\mathbf{B}$ is the model matrix which is $\tilde{\mathbf{P}}^{\mathrm{T}}$ in the PCA model and $\mathbf{e}^*(t)$ is the model residual which contains a measurement noise. A fault will cause the residual $\mathbf{e}(t)$ to increase (Qin and Li, 1998; Qin, 2003).

3.2.1. Sensor fault identification. When a sensor failure is detected, the faulty sensor must be identified. In this paper, we used the structured residual approach with maximized sensitivity (SRAMS) (Qin and Li, 1997). They suggested a structured residual approach wherein a set of residuals is generated of which each residual is most sensitive to a specified subset of faults while being insensitive to others. For the case of a single sensor fault in the ith sensor, Equation (10) becomes

$$\mathbf{e}(t) = \mathbf{e}^*(t) = \mathbf{b}_i\mathbf{f}_i(t) \tag{11}$$

where $\mathbf{b}_i$ is the ith column of matrix $\mathbf{B}$ which represents the fault direction. By pre-multiplying a transformation matrix $\mathbf{W}$ to $\mathbf{e}(t)$, we can generate the following structured residuals $r(t)$:

$$r(t) = \mathbf{W}\mathbf{e}(t) = \mathbf{b}_i\mathbf{f}_i(t) \tag{12}$$

where the matrix $\mathbf{W}$ is designed so that each element of $r(t)$ is insensitive to one particular sensor fault and sensitive to the other faults. $\mathbf{w}_i$ is chosen so that $r_i(t)$ is insensitive to the ith sensor fault but most sensitive to the others. Mathematically, this is equivalent to

$$\max_{\mathbf{w}_i} \sum_{j \neq i} \frac{(\mathbf{w}_i^{\mathrm{T}}\mathbf{b}_j)^2}{\|\mathbf{b}_j\|^2} \tag{13}$$

Geometrically, $\mathbf{w}_i$ is chosen to be orthogonal to $\mathbf{b}_i$ as this minimizes its angle to other fault directions $\mathbf{b}_j$. Four types of sensor fault identification are suggested by Qin and Li (1998), which are (1) an identification index based on exponential weighted moving average (EWMA) filtered squared residuals (I_{FSR}), (2) the generalized likelihood ratio (GLR), (3) the cumulative sum of residuals (Q_{SUM}), and (4) the cumulative variances index (V_{SUM}) (Qin and Li, 1997).

3.2.2. Sensor reconstruction. After a fault is detected, it is important to identify the fault and apply the necessary corrective actions to eliminate the abnormal condition. The procedure to restore normal conditions by applying a corrective change in the data is called data reconstruction. Logically, the procedure for identifying a fault by reconstruction for a given type of faults is called data identification via reconstruction. Reconstruction of the normal data from faulty measurements leads to the estimation of the fault magnitude. Therefore fault reconstruction is presented first, followed by fault identification (Qin, 2003). Given a sensor fault direction, the best reconstruction can be used to estimate the sensor fault magnitude $\mathbf{f}_i(t)$ by minimizing the effect of the fault on $\mathbf{e}(t)$ in the direction $\boldsymbol{\Xi}_i$. Mathematically, this is formulated as the following least-squares problem:

$$J = \|\mathbf{e}^*(t)\|^2 = \|\mathbf{e}(t) - \mathbf{B}\boldsymbol{\Xi}_i\mathbf{f}_i(t)\|^2 \tag{14}$$

DOI: 10.1002/env

The estimated magnitude for the given sensor fault direction is then

$$\hat{\mathbf{f}}_i(t) = (\mathbf{B}\Xi_i)^+ \mathbf{e}(t) \tag{15}$$

where $(\cdot)^+$ is the Moore–Penrose pseudo inverse. The task of sensor reconstruction is to estimate the normal values $\mathbf{x}^*$ by eliminating the effect of a fault $\mathbf{f}_i$. A reconstructed value $\mathbf{x}_i$ is calculated by correcting the effect of a fault on the process data $\mathbf{x}$:

$$\mathbf{x}_i = \mathbf{x} - \mathbf{B}\Xi_i \hat{\mathbf{f}}_i(t) \tag{16}$$

where $\hat{\mathbf{f}}_i$ is an estimate of the actual fault magnitude $\mathbf{f}_i$ along the SRAMS direction. The reconstructed value $(\hat{\mathbf{x}}_i)$ could be used in the monitoring and prediction models instead of the actual faulty measurements (Qin and Li, 1997; Yoo *et al*., 2006).

3.3. Enhanced monitoring scheme

The need for the integration of sensor signals with data quality verification and fault-detection systems is gaining acceptance from plant staff of WWTP (Jetten *et al*., 1997; Olsson, 2002; Jeppsson *et al*., 2003). Figure 3 shows the proposed monitoring scheme which is able to detect and compensate for faulty measurements, thereby enhancing its monitoring usefulness further. The first step in this methodology is the construction of the PCA model using normal historical data in order to construct the common-cause redundancy model. Then, sensor fault identification and reconstruction are executed by the SRAMS sensor validation system in the second step. If any index of a sensor signal exceeds the confidence limit, sensor reconstruction should be executed. The sensor fault magnitude and fault type can be estimated by means of the reconstructed sensor value. When a faulty sensor has been identified, the reconstructed value for the corresponding measurement is used to replace the faulty measurement in the monitoring system. Finally, the statistical monitoring system can discern abnormal events and disturbances from normal operational conditions. The enhanced monitoring approach proposed here therefore gives us the capability to keep the monitoring system running in the presence of faulty measurements. Additionally, the variable contribution plots of conventional monitoring systems can be used as an independent means of interpreting and isolating faults.

4. RESULTS AND DISCUSSION

4.1. Process monitoring in the case of faulty sensors

A process model of the SHARON data set using PCA with two PCs was constructed and an isolation matrix with SRAMS was designed. Filtered SPE and FSR are used to detect the sensor fault and two indices of I_{FSR}, V_{SUM}, and V_{SUM} with the 95% confidence level are monitored to identify faulty sensors. Two types of sensor faults, including complete failure and precision degradation are introduced at time t_{f}, where the abnormal condition is caused by single sensor failure. The remaining measurements are used to reconstruct the faulty sensor based on the redundancy of the measurements.

In order to show the impact of a faulty sensor on the monitoring results, the performance of the monitoring system with and without sensor validation system is compared using two scenarios. Sensors in wastewater treatment are often corrupted due to air bubbles, significant noise, or they may fail completely due to sludge fouling and have to function in very harsh environments (Olsson and Newell, 1999). Four types of sensor faults including bias, drift, complete failure, and precision degradation

DOI: 10.1002/env

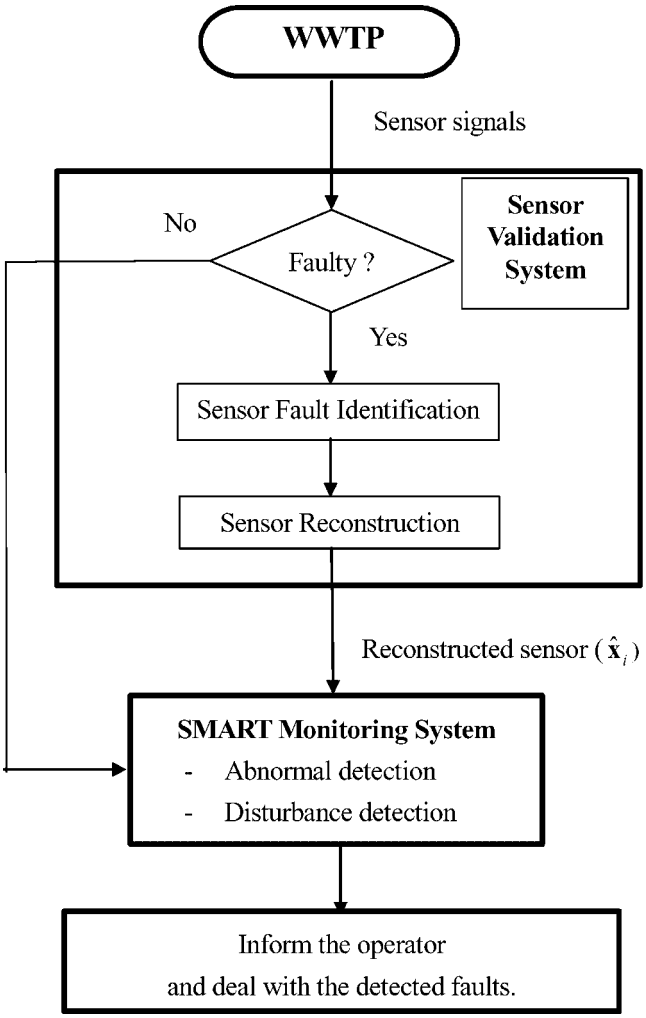

Figure 3. The proposed enhanced monitoring scheme with a sensor reconstruction module

which are common in WWTPs are introduced at time t_f, where the sensor abnormal condition is caused by each failure scenario. Because the DO sensor is the most used sensor in WWTP and represents a high correlation with other variables, we focused on two fault scenarios related to the DO sensor which are the precision degradation due to significant noise and complete failure due to sludge clogging. The remaining measurements are used to reconstruct the faulty sensor based on the redundancy of the measurements (Wang and Chen, 2004). Table 1 summarizes the two types of abnormal conditions detected and lists the fault and detection times. In order to reduce false alarms due to dynamic transients, an EWMA filter with a coefficient $r = 0.90$ was applied to generate the FSR's for both fault cases. The V_{SUM} index is calculated based on the unfiltered structured residuals with a moving window of five samples considering the hydraulic retention time.

In the first test, precision degradation with a noise variance of 2 was added to the DO sensor at sample 50 and last until the end of the data set. EWMA-filtered squared residual (FSR) in Figure 4 can

DOI: 10.1002/env

Table 1. Summary of two sensor failures and their detection results

	Precision degradation	Complete failure
Faulty sensor	DO	DO
Fault expression	$\mathbf{f}_1(t) = \mathrm{DO}(t) + N(0,2)$	$\mathbf{f}_2(t) = c$
Fault size	$\mathrm{DO}(t) + N(0,2)$	$c = 9.0$
Fault time (t_f)	50	50
Detection time ($\hat{t}_f$)	54	51

detect this fault more effectively than the SPE plot. To identify which sensor is faulty, four fault indices of I_{FSR}, GLR, Q_{SUM}, and V_{SUM} are examined for all sensors. Since this is a variance change, V_{SUM} has a value below the control limits for the fourth sensor, which indicates that the DO is a faulty sensor. In Figure 5(a), the reconstructed sensor signal indicates that a difference between normal and reconstructed sensor data is relatively small and the reconstruction can thus be used for replacement of the faulty data, provided the faulty data are compared with the reconstructed data. The estimated fault size in Figure 5(b) shows the result from a precision degradation and how large the fault is. Figure 6 compares the monitoring performances affected by this fault. Although the sensor signal comes from the faulty sensor, this had no effect on the wastewater treatment. With the faulty sensor, the T^2 and SPE charts in Figure 6(a) move up and down many times above the control limits from the beginning of the sensor fault to the end although the operation status of the SHARON process is normal. On the other

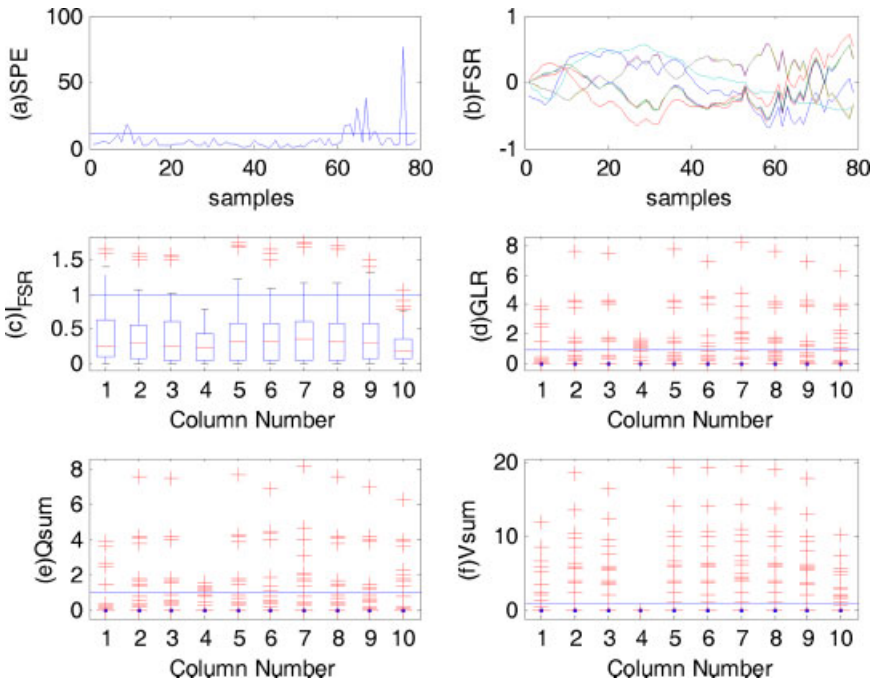

Figure 4. Sensor fault identification of a precision degradation fault in DO sensor, (a) SPE plot, (b) FSR, (c) I_{FSR}, (d) GLR, (e) Q_{SUM}, (f) V_{SUM}

DOI: 10.1002/env

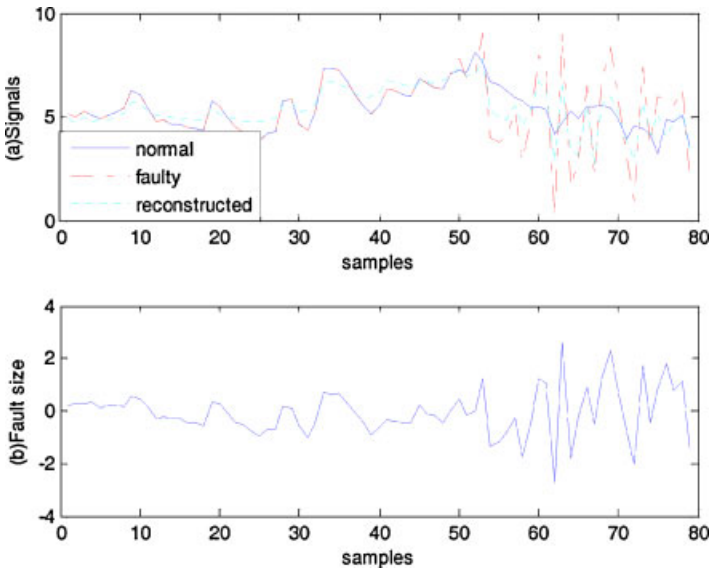

Figure 5. Sensor reconstruction result of a precision degradation fault in DO sensor, (a) normal, faulty, and reconstructed signals, (b) fault size

hand, sensor validation checked the validity of the sensor after it is detected faulty. After the DO measurement is reconstructed, the monitoring results in Figure 6(b) show only three abnormal results, which are close to the case of no sensor faults. The monitoring reliability is improved by validating the measurement data, isolating the failed sensor and finally recovering, by model estimation, the failed sensor measurement. The reconstructed measurements of the faulty sensor result in a more accurate status representation, which in turn leads to a minimized number of false alarms, while true deviations are retained and detected.

Secondly, a complete failure of the DO sensor is tested. The failure is introduced at time 50, where the DO value is assumed to be constant at 9 mg/L and lasts until the end of the data set. As shown in Figure 7, this sensor fault was detected, identified, and reconstructed. The complete failure (sensor 4) is detected in the SPE plot at sample 52 and is effectively detected in the identification indices within a relatively short time. To illustrate the fault identification in detail, four fault indices are shown in Figure 7. Values below 1 indicate faulty situations. FSR can exactly identify two sensors: number 4 (DO) and $NO_{3,e}$ which are below the confidence limit, as $NO_{3,e}$ is strongly correlated with DO in the loading plot (not shown). Since this fault is the result of a complete failure, all fault indices have a smallest value for the fourth sensor (DO) which makes the correct identification of the faulty sensor possible. The estimated fault size in Figure 8(b) shows the result from a complete failure and how large the fault is. The monitoring performances affected by this fault are compared in Figure 9. With the faulty sensor, the SPE charts in Figure 9(a) remain above the control limits from sample 65 to the end although the operation status of the SHARON process is normal and the faulty sensor had no effect on the wastewater treatment. Obviously, the reliability of the multivariate monitoring system is detoriated and makes it subject to unfavorable criticism. When the sensor are reconstructed, the T^2 and SPE values in Figure 9(b) remain within the control limits except for the three abnormal events of extreme acid and base addition, hereby improving the robustness of the monitoring system.

DOI: 10.1002/env

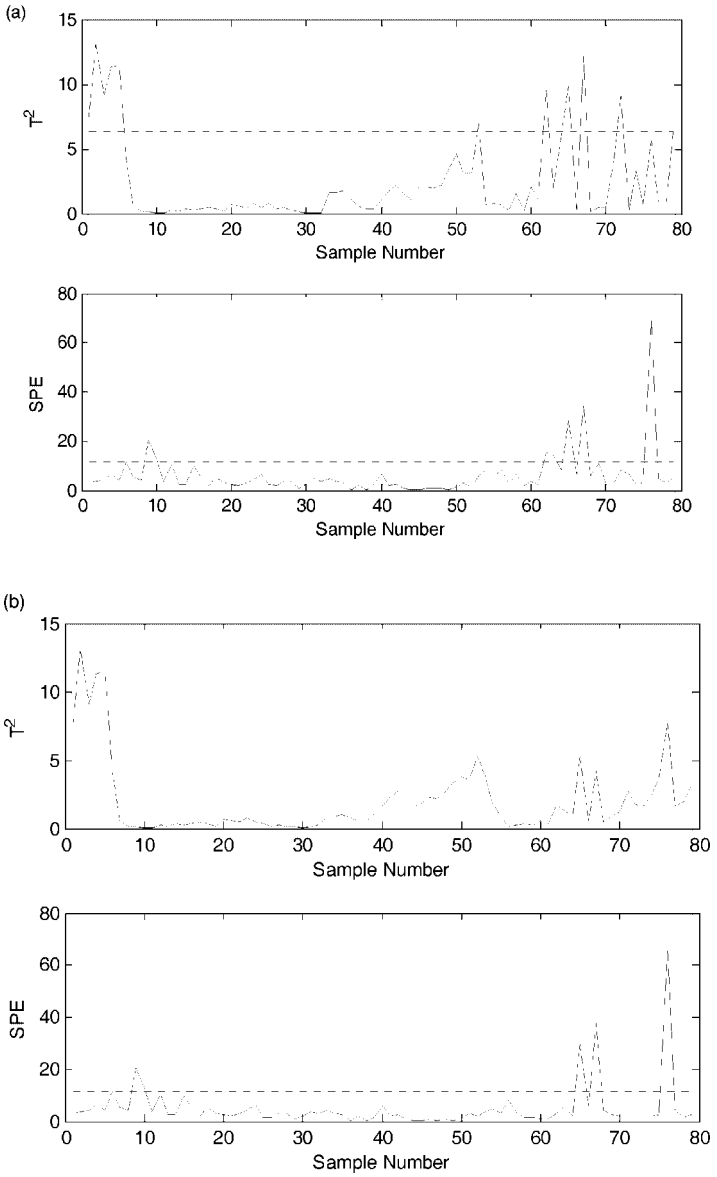

Figure 6. T^2 and SPE plots of the PCA monitoring for a precision degradation in DO sensor using (a) faulty sensor (b) reconstructed sensor

4.2. Contribution plot in the case of faulty sensors

Once a sensor fault or special event has been detected, it is important to diagnose the event to find an assignable cause. For this, the contribution of each measurement variable to the deviations observed in the monitoring metric can be displayed. These contribution or diagnostic charts can be immediately displayed on-line by the operator as soon as the special event is detected. Although these plots do not

DOI: 10.1002/env

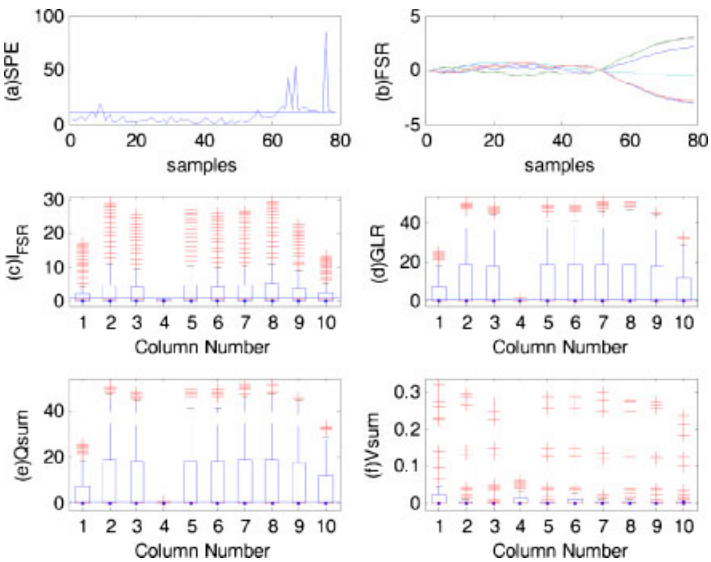

Figure 7. Sensor fault identification of a complete failure in DO sensor, (a) SPE plot, (b) FSR, (c) I_{FSR}, (d) GLR, (e) Q_{SUM}, (f) V_{SUM}

provide direct fault isolation, they show which group of variables is highly correlated with the fault; it is then up to the process operator to use his or her insight to provide feasible interpretations (Wise and Gallagher, 1996; Yoo *et al*., 2004c). In Figure 10, contribution plots of the SPE statistics using the faulty sensor and the reconstructed sensor at sample 75 are shown respectively. The contribution plot of the faulty sensor in Figure 10(a) shows that the variable 4 (DO) primarily contributes to the SPE values

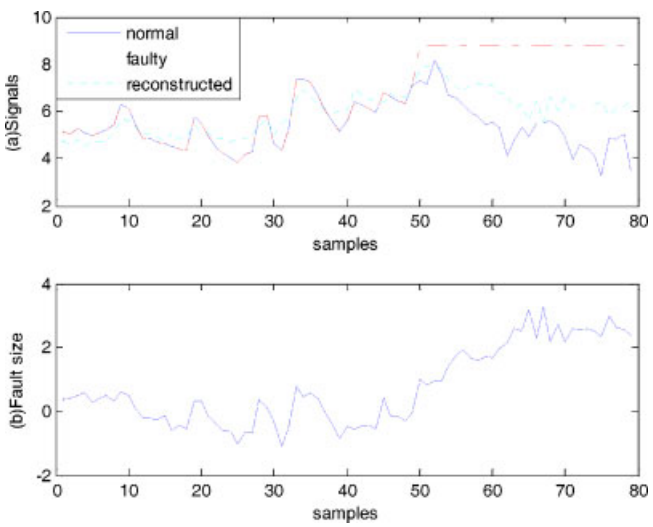

Figure 8. Sensor reconstruction result of a complete failure in DO sensor (a) normal, faulty and reconstructed signals, (b) fault size

DOI: 10.1002/env

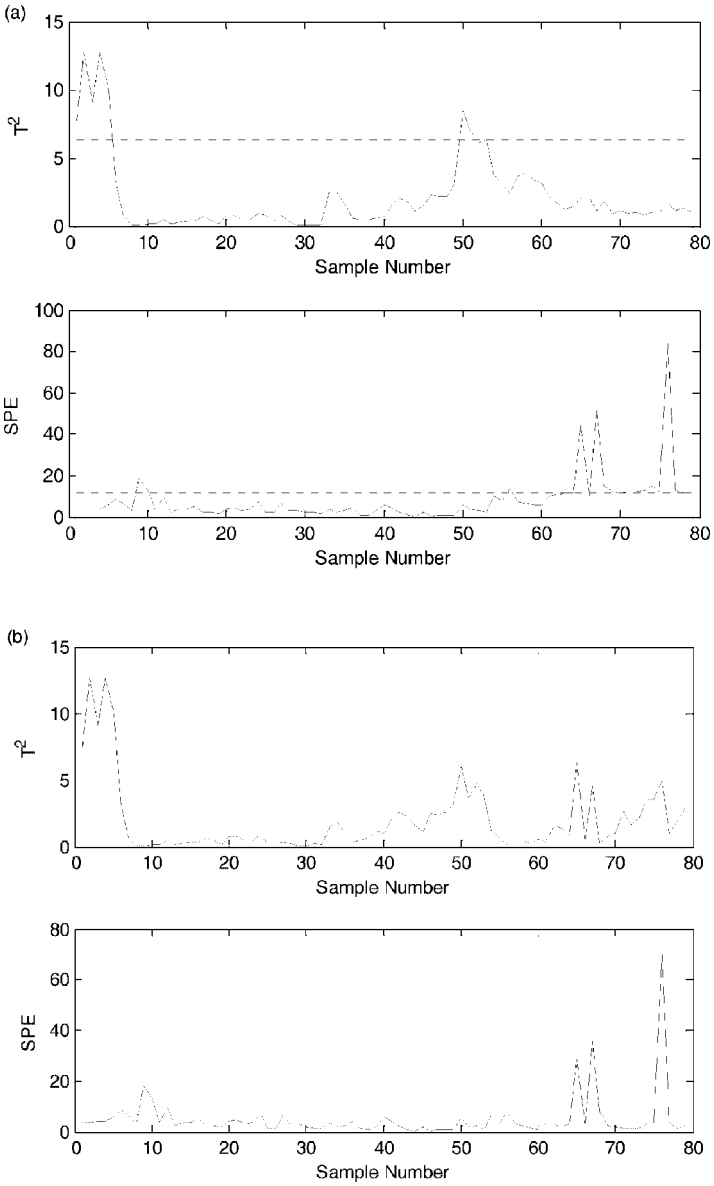

Figure 9. T^2 and SPE plots of the PCA monitoring for a complete failure in DO sensor using (a) faulty sensor (b) reconstructed sensor

but this has no effect on the wastewater treatment performance. This may be misleading to an operator and may eventually cause major process upsets by incorrect control actions. On the other hand, it is notable that the contribution plots for SPE value with the reconstructed sensor in Figure 10(b) show contributions which are relatively smaller than those under the faulty sensor case, thus indicating normal behavior. This difference in magnitude of the contributions are the result of an increased sensor

DOI: 10.1002/env

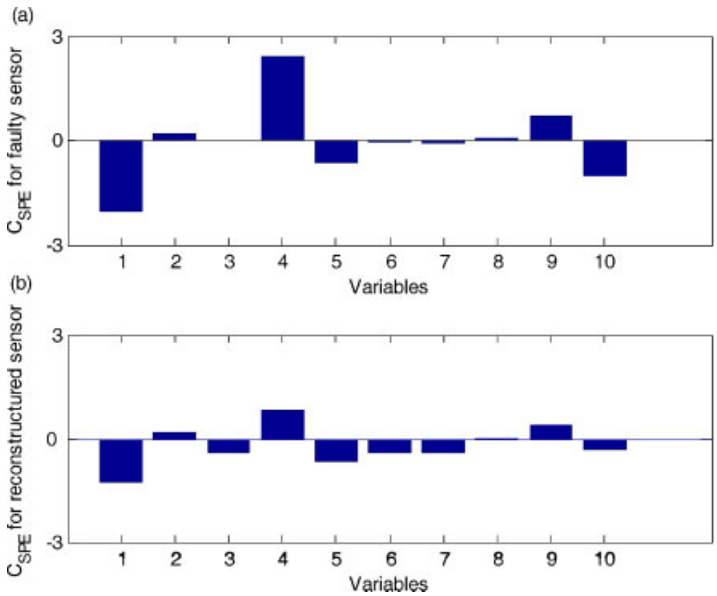

Figure 10. Contribution plots of the PCA monitoring for a complete failure in DO sensor using (a) faulty sensor (b) reconstructed sensor

reliability by considering a complete failure phenomenon such as the sludge fouling which occurs occasionally in WWTP.

5. CONCLUSION

WWTPs are notorious for poor data quality and sensor reliability due to the harsh environment that sensors are put in. The equipment and instrumentation problems in the WWTP therefore pose an interesting challenge for monitoring and supervisory control of systems characterized by faulty sensors, missing data problems, and equipment failures. It motivates the possibility that an enhanced monitoring system with sensor validation modules can be used in WWTPs, in order to allow the process monitoring system and the eventual control system to take the faulty measurement into account and efficiently fulfill its mission with reconstructed data. The application of the proposed monitoring method to the SHARON process has demonstrated its enhanced monitoring performance and highlights its potential for WWTP optimization. In addition, it can be stated that a sensor validation and reconstruction scheme is a crucial element of a robust monitoring system which is able to cope with all possible kinds of sensor failures. Only then, other important tasks in monitoring of biological processes, such as assessment of the process capability and performance, can be performed.

6. NOMENCLATURE

a	the number of latent variables in PCA
$\mathbf{B}$	model matrix

DOI: 10.1002/env

d	the number of variables
$\mathbf{E}$	residual matrix
$\mathbf{e}(k)$	kth row of $\mathbf{E}$
$\mathbf{f}_i$	a vector of the fault magnitude
$\hat{\mathbf{f}}_i$	an estimate of the actual fault magnitude $\mathbf{f}_i$
m	the number of principal components
n	the number of samples
k	sample index
$\mathbf{P}$	loading matrix
PCS	principal component subspace
$r(t)$	the structured residual
RS	residual subspace
SPE	the squared prediction error
$\mathbf{T}$	the score matrix
T^2	Hotelling's T^2 statistic
$\mathbf{X}$	data matrix ($\mathbf{X} \in R^{n \times d}$ in PCA)
$\hat{\mathbf{X}} = \mathbf{T}\mathbf{P}^{\mathrm{T}}$	the model matrix
$\tilde{\mathbf{X}} = \tilde{\mathbf{T}}\tilde{\mathbf{P}}^{\mathrm{T}}$	the residual matrix
$\mathbf{x}$	d-dimensional column vector of data matrix
$\mathbf{x}^*$	a vector of normal sensor values
$\mathbf{x}_j(k)$	the entry in the jth row and kth column of $\mathbf{X}$
$\hat{\mathbf{x}}$	reconstructed value
$\mathbf{W}$	transformation matrix
α	an appropriate level of significance
λ_i	the eigenvalue associated with the ith loading vector
Λ	diagonal matrix of the inverse of the eigenvalues
Ξ	matrix of fault directions.

ACKNOWLEDGEMENTS

This work was supported by the Korea Research Foundation Grant funded by the Korean Government (MOEHRD) (KRF-2007-331-D00089), the Institute for Encouragement of Innovation by means of Science and Technology in Flanders (IWT) and the European Union by means of the IcoN project (EVK1-CT2000-054). P. Vanrolleghem holds the Canada Research Chair on Water Quality Modelling.

REFERENCES

Cox CS, Adgar A, Fletcher I. 2000. Toolkits for advanced sensor validation, condition monitoring and modelling of water treatment systems. Technical report, Control Systems Centre, University of Sunderland.

Hellinga C, Schellen AAJC, Mulder JW, van Loosdrecht MCM, Heijnen JJ. 1998. The SHARON process: an innovative method for nitrogen removal from ammonium-rich waste water. *Water Science and Technology* **37**(9): 135–143.

Jeppsson U, Alex J, Pons MN, Spanjers H, Vanrolleghem PA. 2002. Status and future trends of ICA in wastewater treatment—an European perspective. *Water Science and Technology* **45**(4–5): 485–494.

Jetten MSM, Horn SJ, van Loosdrecht MCM. 1997. Towards a more sustainable wastewater treatment system. *Water Science and Technology* **35**(9): 171–180.

Olsson G. 2002. Lessons learnt at ICA2001. *Water Science and Technology* **45**(4–5): 1–8.

DOI: 10.1002/env

Olsson G, Newell B. 1999. *Wastewater Treatment System: Modelling, Diagnosis and Control*. IWA Publishing: London, UK.
Qin SJ. 2003. Statistical process monitoring: basics and beyond. *Journal of Chemometrics* **17**: 480–502.
Qin SJ, Li W. 1998. Detection, identification, and reconstruction of fault sensors with maximized sensitivity. *AIChE Journal* **45**: 1963–1976.
Rieger L, Alex J, Winkler S, Boehler M, Thomann M, Siegrist H. 2003. Progress in sensor technology—progress in process control? Part I: sensor property investigation and classification. *Water Science and Technology* **47**(2): 103–112.
Rieger L, Thomann M, Joss A, Gujer W, Siegrist H. 2004. Computer-aided monitoring and operation of continuous measuring devices. *Water Science and Technology* **50**(11): 31–39.
Rieger L, Thomann M, Gujer W, Siegrist H. 2005. Quantifying the uncertainty of on-line sensors at WWTPs during field operation. *Water Research* **39**(20): 5162–5174.
Rosen C, Lennox JA. 2001. Multivariate and multiscale monitoring of wastewater treatment operation. *Water Research* **35**: 3402–3410.
Rosen C, Rottorp J, Jeppsson U. 2003. Multivariate on-line monitoring: challenges and solutions for modern wastewater treatment operation. *Water Science and Technology* **47**(2): 171–179.
Seyfried CF, Hippen A, Helmer C, Kunst S, Rosenwinkel KH. 2001. One-stage deammonification: nitrogen elimination at low costs. *Water Science and Technology: Water Supply* **1**(1): 71–80.
Strotmann UJ, Keinath A, Hüttenhain H. 1995. Biological test systems for monitoring the operation of wastewater treatment plants. *Chemosphere* **30**(2): 327–338.
van Dongen U, Jetten MSM, van Loosdrecht MCM. 2001. The SHARON®-Anammox® process for treatment of ammonium rich wastewater. *Water Science and Technology* **44**(1): 153–160.
Van Hulle SWH, Van Den Broeck S, Maertens J, Villez K, Schelstraete G, Volcke E, Vanrolleghem PA. 2005. Construction, start-up and operation of a continuously aerated laboratory-scale SHARON reactor in view of coupling with an Anammox reactor. *Water SA* **31**: 327–334.
Vanrolleghem PA, Lee DS. 2003. On-line monitoring equipment for wastewater treatment processes: state of the art. *Water Science and Technology* **47**(2): 1–34.
Volcke EIP, Van Hulle SWH, Donckels BMR, van Loosdrecht MCM, Vanrolleghem PA. 2005. Coupling the SHARON process with Anammox: model-based scenario analysis with focus on operating costs. *Water Science and Technology* **52**(4): 107–115.
Volcke EIP, van Loosdrecht MCM, Vanrolleghem PA. 2006. Controlling the nitrite: ammonium ratio in a SHARON reactor in view of its coupling with an Anammox process. *Water Science and Technology* **53**(4–5): 45–54.
Wang S, Chen Y. 2004. Sensor validation and reconstruction for building central chilling systems based on PCA. *Energy Conversion and Management* **45**: 673–695.
Wise BM, Gallagher NB. 1996. The process chemometrics approach to process monitoring and fault detection. *Journal of Process Control* **6**(6): 329–348.
Yoo CK, Lee DS, Vanrolleghem PA. 2004a. Application of multiway ICA monitoring method in SBR process. *Water Research* **38**(7): 1715–1732.
Yoo CK, Lee J, Lee I. 2004b. Nonlinear model-based dissolved oxygen control in a biological wastewater treatment process. *The Korean Journal of Chemical Engineering* **21**(1): 14–19.
Yoo CK, Bang YH, Lee I, Vanrolleghem PA, Rosén C. 2004c. Application of fuzzy partial least squares (FPLS) for modeling nonlinear biological processes. *The Korean Journal of Chemical Engineering* **21**(6): 1087–1097.
Yoo CK, Villez K, Lee I, Van Hulle S, Vanrolleghem PA. 2006. Sensor validation and reconciliation for a partial nitrification process. *Water Science and Technology* **53**(4–5): 513–521.

DOI: 10.1002/env