Data-Based Linear Gaussian State-Space Model for Dynamic Process Monitoring

Qiaojun Wen, Zhiqiang Ge and Zhihuan Song
State Key Laboratory of Industrial Control Technology, Institute of Industrial Process Control, Zhejiang University, Hangzhou 310027, Zhejiang, P.R. China

DOI 10.1002/aic.13776
Published online March 19, 2012 in Wiley Online Library (wileyonlinelibrary.com).

This article develops a data-based linear Gaussian state-space model for monitoring of dynamic processes under noisy environment. The Kalman filter is introduced for construction of the linear Gaussian state-space model, and an iterative expectation-maximization algorithm is used for model parameters learning. With the incorporation of the dynamic data information, a new fault detection and identification approach is proposed. The feasibility and effectiveness of the two monitoring statistics in the new method are theoretically evaluated and further confirmed through two case studies. Furthermore, detailed fault smearing effect analysis of the proposed method is provided and compared with other identification methods. Based on the simulation results of two case studies, the superiority of the proposed method is explored. © 2012 American Institute of Chemical Engineers *AIChE J,* 58: 3763–3776, 2012
Keywords: probabilistic, dynamic, linear Gaussian state-space model, fault detection, fault identification

Introduction

Nowadays, industrial processes have widely incorporated the distributed control system, based on which an increasing number of data has been collected. Therefore, process modeling and monitoring on data-based methods have become very popular. Particularly, multivariate statistical process control (MSPC) based methods have been researched for many years, such as principal component analysis (PCA), partial least squares, and so forth.[1–8] By projecting the data into a lower-dimensional space, these MSPC methods can accurately characterize the operation state of the monitored process systems.

However, most traditional MSPC-based monitoring models lack probabilistic interpretation of the process data, since they are constructed through the deterministic manner. In fact, almost all process variables are contaminated by random noises. Hence, the monitored process variables inherently perform through a statistical manner, not the deterministic way. Therefore, it is required that process monitoring should also be carried out through the statistical manner, and the monitoring decisions are made through a probabilistic way. Furthermore, in the probabilistic framework, there are several other advantages which will enhance the performance of process monitoring. First, the combination of a probabilistic model and the expectation-maximization (EM) algorithm allows us to deal with missing values in the dataset, which is very practical in process data. Second, it is more sophisticated to extend the probabilistic model to its mixture form, thus can be used for modeling complex processes, for example, nonlinear process, multimode process, and so forth. What is more, the probabilistic model can naturally exploit Bayesian treatment of the model.[9]

Recently, the traditional PCA method has been extended to its probabilistic counterpart, probabilistic principal component analysis (PPCA), and used for process monitoring.[10–12] Compared to the traditional method, more satisfactory performance has been obtained by the PPCA approach. It has been noticed that the developed PPCA-based monitoring method has a restricted assumption that different process variables share the same noise variance. To relax the assumption of PPCA, the process monitoring method based on factor analysis (FA) has also been proposed.[13,14] In this monitoring method, different noise variances of process variables have been assigned, which is more practical in industry. To our best knowledge, however, an important dynamic characteristic of the process data has been ignored by those probabilistic monitoring methods. Actually, dynamic is a common behavior of many industrial process data. For dynamic process monitoring, several research studies have been carried out. For example, the dynamic form of the PCA method,[15–19] the improved PCA method,[20] the state-space-based methods,[21–23] among others.[24–26] However, the aim of this article is to address the dynamic behavior of the process data among the probabilistic monitoring framework.

Precisely, a dynamic probabilistic model named linear Gaussian state-space model is developed for process monitoring in this article. Because of the efficiency of the Kalman filter in dynamic data processing, it is used for construction of the linear Gaussian state-space model. Similar to the conventional PPCA and FA models, an EM algorithm can also be used to learn the linear Gaussian state-space model. Based on the linear Gaussian state-space model, a dynamic fault detection method is developed. Similar to the

Correspondence concerning this article should be addressed to Z. Ge at zqge@iipc.zju.edu.cn, or Z. Song at zhsong@iipc.zju.edu.cn.

traditional method, two monitoring statistics T^2 and the squared prediction error (SPE) are constructed, which correspond to the latent and residual subspaces, respectively. However, with the incorporation of the dynamic data information, the feasibility and effectiveness of both T^2 and SPE statistics should be re-evaluated, which are presented as two remarks in this article.

When a fault has been detected, an efficient fault identification approach is then developed. To locate the exact root cause of the fault, the smearing effect should be considered. Recently, Alcala and Qin[27] have developed a fault identification method named reconstruction-based contribution (RBC), in which the fault smearing effect is analyzed. Compared with the contribution plots and other reconstruction-based methods, the fault reconstruction and identification performance have been improved by the RBC method. However, as has been explored, there is still smearing effect in the RBC approach, which may lead to misdiagnosis of the process fault. Furthermore, under the dynamic monitoring framework, the fault identification becomes more complicated, which means more significant smearing effect may be cause if the traditional RBC method is used. To this end, we present an improved RBC method, which has no smearing effect if the magnitude/type of the fault is independent with the previous observations. Besides, the smearing effect of the RBC method is theoretically analyzed, and also illustrated in detail by the two case studies.

The contributions of this article can be summarized as follows: (1) a data-based linear Gaussian state-space model is developed for probabilistic monitoring of dynamic processes; (2) the feasibility and effectiveness of both T^2 and SPE statistics for dynamic process monitoring are evaluated and further confirmed through two case studies; (3) an efficient fault identification method is proposed, the smearing effect of which is analyzed and compared with the RBC method.

The layout of this article is given as follows. First, the traditional PCA, PPCA, and FA methods are shown in the section entitled PCA and Its Probabilistic Extensions followed by the detailed description of the proposed dynamic process monitoring method in the next section. Then two case studies are shown in the section entitled Case Studies and conclusions are made in the last section.

PCA and Its Probabilistic Extensions

PCA maps the original measurement variables $\mathbf{x} \in \Re^m$ into a lower dimensional latent space as

$$\mathbf{t} = \mathbf{P}^T\mathbf{x} \tag{1}$$

where $\mathbf{t} \in \Re^l$ are scores in the latent space, and $\mathbf{P} = [p_1, p_2, \cdots, p_l] \in \Re^{m\times l}$ is the loading matrix, which can be calculated by implementing eigenvalue decomposition on the covariance matrix of the measurements. The residual vector $\mathbf{v}$ of $\mathbf{x}$ generated by the PCA model is

$$\mathbf{v} = \mathbf{x} - \tilde{\mathbf{x}} = (\mathbf{I} - \mathbf{P}\mathbf{P}^T)\mathbf{x} \tag{2}$$

where $\tilde{\mathbf{x}}$ is the reconstruction value of the sample $\mathbf{x}$. The goal of PCA is to minimize the squared distance between the measurement points and their reconstructions,[9] which is

$$J = \frac{1}{N}\sum_{i=1}^{N} \|\mathbf{x}_i - \tilde{\mathbf{x}}_i\|^2 \tag{3}$$

where N is the number of samples.

PPCA and FA are termed as two probabilistic extensions of the traditional PCA method, both of which concentrate on the latent variables $\mathbf{t}$ whose distributions are Gaussian, and the original measurement variables $\mathbf{x}$ are treated as linear combinations of $\mathbf{t}$ plus the white noise $\mathbf{v}$ imposed on them. The aim of both PPCA and FA is to find the most probable parameter set $\Theta = \{\mathbf{P}, \Sigma\}$ in the model structure, which is described as follows[9]

$$\mathbf{x} = \mathbf{P}\mathbf{t} + \mathbf{v} \tag{4}$$

where $\mathbf{P}$ is the loading matrix, similar to that in the PCA model and $\mathbf{v} \sim N(0, \Sigma)$.

In the FA model, the variance matrix of measurement noise $\mathbf{v}$ are represented by $\Sigma = \mathrm{diag}\{\lambda_i\}_{1,2,\cdots,m}$, in which different noise levels of measurement variables are assumed. If all of λ are assumed the same value, then FA is equivalent to PPCA. If all of λ are assumed to converge to zero, then FA turns to PCA. Therefore, FA is the general description of Gaussian latent model structure while PPCA and PCA are two special cases of FA.[9,26]

Linear Gaussian State-Space Model for Dynamic Process Monitoring

In this section, we firstly derive the linear Gaussian state-space model with the use of Kalman filter and EM Algorithm, then the dynamic process monitoring scheme is proposed based on the developed model, including fault detection and fault identification.

Derivation of linear Gaussian state-space model

The structure of the linear Gaussian state-space model can be formulated as follows[28,29]

$$\begin{aligned}\mathbf{t}(k+1) &= \mathbf{A}\mathbf{t}(k) + \mathbf{w}(k+1)\\ \mathbf{x}(k+1) &= \mathbf{P}\mathbf{t}(k+1) + \mathbf{v}(k+1)\end{aligned} \tag{5}$$

where $k = 1, 2, \cdots, N-1$ is the sample number of process data, $\mathbf{A} \in \Re^{l\times l}$ is the state space matrix, $\mathbf{P} \in \Re^{m\times l}$ is the loading matrix, and $\mathbf{w} \sim N(0, \Gamma)$ and $\mathbf{v} \sim N(0, \Sigma)$ are both Gaussian residuals with zero mean and variance matrices Γ and Σ. Similarly, the distribution of the latent variables $\mathbf{t}$ is also assumed to be Gaussian. However, the mean and variance values of $\mathbf{t}$ are not restricted to zero and one, thus $\mathbf{t} \sim N(\boldsymbol{\mu}, \mathbf{V})$. Particularly, the initial distribution of the latent variable $\mathbf{t}$ is assumed as $\mathbf{t}(1) \sim N(\boldsymbol{\mu}_1, \mathbf{V}_1)$. Therefore, the model parameter set of the linear Gaussian state-space model can be represented as $\Theta = \{\mathbf{A}, \mathbf{P}, \Sigma, \Gamma, \boldsymbol{\mu}_1, \mathbf{V}_1\}$.

The state variables, $\mathbf{t}$, which are first-order Markov random processes, capture the system dynamics. If the state transition matrix $\mathbf{A}$ of system (5) is set to be zero, then the linear Gaussian state-space model degrades to a static model.[28] In static models such as PCA, PPCA, and FA, since the process measurements are correlated with each other, it is assumed that the process can be explained by latent variables whose dimensionality is lower than the measurements. Similarly, in the linear Gaussian state-space model, the process information can be captured by lower dimensional state variables. Therefore, $l<m$ and the estimation of state variables is also a procedure of dimensionality reduction.

The outputs of model (5), $\mathbf{x}$, are measurements of the system while $\mathbf{t}$ are hidden states. Given the model parameters Θ and measurements $\mathbf{x}(1)$, $\mathbf{x}(2)$, $\cdots$, $\mathbf{x}(k)$, then the hidden states $\mathbf{t}(1)$, $\mathbf{t}(2)$, $\cdots$, $\mathbf{t}(k)$, as well as the later measurements

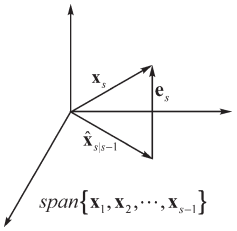

Figure 1. The geometrical illustration of Projection Theorem.

$\mathbf{x}(k+1)$, $\mathbf{x}(k+2)$, $\cdots$, can be estimated or predicted via the Kalman filter. Because the distribution of noises $\mathbf{w}$, $\mathbf{v}$ and the initial distribution of the latent variable $\mathbf{t}$ are all Gaussian, then the model outputs, as well as the hidden states, are Gaussian.

Depending on the maximum likelihood method, the linear Gaussian state-space model can be identified through an EM algorithm. Given the initial distribution of the latent variable, the conditional distributions of the predicted latent variable, and the original variable can be easily formulated, which also follow Gaussian distributions, namely $p(\mathbf{t}_{k+1}|\mathbf{t}_k) = N(\mathbf{A}\mathbf{t}_k, \Gamma)$, $p(\mathbf{x}_{k+1}|\mathbf{t}_{k+1}) = N(\mathbf{P}\mathbf{t}_{k+1}, \Sigma)$. Therefore the log likelihood function of the process data can be given as follows[9]

$$\ln p(\mathbf{X}, \mathbf{T}|\Theta) = \sum_{k=1}^{n} \ln p(\mathbf{x}_k|\mathbf{t}_k, \mathbf{P}, \Sigma) + \sum_{k=2}^{n} \ln p(\mathbf{t}_k|\mathbf{t}_{k-1}, \mathbf{A}, \Gamma) + \ln p(\mathbf{t}_1|\boldsymbol{\mu}_1, \mathbf{V}_1) \quad (6)$$

where $\mathbf{X}$ and $\mathbf{T}$ are datasets of the original and latent variables, respectively. To maximize the above complete-data log likelihood, the model parameter set Θ can be estimated through an iterative manner.

In the E-step, the posterior distribution of the latent variable should be determined. Different from the traditional FA model, one more distribution is needed in the linear Gaussian state-space model, which is the dynamic relationship between latent variables. Using the Kalman filter equations, the expected values of the three statistics can be calculated as follows

$$E(\mathbf{t}_{k+1}) = \boldsymbol{\mu}_{k+1} = \mathbf{A}\boldsymbol{\mu}_k + \mathbf{K}_{k+1}(\mathbf{x}_{k+1} - \mathbf{P}\mathbf{A}\boldsymbol{\mu}_k) \quad (7)$$

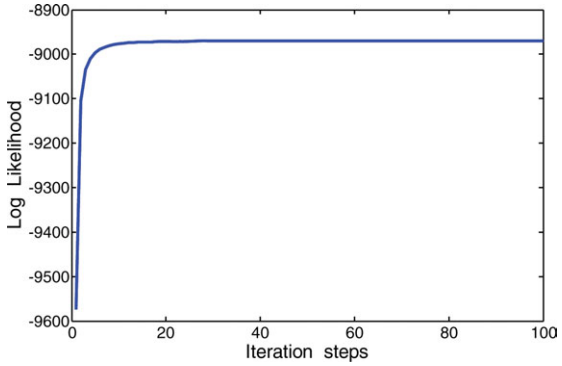

Figure 2. Likelihood value change of the iteration step.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

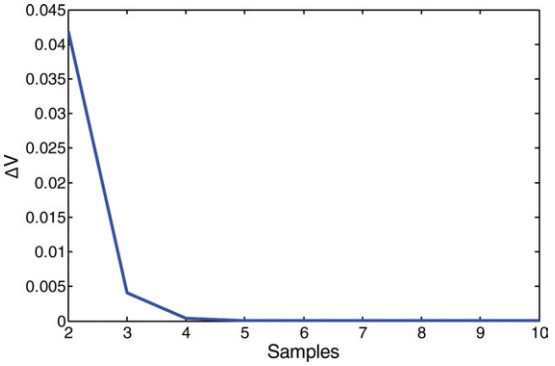

Figure 3. Convergence with the covariance of state variables.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

$$E(\mathbf{t}_{k+1}\mathbf{t}_k^T) = \mathbf{V}_k\mathbf{A}^T(\mathbf{Q}_k)^{-1}\mathbf{V}_{k+1} + \boldsymbol{\mu}_{k+1}\boldsymbol{\mu}_k^T \quad (8)$$

$$E(\mathbf{t}_{k+1}\mathbf{t}_{k+1}^T) = \mathbf{V}_{k+1} + \boldsymbol{\mu}_{k+1}\boldsymbol{\mu}_{k+1}^T \quad (9)$$

where $\mathbf{Q}_k = \mathbf{A}\mathbf{V}_k\mathbf{A}^T + \Gamma$, the updating of the variance $\mathbf{V}_{k+1}$ and the Kalman gain matrix $\mathbf{K}_{k+1}$ are given as

$$\mathbf{K}_{k+1} = \mathbf{Q}_k\mathbf{P}^T(\mathbf{P}\mathbf{Q}_k\mathbf{P}^T + \Sigma)^{-1} \quad (10)$$

$$\mathbf{V}_{k+1} = (\mathbf{I} - \mathbf{K}_{k+1}\mathbf{P})\mathbf{Q}_k \quad (11)$$

In the M-step, depending on the results calculated in the E-step, the parameter set can be re-estimated by maximizing the log likelihood function, which can be calculated as follows

$$\boldsymbol{\mu}_1^{\text{new}} = E(\mathbf{t}_1) \quad (12)$$

$$\mathbf{V}_1^{\text{new}} = E(\mathbf{t}_1\mathbf{t}_1^T) - E(\mathbf{t}_1)E^T(\mathbf{t}_1) \quad (13)$$

$$\mathbf{A}^{\text{new}} = [\sum_{k=1}^{n-1} E(\mathbf{t}_{k+1}\mathbf{t}_k^T)][\sum_{k=1}^{n-1} E(\mathbf{t}_k\mathbf{t}_k^T)]^{-1} \quad (14)$$

$$\mathbf{P}^{\text{new}} = \left[\sum_{k=1}^{n} \mathbf{x}_k E^T(\mathbf{t}_k)\right]\left[\sum_{k=1}^{n} E(\mathbf{t}_k\mathbf{t}_k^T)\right]^{-1} \quad (15)$$

$$\Gamma^{\text{new}} = \frac{1}{n-1}\sum_{k=1}^{n-1}\{E(\mathbf{t}_{k+1}\mathbf{t}_{k+1}^T) - \mathbf{A}^{\text{new}}E(\mathbf{t}_k\mathbf{t}_{k+1}^T) - E(\mathbf{t}_{k+1}\mathbf{t}_k^T)(\mathbf{A}^{\text{new}})^T + \mathbf{A}^{\text{new}}E(\mathbf{t}_k\mathbf{t}_k^T)(\mathbf{A}^{\text{new}})^T\} \quad (16)$$

$$\Sigma^{\text{new}} = \frac{1}{n}\sum_{k=1}^{n}\{\mathbf{x}_k\mathbf{x}_k^T - \mathbf{P}^{\text{new}}E(\mathbf{t}_k)\mathbf{x}_k^T - \mathbf{x}_k E^T(\mathbf{t}_k)(\mathbf{P}^{\text{new}})^T + \mathbf{P}^{\text{new}}E(\mathbf{t}_k\mathbf{t}_k^T)(\mathbf{P}^{\text{new}})^T\} \quad (17)$$

Detailed calculations of Eqs. 12–17 are provided in Appendix. The maximum value of the likelihood can be obtained by carrying out the E-step and the M-step iteratively until all of them are converged.

There are several methods for choosing the dimensionality of the states, l. For instance, l can be chosen according to the average variance explanation of the states, which is similar as the approach mentioned in PPCA modeling.[10] Alternatively, a more simple way is to predetermine a proper dimensionality via PCA modeling, and then adopt the dimensionality in linear

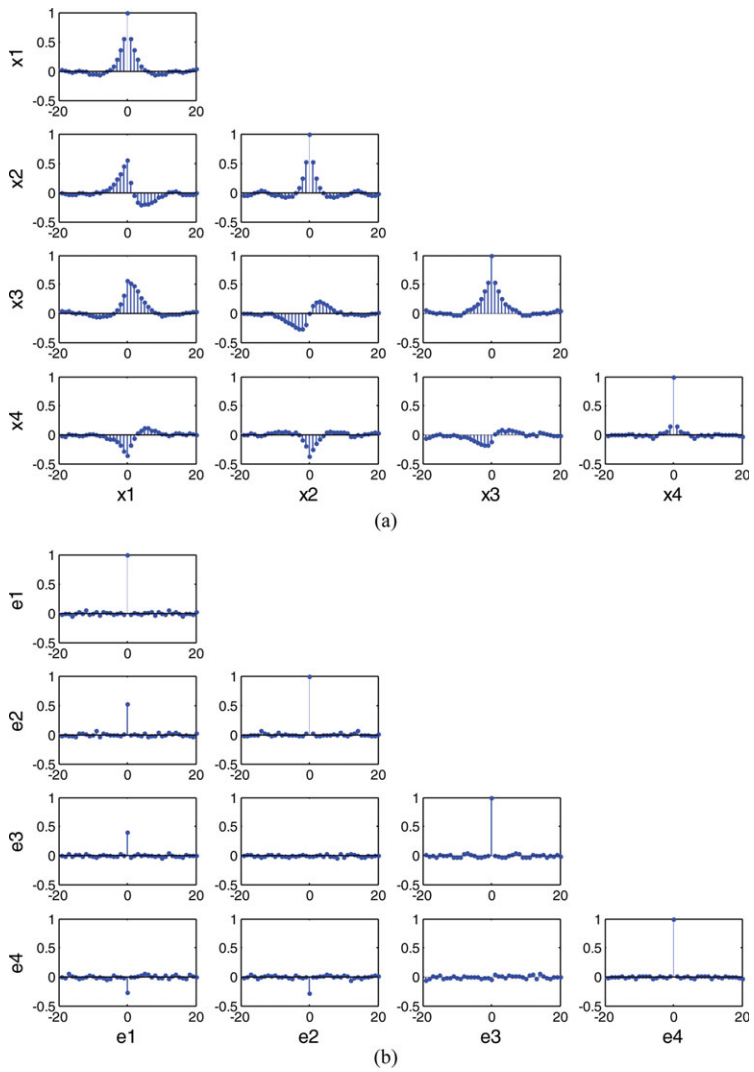

Figure 4. Correlation plots before/after the implementation of the linear Gaussian state-space model.

(a) Process variables; (b) prediction errors. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Gaussian state-space models. In this article, the latter method is used for linear Gaussian state-space modeling.

Fault detection based on linear Gaussian state-space model

Based on the developed linear Gaussian state-space model, a corresponding dynamic fault detection scheme can be formulated. Two statistics, T^2 and SPE are developed here. While the SPE statistic is used for monitoring the prediction error of the linear Gaussian state-space model, the T^2 statistic is used to monitor the variation of the states in the linear Gaussian state-space model.Before the construction of T^2 statistic, the stationarity of state variables should be verified first. Only if the state variables are stationary, the efficiency of T^2 statistics can be guaranteed. Detailed illustration of the stationary property of the state variables is given in Remark 1 as follows.

Table 1. False Alarm Rates in the Numerical Process

Methods/Statistics	T^2	SPE
FA	0.004	0.009
PPCA	0.003	0.01
PCA	0.007	0.005
LGSS	0.007	0.006

Remark 1. *Consider system (5), and suppose that at the initial time* $k = 1$, $\mathbf{t}(1) \sim N(\boldsymbol{\mu}_1, \mathbf{V}_1)$. *Suppose that all eigenvalues of matrix* $\mathbf{A}$ *lie in the unit circle, which means* $|\lambda_i(\mathbf{A})|<1$ *for* $i = 1, 2, \cdots, l$. *Then when* $k \to \infty$, $\{\mathbf{t}_k\}$ *is a stationary process of zero mean and covariance*

$$E[\mathbf{t}_k \mathbf{t}_l^T] = \begin{cases} \mathbf{A}^{k-j}\bar{\mathbf{V}}, k \geq j \\ \bar{\mathbf{V}}(\mathbf{A}^T)^{j-k}, j > k \end{cases} \tag{18}$$

where $\bar{\mathbf{V}}$ is the unique solution of

$$\bar{\mathbf{V}} - \mathbf{A}\mathbf{V}\mathbf{A}^T = \Gamma \tag{19}$$

Furthermore, if $\boldsymbol{\mu}_1 = 0$ *and* $\mathbf{V}_1 = \bar{\mathbf{V}}$, *then* $\{\mathbf{t}_k\}$ *for* $k \geq 1$ *is stationary and have covariance as above.*

Remark 1 comes from the properties of linear time-invariance Kalman filter and it gives a sufficient condition for stationarity, namely, $|\lambda_i(\mathbf{A})|<1$. Anderson and Moore gave a proof for this stationary property of Kalman filter.[28] Let $k = l$, Eq. 18 can be written as

Table 2. Missed Detection Rates in the Numerical Process

	Fault 1		Fault 2		Fault 3		Fault 4		Fault 5	
Methods/Faults	T^2	SPE	T^2	SPE	T^2	SPE	T^2	SPE	T^2	SPE
FA	0.567	**0.023**	0.404	**0.226**	0.943	0.616	0.881	0.988	0.487	0.929
PPCA	0.854	0.06	0.442	0.301	0.983	0.77	0.928	0.948	0.677	**0.818**
PCA	0.599	0.178	**0.332**	0.356	0.951	0.816	0.871	0.974	0.365	0.961
LGSS	**0.092**	0.058	0.365	0.229	**0.937**	**0.581**	**0.816**	**0.926**	**0.249**	0.94

$$E[\mathbf{t}_k\mathbf{t}_k^T] = \bar{\mathbf{V}} \tag{20}$$

Remark 1 shows the stationary property of the state variables when $k \to \infty$. Actually, in the following simulation experiments, the covariance of $\mathbf{t}_k$ will converge to $\mathbf{V}$ in several steps for arbitrarily finite initial mean and covariance. To show the convergence of the state variables covariance, an index for the variance change is defined as following using the Frobenius norm of matrix

$$\Delta V(k) = \|\mathbf{V}_k - \mathbf{V}_{k-1}\|_F = (tr[(\mathbf{V}_k - \mathbf{V}_{k-1})^T(\mathbf{V}_k - \mathbf{V}_{k-1})])^{\frac{1}{2}} \tag{21}$$

where $\|\bullet\|_F$ denotes the Frobenius norm operator and tr($\bullet$) denotes the matrix trace operator. If for arbitrarily small positive number $\varepsilon > 0$, $\Delta V(k) < \varepsilon$ for any $k \geq M$, then the covariance matrix is said to converge before M steps.

Therefore, based on the stationary property of the state variables, the T^2 statistic can be constructed for process monitoring. Denote the new data sample as $\mathbf{x}_s$, and its previous one as $\mathbf{x}_{s-1}$, the expected mean value of the latent variable of $\mathbf{x}_s$can be predicted as

$$\mathbf{Q}_{s-1} = \mathbf{A}\mathbf{V}_{s-1}\mathbf{A}^T + \Gamma \tag{22}$$

$$\mathbf{K}_s = \mathbf{Q}_{s-1}\mathbf{P}^T(\mathbf{P}\mathbf{Q}_{s-1}\mathbf{P}^T + \Sigma)^{-1} \tag{23}$$

$$\boldsymbol{\mu}_s = E(\mathbf{t}_s) = \mathbf{A}\boldsymbol{\mu}_{s-1} + \mathbf{K}_s(\mathbf{x}_s - \mathbf{P}\mathbf{A}\boldsymbol{\mu}_{s-1}) \tag{24}$$

The T^2 monitoring statistic can be constructed as follows

$$T_s^2 = \boldsymbol{\mu}_s^T[\mathrm{var}(\mathbf{T})]^{-1}\boldsymbol{\mu}_s \tag{25}$$

where $\mathbf{T}$ is the dataset of latent variables with the training data and var($\mathbf{T}$) represents the variance of the latent variable in the training dataset.

To construct the SPE monitoring statistic, both of the state transition noise $\mathbf{w}_s$ and the observation noise $\mathbf{v}_s$ should be

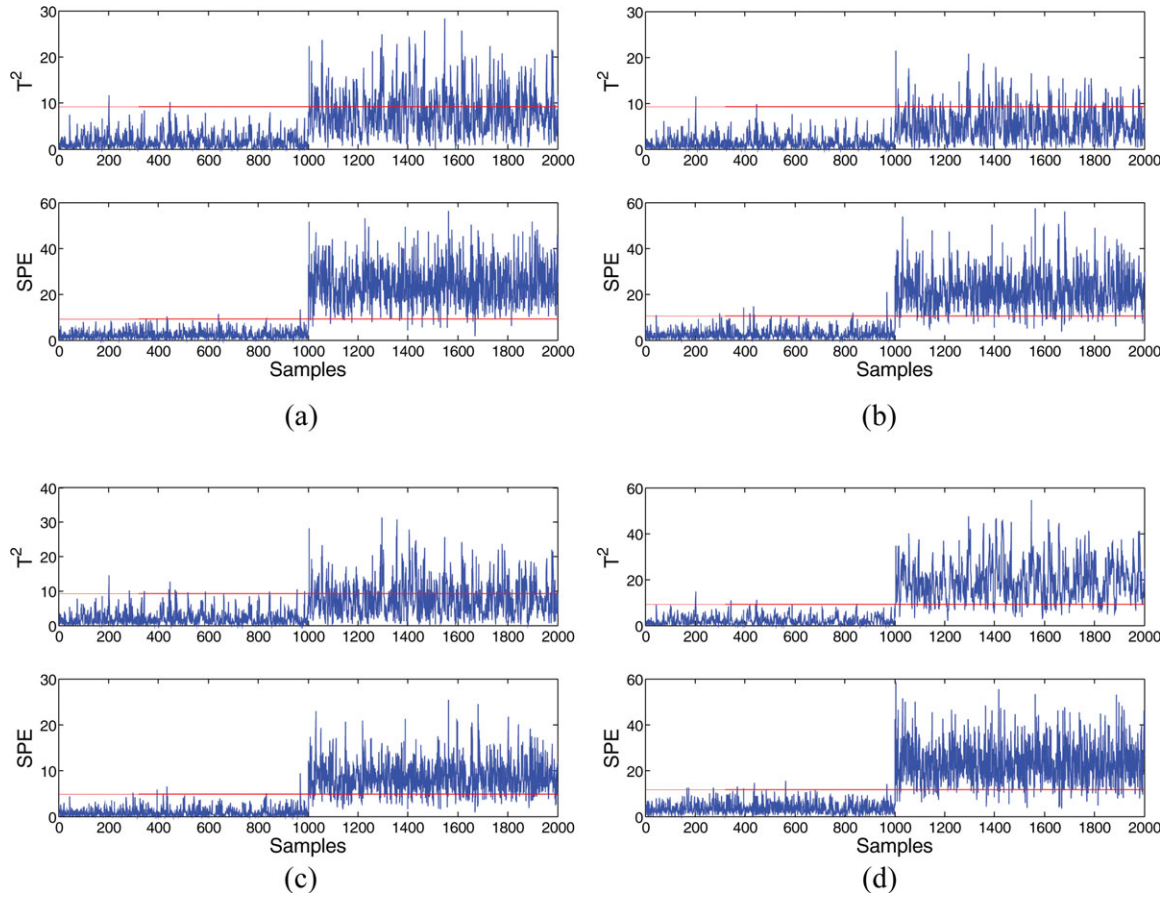

Figure 5. Monitoring performance with the first testing dataset.

(a) FA; (b) PPCA; (c) PCA; and (d) LGSSM. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary. com.]

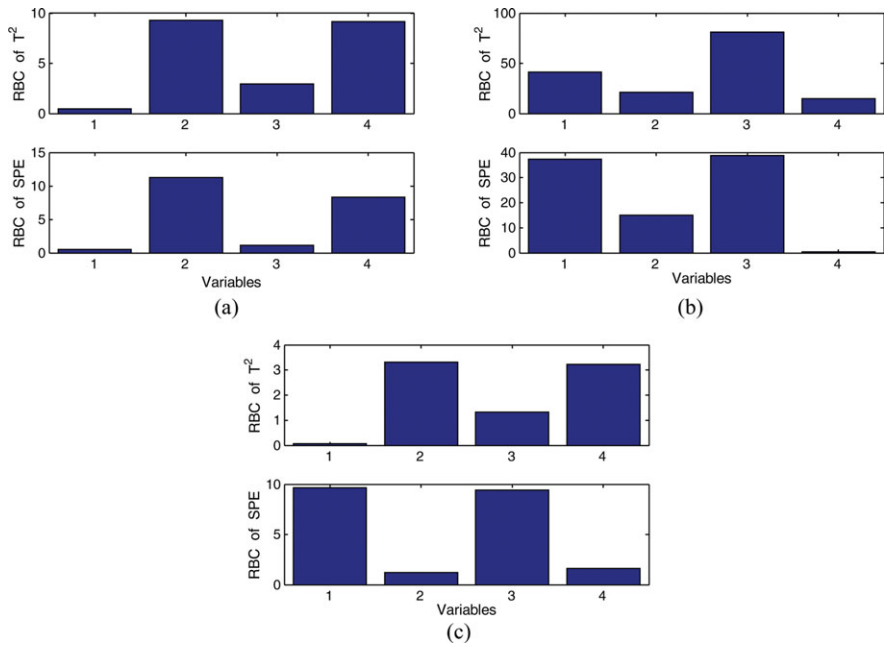

Figure 6. Monitoring performance of RBC fault identification of PCA.

(a) fault 1; (b) fault 2; and (c) fault 3. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

considered. However, since both of $\mathbf{w}_s$ and $\mathbf{v}_s$ influence the estimation and prediction, an alternative way for SPE construction is to combine the two kinds of noises together. Given the distribution of $\mathbf{t}_{s-1|s-1}$

$$\mathbf{t}_{s-1|s-1} \sim N(\boldsymbol{\mu}_{s-1}, \mathbf{V}_{s-1}) \tag{26}$$

the Gaussian parameters of both $\mathbf{t}_{s|s-1}$ and $\mathbf{x}_{s|s-1}$ can be calculated as follows

$$\mathbf{t}_{s|s-1} \sim N(\mathbf{A}\boldsymbol{\mu}_{s-1}, \mathbf{A}\mathbf{V}_{s-1}\mathbf{A}^T + \Gamma) \tag{27}$$

$$\mathbf{x}_{s|s-1} \sim N(\mathbf{P}\mathbf{A}\boldsymbol{\mu}_{s-1}, \mathbf{P}(\mathbf{A}\mathbf{V}_{s-1}\mathbf{A}^T + \Gamma)\mathbf{P}^T + \Sigma) \tag{28}$$

Denote $\hat{\mathbf{x}}_{s|s-1} = \mathbf{P}\mathbf{A}\boldsymbol{\mu}_{s-1}$ and $\Phi_{s|s-1} = \mathbf{P}\mathbf{A}\mathbf{V}_{s-1}\mathbf{A}^T\mathbf{P}^T + \mathbf{P}\Gamma\mathbf{P}^T + \Sigma$. The one-step prediction of the observation can be written as

$$\mathbf{x}_{s|s-1} \sim N(\hat{\mathbf{x}}_{s|s-1}, \Phi_{s|s-1}) \tag{29}$$

Hence, when the new measurement $\mathbf{x}_s$ is obtained from the process, its prediction error can be calculated as follows

$$\mathbf{e}_s = \mathbf{x}_s - \hat{\mathbf{x}}_{s|s-1} \sim N(\mathbf{0}, \Phi_{k|k-1}) \tag{30}$$

Note that $\mathbf{e}_s = \mathbf{x}_s - \hat{\mathbf{x}}_{s|s-1}$ is the innovation of the Kalman filter.The following Remark 2 shows that the prediction error

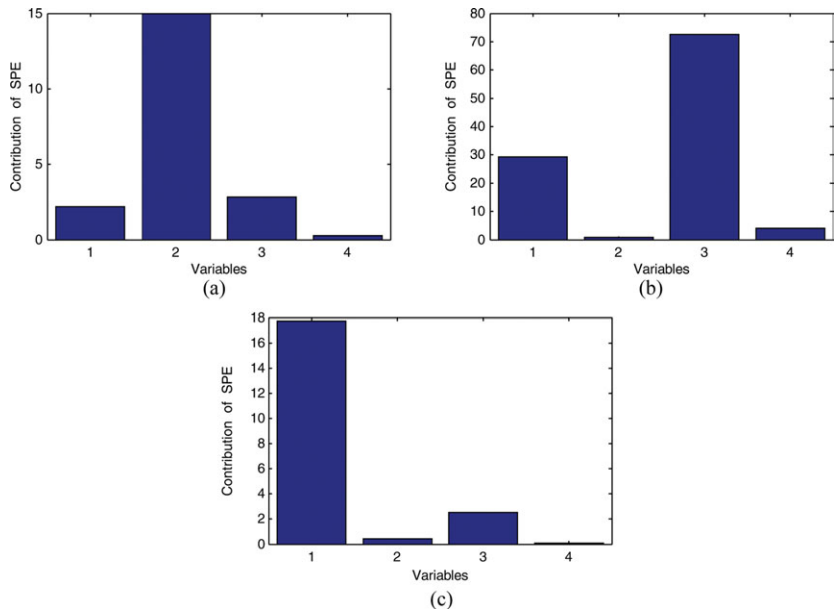

Figure 7. Monitoring performance of fault identification of LGSSM.

(a) fault 1; (b) fault 2; and (c) fault 3. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

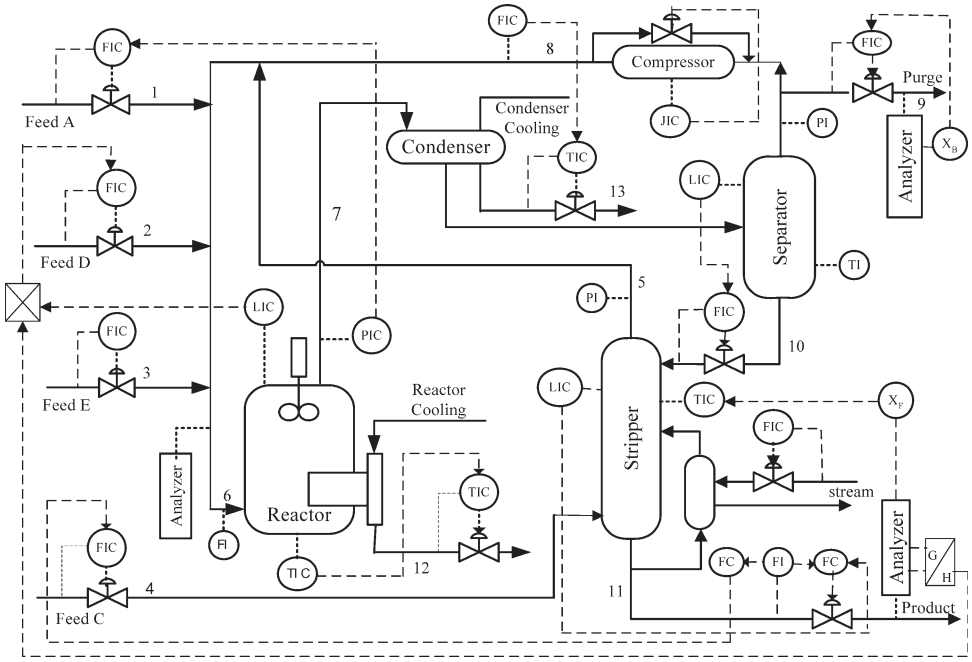

Figure 8. TE benchmark process.

$\mathbf{e}_s$ is independent with the previous observations, which means that $\mathbf{e}_s$ only has correlation with the current observation $\mathbf{x}_s$. Therefore, if there is a fault which is independent with the previous observations, the SPE statistic that is based on $\mathbf{e}_s$ will extract the correct fault information without smearing from previous measurements.

Remark 2. *The prediction error of the linear Gaussian state-space model,* $\mathbf{e}_s$*, is temporally white.*

The proof of this remark is based on the Projection Theorem of Kalman filter.[30] As the linear Gaussian state-space model is actually a linear time-invariant Kalman filter with Gaussian noises, the Projection Theorem still holds in the linear Gaussian state-space model. According to the Projection Theorem, $E[\mathbf{e}_s\mathbf{x}_i^T]=\mathbf{0}$, for $i=1,2,\cdots,s-1$, and $E[\mathbf{e}_s\hat{\mathbf{x}}_{s|s-1}^T]=\mathbf{0}$, $E[\mathbf{e}_s\mathbf{e}_k^T]=\mathbf{0}$. The geometrical illustration of the Projection Theorem is shown in Figure 1. As $\hat{\mathbf{x}}_{s|s-1}$ is the linear estimation based on $\mathbf{x}_1,\mathbf{x}_2,\cdots,\mathbf{x}_{s-1}$, thus $\hat{\mathbf{x}}_{s|s-1}$ lies in the space spanned by $\mathbf{x}_1,\mathbf{x}_2,\cdots,\mathbf{x}_{s-1}$.

The SPE monitoring index is given as follows

$$\mathrm{SPE}_s=\mathbf{e}_s^T\Phi_{s|s-1}^{-1}\mathbf{e}_s \tag{31}$$

As both of the two monitoring statistics are constructed by the Mahalanobis distance, the squared value of which follows χ^2 distribution with appropriate dimension of freedom. Therefore, the control limits of the T^2 and SPE statistics can both be determined by the χ^2 distribution, thus[10]

$$T^2\le T_{\lim}^2=\chi_\alpha^2(l) \tag{32}$$

$$\mathrm{SPE}\le SPE_{\lim}=\chi_\alpha^2(m) \tag{33}$$

where l is the number of states in the linear Gaussian state-space model, m is the number of process variables, and α is the significance level of both monitoring statistics.

For online process monitoring, it can be noticed that there are several matrix inversions; however, it is actually not necessary to calculate matrix inversions online. From the theory of Kalman filter, these matrices which are need to calculate the matrix inversions are independent with the process observations. As a result, the computation of the matrix inversion can be calculated offline and stored for online utilization. Moreover, according to Eqs. 22–31, if the covariance of state variables is known, then all the matrix inversions needed for online monitoring can be calculated offline. Based on Remark 1, it can be known that the covariance of states will converge in several steps. Therefore, only a few matrix inversions are needed for offline computation. The following case studies will show the property of covariance convergence of state variables.

Table 3. Monitoring Variables in the TE Process

No.	Measured Variables	No.	Measured Variables
1	A feed	9	Product separator temperature
2	D feed	10	Product separator pressure
3	E feed	11	Product separator underflow
4	A and C feed	12	Stripper pressure
5	Recycle flow	13	Stripper temperature
6	Reactor feed rate	14	Stripper steam flow
7	Reactor temperature	15	Reactor cooling water outlet temperature
8	Purge rate	16	Separator cooling water outlet temperature

Fault identification

Following the above fault detection method, a sequent procedure is to carry out fault identification, which is also an important issue for process monitoring. Previously, the contribution plots are the most popular approaches to locate the faulty variables.[31] The PCA-based contribution plots decompose the T^2 and SPE statistics into the sum of m contribution terms for the m measurement variables and compare the magnitude of these terms according to both statistics. Later, the reconstruction-based methods have been proposed for fault identification.[32,33] More recently, Alcala and Qin[27] proposed a RBC method to identify faulty variables and

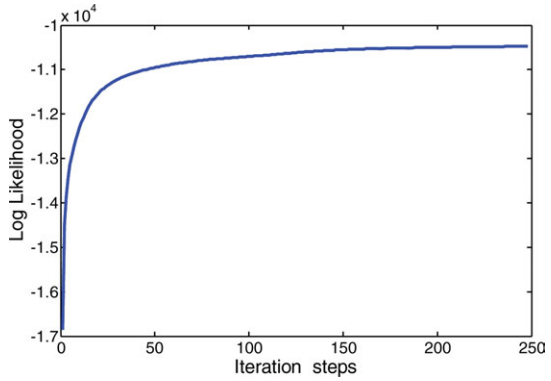

Figure 9. Likelihood value change of the iteration step in TE process.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

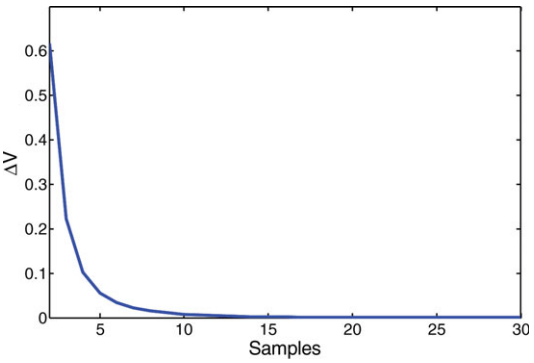

Figure 10. Convergence with the covariance of state variables in TE process.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

compared it with the contribution plots. Chen and Sun[34] extended the reconstruction methods in a probabilistic framework and proposed a missing variable approach.

Particularly, Alcala and Qin have proven that the contribution plot methods fail to guarantee the isolation of the correct faulty variables. They have shown that the RBC approach outperforms the contribution plots and other previously reconstruction-based identification methods. However, Alcala and Qin also explored that there is still smearing effect in the RBC approach, which may lead to misdiagnosis of the process fault. Furthermore, Liu[35] showed that the RBC approach fails to locate faulty variables when the fault is related to multiple sensors in the process.

In this article, a novel fault identification method based on the SPE statistic is proposed under the dynamic monitoring framework. To utilize the process dynamic information, this fault identification approach incorporates one-step prediction of the Kalman filter. According to Remark 2, the prediction error $\mathbf{e}_s$ of the linear Gaussian state-space model is orthogonal to previous observations, therefore, the SPE statistic only involves current information in the process data. Based on the contribution analysis of the SPE statistic, the responsible faulty variables can be located.

Denote ϕ_{sii} as the ith diagonal variable of $\Phi_{s|s-1}$, and e_s^i as the ith variable of vector $\mathbf{e}_s$, the contribution of the ith variable can be calculated as

$$C_{si}^{\mathrm{SPE}} = \phi_{sii}^{-1} e_s^{i^2} \tag{34}$$

As $E[e_s^{i^2}] = \phi_{sii}$ is within normal condition, $C_{si}^{\mathrm{SPE}} \sim \chi^2(1)$ is hold for any, i=1,2,...,m, we can get the statistical control limit of C_{si}^{SPE} as $\gamma_i^2 = \chi_\alpha^2(1)$. When the value of contribution index exceeds the control limit, the corresponding variable is regarded as a faulty variable for the abnormal event. Those variables which have large values of the contribution index should be considered as the most responsible ones for the process fault.

Table 4. Process Fault Description

Fault Number	Process Description	Type
1	A/C Feed ratio, B composition constant (stream 4)	Step
2	B Composition, A/C ratio constant (stream 4)	Step
3	D Feed temperature (stream 2)	Step
4	Reactor cooling water inlet temperature	Step
5	Condenser cooling water inlet temperature	Step
6	A Feed loss (stream 1)	Step
7	C Header pressure loss—reduced availability (stream 4)	Step
8	A, B, C Feed composition (stream 4)	Random variation
9	D Feed temperature (stream 2)	Random variation
10	C Feed temperature (stream 4)	Random variation
11	Reactor cooling water inlet temperature	Random variation
12	Condenser cooling water inlet temperature	Random variation
13	Reaction kinetics	Slow drift
14	Reactor cooling water valve	Sticking
15	Condenser cooling water valve	Sticking
16–20	Unknown	Unknown
21	Valve of stream 4	Fixed

Remark 3. *If the covariance matrix of the prediction error, $\Phi_{s|s-1}$, is diagonal, we will have the following property*

$$\mathrm{SPE}_s = \sum_{i=1}^{m} C_{si}^{\mathrm{SPE}} \tag{35}$$

However, since the variables of the prediction error are usually correlated to each other, which means that $\Phi_{s|s-1}$ is usually not a diagonal matrix, SPE_s is not equal to the sum of C_{si}^{SPE}.

Then, we will show the properties on fault smearing effect in the linear Gaussian state-space model. If the ith variable of $\mathbf{x}_s$, x_{si}, has a fault with magnitude f, we can write $\mathbf{x}_s$ as

$$\mathbf{x}_s = \bar{\mathbf{x}}_s + \Xi_i f \tag{36}$$

where $\bar{\mathbf{x}}_s$ is the fault free portion and Ξ_i is the ith column of an identity matrix with dimensionality m, which represents the direction of the fault with the ith variable. We can also consider the multiple faults in a similar way, in this case, Ξ is a matrix and f is a vector the elements of which represent different fault magnitudes in various directions.

Theorem 1. *If in Eq. 36, the magnitude f of process fault is independent with the previous process observations*

Table 5. Missed Detection Rates in the TE Process

	FA		PPCA		PCA		LGSS	
Fault Number	T^2	SPE	T^2	SPE	T^2	SPE	T^2	SPE
1	0.0025	0.0012	0.01	0.0025	0.0088	0.0025	**0.0025**	**0.0012**
2	0.02	0.0162	0.0337	0.015	0.03	0.015	0.0212	**0.015**
3	0.9987	0.9813	0.9987	0.965	0.9862	0.9825	0.985	0.9775
4	1	0.995	1	0.9738	0.9937	0.9813	0.9813	0.9887
5	0.77	0.7925	0.79	0.7662	0.765	0.8612	**0.7288**	0.895
6	0	0	0	0	0	0	**0**	**0**
7	0.6225	0.72	0.675	0.6887	0.6287	0.7875	**0.5725**	0.8012
8	0.0238	0.0238	0.065	0.0288	0.0337	0.08	**0.02**	0.0325
9	0.99	0.9712	0.9963	0.96	0.9813	0.98	0.9787	0.9775
10	0.3575	0.2025	0.8375	0.615	0.6838	0.7388	**0.185**	**0.1925**
11	0.9837	0.5175	0.955	0.4938	0.8588	0.5312	0.9187	0.5413
12	0.0138	0.0525	0.0275	0.0212	0.0162	0.1488	**0.0012**	0.0463
13	0.055	0.0487	0.0663	0.0475	0.0587	0.0637	**0.055**	0.0525
14	0.3325	0	0.0725	0	0.0112	0	0.395	**0**
15	0.9925	0.9738	1	0.96	0.9762	0.9738	**0.9288**	0.985
16	0.755	0.725	0.905	0.8362	0.7575	0.9187	**0.4163**	0.7588
17	0.2738	0.0288	0.2425	0.0325	0.18	0.0437	**0.0325**	**0.0238**
18	0.1012	0.0987	0.1112	0.095	0.1075	0.0975	**0.0975**	**0.0937**
19	1	0.5137	0.9787	0.8975	0.8238	0.96	0.9538	**0.3712**
20	0.4938	0.4025	0.895	0.595	0.78	0.6587	**0.0975**	**0.145**
21	0.5375	0.615	0.7137	0.5513	0.6325	0.6088	**0.42**	0.7575

$\{\mathbf{x}_k\}_{k=1,2,\cdots,s-1}$, *then there is no smearing effect in fault identification of the proposed method as Eq.* 34.

Proof. From Eqs. 30 and 36

$$\mathbf{e}_s = \mathbf{x}_s - \hat{\mathbf{x}}_{s|s-1} = \mathbf{x}_s - \hat{\mathbf{x}}_{s|s-1} + \Xi_i f \tag{37}$$

The jth variable of $\mathbf{e}_s$ is

$$e_s^j = \begin{cases} \bar{x}_s^j - \hat{x}_{s|s-1}^j, j \neq i \\ \bar{x}_s^j - \hat{x}_{s|s-1}^j + f, j = i \end{cases} \tag{38}$$

Since $E(\bar{x}_s^j f) = 0, if j \neq i$ and f is independent with the previous process observations $\{\mathbf{x}_k\}_{k=1,2,\cdots,s-1}$, which means $E(\bar{x}_{s|s-1}^j f) = 0$. Thus, we have

$$E(e_s^j f) = 0, if j \neq i \tag{39}$$

Since from Eq. 34, $C_{si}^{\mathrm{SPE}}$is only dependent on e_s^i. Therefore, there is no smearing effect in the proposed fault identification approach.

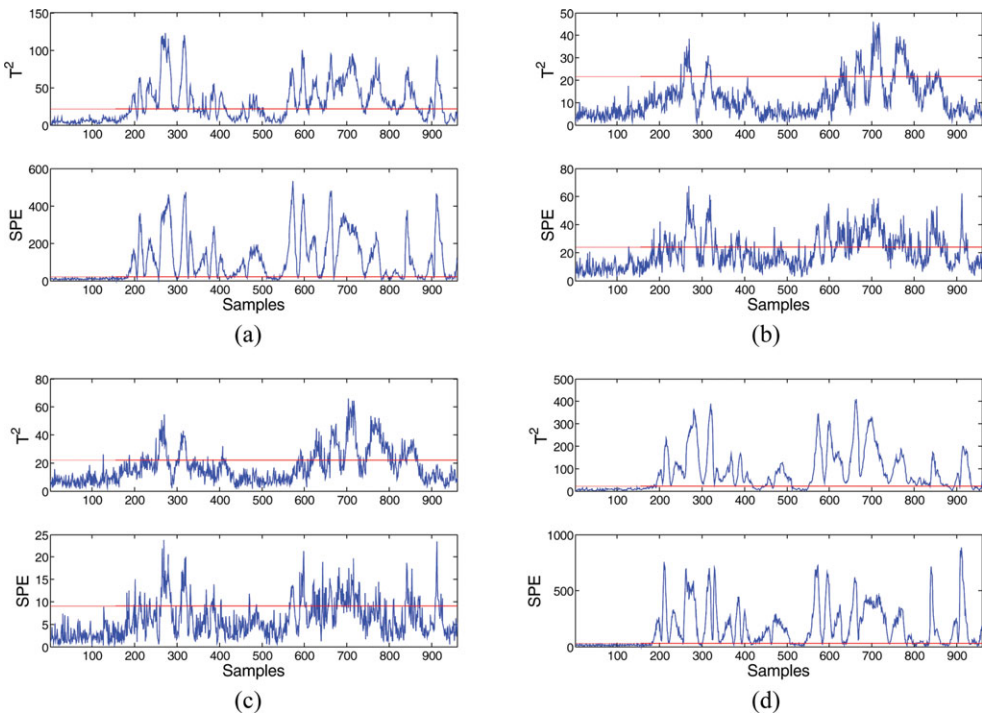

Figure 11. Monitoring performance with fault 10 of TE process.

(a) FA; (b) PPCA; (c) PCA; and (d) LGSSM. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

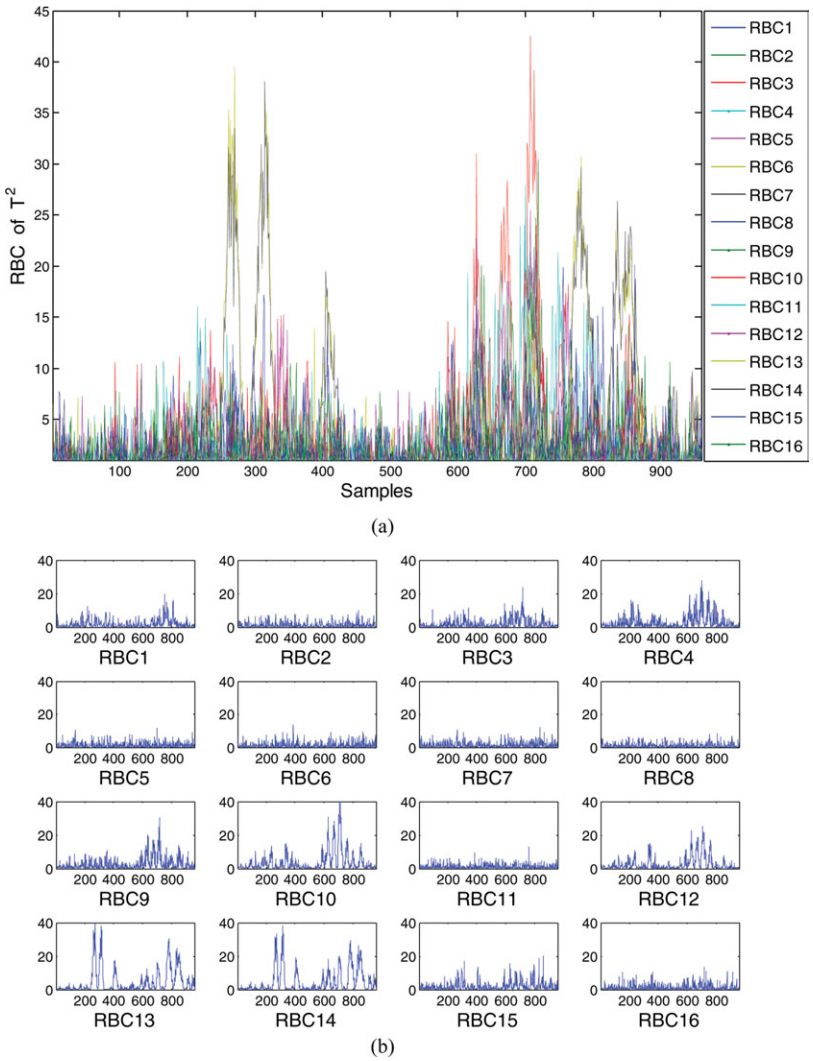

Figure 12. RBC-based fault identification results with T^2 statistics of PCA.

(a) Compared results; (b) specific results. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Remark 4. *If the RBC method is used for fault reconstruction, we can also define a RBC as follows*

$$\min_{f_i} \mathrm{SPE}_s^i = (\mathbf{e}_s + \Xi_i f_i)^T \Phi_{s|s-1}^{-1} (\mathbf{e}_s + \Xi_i f_i) \tag{40}$$

Then calculate

$$RBC_s^i = \|\Xi_i f_i\|^2_{\Phi^{-1}_{s|s-1}} = (\Xi_i f_i)^T \Phi_{s|s-1}^{-1} \Xi_i f_i \tag{41}$$

as the RBC corresponding to the i*th variable. If $\Phi_{s|s-1}$ is a diagonal matrix, this contribution index is the same as Eq. 34. However, since $\Phi_{s|s-1}$ is usually not a diagonal matrix, the reconstruction-based indices defined in Eq. 41 will lead to fault smearing.*

Case Studies

In this section, two case studies are demonstrated to evaluate the efficiency of the proposed monitoring model. The first one is a numerical example, the other one is the Tennessee Eastman (TE) benchmark process.

Numerical example

The numerical example is constructed by the following equations

$$\begin{aligned} \mathbf{z}(k+1) &= \mathbf{B}\mathbf{z}(k) + \xi(k+1) \\ \mathbf{y}(k+1) &= \mathbf{C}\mathbf{z}(k+1) + \zeta(k+1) \end{aligned} \tag{42}$$

where $k = 1, 2, \cdots, n-1$ is the sample number of process data, $\mathbf{z}(k) \in \Re^2$ is the Gaussian distributed latent variable, and $\mathbf{y}(k) \in \Re^4$ is the measured process variables. $\xi = [\,\xi_1 \quad \xi_2\,]^T \in \Re^2$ and $\zeta \in \Re^4$ are Gaussian residuals with zero mean and variance 0.01. The state transfer matrix and the observation matrix are given as

$$\mathbf{B} = \begin{bmatrix} 0.8 & -0.1 \\ 0.4 & 0.6 \end{bmatrix}$$

$$\mathbf{C} = \begin{bmatrix} -0.398 & -1.446 & 0.443 & 0.233 \\ 1.382 & 1.114 & 0.662 & -0.304 \end{bmatrix}^T$$

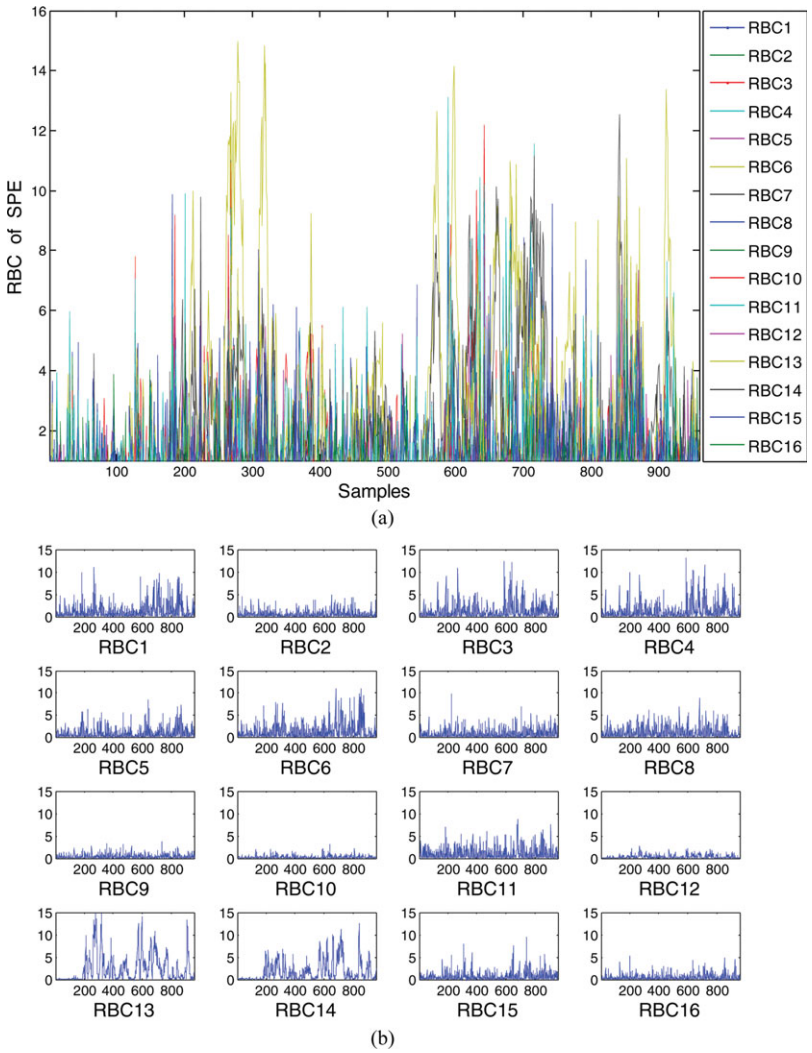

Figure 13. RBC-based fault identification results with SPE statistics of PCA.

(a) Compared results; (b) specific results. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

For model construction, a total of 2000 data samples have been generated through Eq. 42, and two latent variables are selected for developing the models. The change of the log likelihood value through the linear Gaussian state-space model modeling process is shown in Figure 2. One can find that the iteration process has already converged after about 30 steps. For performance comparison, three monitoring models which are based on PCA, PPCA, and FA are also constructed using the same training data samples, respectively. The number of latent variables in each of these three models is also selected as 2.

To test the performance of the proposed monitoring approach, five different testing datasets (each contains 2000 data samples) have also been generated, which are listed as follows

1. A step change of the second process variable by 1 was introduced starting from sample 1001.

2. A ramp change of the third process variable from sample 1001 to 2000 by adding $0.002(i-1000)$ to each data sample, where i is the sample number.

3. A random change of the first variable from sample 1001 to 2000 by adding a random variable with zero mean and variance 0.25.

4. Process structure change from sample 1001 to 2000 by changing matrix $\mathbf{B}$ to $\mathbf{B}=\begin{bmatrix}0.8 & -0.1\\ -0.4 & 0.6\end{bmatrix}$.

5. Abnormal state transfer noises from sample 1001 to 2000 by setting the mean of first state transfer noise ξ_1 to 0.15.

The first three testing datasets (faults 1, 2, and 3) represent some abnormal events with the process sensors: fault 1 represents sensor bias, fault 2 represents sensor drift, and fault 3 represents precision degradation of the sensor. The last two testing datasets (faults 4 and 5) represent two others types of process fault, which are changes of process structure and noise level, respectively.

The convergence of the state variables covariance in testing datasets is shown as in Figure 3. One can see that the convergence index $\Delta V(k)$ will become very small after about five samples have been updated in the model. Therefore, the covariance of state variables converges in only a few steps. Figures 4a, b shows the auto- and cross-correlations of different process variables and prediction errors before/after the implementation of the linear Gaussian state-space model. From this figure, one can see that the dynamic modeling performance of the linear Gaussian state-space model is significant, as the prediction errors are temporally white, based on which Remark 2 is also confirmed.

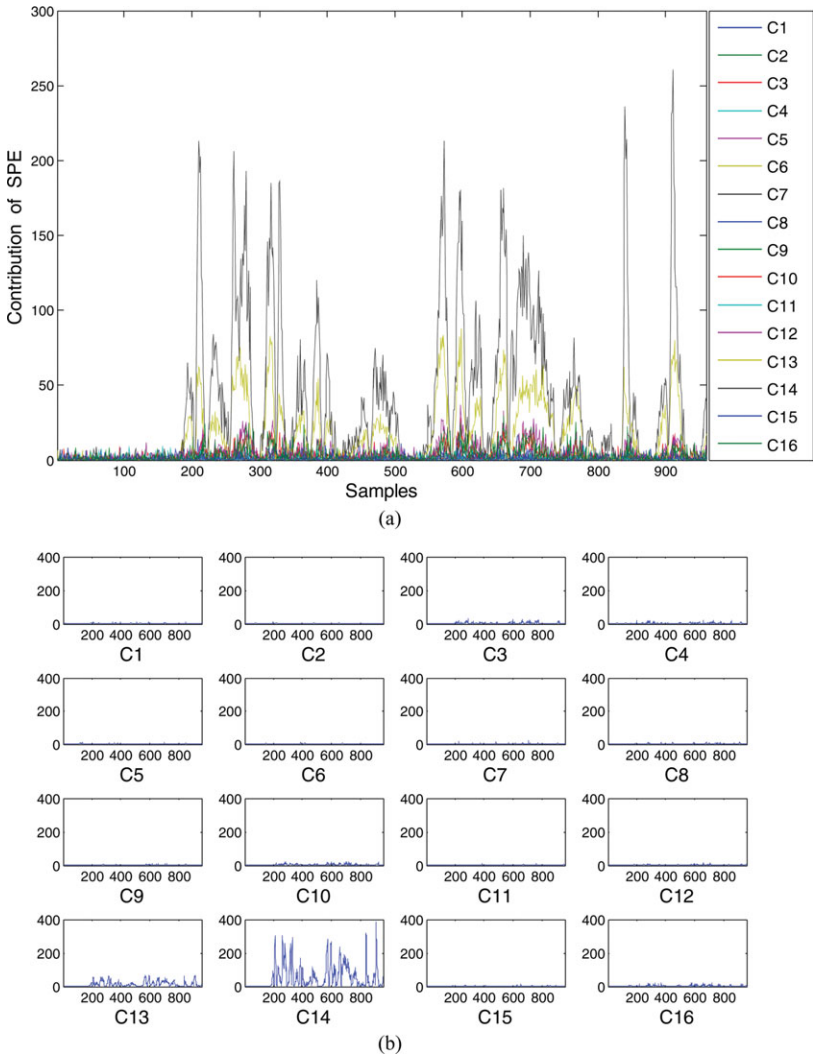

Figure 14. LGSS model-based fault identification results with SPE statistics.

(a) Compared results; (b) specific results. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

A general view of the monitoring performance of these four models is showed in Tables 1 and 2. Table 1 summarizes the false alarm rate (Type I error) of each monitoring statistic while Table 2 provides the detailed missed detection rate (Type II error) of the corresponding monitoring statistic under different cases. As can be seen in Table 1, the false alarm rate of each monitoring statistic in all of the four methods is acceptable. However, Table 2 shows that linear Gaussian state-space model outperforms other methods in detecting the five faults in the process. To be clear, the minimum false alarm rates with each monitoring statistic are marked in bold in Table 2, and we can see that linear Gaussian state-space model has the smallest missed detection rates in most cases. Specifically, the monitoring results for fault 1 obtained by the FA, PPCA, PCA, and linear Gaussian state-space models are given in Figures 5a–d. The $\boldsymbol{T^2}$ monitoring statistic of linear Gaussian state space model has successfully detected almost all the faulty samples while other models fails with more than half of the data samples.

Next, the fault identification performance of the linear Gaussian state-space model is verified through the first three testing datasets. For comparison, the RBC-based fault identification method with PCA is also carried out. The identification results of the last faulty measurements in each dataset are demonstrated as Figures 6 and 7. The fault identification results of RBC-based method are showed in Figure 6. Note that the faulty variables of fault 1, fault 2, and fault 3 are the 2nd, the 3rd and the 1st variable, respectively, in this system. Obviously, there is some fault smearing in the RBC-based identification approach, which has lead to misdiagnosis for the faults. In Figure 6a, particularly, the T^2-based RBC contributions of the 2nd and 4th variable are quite close to each other while in Figure 6b, the SPE-based RBC contributions of the 1st and 3rd variable are similar. In Figure 6c, even worse, the T^2-based RBC contribution method has generated a wrong identification result for the fault. In contrast, based on the proposed fault identification method in this article, an obvious and correct identification results have been provided for all of the three faults, which are shown in Figures 7a–c.

TE Benchmark process

As a benchmark simulation process, TE process has been widely used for algorithm testing and evaluation in the past decades. This process consists of five major unit operations: a reactor, a condenser, a compressor, a separator, and a stripper.

There are four reactants A, C, D, and E and one inert B which are fed to the reactor, and in which the products G and H are formed, but at the same time, a byproduct F is also produced. A total of 41 process variables and 12 manipulated variables are measured, and a set of 21 programmed faults can be simulated, which include seven step faults, five random faults, three sticking and slow change fault, and six unknown process faults. The flowchart of this process is given in Figure 8. More detailed description of this process can be found in reference.[36] In this article, 16 continuous process variables are selected for monitoring purpose, which are tabulated in Table 3.

To develop the linear Gaussian state-space model and traditional FA, PPCA, PCA monitoring models, 960 data samples have been collected for training purpose. Each of the monitoring models uses nine latent components in the model structure. Similarly, the change of the log likelihood value through the linear Gaussian state-space model modeling process is shown in Figure 9. It takes about 150 iteration steps for convergence. To test the monitoring performance of the linear Gaussian state-space model, all of the 21 faults of the TE process have been used. The descriptions of these faults are detailed in Table 4. Figure 10 shows the convergence of the covariance with the state variables in TE process. One can see that the convergence index $\Delta \boldsymbol{V}(\boldsymbol{k})$ will converge after only less than 20 samples are updated in the model.

The monitoring results of the four models for the 21 faults are showed together in Table 5. The best monitoring results obtained by the linear Gaussian state-space model monitoring method have been highlighted. From this table, generally, one can see that the overall performance of the linear Gaussian state-space model is much better than the other three methods. Particularly, significant improvements have been made by the T^2 statistic for fault 7, 10, 16, 17, 20, 21 and the SPE statistic for fault 19, 20 of linear Gaussian state-space model. Detailed monitoring results of different methods for fault 10 are given in Figure 11. Fault 10 is a random change of the temperature of stream 4 (C feed) in the TE process. From the flowchart of TE process in Figure 7, it is easy to find that stream 4 is the inlet stream of the Stripper. Therefore, the most relevant variables among the monitoring variables are the temperature (the 13th monitoring variable) and stream flow (the 14th monitoring variable) of the Stripper. Based on the results presented in Figure 11, one can see that the linear Gaussian state-space model outperforms all the other models in both T^2 and SPE statistics.

The fault identification results of RBC and linear Gaussian state-space model are shown in Figures 12–14. Figures 12 and 13 show the RBC-based fault identification results with T^2 statistic and SPE statistic, respectively. While Figures 12a and 13a give the compared contribution results of different process variables for fault 10, Figures 12b and 13b provide the detailed contribution result of each process variable responsible for the fault. Figure 14 shows the contribution results of the linear Gaussian state-space model-based method, which is based on the SPE statistic. Similarly, different contributions of the process variables are plotted together in Figure 14a, and the specific contribution of each process variable is demonstrated in Figure 14b. Based on Figure 11, it can be seen that variable 4, 9, 10, 12, 13, and 14 have larger contribution than other variables. Similarly, according to the results of Figure 10, the significant contribution variables are variable 1, 3, 4, 6, 13, and 14 listed in Table 4. Although the most responsible variables 13 and 14 have been identified, it is still very difficult to determine the root cause of the fault. However, based on the results given in Figures 14a, b, it is easy to find that variable 13 and 14 are the most relevant variables of the fault. Compared to these two variables, the contributions of other process variables are relatively small. Therefore, the most fault responsible variables should be determined as these two variables, which is the real case in practice.

Conclusions

In this article, a linear Gaussian state-space model-based monitoring approach, including fault detection and fault identification, has been proposed, which is not only probabilistic but also dynamic. Therefore, it is suitable for monitoring time-series process data under noisy environment. To evaluate the performance of the proposed method, two examples are studied, both of which have demonstrated the efficiency of the linear Gaussian state-space model. Compared to the traditional data-based methods such as PCA, PPCA and FA, the fault detection performance has been improved by the linear Gaussian state-space model. Furthermore, the superiority of the new fault identification method has also been illustrated, and detailed analyses of the fault smearing effect has been provided and compared to the recently developed RBC approach.

Acknowledgments

This work was supported in part by the National Basic Research Program of China (973 Program) Grant Number 2012CB720500, the National Natural Science Foundation of China (60974056, 61004134), and Fundamental Research Funds for the Central Universities.

Literature Cited

1. Kano M, Nagao K, Hasebe H, Hashimoto I, Ohno H, Strauss R, Bakshi BR. Comparison of multivariate statistical process monitoring methods with applications to the Eastman challenge problem. *Comput Chem Eng*. 2002;26:161–174.
2. Qin SJ. Statistical process monitoring: basics and beyond. *J Chemom* 2003;17:480–502.
3. Wang XZ, Medasani S, Marhoon F, Albazzaz H. Multidimensional visualization of principal component scores for process historical data analysis. *Ind Eng Chem Res*. 2004;43:7036–7048.
4. MacGregor JF, Yu HL, Munoz SG, Flores-Cerrillo J. Data-based latent variable methods for process analysis, monitoring and control. *Comput Chem Eng*. 2005;29:1217–1223.
5. Singhai A, Seborg DE. Evaluation of a pattern matching method for the Tennessee Eastman challenge process. *J Process Control*. 2006;16:601–613.
6. Doan XT, Srinivasan R. Online monitoring of multi-phase batch processes using phase-based multivariate statistical process control. *Comput Chem Eng*. 2008;32:230–243.
7. He QP, Wang J. Statistics pattern analysis: a new process monitoring framework and its application to semiconductor batch processes. *AIChE J*. 2011;57:107–121.
8. Yu J. Nonlinear bioprocess monitoring using multiway kernel localized fisher discriminant analysis. *Ind Eng Chem Res*. 2011;50:3390–3402.
9. Bishop CM. *Pattern Recognition and Machine Learning*. Springer, Singapore, 2006.
10. Kim D, Lee IB. Process monitoring based on probabilistic PCA. *Chem Intell Lab Syst*. 2003;67:109–123.
11. Narasimhan S, Shah SL. Model identification and error covariance matrix estimation from noisy data using PCA. *Control Eng Pract*. 2008;16:146–155.
12. Feital T, Kruger U, Xie L, Schubert U, Lima EL, Pinto JC. A unified statistical framework for monitoring multivariate systems with unknown source and error signals. *Chem Intell Lab Syst*. 2010;104:223–232.
13. Kim DS, Yoo CK, Kim YI, Jung JK, Lee IB. Calibration, prediction and process monitoring model based on factor analysis for incomplete process data. *J Chem Eng Jpn*. 2005;38:1025–1034.

14. Ge ZQ, Xie L, Song ZH. A novel statistical-based monitoring approach for complex multivariate processes. *Ind Eng Chem Res*. 2009;48:4892–4898.
15. Ku W, Storer RH, Georgakis C. Disturbance rejection and isolation by dynamic principal component analysis. *Chem Intell Lab Syst*. 1995;30:179–196.
16. Chen J, Liu K. On-line batch process monitoring using dynamic PCA and dynamic PLS models. *Chem Eng Sci*. 2002;14:63–75.
17. Choi SW, Lee IB. Nonlinear dynamic process monitoring based on dynamic kernel PCA. *Chem Eng Sci*. 2004;59:5897–5908.
18. Jia MX, Chu F, Wang FL, Wang W. On-line batch process monitoring using batch dynamic kernel principal component analysis. *Chem Intell Lab Syst*. 2010;101:110–122.
19. Zhang YW, Li ZW, Zhou H. Statistical analysis and adaptive technique for dynamic process monitoring. *Chem Eng Res Des*. 2010;88:1381–1392.
20. Kruger U, Zhou YQ, Irwin GW. Improved principal component monitoring of large-scale processes. *J Process Control*. 2004;14: 879–888.
21. Xie L, Kruger U, Lieftucht D, Littler T, Chen Q, Wang SQ. Statistical monitoring of dynamic multivariate processes part1 modeling autocorrelation and cross-correlation. *Ind Eng Chem Res*. 2006;45: 1659–1676.
22. Yao Y, Gao FR. Subspace identification for two-dimensional dynamic batch process statistical monitoring. *Chem Eng Sci*. 2008; 63:3411–3418.
23. Odiowei PP, Cao Y. State-space independent component analysis for nonlinear dynamic process monitoring. *Chem Intell Lab Syst*. 2010;103:59–65.
24. Alabi SI, Morris AJ, Martin EB. On-line dynamic process monitoring using wavelet-based generic dissimilarity measure. *Chem Eng Res Des*. 2005;83:698–705.
25. Choi SW, Morris J, Lee IB. Dynamic model-based batch process monitoring. *Chem Eng Sci*. 2008;63:622–636.
26. Ge ZQ, Kruger U, Lamont L, Xie L, Song ZH. Fault detection in non-Gaussian vibration systems using dynamic statistical-based approaches. *Mech Syst Signal Process*. 2010;24:2972–2984.
27. Alcala CF, Qin SJ. Reconstruction-based contribution fro process monitoring. *Automatica*. 2009;45:1593–1600.
28. Roweis S, Ghahramani Z. A unifying review of linear Gaussian models. *Neural Comput*. 1999;11:305–345.
29. Barber D, Chiappa S. *Unified inference for variational Bayesian linear Gaussian state-space model*. In: Schölkopf B, Platt J, Hoffman T, editors. *Advances in Neural Information Processing Systems 19*. Cambridge, MA: MIT Press, 2007:81–88.
30. Anderson BDO, Moore JB. *Optimal Filtering*. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1979.
31. Miller P, Swanson RE, Heckler CF. Contribution plots: the missing link in multivariate quality control. In: *Fall Conf. of the ASQC and ASA*. Milwaukee, WI, 1993.
32. Dunia R, Qin SJ, Edgar T, McAvoy T. Identification of faulty sensors using PCA. *AIChE J*. 1996;42:2797–2812.
33. Yue H, Qin SJ. Reconstruction based fault identification using a combined index. *Ind Eng Chem Res*. 2001;40:4403–4414.
34. Chen T, Sun Y. Probabilistic contribution analysis for statistical process monitoring: a missing variable approach. *Control Eng Pract*. 2009;17:469–477.
35. Liu J. Data-driven fault detection and isolation for multimode processes. *Asia-Pacific J Chem Eng*. 2011;6:470–483.
36. Downs JJ, Vogel EF. A plant-wide industrial process control problem. *Comput Chem Eng*. 1993;17:245–255.

Appendix: Derivation of EM Algorithm for Linear Gaussian State-Space Model

Given the three expected statistics obtained in the E-step, we aim to update the model parameters in the linear Gaussian state-space model. First, the expectation value of the complete-data log likelihood which related to $\boldsymbol{\mu}_1$ and $\mathbf{V}_1$ can be written as[9]

$$E[\ln p(\mathbf{X},\mathbf{T}|\Theta)]|_{\boldsymbol{\mu}_1,\mathbf{V}_1} = -\frac{1}{2}\ln(|\mathbf{V}_1|) - E[\frac{1}{2}(\mathbf{t}_1-\boldsymbol{\mu}_1)^T\mathbf{V}_1^{-1}(\mathbf{t}_1-\boldsymbol{\mu}_1)] \quad \text{(A1)}$$

By maximizing the above equation according to the parameters $\boldsymbol{\mu}_1$ and $\mathbf{V}_1$, the updated values can be calculated as

$$\boldsymbol{\mu}_1^{\text{new}} = E(\mathbf{t}_1) \quad \text{(A2)}$$

$$\mathbf{V}_1^{\text{new}} = E(\mathbf{t}_1\mathbf{t}_1^T) - E(\mathbf{t}_1)E^T(\mathbf{t}_1) \quad \text{(A3)}$$

Similarly, the expectation value of the complete-data log likelihood which related to $\mathbf{A}$, Γ, and $\mathbf{P}$, Σcan be written as follows

$$E[\ln p(\mathbf{X},\mathbf{T}|\Theta)]|_{\mathbf{A},\Gamma} = -\frac{n-1}{2}\ln(|\Gamma|) - E\{\frac{1}{2}\sum_{k=1}^{n-1}[(\mathbf{t}_{k+1}-\mathbf{A}\mathbf{t}_k)^T\Gamma^{-1}(\mathbf{t}_{k+1}-\mathbf{A}\mathbf{t}_k)]\} \quad \text{(A4)}$$

$$E[\ln p(\mathbf{X},\mathbf{T}|\Theta)]|_{\mathbf{P},\Sigma} = -\frac{n}{2}\ln(|\Sigma|) - E\{\frac{1}{2}\sum_{k=1}^{n}[(\mathbf{x}_k-\mathbf{P}\mathbf{t}_k)^T\Sigma^{-1}(\mathbf{x}_k-\mathbf{P}\mathbf{t}_k)]\} \quad \text{(A5)}$$

By setting the derivative of $E[\ln p(\mathbf{X},\mathbf{T}|\Theta)]|_{\mathbf{A},\Gamma}$ respect to $\mathbf{A}$ and Γ to zeros, the following updated equations can be obtained

$$\mathbf{A}^{\text{new}} = [\sum_{k=1}^{n-1}E(\mathbf{t}_{k+1}\mathbf{t}_k^T)][\sum_{k=1}^{n-1}E(\mathbf{t}_k\mathbf{t}_k^T)]^{-1} \quad \text{(A6)}$$

$$\Gamma^{\text{new}} = \frac{1}{n-1}\sum_{k=1}^{n-1}\{E(\mathbf{t}_{k+1}\mathbf{t}_{k+1}^T) - \mathbf{A}^{\text{new}}E(\mathbf{t}_k\mathbf{t}_{k+1}^T) - E(\mathbf{t}_{k+1}\mathbf{t}_k^T)(\mathbf{A}^{\text{new}})^T + \mathbf{A}^{\text{new}}E(\mathbf{t}_k\mathbf{t}_k^T)(\mathbf{A}^{\text{new}})^T\} \quad \text{(A7)}$$

Then, set the derivative of $E[\ln p(\mathbf{X},\mathbf{T}|\Theta)]|_{\mathbf{P},\Sigma}$ respect to $\mathbf{P}$ and $\boldsymbol{\Sigma}$ to zeros, these two parameter matrices can be updated as

$$\mathbf{P}^{\text{new}} = [\sum_{k=1}^{n}\mathbf{x}_kE^T(\mathbf{t}_k)][\sum_{k=1}^{n}E(\mathbf{t}_k\mathbf{t}_k^T)]^{-1} \quad \text{(A8)}$$

$$\boldsymbol{\Sigma}^{\text{new}} = \frac{1}{n}\sum_{k=1}^{n}\{\mathbf{x}_k\mathbf{x}_k^T - \mathbf{P}^{\text{new}}E(\mathbf{t}_k)\mathbf{x}_k^T - \mathbf{x}_kE^T(\mathbf{t}_k)(\mathbf{P}^{\text{new}})^T + \mathbf{P}^{\text{new}}E(\mathbf{t}_k\mathbf{t}_k^T)(\mathbf{P}^{\text{new}})^T \quad \text{(A9)}$$

Manuscript received Jun. 2, 2011, revision received Oct. 14, 2011, and final revision received Feb. 9, 2012.