A Mixture of Variational Canonical Correlation Analysis for Nonlinear and Quality-relevant Process Monitoring

Yiqi Liu, Bin Liu, Xiujie Zhao,
and Min Xie, *Fellow IEEE*

***Abstract*—Proper monitoring of quality-related variables in industrial processes is nowadays one of the main worldwide challenges with significant safety and efficiency implications. Variational Bayesian mixture of Canonical correlation analysis (VBMCCA)-based process monitoring method was proposed in this paper to predict and diagnose these hard-to-measure quality-related variables simultaneously. Use of Student's *t*-distribution, rather than Gaussian distribution, in the VBMCCA model makes the proposed process monitoring scheme insensitive to disturbances, measurement noises and model discrepancies. A sequential perturbation method together with derived parameter distribution of VBMCCA** is **employed to approach the uncertainty levels, which is able to provide confidence interval around the predicted values and give additional control line, rather than just a certain absolute control limit, for process monitoring. The proposed process monitoring framework has been validated in a Wastewater Treatment Plant (WWTP) simulated by Benchmark Simulation Model (BSM) with abrupt changes imposing on a sensor and a real WWTP with filamentous sludge bulking. The results show that the proposed methodology is capable of detecting sensor faults and process faults with satisfactory accuracy.**

***Index Terms*—Canonical correlation analysis; Process monitoring; Soft-sensor; Wastewater; Uncertainty**

Nomenclature

N	The number of samples for the process data, $k = 1, \cdots N$
d_1	The number of variables for the input data
d_2	The number of variables for the response data
d	Sum of d_1 and d_2
X_1	Input data matrix (process variables) $X_1 \in \mathbb{R}^{d_1 \times N}$
X_2	Response data matrix (quality variables) $X_2 \in \mathbb{R}^{d_2 \times N}$
U	Left singular matrix
V	Right singular matrix
w	The singular vectors of U
v	The singular vectors of V
Λ	Matrix of eigenvalues
l	the number of nonzero eigenvalues
Λ_l	Matrix of nonzero eigenvalues
I	Identity matrix
$x_{2,new}$	New coming response values
$\hat{x}_{2,new}$	The predicted new coming response values
Ψ_i	Precision matrices of X_i, *i*=1,2
t	The latent variables
W_i	The projection matrices of X_i, *i*=1,2
μ_i	mean values for the matrix X_i, *i*=1,2
g	gamma distribution
$\mathbb{W}$	Wishart distribution
$\mathbb{N}$	Gaussian distribution
S	The number of CCA models
μ_{t_s}	mean values of the t_s (*s*-th *t*, $s = 1,2, \cdots, S$)
Σ_{t_s}	Covariance matrix of the t_s
$\mu_{\mu_i^s}$	mean values of the μ_i^s (*s*-th *t*, $s = 1,2, \cdots, S$)
$\Sigma_{\mu_i^s}$	Covariance matrix of the μ_i^s
$\mu_{W_i^s}$	mean values of the W_i^s (*s*-th *t*, $s = 1,2, \cdots, S$)
$\Sigma_{W_i^s}$	Covariance matrix of the W_i^s
γ_i^s	Wishart distribution parameter for the *i*-th X and *s*-th CCA model
Φ_i^s	Wishart distribution parameter for the *i*-th X and *s*-th CCA model
$\alpha_{1,i}^s$	Parameter to translate *S* distribution to Normal distribution for the *i*-th X and *s*-th CCA model
a_i^s	Gamma distribution parameter for $\alpha_{1,i}^s$ related to the *i*-th X and *s*-th CCA model
b_i^s	Gamma distribution parameter for $\alpha_{1,i}^s$ related to the *i*-th X and *s*-th CCA model
u_n	Scale vector to translate *Student* distribution to Normal distribution
α_2	Gamma distribution parameter for u_n related to the *i*-th X and *s*-th CCA model
α_3	Significance level
β	Gamma distribution parameter for u_n related to the *i*-th X and *s*-th CCA model
$(\cdot)^+$	Upper uncertainty level with respect to $(\cdot)$
$(\cdot)^-$	Lower uncertainty level with respect to $(\cdot)$
$K(\cdot)$	Kernel function
h	Bandwidth of kernel function
δx	The uncertainty level with respect to variable *x* (such as *t* , *W* and so on)
U_{T^2}	Total standard variance with respect to T^2 uncertainty
U_{SPE}	Total standard variance with respect to SPE uncertainty

Manuscript received XXX, XXX. This work was supported by the National Natural Science Foundation of China (61673181), the Natural Science Foundation of Guangdong Province, China (2015A030313225), the Science and Technology Planning Project of Guangdong Province, China (2016A020221007).

Y.Liu is with the School of Automation Science & Engineering, South China University of Technology, Wushan Road, 510640, P. R. China (phone: 86-15814826119; fax: 303-555-5555; e-mail: aulyq@ scut.edu.cn).

Bin Liu, Xiujie Zhao and Min Xie are with the Department of Systems Engineering and Engineering Management, City University of Hong Kong, Kowloon, Hong Kong (e-mail: minxie@cityu.edu.hk).

I. Introduction

During recent decades, industrial process monitoring has gained significant attentions in industries and academia due to the increased awareness to ensure safer operation and better product qualities [1]. Multivariate statistical process monitoring (MSPM) is one of most commonly used strategies to deal with these issues [2-3]. These methods tend to explore the data by building an empirical model, which in turn acts as a reference to justify the desired process behavior of the new data by Hotelling's T^2 or squared predicted errors (SPE). Among the approaches, Principal Component Analysis (PCA) and Factor Analysis (FA) are investigated typically and applied intensively [4]. However, they focus only on the process variables (X_1) without any information about the quality variables (X_2), therefore always leading to false alarms. By comparison, Partial Least Squares (PLS) is able to decompose a data set into two subspaces by maximizing the covariance between X_1 and X_2. Nonetheless, PLS components may contain useless information to predict X_2 due to variations orthogonal to X_2. Moreover, maximum covariance criterion of PLS to extract principal components usually makes the residuals of X_1 or X_2 not necessarily small. This further results in the residual space monitoring with *SPE* statistic inappropriate. In contrast, Canonical correlation analysis (CCA) maximizes the correlation between X_1 and X_2, thus being able to correct within-set covariance prior to the decomposition [5-7]. CCA-based process monitoring methods, however, adhere to the assumption that quality variables are on-line measurable or measurable without a large time delay. To the best of the authors' knowledge, quality-based process monitoring is still far from sufficient investigation.

A further complicated characteristic of industrial process observations is that the data may be nonlinear in the time domain or may involve nonlinear interactions between variables. To handle the nonlinearity of process data, several nonlinear PCA approaches have been developed [8]. Kernel PCA (KPCA) is able to deal with a wide range of nonlinearities using different kernels without resorting to nonlinear optimization necessarily [9]. Also, a nonlinear approach can be obtained by postulating a finite mixture of linear sub-models for the Gaussian distribution of the full observation vector, yielding the so-called mixture of statistical models, such as mixture of PCA, mixture of CCA and mixture of Factor Analysis (FA) [10-12]. In general, these methods premise the assumption that the process is nonlinear with the operation mode being separable. Also, mixture-based models commonly suffer from model parameter estimation instability adversely resulting from outliers in the training data, therefore leading to unreliable model-based diagnosis systems. Student's *t*-distribution has been proposed recently against parameter estimation instability due to its bell-shaped distribution with heavier tails compared with the normal distribution [13]. Additionally, few papers devoted to a mixture of CCA for nonlinear process monitoring [14].

In this paper, we exploit a robust mixture of CCA (MCCA) model with latent variables and errors adhering to multivariate Student's *t*-distributions. Generally, maximum likelihood estimation for MCCA can be achieved by expectation maximization (EM) algorithm [15], and then, the estimated models can be cross-validated to choose an appropriate model. However, the task of model estimation and validation for various combinations of cardinality and local dimensionality could be tedious. Bayesian estimation can simplify the task as it penalizes complex models and allows for model selection without cross-validation [16-17]. However, a complete Bayesian analysis for MCCA may be infeasible. For finite mixture models, Variational Bayesian algorithm is a deterministic alternative to Markov chain Monte Carlo (MCMC) algorithms for Bayesian inference [18], with better scalability in terms of computational cost, thereby avoiding the singularity and over-fitting problems of maximum-likelihood approaches. Even though Variational Bayesian algorithm has been extended to Independent Component Analysis (ICA), PCA, Mixture of PCA and mixture of FA [19-20], few papers devoted them to process monitoring. Variational Bayesian algorithm is used to learn the parameters and model structure of mixture of CCA (VBMCCA) in this paper. The derived VBMCCA model in this paper further serves as the process monitoring model for abnormal process conditions identification. Due to use of Student's *t*-distribution in VBMCCA, the process monitoring scheme is more robust, implying that the trained model favors the trained pattern (normal condition) and is insensitive to dissimilar patterns (disturbances, noises or outliers). Since the residuals of process monitoring are generated by subtracting the predicted values from the true values, insensitive prediction values will make residual more obvious and process monitoring strategy more sensitive to the faults. Also, different from the assumption that hard-to-measure quality variables can be derived on-line directly in the typical process monitoring methods, VBMCCA is able to predict multiple hard-to-measure quality variables simultaneously, thus resulting in off-line measurement for process monitoring conveniently.

Along with the standard routes in the process monitoring, a process monitoring model is typically obtained by selecting the proper structure and then estimating the unknown parameters with available historical data or on-line monitored information [21]. With respect to uncertainty construction, two sources of information are necessary in the present standard identification process: the priori information on model structure, and the posteriori information (the data). Model structure can resort to the true system structure. The process and measurement uncertainty from data are typically assumed to be Gaussian. The mean and covariance for a Gaussian distribution can be inferred from historical data by statistical methods. However, few researches are devoted to analyzing the prediction accuracy. The uncertainty of industrial processes (such as lack of knowledge of behaviors) and unstable model parameters add further complexity for uncertainty description. A practical way of assigning the accuracy is to provide a confidence interval around the predicted value, to which the future output is guaranteed to belong with a certain probability. In this paper, we proposed to use a sequential perturbation method together with derived parameter distribution of VBMCCA to approach the uncertainty levels of the predicted results, therefore providing a confidence interval around the predicted value and giving additional control lines for fault detection, rather than just certain absolute values, for process monitoring.

The remainder of this paper is organized as follows. Section II gives some basic theories on the CCA-based process monitoring. Section III discusses the implementation of proposed VBMCCA-based process monitoring strategy. The corresponding uncertainty levels from the SP method is described in Section IV. Section V presents the performance of the proposed methodology through case studies of wastewater treatment processes. Finally, the work ends with a discusion section.

II. Preliminaries

A. Canonical correlation analysis (CCA)

Canonical correlation analysis is a method to find linear combinations of two vectors by maximizing correlations among them [7]. Suppose that N samples of the process data are collected, and X_1 and X_2 represent input variables and response variables, respectively, where $X_1 \in \mathbb{R}^{d_1 \times N}$ and $X_2 \in \mathbb{R}^{d_2 \times N}$, $d = d_1 + d_2$. Let X_1 and X_2 be centered by mean, then

$$\begin{bmatrix} \Sigma_{X_1X_1} & \Sigma_{X_1X_2} \\ \Sigma_{X_2X_1} & \Sigma_{X_2X_2} \end{bmatrix} \approx \frac{1}{N}\begin{bmatrix} X_1{X_1}^T & X_1{X_2}^T \\ X_2{X_1}^T & X_2{X_2}^T \end{bmatrix} \tag{1}$$

Let the matrix $\sum_T$

$$\Sigma_T = \Sigma_{X_1X_1}^{-1/2}\Sigma_{X_1X_2}\Sigma_{X_2X_2}^{-1/2} \tag{2}$$

By performing the singular value decomposition (SVD) on the matrix,

$$\Sigma_T = U\Lambda V^T \tag{3}$$

where $U = (\omega_1, \cdots, \omega_{d_1})$, $V = (v_1, \cdots, v_{d_2})$ and

$$\Lambda = \begin{bmatrix} \Lambda_l & 0 \\ 0 & 0 \end{bmatrix} \tag{4}$$

where $\Lambda_l = diag(\lambda_1, \cdots, \lambda_l)$, $\lambda_1 \geq \lambda_2 \cdots \geq \lambda_l > 0$ are the singular values. w_i and v_j represent the corresponding singular vectors, where $i = 1, \cdots d_1$ and $j = 1, \cdots d_2$. Two canonical correlation vectors or weighted matrixes with respect to X_1 and X_2 can be defined as $J = \Sigma_{X_1X_1}^{-1/2}U(:,1:l)$ and $L = \Sigma_{X_2X_2}^{-1/2}V(:,1:l)$, where $l \leq min(d_1, d_2)$.

B. CCA-based process monitoring

It is reasonable to define the residual vectors as follows [6]:

$$e(k) = L^T X_2(k) - \Lambda_l J^T X_1(k) \tag{5}$$

According to a quadratic form with the k-th sample, the T^2 statistic of CCA is derived as follows:

$$T^2 = (N-1)e^T(k)(I - \Lambda_l^2)^{-1}e(k) \tag{6}$$

Also, SPE can be defined accordingly:

$$SPE = e^T(k)e(k) \tag{7}$$

Based on [4],

$$SPE_{\lim} = \varsigma\chi^2_{1-\alpha_3}(\tau) \tag{8}$$

where $\varsigma = \Sigma_{SPE}/2\mu_{SPE}$, $\tau = 2\mu^2_{SPE}/\Sigma_{SPE}$, Σ_{SPE} and μ_{SPE} are estimated as

$$\mu_{SPE} = \frac{1}{N}\sum_{j=1}^{N} SPE(j), \Sigma_{SPE} = \frac{1}{N-1}\sum_{j=1}^{N}(SPE(j) - \mu_{SPE})^2$$

and the control limit of T^2 can be defined as

$$T^2_{\lim} = \frac{d(N^2-1)}{N(N-d)}F_{\alpha_3}(d, N-d) \tag{9}$$

Given the new sample $x_{1,new}$, the predicted values can be derived as follows:

$$\hat{x}_{2,new} = (x_{1,new})^T J\Lambda_l L^T(LL^T)^{-1} = (x_{1,new})^T \Sigma^{-1}_{X_1X_1}\Sigma_{X_1X_2} \tag{10}$$

where $\hat{x}_{2,new}$ denotes the predicted values of responses $x_{2,new}$ [22].

III. VBMCCA-based Process monitoring

In the present work, a VBMCCA-based process monitoring strategy is proposed. The detailed procedures to implement VBMCCA for process monitoring and prediction is presented in Fig.1. This will perform as follows: (1) For the first stage as illustrated in Fig.1, VBMCCA model is trained and learned by Variational Bayesian learning. (2) At the arrival of a new sample, the local prediction values, local prediction uncertainties, local T^2 and local SPE are derived by the local CCA; (3) the local prediction values and local prediction uncertainties are combined to derive the global quantities; (4) local T^2 and local SPE are integrated for global process monitoring. In the fault detection procedure, once the upper uncertainty values of SSPE and ST^2 reach control limits, the system will trigger pre-caution to find and remove the assignable causes. If SSPE or ST^2 exceeds their control limit, inspection and replacement of faulty components are necessary.

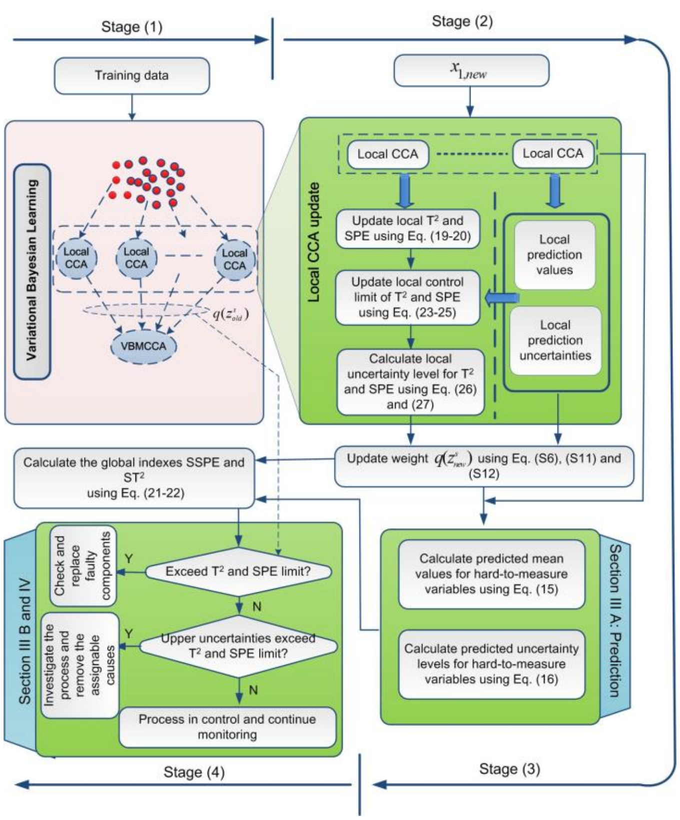


Fig.1 Schematic diagram to implement VBMCCA for process monitoring and prediction

A. VBMCCA

To alleviate a strict assumption of Gaussian noise in the Bayesian MCCA model [23], which is sensitive to outliers, both the Gaussian noise and the Gaussian latent variables were replaced with Student's *t*-distribution. Generally, the Student's *t*-distribution is a bell-shaped with heavier tails but with one more parameter, degree of freedom (DOF), compared to Gaussian distribution, therefore leading to a more robust approach to deal with aforementioned issues [13]. For efficient inference, we exploit the latent infinite scale-mixture formulation of the *t*-distribution using Variational Bayesian learning, leading to a robust Bayesian CCA. Using this formulation, we can further write the robust Bayesian CCA model by adding an extra level of hierarchy

$$u_n \sim g(u_n \mid \alpha_2, \beta) \tag{11}$$

$$t \mid u_n \sim \mathbb{N}(t \mid \mu_t, u_n \Sigma_t) \tag{12}$$

$$X_1 \mid t \sim \mathbb{N}(X_1 \mid W_1 t + \mu_1, u_n \Psi_1) \tag{13}$$

$$X_2 \mid t \sim \mathbb{N}(X_2 \mid W_2 t + \mu_2, u_n \Psi_2) \tag{14}$$

where Ψ_i denotes the precision matrices of the normal distribution. The latent variables *t* encodes the low-dimensional statistically dependent part while the projection matrices W_i specify how this dependency is manifested in each data source. μ_i represents the mean values, *i*=1,2.

Remark 1: Due to the convenience and equality between the Normal distribution and Student's *t*-distribution, Student's *t*-distribution of *X* can be transformed to Normal distribution accordingly, which is proved by reference [24].

Fig.2 further illustrates the relationship between the variables:

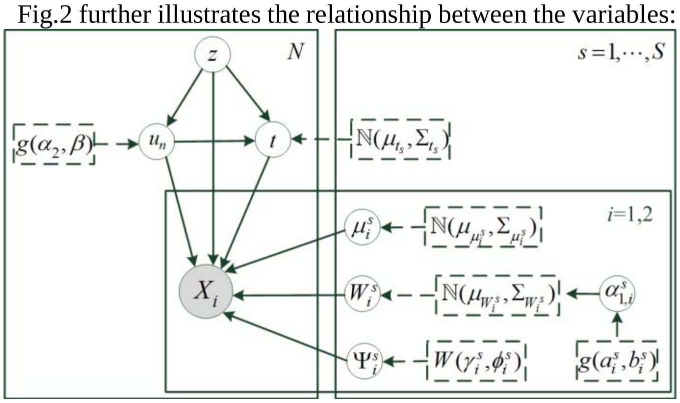


Fig.2 Schematic diagram of VBMCCA

where $g(\cdot)$ is the gamma distribution, $\mathbb{W}(\cdot)$ denotes the Wishart distribution and $\mathbb{N}(\cdot)$ represents the Gaussian distribution. In this work, we introduce a probabilistic mixture of the aforementioned robust Bayesian CCA models, letting each mixture cluster to model different kinds of dependencies between the signals by replacing it with piecewise stationarity. The robust mixture CCA model is therefore obtained by adding the latent variable $z \sim \text{Multinomial}(z|r)$, where r is the parameter of *Multinomial* distribution. The core of the inference process lies in learning the posterior distribution $P(H|X_1, X_2, \Theta)$ of both the latent variables and the model parameters, denoted collectively as $H = \{\mathrm{z}, \mathrm{u_n}, \mathrm{t}, \mathrm{W_1}, \mathrm{W_2}, \mu_1, \mu_2, \Psi_1, \Psi_2\}$, $\Theta = \{\mathrm{a_i^s}, \mathrm{b_i^s}, \gamma_i^s, \Phi_i^s, \alpha_2, \beta_2, \beta_1\}$.

Remark 2: Usually, the ad hoc values for hyper-parameters treated under a Variational Bayes' framework are chosen such that broad prior distributions are derived. Therefore, a good selection for the hyper-parameters of the priors on Ψ_1, Ψ_2. $\gamma_i^s = d_i + 1$, $\Phi_i^s = 10^2 \mathrm{I}$ is to obtain broad distributions. We can also choose broad priors for the shape hyper-parameters a_i^s, α_2 and the inverse-scale hyper-parameter b_i^s, β_2, which can be set to $a_i^s = b_i^s = 0.1$ and $\alpha_2 = \beta_2 = 1$. Finally, concerning the prior over the mixing proportions vector, we obtain broad priors for the β_1 being 10^{-3} [10].

The procedures of the Variational Bayesian algorithm to search for the model parameters are shown in the Supplementary Information B. Given the inferred latent variables and model parameters, the predictions for the incoming new data can be inferred using the following equation:

$$\mu_{\hat{x}_{2,new}} = \sum\nolimits_{s=1}^{S} q(z_{new}^s)(\mu_{W^s}\mu_{t_{new}^s} + \mu_{\mu^s}) \tag{15}$$

$$\Sigma_{\hat{x}_{2,new}}^{-1} = \sqrt{\sum\nolimits_{s=1}^{S}[q(z_{new}^s)((u_n\Psi^s)^{-1} + W^s\Sigma_{t_{new}^s}(W^s)^T)\ ]^2} \tag{16}$$

where $\mu_{\hat{x}_{2,new}}$ and $\Sigma_{\hat{x}_{2,new}}^{-1}$ denote the predicted mean values and precision matrix, respectively; $z = \{z^s\}_{s=1}^S$ is the set of label indicator vectors, with $z^s \in \{0,1\}$. For the *i*th data sample, if x_i is attributed as the *s*-th mixture component, $z_i^s = 1$, otherwise, $z_i^s = 0$. $q(z_{new}^s)$ represents the contribution of each local Bayesian CCA to the global model with respect to the incoming new data points. By deriving the contribution of each Bayesian CCA, they can be combined using average.

B. VBMCCA-based process monitoring

1) Calculation of T^2 and SPE

To perform process monitoring, monitoring statistic T^2 and SPE are constructed as the first step. The corresponding monitoring statistics for the *s*-th local Bayesian CCA can be represented as follows:

$$T_s^2 = t_s^T \Sigma_{t_s}^{-1} t_s = \mu_{t_s}^T \Sigma_{t_s}^{-1} \mu_{t_s} \tag{17}$$

with

$$\Sigma_{t_s}^{-1} = \sum_{i=1}^{2}((\mu_{W_i^s})^T \Psi_i^s \mu_{W_i^s} + \sum_{j=1}^{d_2} \Psi_{ij}^s \mu_{W_{ij}^s}) + I_d$$

$$\mu_{t_s} = \Sigma_{t_s}(\sum_{i=1}^{2}(\mu_{W_i^s})^T \Psi_i^s (x_i - \mu_{\mu_i^s}))$$

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

$$SPE^s = e_s^T \Psi_s^{-1} e_s \tag{18}$$

with

$$e_s = x - \mu_{W^s}\mu_{t_s} - \mu_{\mu^s}$$

where $s = 1,2,\cdots,S$. l^s is the number of retained factors in the *s*-th local VBCCA model. As for a new coming sample data x_{new}, both statistics are updated as follows:

$$T_{s,new}^2 = t_s^T \Sigma_s^{-1} t_s = \mu_{t_s,new}^T \Sigma_s^{-1} \mu_{t_s,new} \tag{19}$$

$$SPE_{new}^s = e_{s,new}^T \Psi_s^{-1} e_{s,new} \tag{20}$$

where

$$\mu_{t_s,new} = \Sigma_{t_s}(\sum_{i=1}^{2}(\mu_{W_i^s})^T \Psi_i^s (x_{new} - \mu_{\mu_i^s}))$$

$$e_{s,new} = x_{new} - \mu_{W^s}\mu_{t_s,new} - \mu_{\mu^s}$$

Remark 3: In the case that on-line quality measurement is unavailable, x_{new} is replaced by $\mu_{\hat{x}_2}$.

However, it is cumbersome to monitor the process with all the local monitoring charts, which also potentially frustrates the decision-making for process monitoring with so many monitoring charts. Therefore, coordination of all the local monitoring charts is imperative. We can combine them through the estimated posterior probabilities. Both of the synthetic monitor statistics T^2 and SPE can be obtained as follows:

$$ST_{new}^2 = \sum_{s=1}^{S} q(z_{new}^s) T_{s,new}^2 \tag{21}$$

$$SSPE_{new} = \sum_{s=1}^{S} q(z_{new}^s) SPE_{new}^s \tag{22}$$

2) Control limit of CCA-based and VBMCCA-based process monitoring

For the standard multivariate process monitoring, control limit of T^2or SPE is premised on the assumption of following χ^2 distribution. To relax this assumption, kernel density function estimation is proposed in this paper. Given N samples of derived SPE, the density function estimation with respect to the sth SPE can be defined as follows

$$\hat{f}(SPE^s) = \frac{1}{Nh} \sum_{i=1}^{N} K\left(\frac{SPE_i^s - \mu_{spe}^s}{h}\right) \tag{23}$$

where K is the kernel function and h is the bandwidth, respectively. μ_{spe}^s denotes the mean values of the s-th SPE. SPE_i^s is the i-th component in the diagonal of the matrix and can be derived by Eq. (18). On the basis of the density function $\hat{f}(SPE^s)$, the control limit "SPE_{limit}^s" at a particular confidence level α_3 is defined by

$$P(SPE < SPE_{\mathrm{limit}}) = \int_{-\infty}^{SPE_{\mathrm{limit}}} \hat{f}(SPE) d(SPE) = \alpha_3 \tag{24}$$

The SPE_{limit} can be obtained by calculating the PDF (Probability Density Function) of Eq. (24). α_3 is the significance level of both monitoring statistics.

For the VBMCCA-based process monitoring strategy, it is necessary to combine the control limits. $q(z^s)$ is used to weight the associated SPE control limit as follows:

$$SPE < SPE_{\mathrm{limit}} = \sum_{s=1}^{S} q(z^s) SPE_{\mathrm{limit}}^s \tag{25}$$

The calculation of SPE control limit can be generalized to T^2 counterpart of both the CCA-based and VBMCCA-based strategy.

Remark 4: The purpose of mixture of local Bayesian CCA is to combine the decisions of local models, in such way that the correct decisions are amplified, and the incorrect ones cancelled out. However, to simplify the process monitoring and avoid false alarms, we learn the control limit for each Bayesian CCA in the training set and combine them as a constant control limit for the testing set. Therefore, $q(z^s)$ is used to average the monitoring statistic instead of $q(z_{new}^s)$.

IV. Uncertainty Intervals Calculation for VBMCCA-based Process Monitoring

In process monitoring engineering, process monitoring results are typically determined through a functional relationship with the calculated values, such as t, μ. Clearly, small errors in estimating these values will lead to large differences in the outputs. Thus, it is imperative to identify, quantify and combine the errors in the final results, so as to measure the 'goodness' of a result, which is always defined as uncertainty analysis. Uncertainty analysis involves determining the uncertainty in model predictions that results from imprecisely known variables or parameters. The commonly used uncertainty analysis methods consist of analytical methods and numerical methods. This paper tends to use the Sequential Perturbation (SP) technique as a tool to qualify the uncertainty level for process monitoring statistical indexes. The procedure to calculate uncertainties using SP is shown in the Supplementary Information C.

TABLE I
Upper and Lower Uncertainty Level Calculation

		Upper and Lower limits	
T^2	t_s	$t_s^+ = f_T(\mu_{t_s} + \Sigma_{t_s})$; $t_s^- = f_T(\mu_{t_s} - \Sigma_{t_s})$	$\delta t = \frac{\lvert t_s^+\rvert - \lvert t_s^-\rvert}{2}$
SPE	W_s	$W_s^+ = f_{spe}(\mu_{W_s} + \Sigma_{W_s}, \mu_{t_s}, \mu_s)$; $W_s^- = f_{spe}(\mu_{W_s} - \Sigma_{W_s}, \mu_{t_s}, \mu_s)$	$\delta W = \frac{\lvert W_s^+\rvert - \lvert W_s^-\rvert}{2}$
	t_s	$t_s^+ = f_{spe}(\mu_{t_s} + \Sigma_{t_s}, \mu_{W_s}, \mu_s)$; $t_s^- = f_{spe}(\mu_{t_s} - \Sigma_{t_s}, \mu_{W_s}, \mu_s)$	$\delta t = \frac{\lvert t_s^+\rvert - \lvert t_s^-\rvert}{2}$
	μ_s	$\mu_s^+ = f_{spe}(\mu_{\mu_s} + \Sigma_{\mu_s}, \mu_{W_s}, \mu_{t_s})$; $\mu_s^- = f_{spe}(\mu_{\mu_s} - \Sigma_{\mu_s}, \mu_{W_s}, \mu_{t_s})$	$\delta \mu_s = \frac{\lvert \mu_s^+\rvert - \lvert \mu_s^-\rvert}{2}$

\+ is upper limit and - is the lower limit

SP method is a numerical approach to estimate the propagation of uncertainty and especially suitable for the situation where direct partial differentiation is too cumbersome or extremely complex or the number of related variables is large. Based on the measurements of the independent variables under specific operating condition, the first step is to increase the independent variables by their respective uncertainty and recalculate the result based on each of these new values, then, in the similar manner, decrease the independent variables by their respective uncertainties and recalculate the result based on each of these new values. Finally, all the derived uncertainties are aggregated by the root mean square.

Firstly, we let T^2 function be equal to f_T and SPE function be equal to f_{spe} , respectively. Table I summarizes the corresponding elements of the SP method. To avoid neutralization of the positive and negative differences, we sum the total differential by using the absolute value of each difference. With the table fully filled out, we can now calculate the total error by summing the values of δx and taking the square root of the value, shown as follows:

$$U_{T^2} = \pm\sqrt{\delta t^2} \tag{26}$$

$$U_{SPE} = \pm\sqrt{\delta W^2 + \delta t^2 + \delta\mu_s^2} \tag{27}$$

where U_{T^2} and U_{SPE} represent a confidence level of one standard deviation of T^2 and SPE, respectively, equivalent to a probability of 68% for a normal distribution. The intervals defined such uncertainties at 95% and 99% probability level are written as $\pm 2U$ and $\pm 3U$.

V. Case Studies

A. Benchmark Simulation Model (BSM) to simulate a WWTP operation

1) Background

Benchmark Simulation Model (BSM) was proposed by the International Water Association (IWA) Task Group, which offers a platform to evaluate diverse control strategies without necessarily resorting to a particular facility. The simulated plant mainly comprises of five sequentially connected reactors along with a 10-layer secondary settling tank, the aim of which is to remove nitrogen by pre-denitrification (1st and 2nd tanks not to be aerated) and nitrification reaction (3rd, 4th and 5th tanks to be aerated) (Fig. S1 in the Supplementary Information A). For more details, one can see reference [25].

This case study is to develop a VBMCCA-based process monitoring methodology to ensure safety operation of a WWTP while monitoring the hard-to-measure quality-related variables, such as BOD_5 (Biological oxygen demand for five days), TN (Total nitrogen) and others (Variable 23-29 in Table S1 and Fig. S2 in the Supplementary Information A). Even though some expensive sensors are available in certain WWTPs, the reliability of the sensor always compromises the decision making for a WWTP management. In this case, all input variables for model construction are sampled every 15 min and tabulated in Table S1 in Supplementary Information A. A plant under dry weather over 14 days was simulated in the BSM1 platform. In this case study, 1344 samples were collected. 700 samples were used for model training, while the remaining was for testing.

To demonstrate the effectiveness of the proposed method, we conducted comparison experiments with existing CCA-based and PLS-based schemes. By crossing validation, the dimension of t was set up as 5, 6 and 6 for CCA, PLS and KPCA, respectively. The 'Gaussian' kernel was selected for KPCA. For VBMCCA, we fixed the hyper-parameters of VBMCCA corresponding to broad priors as Remark 2 and consequently let the data determine the model parameters. By performing the crossing validation, S is equal to 2. Process monitoring schemes were applied to a process data set with an abrupt change fault (35% positive bias in S_O sensor) and their performances were analyzed in terms of the overall error rates including type I error and type II error. Abrupt change one type of common sensor faults in the process control. The Root Mean Square Error (*RMSE*) and coefficient (*R*) were used to access the prediction performance of inferential model for each response. Root-Mean Sum of Squares of the Diagonal (*RMSSD*) was used as a criterion to assess multiple prediction. Details of these indexes are shown in Supplementary Information D.

2) Prediction performance

In the WWTP, the effluent variables are always hard-to-measure due to time delay or unavailability of hardware sensors.

The proposed VBMCCA prediction model was validated and compared with CCA model under the S_O sensor fault in the fifth tank. The efficiency of the proposed strategy for prediction is shown in Fig.3, indicating that, during normal conditions, the proposed soft-sensor achieved better performance with respect to the RMSE and R for S_{NO} and COD compared with CCA. Prediction performance of other variables can be seen in Table S2 in the Supplementary Information A. This is mainly due to the nonlinear format of VBMCCA to capture the peak and valley of the variables. Conversely, the VBMCCA is insensitive to ill-situations under the faulty conditions, thus failing to track the state variation. However, this is in turn able to enlarge the discrepancy between the true and predicted values and further amplify the residuals for process monitoring.

Table II further displays the summarized prediction performance of VBMCCA in terms of RMSSD, suggesting that VBMCCA can achieve the best prediction performance during the normal state but show the significant difference from the true values during abnormal state. Due to deterioration of predicted results of VBMCCA in the abnormal state, the uncertain intervals become wider accordingly to suggest the lower confidence for predicted results (2×standard variance, 90%).The standard variance can be calculated by Eq. (16) and is able to serve the basis for the uncertainty calculation. For the SNO prediction, it is also perceived that VBMCCA-based prediction is closer to the mean values without faults influence than the CCA-based counterpart, demonstrating that VBMCCA-based model is more robust to disturbances. However, since the fault residuals are generated by subtracting SNO prediction to true values, VBMCCA-based fault detection strategy is more sensitive to the faults.

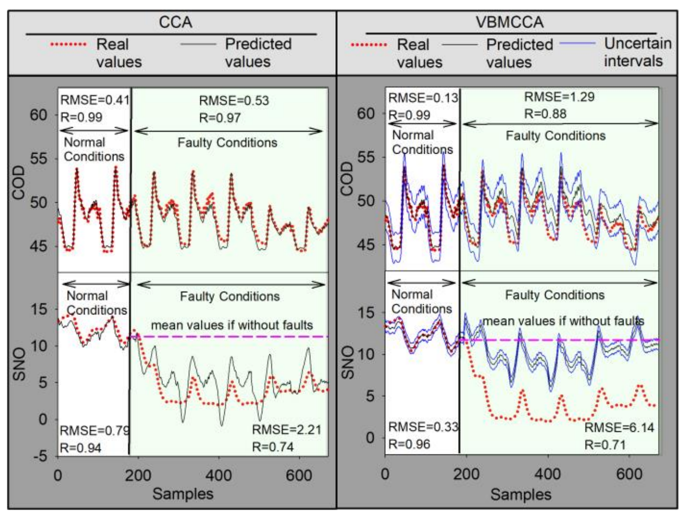


Fig.3 Prediction results of a wastewater plant under BSM1 using CCA and VBMCCA in case of the So fault

3) Process monitoring

To validate the performance of VBMCCA-based process monitoring, the monitoring results of the four schemes for the

TABLE II
ERROR RATES (TYPE I/TYPE II) OF EACH MONITORING SCHEME IN THE BSM PROCESS (%) AND RMSSD

Methods	T^2		SPE		RMSSD	
	I	II	I	II	Normal	So fault
CCA	0.6	16	16	14.4	1.4	4.1
PLS	3	18.5	18	13.6	1.5	3.6
KPCA	2.3	12.3	8.8	6.9	\	\
VBMCCA	0	4.6	5.5	2	0.74	12.5

I is type I error and II is the Type II error testing data set are summarized in Table II. The type I error rate was estimated by the rate of misclassified normal samples to entire normal samples from observation 1 to 165 and the type II error rate was the rate of misclassified fault samples to entire fault samples from observation 166 to 644. Due to the abrupt change fault in the S_O sensor inside a closed-loop, this fault will propagate to the hard-to-measure quality-related variables (S_{NO}) in the discharge of the WWTP, leading to regulation violation. As tabulated in Table II, both of the linear schemes (CCA and PLS) show very similar performances for T^2 and SPE with high type I and type II error rates. On the contrary, VBMCCA achieved the best performance with the lowest type I and type II error rates. Even though KPCA can achieve better performance than linear models, its performance was still poorer than VBMCCA. The reason is mainly due to the fact that the VBMCCA model makes a better prediction during normal state, thus potentially leading to a smaller residuals and less false alarms on one hand. On the other, the predicted model is insensitive to ill-situations due to involvement of Student's *t*-distribution, which can further enlarge the residuals. Therefore, all the hard-to-measure quality-related variables can be predicted properly in the normal state, whereas significant deviation can be achieved from the true values.

Fig. 4 further profiles that VBMCCA declares the fault over 34 steps ahead of CCA-based strategy approximately. Upper uncertainty bounds of VBMCCA-based methodology are generated from SP algorithm. This additional control limit has recognized the incipient variations and can provide a pre-caution in advance. Since the state of a potential fault is still less than absolute control limit before 166, the corresponding faults can be investigated and maintained in time at two hours ahead approximately.

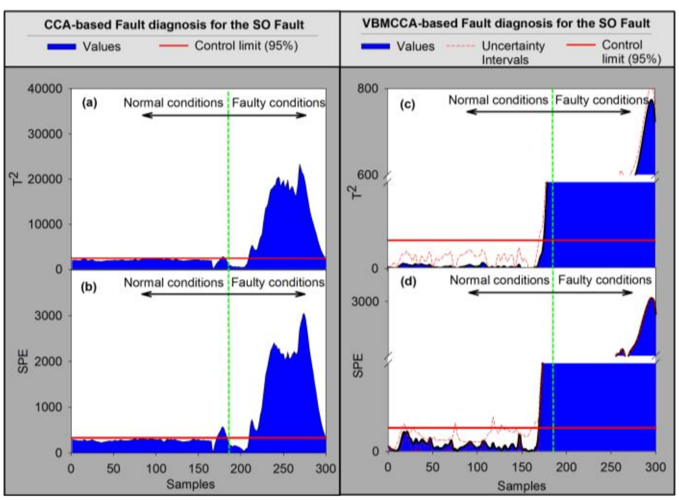


Fig.4 Monitored results of a wastewater plant under BSM1 using CCA and VBMCCA in case of the So fault

B. *A full-scale WWTP*

1) *Background*

The present case is a full-scale WWTP (Beijing, China, 480,000 population equivalents), designed to treat municipal wastewater with an Oxidation ditch (OD) process. OD process is an enhanced activated sludge biological treatment process, which aims to make solids retention time (SRT) longer to better nitrogen removal performance. Filamentous bulking sludge, a term used to describe the excess proliferation of filamentous bacteria, often results in slower settlement, poorer operational performance and higher treatment cost [26-27]. The selected monitored variables for model construction are shown as Table S3 in the Supplementary Information. 213 data points were sampled from the field at day interval. Data for the first 80 days was used for training, while the remaining was for testing. From the 20th day, Filamentous bulking sludge occurred due to the low COD of influent. The phenomenon of bulking sludge lasted for about half a year. These data was used to develop and validate the model in this study.

2) *Performance of process monitoring and quality-variables prediction*

Different from the abrupt changes fault in the first case study, filamentous bulking sludge is typical drifting errors, which vary in small magnitude and slow frequency. By cross-validation, the dimension of *t* was set up as 4, 5 and 5 for CCA, PLS and KPCA, respectively. The 'Gaussian' kernel was selected for KPCA. In the VBMCCA model, we fixed the hyper-parameters of VBMCCA corresponding to broad priors as Remark 2. By performing the crossing validation, *S* is equal to 4. The projected components of data sets were set up as 2 and 3, i.e., the dimension of *t*, for CCA and PLS, respectively.

The prediction performance was validated firstly. Table II indicates that VBMCCA can achieve the best prediction performance in terms of RMSSD during the normal state but gain the worst performance during the faulty stage. Best fitting under the normal conditions is able to alleviate the Type I error, whereas poorest fitting under the faulty conditions is capable of enlarging the discrepancies of faulty signals and predicted signals, therefore decrease the Type II error.

The monitoring performance for filamentous sludge bulking is summarized in Table III and Fig. S3, suggesting that VBMCCA-based process monitoring strategy achieved the best performance in terms of type I error rate and type II error rate for both of T^2 and SPE statistics. This mainly lies in the fact that a probabilistic mixture of robust Bayesian CCA models can characterize different kinds of dependencies between the signals with piecewise stationarity. The piecewise stationarity is able to approach the nonlinear relationship between the variables properly. KPCA can approach nonlinear relationship, but it is tedious to select the kernel function. Even though a proper kernel can be selected, KPCA is unsuitable to deal with wide range of nonlinearity.

TABLE III
ERROR RATES (TYPE I/TYPE II) OF EACH MONITORING SCHEME IN THE FULL-SCALE PROCESS (%) AND RMSSD

Methods	T^2		SPE		RMSSD	
	I	II	I	II	Normal	Fault
CCA	0	76	0	87	12	36
PLS	0	80	0	83	11.9	38
KPCA	0	39	18	26	/	/
VBMCCA	0	32	15	19	10.2	41

I is type I error and II is the Type II error

To further illustrate the efficiency of the proposed strategy for process monitoring, VBMCCA-based and CCA-based process monitoring schemes are shown in Fig. 5. Fig. 5 suggests that CCA-based strategy failed to identify the slow variations of filamentous sludge bulking for both of T^2 and SPE until the

70th day. On the contrary, VBMCCA-based strategy is able to identify the filamentous sludge bulking from the 21th day. As aforementioned, due to the use of Student's *t*-distribution, rather than Gaussian, the residuals between the true and predicted values can be enlarged significantly and thus making the drifting errors more obvious herein.

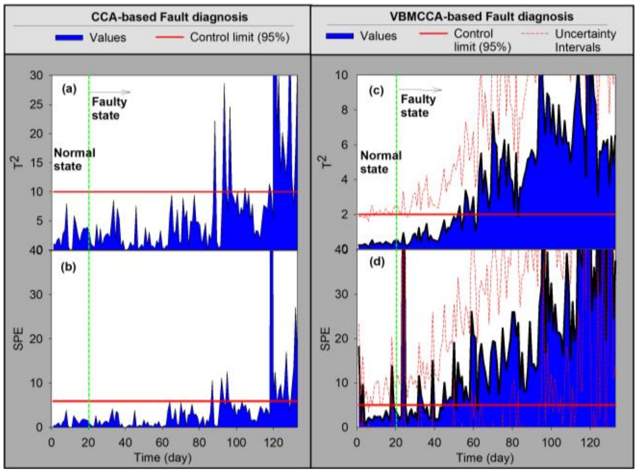


Fig.5 Monitored results of a wastewater plant using CCA and VBMCCA in case of filamentous sludge bulking

Notice that additional control limit formulated by upper uncertainty level is able to indicate the confidence of diagnosis results (90%) and further claim the fault in advance. Since the state of a potential fault is still less than absolute control limit, the corresponding faults can be investigated and maintained at an acceptable level at twelve days ahead approximately.

VI. Discussions

This study develops a VBMCCA-based process monitoring tool for diagnosis and estimation of multiple hard-to-measure quality-related variables simultaneously for a simulated and a full-scale WWTP. Simulation study results show that the proposed method can achieve satisfactory process monitoring performance in terms of Type I and II errors and better unforeseen variables prediction in terms of RMSE, R and RMSSD. This study further illustrates that the derived uncertainty interval not only provides confident description of the process monitoring and prediction results, but also offers double control limits for process monitoring, thereby leading to less false alarms and more effective maintenance strategies.

Different from standard mixture of models, the MCCA is learned by the Variational Bayesian methods which allows for optimization by using the entire training set in a single pass, rather than cross-validation as the case of maximum-likelihood approaches. Also, the standard Gaussian distribution is replaced by the Student's *t*- distribution, which results in a more robust model. Consequently, the derived VBMCCA model can make a better prediction for unforeseen variables during the normal state, thus potentially leading to smaller residuals and less false alarms on one hand. On the other, the predicted model is insensitive to abnormal conditions due to the involvement of Student's distribution, which can further enlarge the residuals to process monitoring significantly. However, this could in turn make the derived model only be adhered to the trained patterns and be difficult to be generalized into exclusive scenarios. This could be solved by using the online optimization algorithm to enhance VBMCCA for full-scale adaptive process monitoring. In this paper, we derived the number of sub-CCA models by crossing validation. The main purpose is to simplify the calculation procedure and can be further addressed by automatic relevant determination (ARD) [10].

In this study, we demonstrate the performance of VBMCCA through simulation studies. The first case study represents a highly instrumented WWTP system with an abrupt fault and a lowly instrumented system with a drifting error. Although the proposed methodologies achieve satisfactory performance in both the simulation studies, they require further verification through application to real WWTPs. Further, it is of importance to notice that the uncertainty intervals of the first case are more obvious than the second one. In the drifting errors, the onset of drifting errors is slight and recognized by the normal state firstly. In the present study, even though wastewater processes are used for validation, abrupt changes and drifting errors are very common in industrial processes. Especially in the microbial system including pharmaceutical industry and food systems, the online analyzer is unreliable and numerous variables are needed for prediction and monitoring.

References

[1] S. X. Ding, *et al.*, "A Novel Scheme for Key Performance Indicator Prediction and Diagnosis With Application to an Industrial Hot Strip Mill," *IEEE Transactions on Industrial Informatics*, vol. 9, pp. 2239-2247, 2013.

[2] Z. Ge, "Review on data-driven modeling and monitoring for plant-wide industrial processes," *Chemometrics and Intelligent Laboratory Systems*, vol. 171, pp. 16-25, 2017.

[3] Z. Ge, *et al.*, "Data Mining and Analytics in the Process Industry: The Role of Machine Learning," *IEEE Access*, vol. 5, pp. 20590-20616, 2017.

[4] J. Zhu, *et al.*, "Distributed Parallel PCA for Modeling and Monitoring of Large-Scale Plant-Wide Processes With Big Data," *IEEE Transactions on Industrial Informatics*, vol. 13, pp. 1877-1885, 2017.

[5] Z. Chen, *et al.*, "Canonical correlation analysis-based fault detection methods with application to alumina evaporation process," *Control Engineering Practice*, vol. 46, pp. 51-58, 2016.

[6] Z. Chen, *et al.*, "Improved canonical correlation analysis-based fault detection methods for industrial processes," *Journal of Process Control*, vol. 41, pp. 26-34, 2016.

[7] S. X. Ding, "Canonical Variate Analysis Based Process Monitoring and Fault Diagnosis," in *Data-driven Design of Fault Diagnosis and Fault-tolerant Control Systems*, ed London: Springer London, 2014, pp. 117-131.

[8] X. Deng and X. Tian, "Nonlinear process fault pattern recognition using statistics kernel PCA similarity factor," *Neurocomputing*, vol. 121, pp. 298-308, 2013.

[9] R. T. Samuel and Y. Cao, "Fault detection in a multivariate process based on kernel PCA and kernel density estimation," in *Automation and Computing (ICAC), 2014 20th International Conference on*, 2014, pp. 146-151.

[10] S. P. Chatzis, *et al.*, "Signal Modeling and Classification Using a Robust Latent Space Model Based on t distribution," *IEEE Transactions on Signal Processing*, vol. 56, pp. 949-963, 2008.

[11] J. Chen and J. Liu, "Mixture Principal Component Analysis Models for Process Monitoring," *Industrial & Engineering Chemistry Research,* vol. 38, pp. 1478-1488, 1999.

[12] J. Viinikanoja, *et al.*, "Variational Bayesian Mixture of Robust CCA Models," *European Conference* in *Machine Learning and Knowledge Discovery in Databases, Barcelona, Spain, September 20-24, 2010, Part III*, J. L. Balcázar, *et al.*, Eds., ed, Springer Berlin Heidelberg, , pp. 370-385, 2010.

[13] S. Shoham, "Robust clustering bydeterministic agglomeration EM of mixtures of multivariate t-distributions," *Pattern Recognition,* vol. 35, pp. 1127-1142, 2002.

[14] P. E. P. Odiowei and Y. Cao, "Nonlinear Dynamic Process Monitoring Using Canonical Variate Analysis and Kernel Density Estimations," *IEEE Transactions on Industrial Informatics,* vol. 6, pp. 36-45, 2010.

[15] U. Thissen, *et al.*, "Multivariate statistical process control using mixture modelling," *Journal of Chemometrics,* vol. 19, pp. 23-31, 2005.

[16] B. Cai, *et al.*, "Bayesian Networks in Fault Diagnosis," *IEEE Transactions on Industrial Informatics,* vol. 13, pp. 2227-2240, 2017.

[17] B. Cai, *et al.*, "A real-time fault diagnosis methodology of complex systems using object-oriented Bayesian networks," *Mechanical Systems and Signal Processing,* vol. 80, pp. 31-44, 2016.

[18] P. Bunch, *et al.*, "Bayesian Learning of Degenerate Linear Gaussian State Space Models Using Markov Chain Monte Carlo," *IEEE Transactions on Signal Processing,* vol. 64, pp. 4100-4112, 2016.

[19] K. Watanabe, *et al.*, "Variational Bayesian Mixture Model on a Subspace of Exponential Family Distributions," *IEEE Transactions on Neural Networks,* vol. 20, pp. 1783-1796, 2009.

[20] A. K. David M. Blei, Jon D. McAuliffe, "Variational Inference: A Review for Statisticians," *Journal of the American Statistical Association* vol. 112, pp. 859-877, 2017.

[21] S. Simani, *et al.*, "Fault Diagnosis of a Wind Turbine Benchmark via Identified Fuzzy Models," *IEEE Transactions on Industrial Electronics,* vol. 62, pp. 3775-3782, 2015.

[22] Z. Chen, *et al.*, "Fault Detection for Non-Gaussian Processes Using Generalized Canonical Correlation Analysis and Randomized Algorithms," *IEEE Transactions on Industrial Electronics,* vol. PP, pp. 1-1, 2017.

[23] D. R. Hardoon, *et al.*, "Canonical Correlation Analysis: An Overview with Application to Learning Methods," *Neural Computation,* vol. 16, pp. 2639-2664, 2004.

[24] C. Liu and D. Rubin, "ML estimation of the t distribution using EM and its extensions, ECM and ECME," *Statistica Sinica,* vol. 5, pp. 19-39, 1995.

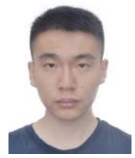
[25] I. Nopens;, *et al.*, "Benchmark Simulation Model No 2: finalisation of plant layout and default control strategy," *Water Science & Technology,* vol. 62, pp. 1967-1974, 2010.

[26] G. Olsson, "ICA and me – A subjective review," *Water Research,* vol. 46, pp. 1585-1624, 2012.

[27] A. M. P. Martins, *et al.*, "Filamentous bulking sludge—a critical review," *Water Research,* vol. 38, pp. 793-817, 2004.

Yiqi Liu was born in Haikou, China, in 1983. He received the B.S. and M.S. degrees in control engineering from the Chemical University of Technology, Beijing, in 2009 and the Ph.D. degree in control engineering from South China University of Technology, Guangzhou, China, in 2013. From 2013 to 2016, he was a Lecturer with the South China University of Technology. Since 2016, he has been an Associate Professor with Department of Automation, South China University of Technology. He is the author of more than 44 articles. His research interests include soft-sensors, fault diagnosis, and wastewater treatment and holds three patents. He was a recipient of the Hongkong Scholar Award in 2016, Chinese Scholarship Council Award in 2011, and the Deutscher Akademischer Austausch Dienst Award in 2015.

Bin Liu received the B.S. degree from the Department of Automation, Zhejiang University, Zhejiang, China, in 2013. He is currently working toward the Ph.D. degree in the Department of Systems Engineering and Engineering Management, City University of Hong Kong, Kowloon, Hong Kong. His current research interests include reliability and maintenance modeling and importance measures with application to complex systems and data analysis.



Xiujie Zhao received the B.E. degree in Industrial Engineering from Tsinghua University, Beijing, China, in 2013, the M.S. degree in Industrial Engineering from the Pennsylvania State University, University Park, PA, USA, in 2015. He is currently pursuing the Ph.D. degree in Industrial Engineering in the Department of Systems Engineering and Engineering Management, City University of Hong Kong. His research interests include accelerated reliability testing, degradation modeling, maintenance modeling and design of experiments.

Min Xie (M'91–SM'94–F'06) received the Ph.D. degree in quality technology from Linkoping University, Linkoping, Sweden, in 1987.He is currently a Chair Professor at the City University of Hong Kong, Shenzhen, China. He has authored or coauthored numerous refereed journal papers, and some books on quality and reliability engineering, including *Software Reliability Modeling* (World Scientific), *Weibull Models* (Wiley), *Computing Systems Reliability* (Kluwer), and *Advanced QFD Applications* (ASQ Quality Press). He is a Department Editor of *IIE Transactions* and an Associate Editor of *Reliability Engineering and System Safety*. Prof. Xie was awarded the prestigious LKY Research Fellowship in 1991. He is on the Editorial Board of a number other international journals.