

Research Article

Semiadaptive Fault Diagnosis via Variational Bayesian Mixture Factor Analysis with Application to Wastewater Treatment

Hongjun Xiao, Yiqi Liu, and Daoping Huang

School of Automation Science and Engineering, South China University of Technology, Guangzhou 510641, China

Correspondence should be addressed to Yiqi Liu; liuyiqi769@sina.com and Daoping Huang; audhuang@scut.edu.cn

Received 11 June 2015; Accepted 17 February 2016

Academic Editor: Seiichiro Katsura

Copyright © 2016 Hongjun Xiao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Mainly due to the hostile environment in wastewater plants (WWTPs), the reliability of sensors with respect to important qualities is often poor. In this work, we present the design of a semiadaptive fault diagnosis method based on the variational Bayesian mixture factor analysis (VBMFA) to support process monitoring. The proposed method is capable of capturing strong nonlinearity and the significant dynamic feature of WWTPs that seriously limit the application of conventional multivariate statistical methods for fault diagnosis implementation. The performance of proposed method is validated through a simulation study of a wastewater plant. Results have demonstrated that the proposed strategy can significantly improve the ability of fault diagnosis under fault-free scenario, accurately detect the abrupt change and drift fault, and even localize the root cause of corresponding fault properly.

1. Introduction

Due to stringent environmental regulations, wastewater treatment plants (WWTPs) are always challenged to meet new effluent quality requirements [1]. In such context, the use of online sensors for control and optimization in the wastewater treatment plants is indispensable. Measurements from instrumentation shall be available 24 hours a day and 7 days a week. However, even reliable instrumentation can fail during operation, which can have serious consequences if the instrumentation is used in the closed form. The huge amount of correlated data adds more complexity to this problem, thus making manual expert-based quality evaluation infeasible [2]. The development of methods that allow us to extract useful information for automatic detection of faults and identifying their root cause is urgently needed.

This vision has motivated inspiring efforts toward using multivariate statistical process monitoring (MSPM) methods, such as principal component analysis (PCA) [3, 4] and partial least-squares (PLS) [5, 6], for fault diagnosis in the wastewater treatment process, even though they had been used for many years in the chemical process industries [3, 7, 8]. The goal of such methods is to project a high dimensional measurement space onto a space with significantly fewer dimensions. In

PCA, Hotelling statistics, T^2 , and the sum of squared residuals, SPE, or Q statistic are formulated as tools to identify the deviations of normal process. The T^2 statistic is a measure of the variation in the PCA model, whereas the Q statistic is a measure of the amount of variation not captured by the PCA model [9]. However, these methods adhere to an assumption that process noises for each variable are normally distributed. The presence of variations such as shifting operational conditions, feed flow fluctuations, control strategy changes, and non-Gaussian disturbances in the WWTPs often deviates from such an assumption. Independent component analysis (ICA) provides an alternative to make data more comprehensible and useful for non-Gaussian fault diagnosis [10–12]. Other complementary MSPM approaches include Fisher discriminant analysis (FDA) [13], canonical variable analysis (CVA) [14], and support vector machines (SVM) [15]. Nevertheless, WWTPs pose an additional challenge to the development and application of these methods. The problem is highly uncertain environment. To alleviate such problem, a probabilistic PCA based method was proposed with the assumption that both of process variables and noises follow Gaussian distribution [16, 17]. The issue of PPCA is that it assumes the same noise level for all measurement variables with potentially underfitting the fault diagnosis model. Factor

analysis provides a potential way to address such a problem using different noise variance for each variable [18]. Unfortunately, parameter identification of FA using the least-squares method or maximum likelihood for optimization potentially results in an overfitting problem. Therefore, variational Bayesian learning was introduced to identify corresponding parameters for FA or PCA and further used to process monitoring [19]. These methods work reasonably well with synthetic data sets but they have difficulties in handling wastewater plant dynamics (e.g., seasonal or diurnal changes) and nonlinearity (e.g., temperature-dependent kinetics). One of common and plausible ways is to extend the conventional methods to a mixture form with the capability of approaching global nonlinearity by some local linearity models. However, there are several drawbacks of the mixture model based methods [20]: (1) they cannot determine the effective factor number for each factor analyzer automatically through the modeling process; (2) the importance of each retained factor in the model structure cannot be easily differentiated; (3) singularity and overfitting problems maybe caused when the number of modeling data samples is limited. Variational Bayesian mixture factor analysis offers a good alternative to overcome these problems. Additionally, the VBMFA method is not only able to determine the dimensionality automatically, but also capable of avoiding overfitting problem [21].

It is common sense that batch processes are in fact a special type of nonlinear processes. Therefore, in this paper, unlike other mixture models for batch multimode process monitoring in general [22–25], the VBMFA-based process monitoring method is proposed to address the significant nonlinear problems in the WWTPs. This is more practical because systems are always nonclassifiable clearly for most cases. By doing so, the corresponding T^2 and SPE statistics are formulated as tools for fault diagnosis and identification. T^2 and SPE statistics from local factor analyzers are assimilated by weights, respectively. In order to account for changing process behavior, such weights can change over time, thereby offering more reliable way for fault diagnosis and identification. Therefore, the implementation of adaptive mechanism for the proposed method is not by their mean and variance but rather depending on their weight update. In other words, the resulting online VBMFA model is obtained by weighting adaptively several factor analyzers that are already trained offline. Therefore, the resulting method is termed as a semi-adaptive fault diagnosis method. In summary, the contributions of this paper is given as follows: (1) typical variables are generated to account for the fluctuations of diurnal wastewater changes in workdays and weekends, serving as data pretreatment and an initial step to break up dynamics of the wastewater process; (2) the VBMFA is introduced for fault diagnosis; (3) an adaptive probabilistic weighted strategy is adapted to mix different local factor analyzers; (4) an adaptive control limit is proposed to monitor a WWTP with strong nonlinearity and dynamic feature.

The rest of this paper is organized as follows. In Section 2, an overview of issues in the WWTPs is given, together with a description of the simulation systems involved in this work. Section 3 will deal with the description of the FA and VBMFA. Section 4 provides a detailed demonstration for the

development of the VBMFA-based fault diagnosis strategy, which is followed by the case study of the corresponding process monitoring scheme in the next section. Finally, some conclusions are made.

2. A Simulated Platform for Process Monitoring

2.1. WWTP. A wastewater treatment plant is designed to degrade wastewater and return clean water to the river. One of the most popular processes available for dealing with wastewater is the activated sludge. However, the complexity of the physical, chemical, and biological phenomena associated with treatment units always frustrates the reliability of WWTPs dramatically. Moreover, the reactors exhibit strong nonlinear dynamics and have coupling effects among the variables. More care should also be taken for fault diagnosis model construction; more details are shown as follows:

- (1) *Modeling Challenges for Process Monitor.* The overall WWTP model consists of the hydraulic model and the biochemical model. The hydraulic model is related to reactor behavior, flow rates, and recirculation, whereas the biochemical model is mainly associated with the activated sludge and describes microbial growth, death, and nutrient consumption. Such huge number of biological processes discourage the process model construction.
- (2) *Uncertainties.* The main goal of a WWTP is to maximize the reduction of wastewater while meeting the environmental requirement from government. However, uncertainties such as influent flow rate, water quality, and contamination loads add more complexity for process monitor model building.
 - (a) *Uncertainty about the Initial Conditions.* The large variety and complexity of all present biological species in WWTPs make it identify the role of each one in a process. Moreover, even if it is possible to determine a total initial biomass concentration, this does not generally correspond to the concentration of the effectively active species.
 - (b) *Uncertainty in Process Kinetics.* Again, the large variety complexity of biological species presented in WWTPs makes it impossible to determine all kinetic parameters for all biological reactions. Moreover, bacteria can be sensitive to environmental changes and even if the kinetics could be globally described, their parameters vary with time as a function of the time-varying distribution of the biological population.
 - (c) *Uncertainty in the Input Contractions.* In contrast with food and pharmaceutical biological processes [26, 27], where feed concentrations are usually well known, the inputs of a WWTP have to be considered as disturbances. Input concentrations depend on several random factors. For instance, in municipal WWTPs, the

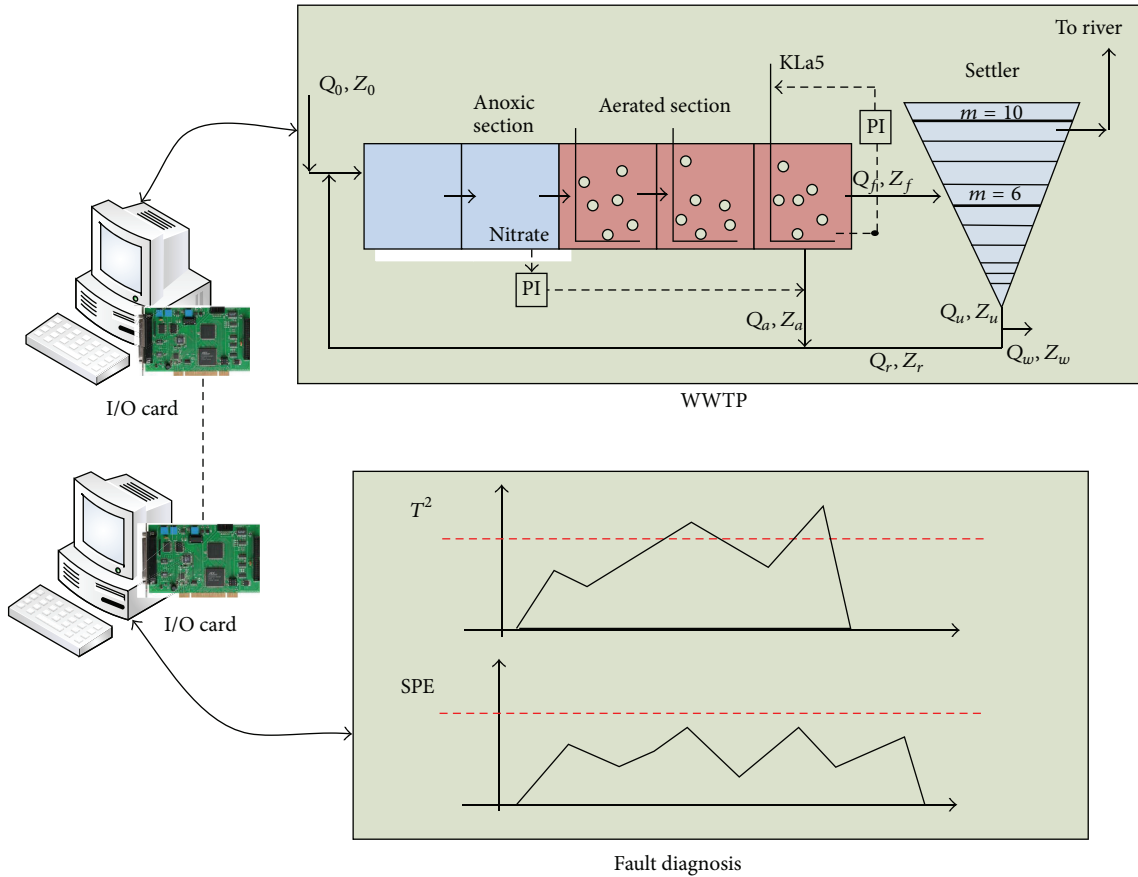


FIGURE 1: The schematic of wastewater plant for BSM1.

organic load varies during the day according to the level of human activity. Also, they themselves are strongly affected by weather conditions and seasonal change. Such fluctuations often result in the degradation of the performance or even plant failure.

2.2. BSM1 (Benchmark Simulation Model 1) Platform. The simulation platform used is the BSM1 developed by the second IWA Respirometry Task Group, offering an unbiased benchmarking system for comparing various control strategies (<http://www.benchmarkwwtp.org/>) [28]. In this study, a relatively simple plant layout, which is most commonly used for removal of organic matter, nitrification, and denitrification, is selected in the simulation benchmark (see Figure 1). The plant consists of five compartment biological tanks (5999 m^3) and a secondary settler (6000 m^3), designing to treat an average flow of $20,000 \text{ m}^3/\text{day}$ with an average biodegradable chemical oxygen demand (COD) concentration of 300 mg/L . The last three compartments of the bioreactor are aerated, whereas the others are not. The nonreactive secondary settler is modeled with a series of 10 layers with an area of 1500 m^2 and a depth of 4 m . There are two internal recycles: the nitrate internal recycle from the last tank to the first tank and the activated sludge recycle from the underflow of the secondary settler to the front end of the plant. The feed

from the final biological reactor is connected to the sixth layer from the bottom of the secondary settler. The scenario with dry weather condition is tested in this case study. The simulated BSM1 comprises 14 days of influent data recorded at 15-minute intervals from a real plant. In order to make sure that the methodology is able to translate to a real WWTP easily in the future, the BSM1 model and proposed method were run in separate computer and connected by two I/O cards.

3. Factor Analysis and Mixture Factor Analysis Model Based on Variational Bayesian Learning

3.1. Factor Analysis. Factor analysis is a statistical method with the capability to describe variability among observed correlated variables in terms of a potentially lower number of unobserved variables called factors. The original observed measurement variables $x \in R^m$ are treated as linear combination of factors t plus the noise $e \in R^m$ and mean vector $\mu \in R^m$; that is,

$$x = \Lambda t + \mu + e, \quad (1)$$

where $\Lambda \in R^{(m \times k)}$ is the linear transformation known as the factor loading matrix and e is the zero-mean Gaussian noise

with diagonal covariance matrix Ψ . Given the observed variables $x \in R^m$, the parameters $\theta = \{\mu, \Lambda, \Psi\}$ in the model are needed to be identified. One of common ways is to use the expectation and maximization (EM) algorithm, also termed as the iterative likelihood maximization algorithm.

3.2. Mixture Factor Analysis Model Based on Variational Bayesian Learning (VBMFA). It is a common sense that high dimensional data in fact lie on a low dimensional manifold generally [29]. In factor analysis, each factor dictates the amount of each linear transformation on the observed variables. However, such transformation can only well explain a small region of the manifold in which it is locally linear, even though the manifold is globally nonlinear.

One way to deal with this problem is to use mixture models to tile the data manifold. A mixture of factor analyzers models the density for a data point x_i as a weighted average of factor analyzer densities as follows:

$$p(x_i | \pi, \Lambda, \mu, \Psi) = \sum_{s_i} p(s_i | \pi) p(x_i | s_i, \Lambda, \mu, \Psi), \quad (2)$$

where S is the number of mixture components in the model, π is the vector of mixing proportions, s_i is the discrete indicator variable for the mixture component selected for modeling the data point i , and $\Lambda = \{\Lambda^s\}_{s=1}^S$ is a set of factor loadings with the factor matrix Λ^s for analyzer s .

Due to the complexity of integrating out all those parameters, it is necessary to choose priors that the likelihood terms greatly simplify inference and interpretability. Therefore, the Dirichlet prior for mixing proportion π , with strength α^* , is chosen as follows:

$$p(\pi | \alpha^* m^*) = \text{Dir}(\pi | \alpha^* m^*) \quad (3)$$

with $m^* = [1/S, \dots, 1/S]$. Thus, only a single hyperparameter is needed regardless of the dimensionality of π . The column of each factor loading matrix has a Gaussian prior with zero mean and a different precision parameter (drawn from a gamma distribution with fixed hyperparameters):

$$p(\Lambda | \nu) = \prod_{s=1}^S p(\Lambda^s | \nu^s) = \prod_{s=1}^S \prod_{j=1}^{k_s} p(\Lambda^s_{\cdot j} | \nu_j^s) \quad (4)$$

$$= \prod_{s=1}^S \prod_{j=1}^{k_s} p\left(\Lambda^s_{\cdot j} | 0, \frac{I}{\nu_j^s}\right),$$

where $\Lambda^s_{\cdot j}$ denotes the vector of entries in the j th column of the s th analyzer in the mixture and ν_j^s is the same scalar precision for each entry in the corresponding column.

Since the number of hyperparameters in $\nu = \{\{\nu_j^s\}_{j=1}^{k_s}\}$ increases with the number of analyzers, a hyperprior on every element of each ν^s precision vector is as follows:

$$p(\nu | a^*, b^*) = \prod_{s=1}^S \prod_{j=1}^{k_s} \text{Ga}(\nu_j^s | a^*, b^*), \quad (5)$$

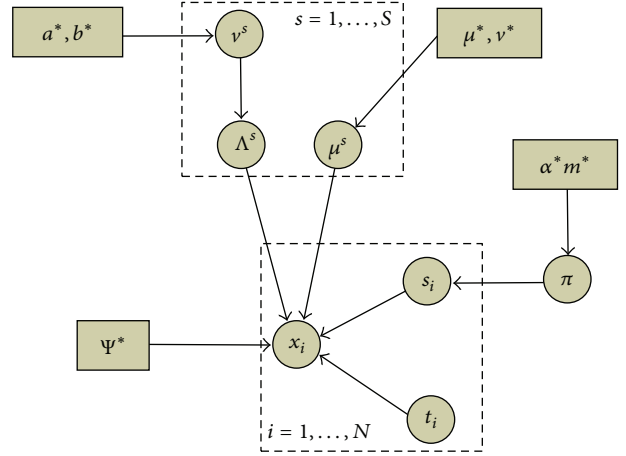


FIGURE 2: The formulation for VBMFA.

where a^* and b^* are shape and inverse-scale hyperparameters for a gamma distribution. Finally, the means of each analyzer in the mixture need to be integrated out. A Gaussian prior with mean μ^* and axis-aligned precision $\text{diag}(\nu^*)$ is placed on each mean μ^* :

$$p(\mu | \mu^*, \nu^*) = \prod_{s=1}^S N(\mu^s | \mu^*, \text{diag}(\nu^s)^{-1}). \quad (6)$$

The directed acyclic graph for the generative model for VBMFA is shown graphically in Figure 2.

The aim of VBMFA is to optimize

$$L = \ln x = \ln \left(\int d\pi p(\pi | \alpha^* m^*) \int d\nu p(\nu | a^*, b^*) \cdot \int \Lambda p(\Lambda | \nu) \int \mu p(\mu | \mu^*, \nu^*) \prod_{i=1}^N \left[\prod_{s_i=1}^S p(s_i | \pi) \cdot \int dt_i p(t_i) p(x_i | s_i, \Lambda, \mu, \Psi) \right] \right). \quad (7)$$

An arbitrary distribution $q(\pi, \nu, \Lambda, \mu)$ is introduced to lower bound. By using the Kullback-Leibler divergence [30], L can be reformulated as follows:

$$L \geq \int d\pi \ln \frac{p(\pi | \alpha^* m^*)}{q(\pi)} + \sum_{s=1}^S \int d\nu^s q(\nu^s) \cdot \left[\ln \frac{p(\nu^s | a^*, b^*)}{q(\nu^s)} + \int d\tilde{\Lambda}^s q(\tilde{\Lambda}^s) \ln \frac{p(\tilde{\Lambda}^s | a^*, b^*)}{q(\tilde{\Lambda}^s)} \right] + \sum_{i=1}^N \sum_{s_i} q(s_i) \left[\int d\pi \ln \frac{p(s_i | \pi)}{q(s_i)} + \int dt_i q(t_i | s_i) \cdot \ln \frac{p(t_i)}{q(t_i | s_i)} \right] + \int \tilde{\Lambda}^s q(\tilde{\Lambda}^s) \int dt_i q(t_i | s_i) \cdot \ln p(x_i | s_i, t_i, \tilde{\Lambda}, \Psi) = F(q(\theta), q(s, t), \Theta), \quad (8)$$

where $\Theta = (\alpha^* m^*, a^*, b^*, \mu^*, v^*, \Psi)$. Thus, the lower bound is a sum of a foundational of variational posterior distribution over the parameters, denoted by $q(\theta)$, and a foundational of variational posterior distribution over the hidden variables of each data point.

In order to optimize the lower bound, functional derivatives with respect to each $q(\cdot)$ distribution are taken and are equal to zero to find the distributions. By doing so, the corresponding $q(\cdot)$ can be obtained as follows:

$$q(\pi) = \text{Dir}(\pi | \alpha m), \quad (9)$$

where each element of the variational parameter αm is given by

$$\alpha m_s = \alpha^* m_s^* + \sum_{i=1}^N q(s_i) \quad (10)$$

with $\alpha = \alpha^* + n$, $m_s^* = 1/S$, and $\sum_{s=1}^S m_s = 1$:

$$q(v_i^s) = \text{Ga}\left(v_i^s | a^* + \frac{p}{2}, b^* + \frac{1}{2} \sum_{j=1}^p \langle \Lambda_{jl}^s \rangle_{q(\Lambda^s)}\right), \quad (11)$$

$$q(\tilde{\Lambda}^s) = \prod_{j=1}^p q(\tilde{\Lambda}_j^s) = \prod_{j=1}^p N\left(\tilde{\Lambda}_j^s | \tilde{\Lambda}_j^s, \tilde{\Gamma}_j^s\right),$$

where $\tilde{\Lambda}_j^s$ denotes the column vector associated with the j th row of $\tilde{\Lambda}^s$, which has $k_s + 1$ dimensions. These variational poster parameters are given by

$$\tilde{\Gamma}_j^s = \begin{bmatrix} \Sigma_{\Lambda\Lambda}^{j,s-1} & \Sigma_{\Lambda\mu}^{j,s-1} \\ \Sigma_{\mu\Lambda}^{j,s-1} & \Sigma_{\mu\mu}^{j,s-1} \end{bmatrix}^{-1}, \quad (12)$$

$$\tilde{\Lambda}_j^s = \begin{bmatrix} \tilde{\Lambda}_j^s \\ \tilde{\mu}_j^s \end{bmatrix}$$

with

$$\Sigma_{\Lambda\Lambda}^{j,s-1} = \text{diag}\langle v^s \rangle_{q(v^s)} + \Psi_{jj}^{-1} \sum_{i=1}^N q(s_i) \langle t_i t_i^T \rangle_{q(t_i|s_i)},$$

$$\Sigma_{\mu\mu}^{j,s-1} = v_j^* + \Psi_{jj}^{-1} \sum_{i=1}^N q(s_i),$$

$$\Sigma_{\mu\Lambda}^{j,s-1} = \Psi_{jj}^{-1} \sum_{i=1}^N q(s_i) \langle t_i \rangle_{q(t_i|s_i)} = \Sigma_{\mu\Lambda}^{j,s-1}, \quad (13)$$

$$\tilde{\Lambda}_j^s = [\tilde{\Gamma}_j^s]_{\Lambda\Lambda} \left(\Psi_{jj}^{-1} \sum_{i=1}^N q(s_i) x_{ij} \langle t_i \rangle_{q(t_i|s_i)} \right),$$

$$\tilde{\mu}_j^s = [\tilde{\Gamma}_j^s]_{\mu\mu} \left(\Psi_{jj}^{-1} \sum_{i=1}^N q(s_i) x_{ij} + v_j^* \mu_j^* \right),$$

$$q(t_i | s) = N(t_i | \tilde{t}_i^s, \Sigma^s)$$

with $|\Sigma^s|^{-1} = I + \langle \Lambda^s \Psi^{-1} \Lambda^s \rangle_{q(\tilde{\Lambda}^s)}$, $\tilde{t}_i^s = \Sigma^s \langle \Lambda^s \Psi^{-1} (x_i - \mu^s) \rangle_{q(\tilde{\Lambda}^s)}$.

By using a result of Dirichlet distributions in [21], $q(s_i)$ is computed by

$$q(s_i) = \frac{1}{Z_i} \exp \left[\Psi(\alpha m_{s_i}) - \Psi(\alpha) + \frac{1}{2} \ln |\Sigma^s| - \frac{1}{2} \cdot \text{tr} \left[\Psi^{-1} \left\langle \left(x_i - \bar{\Lambda}^{s_i} \begin{bmatrix} t_i \\ 1 \end{bmatrix} \right) \left(x_i - \bar{\Lambda}^{s_i} \begin{bmatrix} t_i \\ 1 \end{bmatrix} \right)^T \right\rangle_{q(\tilde{\Lambda}^{s_i})q(t_i|s_i)} \right] \right], \quad (14)$$

where Z_i is a normalization constant for each data point, such that $\sum_{s_i=1}^S q(s_i) = 1$:

$$\Psi^{-1} = \text{diag} \left[\frac{1}{N} \cdot \sum_{i=1}^N \left\langle \left(x_i - \bar{\Lambda}^{s_i} \begin{bmatrix} t_i \\ 1 \end{bmatrix} \right) \left(x_i - \bar{\Lambda}^{s_i} \begin{bmatrix} t_i \\ 1 \end{bmatrix} \right)^T \right\rangle_{q(\tilde{\Lambda}^{s_i})q(s_i)q(t_i|s_i)} \right],$$

$$\psi(a^*) = \ln b^* + \frac{1}{Sk} \sum_{s=1}^S \sum_{j=1}^k \langle \ln v_i^s \rangle_{q(v_i^s)}, \quad (15)$$

$$b^{*-1} = \frac{1}{\alpha^* Sk} \sum_{s=1}^S \sum_{j=1}^k \langle v_i^s \rangle_{q(v_i^s)},$$

$$\psi(\alpha^*) - \psi\left(\frac{\alpha^*}{S}\right) = \frac{1}{S} \sum_{s=1}^S [\psi(\alpha) - \psi(\alpha m_s)].$$

The solution for μ^* and v^* can be found by maximizing (8):

$$\mu^* = \frac{1}{S} \sum_{s=1}^S \langle \mu^s \rangle_{q(\mu^s)}, \quad (16)$$

$$v^* = [v_1^*, \dots, v_p^*]$$

with

$$v_j^* = \frac{1}{S} \sum_{s=1}^S \langle (\mu_j^s - \mu_j^*) (\mu_j^s - \mu_j^*) \rangle_{q(\mu^s)}, \quad (17)$$

where the update for v^* uses the already updated μ^* .

4. Units Process Monitoring Based on VBMFA

4.1. Fault Detection. In order to perform process monitoring, monitoring statistic T^2 and SPE is constructed first of all. Similar to PCA method, corresponding monitoring statistics for each local FA can be represented as follows:

$$T_s^2 = \tilde{t}^{sT} (\text{var}(\tilde{t}^s))^{-1} \tilde{t}^s \quad (18)$$

with

$$\begin{aligned}\hat{t}^s &= E(t^s | x, s, \theta) \\ &= \Sigma^s \left\langle \Lambda^{sT} \Psi^{-1} (x - \mu^s) \right\rangle_{q(\bar{\Lambda}^s)}, \\ \text{var}(\hat{t}^s) &= \Sigma^s = \left(I + \left\langle \Lambda^{sT} \Psi^{-1} \Lambda^s \right\rangle_{q(\bar{\Lambda}^s)} \right)^{-1}.\end{aligned}\quad (19)$$

Similarly, the associated SPE statistic can be given as follows:

$$\begin{aligned}\hat{e}^s &= x - \mu^s - \Lambda^s \hat{t}^s, \\ \text{SPE}_s &= \hat{e}^{sT} \Psi^{-1} \hat{e}^s.\end{aligned}\quad (20)$$

The control limits of the T^2 and SPE statistic can be both determined by the χ^2 distribution; thus,

$$\begin{aligned}T_s^2 &\leq T_{\text{lim}}^2 = \chi_\alpha^2(k_s), \\ \text{SPE}_s &\leq \text{SPE}_{\text{lim}} = \chi_\alpha^2(p),\end{aligned}\quad (21)$$

where $s = 1, 2, \dots, S$. k is the number of retained factors in each local FA model, p is the number of process variables, and

$$q(s_{\text{new}}) = \frac{1}{Z_{\text{new}}} \exp \left[\Psi(\alpha m_s) - \Psi(\alpha) + \frac{1}{2} \ln |\Sigma^s| - \frac{1}{2} \text{tr} \left[\Psi^{-1} \left\langle \left(x_{\text{new}} - \bar{\Lambda}^s \begin{bmatrix} t^s \\ 1 \end{bmatrix} \right) \left(x_{\text{new}} - \bar{\Lambda}^s \begin{bmatrix} t^s \\ 1 \end{bmatrix} \right)^T \right\rangle_{q(\bar{\Lambda}^s)q(t^s)} \right] \right] \quad (23)$$

which represents the responsibility of the components for the new coming data point x_{new} . Both of synthetic monitor statistics T^2 and SPE can be obtained as follows:

$$\begin{aligned}ST_{\text{new}}^2 &= \sum_{s=1}^S q(s_{\text{new}}) T_{s,\text{new}}^2, \\ \text{SSPE}_{\text{new}} &= \sum_{s=1}^S q(s_{\text{new}}) \text{SPE}_{s,\text{new}}.\end{aligned}\quad (24)$$

Since the control limit of the local T^2 is correlated with the number of factors k_s in each FA model, there is a need to combine the control limit as well. Thus, $q(s_{\text{new}})$ is used to weight associated T^2 control limit as follows:

$$T^2 \leq T_{\text{lim}}^2 = \sum_{s=1}^S q(s_{\text{new}}) \chi_\alpha^2(k_s). \quad (25)$$

Note that, due to constant control limit of each local SPE $\chi_\alpha^2(p)$, it is not necessary to calculate the combined SPE control limit any more.

Even though the structure for each local FA model is unchanged after model training, our process monitoring model can be adjusted by updating mixture proportions in an adaptive way. Hence, process behaviors can be judged by simply checking both of statistics accordingly if they exceed predefined control limits.

α is the significance level of both monitoring statistics. As for a new coming sample data x_{new} , both of statistics are updated as follows:

$$\begin{aligned}\hat{t}_{\text{new}}^s &= E(t_{\text{new}} | x_{\text{new}}, s, \theta) \\ &= \Sigma^s \left\langle \Lambda^s \Psi^{-1} (x_{\text{new}} - \mu^s) \right\rangle_{q(\bar{\Lambda}^s)}, \\ T_{s,\text{new}}^2 &= \hat{t}_{\text{new}}^{sT} \left(\text{var}(\hat{t}^s) \right)^{-1} \hat{t}_{\text{new}}^s, \\ \hat{e}_{\text{new}}^s &= x_{\text{new}} - \mu^s - \Lambda^s \hat{t}_{\text{new}}^s, \\ \text{SPE}_{s,\text{new}} &= \hat{e}_{\text{new}}^{sT} \Psi^{-1} \hat{e}_{\text{new}}^s.\end{aligned}\quad (22)$$

However, it is cumbersome to monitor the process with all of the local monitoring charts, which also potentially frustrates the decision making for fault diagnosis with such many monitoring charts. Therefore, coordination of all the local monitoring charts is imperative. Different from the hard assignment method employed by [31], an adaptive method is implemented in this section. As the monitoring results of the new sample x_{new} in each local subspace have been identified, we can combine them through the estimated posterior probabilities, which are given as follows:

4.2. Fault Identification. Once a fault is detected, more information is necessarily obtained to find the root cause of the corresponding faults. A contribution plot is a common way to fill this gap, which is able to show the contribution of each process variable to the statistic calculated. A high contribution of a process variable usually indicates a problem with this specific variable. Plots of the contribution to the associated statistic are similar to standard plots of squared residuals obtained, but the contribution plots presented here also have control limits. These control limits are used for comparing the residuals of the new values for the present to the residuals of the fault-free data. If in the fault-free data a certain process variable had high residuals, it can also be expected to have high residuals in the present. However, if there are high residuals for a certain process variable that had low residuals in the fault-free data; this probably is due to a special event in the new process. If the contributions of a large group of process variables are studied, it is usually found that several process variables have high contributions. Using the control limits, it is easier to find those process variables that are really different, compared to the fault-free data. In this paper, a control limit for the contribution for each process variable is calculated as the mean of the contributions plus three times the standard deviation of the contributions for each process variable at each time period. That means contribution of each variable follows Gaussian distribution in a long period, but the matching mean and variance will

TABLE 1: Selected sensors for process monitoring.

Variables	Comments	Symbol
1	Total nitrogen concentration in effluent (mg N/L)	TN
2	Total chemical oxygen demand in the effluent (mg COD/L)	TCOD
3	NH ₄ ⁺ + NH ₃ nitrogen in effluent (mg N/L)	S _{NH}
4	Suspended solids concentration in effluent (mg SS/L)	SS
5	Biological oxygen demand concentration in effluent [g COD m ⁻³]	BOD
6	Suspended solids concentration in influent (mg SS/L)	SS
7	NH ₄ ⁺ + NH ₃ nitrogen in influent (mg N/L)	S _{NH}
8	Oxygen in the third reactor g (-COD)·m ⁻³ ;	S _O
9	Nitrate and nitrite nitrogen in the third reactor (mg N/L)	S _{NO}
10	NH ₄ ⁺ + NH ₃ nitrogen in the third reactor (mg N/L)	S _{NH}
11	Oxygen in the fifth reactor g (COD)/m ³ ;	S _O
12	Nitrate and nitrite nitrogen in the fifth reactor (mg N/L)	S _{NO}
13	NH ₄ ⁺ + NH ₃ nitrogen in the fifth reactor (mg N/L)	S _{NH}
14	Flowrate in the third reactor m ³ /d	Flowrate
15	Oxygen transfer coefficient in the third reactor	KLa
16	Flowrate in the fifth reactor m ³ /d	Flowrate

be different. Control limits for contributions are studied by many researchers [32, 33]. Here, the control limits are determined according to the 3σ method [32]. The reason why we used the 3σ method is because of its efficiency and simplicity.

5. Experimental Study

5.1. Data Pretreatment. In this case, 16 important variables shown in Table 1 were selected for monitoring, covering almost all the wastewater plant qualities required by engineers and the government.

As depicted in Figures 3(a) and 3(b), the historical data sets reveal that, during dry weather periods, daily SS and S_{NH} concentration follow a very similar pattern, which is also observed in other variables. In this respect, the typical data profiles for each variable under dry weather conditions were generated. This was done by stacking all the historical data in one set (from 0 to 1 day), and from it calculating the average flow at each time step. It is worth noting that daily data from Monday to Friday are very similar, whereas a slightly different pattern (shift in amplitude) is observed on the weekend profiles. Therefore, two typical data profiles were generated for each variable: one for weekdays and one for weekends. These patterns are consistent with the water consumptions by inhabitants during daily life. Results about SS and S_{NH} concentrations are presented in Figures 3(c) and 3(d). By subtracting typical data from historical data, dynamic behaviors during the wastewater plant dataset are capable of being broken up, thereby providing better data for fault detection, identification, and even prediction. Moreover, to make sure all data from different variables are consistent mutually, they were centered and scaled accordingly. We denote all these procedures as “data whitening.” Finally, the whitening data were further used for ensuing VBMFA model building.

5.2. Fault Diagnosis

5.2.1. Fault Detection for the Fault-Free Scenario. 1344 samples in two weeks under normal operation condition were collected to assess the proposed methodology. 900 data points were used for VBMFA model training; the remaining was for testing. As depicted in Figures 4(a) and 4(c), the FA model performed well with most of points under the 95% control limit in terms of T^2 and SPE statistics. However, it is worthy to emphasize that unacceptable numbers of false alarm occurred during this process. This is due to the fact that, when the measurements reach the peak or the valley, FA model is incapable of capturing such strong nonlinearity properly, thus leading to too many false alarms frequently. On the contrary, better performance was achieved by VBMFA with almost all of test points under 95% control limit. It is obvious that the control limit changed along with new data points coming, resulting in a more compact control limit and less probability of false alarms.

Table 2 depicts the performance of four different monitor strategies, including PCA, FA, VBPCA, and VBMFA, under fault-free scenario in dry weather conditions. The frequency of false alarm for PCA and FA was very similar in terms of T^2 , but the SPE for both of strategies were quite different. The FA relying on different variance for different variables achieved less false alarms in comparison to the PCA. Furthermore, variational Bayesian learning is incorporated into PCA to learn its parameters. The number of false alarms was reduced as expected. Since variational Bayesian learning is the sensitive posterior probability mass rather than the posterior probability density, thus making it more resistant against overfitting compared to other estimation methods (least-squares, point estimation). Furthermore, because uncertainty information for unknown quantities can be obtained by this variational Bayesian learning, it is useful to detect unreliable results. However, due to strong nonlinearity and dynamics,

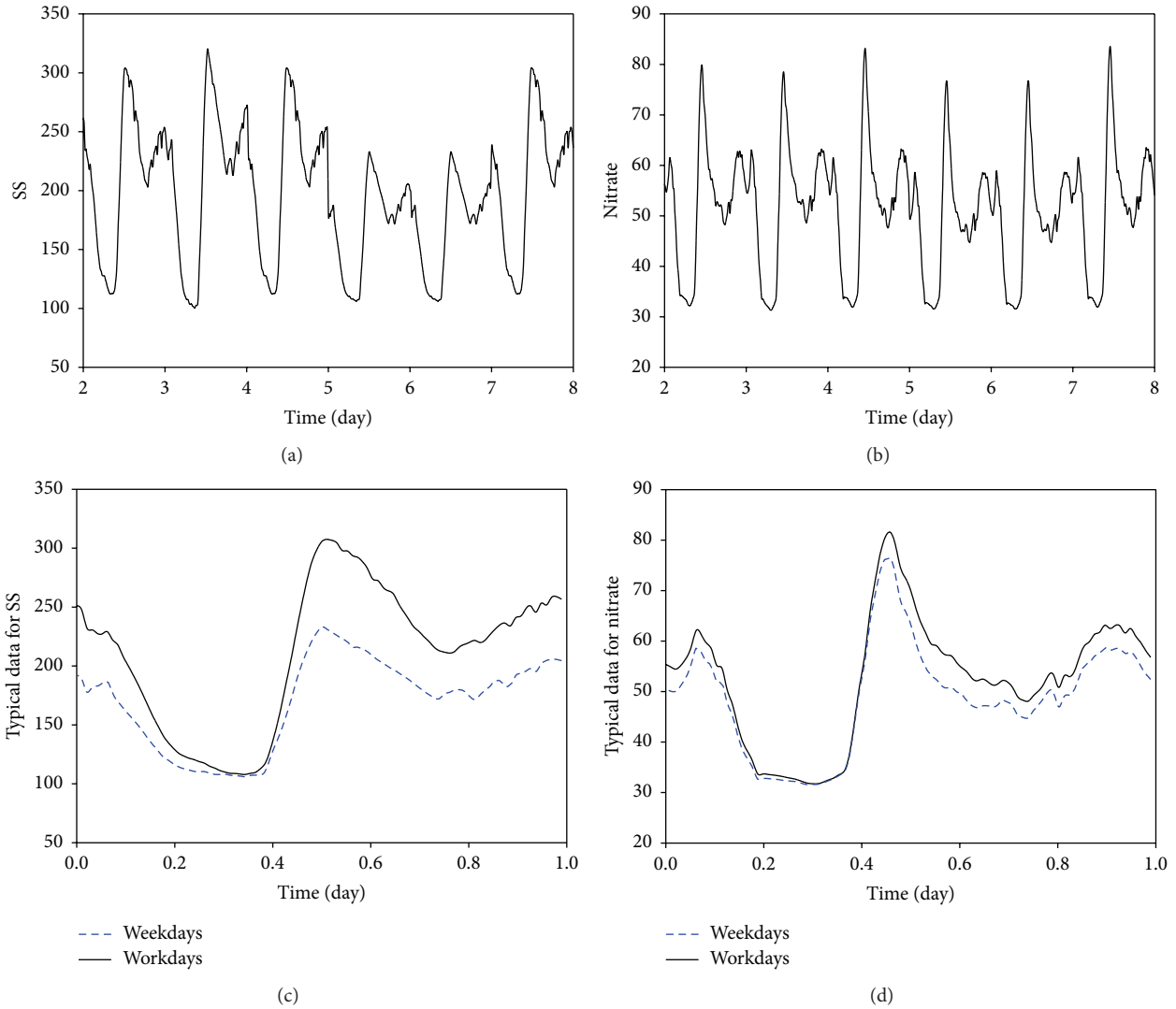


FIGURE 3: Typical variables for workdays and weekdays.

TABLE 2: Selected sensors for process monitoring.

Statistics\methods	PCA	FA	VBPCA	VBMFA
T^2	26/444	24/444	16/444	0/444
SPE	21/444	15/444	13/444	1/444
Parameters	PC: 4	Factors: 4	PC: 3	Maximum factors: 14 Local factors: 2, 3, 1

VBPCA failed to capture all the characteristics properly. Obviously, VBMFA achieved the best performance. This is an important improvement as the peaks or the valleys that frequently occurred in the concentrations of different variables were described accurately, thus leading to better monitor effectiveness. It is necessary to emphasize that all model parameters excluding VBMFA were obtained somehow with the help of manual operation. This would potentially frustrate the use of these methods in practice. Conversely, the number of factors for each factor analyzer in VBMFA was determined automatically.

5.2.2. Fault Detection for an Abrupt Change and a Drift Fault. Abrupt change and drift fault are two types of common faults in the process control. Thus, to access the fault detection ability of VBMFA, two cases with different faults were generated during the testing data on the second sensor (TCOD) and the fourth sensor (SS), respectively, tabulated as Table 3. Due to cost saving demand and stricter regulations from governments, TCOD and SS sensors are fairly important for WWTP monitoring.

Therefore, it is monitored not only by the engineers from WWTP, but also by the government using a remote

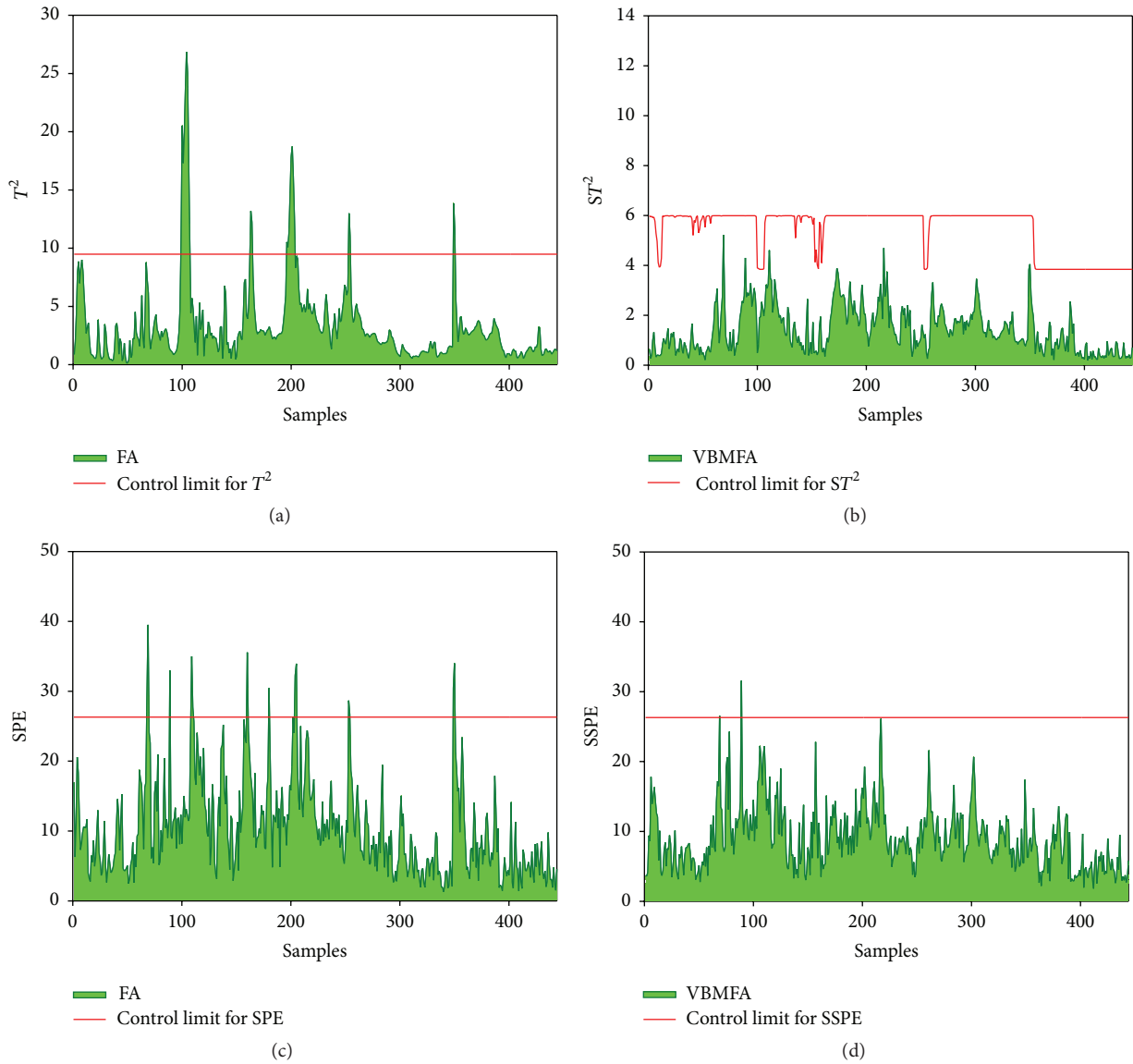


FIGURE 4: Fault diagnosis under normal condition using FA and VBMFA.

TABLE 3: Faults generated on two important sensors.

Fault scenarios	Comments on the fault type
Case 1 (the 2nd sensor)	30% positive bias on the second sensor (an abrupt change fault)
Case 2 (the 4th sensor)	35% positive bias of standard deviation (drift fault)

monitoring system. In the first type of fault, normal data was corrupted by a 30% abrupt change from day from the sample 200 to 300. Results from Figures 5(a) and 5(c) suggest that ST^2 and SSPE index can recognize the fault accurately. Even if a valley occurred in Figure 5(a) between data points 200 and 300, which potentially leads to a real fault being failed to be identified, VBMFA was able to adjust its control limit to adapt to such variation. To further illustrate the efficiency of VBMFA, the second scenario with a drift fault exerting on the fourth sensor was also simulated, showing in Figures 5(b)

and 5(d) that it achieved a good performance with accurate detection when the drift fault was happening. It is important to notice that the control limit of ST^2 decreased to lower values once it recognized significant changes in the updating data, thereby providing more potentials to identify the drift fault earlier.

5.3. *Fault Identification.* In order to find the root cause of the corresponding fault, the contribution plot was employed in this paper. The contribution plot of SSPE for the abrupt

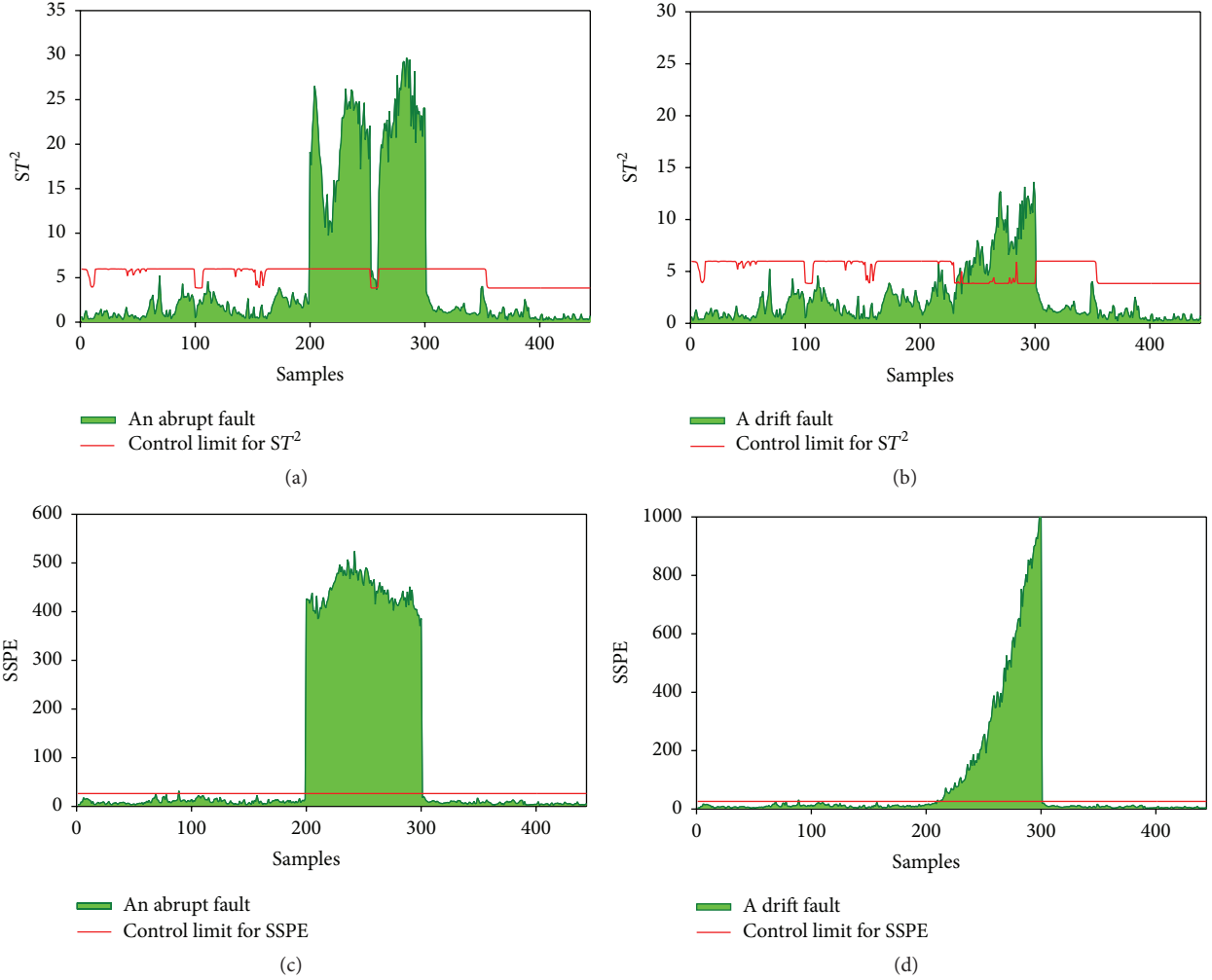


FIGURE 5: Fault detection for an abrupt change and a drift fault using VBMFA.

change can discriminate between the faulty sensor and the normal sensors (Figure 6(a)). It is also obvious in Figure 6(a) that the faulty sensor (TCOD) deviated from the major pattern and exceeded its control limit due to the negative influence of abrupt changes. Similarly, the contribution plot for the drift fault was shown in Figure 6(b). The result clearly shows that the contribution plot of the VBMFA strategy successfully separated the out of control data from normal operation ones.

5.4. Parameters Analysis. When optimizing (8), some $q(s_i)$ could be zero. This can be attributed to the fact that local data is insufficient to support the dimensional complexity on the factor loading matrices. Such redundant components should be removed to increase F , because they are no longer needed. This is the component death.

Component birth does not happen spontaneously; thus, a heuristic is introduced. A parent-component stochastically has a birth with probability proportional to $e^{-\beta F}$, and we attempt to split it into two. The parameter distributions of the

two Gaussians created from the split are initialized by partitioning the responsibilities for the data, $q(s_i)$. This usually causes F to decrease, so by monitoring the progress of F , this attempted birth can be rejected if F does not recover.

Furthermore, in the VBMFA model, the dimensionality of each factor analyzer is determined automatically by automatic relevance determination (ARD) with the potentials to improve VBMFA model performance dramatically [21, 34]. As such, the corresponding weight of each factor analyzer can be obtained in an automatic manner accordingly without necessarily resorting to trial and error method. The ST^2 and SSPE index can be weighted and updated adaptively.

As for the maximum factors for each factor analyzer, it represents the freedom of parameters the ARD method is able to learn. Thus, this parameter is just set to $p - 1$ or $p - 2$.

Each of the mixture components takes for a particular data point x_{new} and on the parameters θ of the mixture components, not on the values of other data points. Also, due to the parameters θ having been identified in the training stage

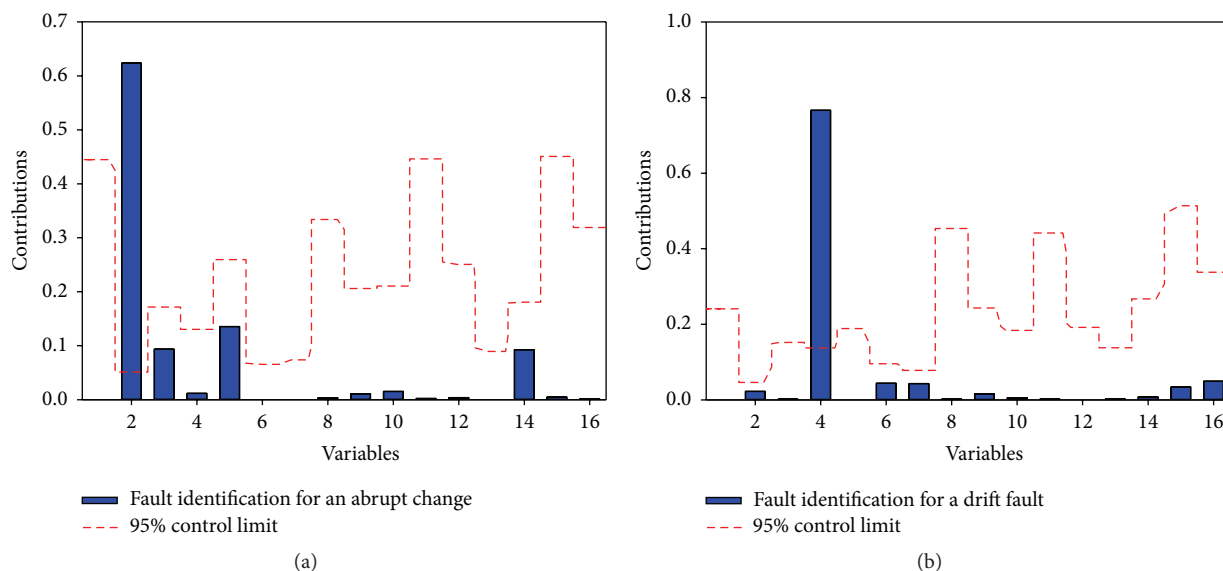


FIGURE 6: Fault identification for an abrupt change and a drift fault using VBMFA.

already, the update of $q(s_i)$ is only related to the data point x_{new} in fact.

5.5. Discussions. The present work describes the development of a VBMFA model to monitor a WWTP. This is an attempt to provide water utilities with a tool to improve the process monitor ability and formulate a guideline for system operation.

The VBMFA model has been successfully applied to a simulated WWTP with strong nonlinearity and dynamics in the dry weather. Results showed that false alarms can be avoided significantly under fault-free scenario. Also, accurate fault detection and identification can be achieved if an abrupt change or a drift fault occurred. Due to the mixture of factor analyzers for modeling, VBMFA is able to describe a small region of process in a locally linear way and the entire process in a globally nonlinear manner and therefore facilitate catching the characteristic of process even in the peaks or the valleys in terms of results in the previous sections.

In this paper, we show the benefits of VBMFA model for a WWTP process monitor. The results clearly demonstrate the superior performance of VBMFA model compared with other methods. However, future work should update not only the weights of each factor analyzer and corresponding monitor indexes, but also the associated variance and mean. Also, future work should aim to test and validate this methodology in a real WWTP system, which incurs us to implement our work in separate computer as shown in Figure 2. Additionally, due to the demands of cost saving in the water communities, it is imperative to assimilate economic index into the process monitoring in our future work.

Additional Points

A methodology for fault diagnosis in WWTPs has been successfully developed based on VBMFA, where estimated

posterior probabilities have been used to weight different local factor analyzers adaptively. Resulting monitoring statistics as well as associated control limit are reformulated as an adaptive manner. The results show that the proposed method achieves the best performance under fault-free scenario. This study further demonstrates the effectiveness of detecting two typical faults and localizing the associated faulty sensor.

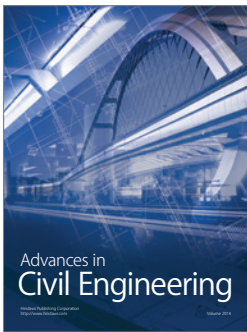
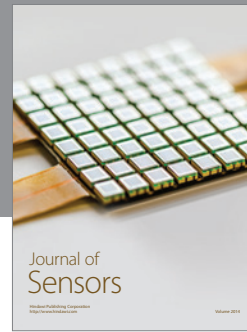
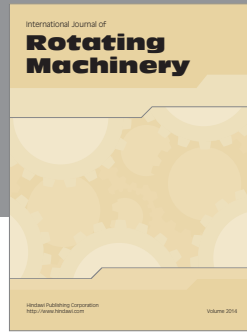
Competing Interests

The authors declare that they have no competing interests.

References

- [1] G. Olsson, "ICA and me—a subjective review," *Water Research*, vol. 46, no. 6, pp. 1585–1624, 2012.
- [2] F. Corona, M. Mulas, H. Haimi, L. Sundell, M. Heinonen, and R. Vahala, "Monitoring nitrate concentrations in the denitrifying post-filtration unit of a municipal wastewater treatment plant," *Journal of Process Control*, vol. 23, no. 2, pp. 158–170, 2013.
- [3] G. Li, C. F. Alcala, S. J. Qin, and D. Zhou, "Generalized reconstruction-based contributions for output-relevant fault diagnosis with application to the Tennessee Eastman process," *IEEE Transactions on Control Systems Technology*, vol. 19, no. 5, pp. 1114–1127, 2011.
- [4] S. P. N. G. G. Kaur, "Data driven multivariate technique for fault detection of waste water treatment plant," *International Journal of Engineering and Advanced Technology*, vol. 1, no. 4, pp. 45–50, 2012.
- [5] G. Lee, C. Han, and E. S. Yoon, "Multiple-fault diagnosis of the tennessee eastman process based on system decomposition and dynamic PLS," *Industrial & Engineering Chemistry Research*, vol. 43, no. 25, pp. 8037–8048, 2004.
- [6] T. Næs, O. Tomic, B.-H. Mevik, and H. Martens, "Path modelling by sequential PLS regression," *Journal of Chemometrics*, vol. 25, no. 1, pp. 28–40, 2011.

- [7] L. Qiang, S. J. Qin, and C. Tianyou, "Decentralized fault diagnosis of continuous annealing processes based on multilevel PCA," *IEEE Transactions on Automation Science and Engineering*, vol. 10, no. 3, pp. 687–698, 2013.
- [8] M. Ram Maurya, R. Rengaswamy, and V. Venkatasubramanian, "Application of signed digraphs-based analysis for fault diagnosis of chemical process flowsheets," *Engineering Applications of Artificial Intelligence*, vol. 17, no. 5, pp. 501–518, 2004.
- [9] I. T. Jolliffe, *Principal Component Analysis*, Springer, New York, NY, USA, 2002.
- [10] J.-M. Lee, C. Yoo, and I.-B. Lee, "Statistical process monitoring with independent component analysis," *Journal of Process Control*, vol. 14, no. 5, pp. 467–485, 2004.
- [11] C. K. Yoo, D. S. Lee, and P. A. Vanrolleghem, "Application of multiway ICA for on-line process monitoring of a sequencing batch reactor," *Water Research*, vol. 38, no. 7, pp. 1715–1732, 2004.
- [12] L. J. Cao, K. S. Chua, W. K. Chong, H. P. Lee, and Q. M. Gu, "A comparison of PCA, KPCA and ICA for dimensionality reduction in support vector machine," *Neurocomputing*, vol. 55, no. 1-2, pp. 321–336, 2003.
- [13] H. Wu and D. Huang, "Kernel Fisher discriminant analysis using feature vector selection for fault diagnosis," in *Proceedings of the 2nd International Symposium on Intelligent Information Technology Application (IITA '08)*, pp. 109–113, Shanghai, China, December 2008.
- [14] E. L. Russell, L. H. Chiang, and R. D. Braatz, "Fault detection in industrial processes using canonical variate analysis and dynamic principal component analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 51, no. 1, pp. 81–93, 2000.
- [15] K. Salahshoor, M. Kordestani, and M. S. Khoshro, "Fault detection and diagnosis of an industrial steam turbine using fusion of SVM (support vector machine) and ANFIS (adaptive neuro-fuzzy inference system) classifiers," *Energy*, vol. 35, no. 12, pp. 5472–5482, 2010.
- [16] Q. Jiang and X. Yan, "Probabilistic monitoring of chemical processes using adaptively weighted factor analysis and its application," *Chemical Engineering Research and Design*, vol. 92, no. 1, pp. 127–138, 2014.
- [17] Z. Ge, F. Gao, and Z. Song, "Mixture probabilistic PCR model for soft sensing of multimode processes," *Chemometrics and Intelligent Laboratory Systems*, vol. 105, no. 1, pp. 91–105, 2011.
- [18] T. Chen, J. Morris, and E. Martin, "Probability density estimation via an infinite Gaussian mixture model: application to statistical process monitoring," *Journal of the Royal Statistical Society—Series C: Applied Statistics*, vol. 55, no. 5, pp. 699–715, 2006.
- [19] Y. Liu, Y. Pan, Z. Sun, and D. Huang, "Statistical monitoring of wastewater treatment plants using variational Bayesian PCA," *Industrial & Engineering Chemistry Research*, vol. 53, no. 8, pp. 3272–3282, 2014.
- [20] Z. Ge and Z. Song, "Maximum-likelihood mixture factor analysis model and its application for process monitoring," *Chemometrics and Intelligent Laboratory Systems*, vol. 102, no. 1, pp. 53–61, 2010.
- [21] M. J. Beal, *Variational algorithms for approximate Bayesian inference [Ph.D. thesis]*, Gatsby Computational Neuroscience Unit, University College London, London, UK, 2003.
- [22] W. C. Sang, E. B. Martin, A. J. Morris, and I.-B. Lee, "Fault detection based on a maximum-likelihood principal component analysis (PCA) mixture," *Industrial & Engineering Chemistry Research*, vol. 44, no. 7, pp. 2316–2327, 2005.
- [23] U. Thissen, H. Swierenga, A. P. de Weijer, R. Wehrens, W. J. Melssen, and L. M. C. Buydens, "Multivariate statistical process control using mixture modelling," *Journal of Chemometrics*, vol. 19, no. 1, pp. 23–31, 2005.
- [24] J. Yu and S. J. Qin, "Multimode process monitoring with bayesian inference-based finite Gaussian mixture models," *AIChE Journal*, vol. 54, no. 7, pp. 1811–1829, 2008.
- [25] J. Yu, "A Bayesian inference based two-stage support vector regression framework for soft sensor development in batch bioprocesses," *Computers and Chemical Engineering*, vol. 41, pp. 134–144, 2012.
- [26] M. A. L. Traple, A. M. Saviano, F. L. Francisco, and F. R. Lourenço, "Measurement uncertainty in pharmaceutical analysis and its application," *Journal of Pharmaceutical Analysis*, vol. 4, no. 1, pp. 1–5, 2014.
- [27] N. Chotyakul, G. Velazquez, and J. A. Torres, "Assessment of the uncertainty in thermal food processing decisions based on microbial safety objectives," *Journal of Food Engineering*, vol. 102, no. 3, pp. 247–256, 2011.
- [28] J. Alex, L. Benedetti, J. Copp et al., *Benchmark Simulation Model No. 1 (BSMI)*, Department of Industrial Electrical Engineering and Automation, Lund University, Lund, Sweden, 2008.
- [29] L. K. Saul and S. T. Roweis, "Think globally, fit locally: unsupervised learning of low dimensional manifolds," *Journal of Machine Learning Research*, vol. 4, pp. 119–155, 2003.
- [30] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, New York, NY, USA, 2006.
- [31] S. W. Choi, J. H. Park, and I.-B. Lee, "Process monitoring using a Gaussian mixture model via principal component analysis and discriminant analysis," *Computers and Chemical Engineering*, vol. 28, no. 8, pp. 1377–1387, 2004.
- [32] E. B. Martin and A. J. Morris, "Non-parametric confidence bounds for process performance monitoring charts," *Journal of Process Control*, vol. 6, no. 6, pp. 349–358, 1996.
- [33] J. A. Westerhuis, S. P. Gurden, and A. K. Smilde, "Generalized contribution plots in multivariate statistical process monitoring," *Chemometrics and Intelligent Laboratory Systems*, vol. 51, no. 1, pp. 95–114, 2000.
- [34] M. I. Jordan, Z. Ghahramani, T. S. Jaakkola, and L. K. Saul, "Introduction to variational methods for graphical models," *Machine Learning*, vol. 37, no. 2, pp. 183–233, 1999.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

