Mixture Bayesian Regularization Method of PPCA for Multimode Process Monitoring

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

DOI 10.1002/aic.12200
Published online March 10, 2010 in Wiley Online Library (wileyonlinelibrary.com).

This article intends to address two drawbacks of the traditional principal component analysis (PCA)-based monitoring method: (1) nonprobabilistic; (2) single operation mode assumption. On the basis of the monitoring framework of probabilistic PCA (PPCA), a Bayesian regularization method is introduced for performance improvement, through which the effective dimensionality of the latent variable can be determined automatically. For monitoring processes with multiple operation modes, the Bayesian regularization method is extended to its mixture form, thus a mixture Bayesian regularization method of PPCA has been developed. To enhance the monitoring performance, a novel probabilistic strategy has been proposed for result combination in different operation modes. In addition, a new mode localization approach has also been developed, which can provide additional information and improve process comprehension for the operation engineer. A numerical example and a real industrial application case study have been used to evaluate the efficiency of the proposed method. *AIChE J,* 56: 2838–2849, 2010
Keywords: multimode process monitoring, Bayesian regularization, principal component analysis, model localization

Introduction

Along last several decades, multivariate statistical process control (MSPC)-based methods such as principal component analysis (PCA) and partial least squares (PLS) have gained much attentions in the process monitoring area.[1–6] However, traditional MSPC-based monitoring methods have assumed that process variables are linear, deterministic, normally distributed, and operated under single mode. In reality, those restrictions can be easily violated. In other words, data obtained from complex processes are not always linear, Gaussian distributed, and they often performs through a random manner. Besides, they may come from different operation conditions.

The main focus of this article is to address the following two restrictions of the traditional method: (1) nonprobabilistic; (2) single operation mode assumption. To address the random manner of the process data, the traditional PCA method has been extended to the probabilistic PPCA (PPCA),[7] which has recently been introduced for process monitoring.[8] For monitoring processes with multiple operation conditions under the noisy environment, the PPCA-based method has been extended to its mixture form.[9–13] However, there are several drawbacks of the (mixture) PPCA-based methods. First, they cannot determine the effective principal component number automatically through the modeling process. Second, the importance of each retained principal component in the model structure cannot be easily differentiated. Besides, singularity and over-fit problems may be caused when the number of modeling data samples is limited.

In fact, the selection of the latent variable dimensionality can be termed as a model complexity problem, which can be addressed by the Bayesian method. In the present article, a Bayesian regularization of PCA (Bayesian PCA, BPCA) is introduced, which was originally proposed by Bishop.[14]

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

Later, a more general modeling framework of the BPCA algorithm has been proposed, in which different aspects among the Bayesian model structure has been detailed, such as the prior density function, the loss function, the model rank, etc.[15] In this article, a special form of Bayesian PCA has been used, which has a Gaussian prior. Depending on the Bayesian method, the singularity and over-fit problems can be alleviated. Through the introduction of a hyperparameter in the model structure, BPCA can successfully control the dimensionality of the latent variable space. Because of the advantages of the BPCA method, it has been widely used in different application areas, such as image processing, pattern recognition, machine learning, etc.[16] However, to our best knowledge, the utilization of this method has not yet been reported in the process monitoring area. Different from the traditional PPCA method, BPCA can determine the effective number of principal components automatically. Furthermore, the importance of each principal component can also be measured by the introduced hyperparameter.

However, when the process runs in multiple operation modes, a single BPCA model may not function well. To this end, the basic BPCA model is extended to its mixture form, thus a mixture Bayesian regularization model of PPCA (MBPCA) is constructed in this article. On the basis of the developed MBPCA model, a corresponding multimode monitoring scheme can be formulated. For monitoring a new process data sample, the traditional mixture PPCA method first calculates its posterior probability in different operation modes, then the sub-model with the largest posterior probability value is selected for monitoring. However, when two or several posterior probability values are comparative, the model selection will be confused, which may decrease the monitoring performance. To address this issue, this article proposes a probabilistic combination strategy to enhance the multimode monitoring performance. Precisely, a weighted monitoring scheme is proposed, that is, the operation mode with a higher posterior probability value will be given a larger weight through the combination step. Therefore, the confused model selection problem is avoided. Instead, all monitoring results in different operation modes have been utilized to make the final decision, which is more reliable. Besides, another important aspect considered in this article is the mode localization problem. Precisely, one may want to know the mode information of the monitored data sample, thus which operation mode it belongs to. Depending on the proposed MBPCA-based monitoring approach, a new mode localization approach is also developed, which can successfully locate the monitored data sample to its correct operation mode.

In summary, contributions of the present article can be given as follows: (1) The Bayesian regularization method of PPCA is introduced for process monitoring; (2) A mixture form of the Bayesian PCA method is developed for multimode process monitoring; (3) A new probabilistic combination strategy is proposed for monitoring results integration in different operation modes; (4) A new mode localization approach is provided. The rest of this article is organized as follows. In Section "PPCA-Based Method for Process Monitoring," a brief description of the PPCA-based process monitoring method and its deficiency are given. Section "Mixture Bayesian Regularization of PPCA" provides a detailed demonstration for the development of the mixture Bayesian regularization method of PPCA, which is followed by the formulation of the corresponding multimode process monitoring scheme in the next section. Two case studies are provided in Section "Case Studies" to evaluate the efficiency of the proposed method. Finally, some conclusions are made.

PPCA-Based Method for Process Monitoring

As a probabilistic counterpart of PCA, PPCA was first proposed by Tipping and Bishop,[7] and later adopted for process monitoring purpose.[8] Through a generative model structure, the formulation of PPCA can be given as:

$$\mathbf{x} = \mathbf{Pt} + \mathbf{e} \tag{1}$$

where $\mathbf{x} \in R^m$ represents the process variable, $\mathbf{t} \in R^k$ is the latent variable, $\mathbf{P} \in R^{m\times k}$ is the loading matrix, $\mathbf{e} \in R^m$ is a zero mean white noise term with variance $\beta^{-1}\mathbf{I}$, thus $p(\mathbf{e}) = N(\mathbf{e}|0,\beta^{-1}\mathbf{I})$. In the PPCA model, the prior distribution of the latent variable $\mathbf{t}$ is also assumed to be a Gaussian distribution with zero mean and one variance $p(\mathbf{t}) = N(\mathbf{t}|0,\mathbf{I})$. Therefore, the marginal likelihood of $\mathbf{x}$ can be calculated as:

$$p(\mathbf{x}|\mathbf{P},\beta) = \int p(\mathbf{x}|\mathbf{t},\mathbf{P},\beta)p(\mathbf{t})d\mathbf{t} \tag{2}$$

.

For a given dataset $\mathbf{X} = (\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_n)$ of n data samples, $\mathbf{P}$ and β can be determined by maximizing the following likelihood function through a Expectation Maximization (EM) algorithm

$$L(\mathbf{P},\beta) = \ln\prod_{i=1}^{n} p(\mathbf{x}_i|\mathbf{P},\beta) \tag{3}$$

.

After the parameter set of PPCA has been determined, the corresponding monitoring scheme can be developed. For a new process data sample $\mathbf{x}_{\text{new}}$, the latent variable of this sample can be calculated as:

$$\mathbf{t}_{\text{new}} = \mathbf{Q}\mathbf{x}_{\text{new}} = \mathbf{P}^T(\mathbf{PP}^T + \Sigma)^{-1}\mathbf{x}_{\text{new}} \tag{4}$$

.

The estimated variance of the latent variable is given in Eq. 5, through which we can find that it has no relationship with the current data sample $\mathbf{x}_{\text{new}}$.

$$\text{var}(\mathbf{t}_{\text{new}}) = \mathbf{Q}(\mathbf{PP}^T + \Sigma)\mathbf{Q}^T \tag{5}$$

.

Therefore, the T^2 statistic can be constructed as:

$$T^2_{\text{new}} = \mathbf{t}^T_{\text{new}}(\text{var}(\mathbf{t}_{\text{new}}))^{-1}\mathbf{t}_{\text{new}} \tag{6}$$

.

Similarly, the SPE statistic can be constructed as:

$$\text{SPE}_{\text{new}} = \mathbf{e}^T_{\text{new}}(\beta^{-1}\mathbf{I})^{-1}\mathbf{e}_{\text{new}} \tag{7}$$

where $\mathbf{e}_{\text{new}} = \mathbf{x}_{\text{new}} - \mathbf{Pt}_{\text{new}} = (\mathbf{I}-\mathbf{PQ})\mathbf{x}_{\text{new}}$ is the error term of the new data sample. The control limits of both monitoring statistics can be determined by the χ^2 distribution with appropriate dimensions of freedom, thus $T^2_{\lim} = \chi^2_\gamma$, $\mathrm{SPE}_{\lim} = \chi^2_\gamma(m)$, where γ is the significant level.[8]

It is noted that an important assumption of PPCA is that the dimension of the latent variable k is known beforehand. In fact, if the number of process data samples is limited, the selection of principal component number will become problematic. This is because the PPCA method itself does not provide any mechanism to determine the effective latent variable dimensionality. If there are not enough data samples available for cross validation, it is difficult to determine this important number. Therefore, it is desired that the number of effective latent variables could be determined automatically through the model development step, especially when training data samples are limited. Fortunately, this problem can be well solved by the Bayesian regularization method, which is elaborately demonstrated in the next section.

Mixture Bayesian Regularization of PPCA

This section first provides a Bayesian regularization method for PPCA, which is termed as Bayesian PCA (BPCA) in the present article. Through the Bayesian treatment, the dimensionality of the latent variable space can be determined automatically, which is controlled by a hyperparameter in the BPCA model. Then, the single BPCA model is extended to its mixture form, based on which a special multimode monitoring scheme is developed.

Bayesian regularization of PPCA

Depending on the probabilistic formulation of PPCA defined in Eq. 1, the Bayesian treatment of PPCA can be obtained by introducing a prior distribution over the loading matrix **P**. Then, the posterior distribution of the loading matrix $p(\mathbf{P}|\mathbf{X})$ can be calculated by multiplying the prior by the likelihood function, whose logarithm is given in Eq. 3. To implement the Bayesian PCA algorithm, the dimensionality of the latent space is set to its maximum value $d = m-1$. Then, a hyperparameter vector $\alpha = \{\alpha_1,\alpha_2,\ldots\alpha_d\}$ is introduced to control the dimensionality of the latent space. Though a conditional Gaussian distribution, each hyperparameter controls one column of the loading matrix **P**, which is defined as[14]:

$$p(\mathbf{P}|\alpha) = \prod_{i=1}^{d} \left(\frac{\alpha_i}{2\pi}\right)^{m/2} \exp\left\{-\frac{1}{2}\alpha_i\|\mathbf{p}_i\|^2\right\} \tag{8}$$

where $\mathbf{p}_i$ is the ith column of the loading matrix **p**. Since each α_i controls the inverse variance of $\mathbf{p}_i$, if it has a large value, the corresponding $\mathbf{p}_i$ will tend to be very small, and would be effectively removed from the latent space loading matrix. In practice, a threshold value of α_i can be set up for the selection of the effective latent variable dimensionality.

To estimate an optimal value of **P** and β, an EM algorithm can also be developed. Thus, by maximizing **P** and β in the log posterior distribution function given in Eq. 9, the Bayesian regularization value of **P** and β can be obtained.

$$\ln p(\mathbf{P}|\mathbf{X}) = \ln\frac{p(\mathbf{X}|\mathbf{P})p(\mathbf{P})}{\int p(\mathbf{X},\mathbf{P})d\mathbf{P}} = L - \frac{1}{2}\sum_{i=1}^{m-1}\alpha_i\|\mathbf{p}_i\|^2 + \mathrm{const} \tag{9}$$

where "const" represents a constant value. The EM algorithm can be calculated as follows. In the E-step, the expected sufficient statistics of the latent variable can be evaluated as:

$$\begin{aligned} &E(\hat{\mathbf{t}}|\mathbf{x}) = \mathbf{M}^{-1}\mathbf{P}^T\mathbf{x} \\ &E(\hat{\mathbf{t}}\hat{\mathbf{t}}^T|\mathbf{x}) = \beta^{-1}\mathbf{M}^{-1} + E(\hat{\mathbf{t}}|\mathbf{x})E^T(\hat{\mathbf{t}}|\mathbf{x}) \end{aligned} \tag{10}$$

where $\mathbf{M} = \mathbf{P}^T\mathbf{P} + \beta^{-1}\mathbf{I}$. The M-step then updates the model parameters as:

$$\begin{aligned} &\hat{\mathbf{P}} = \left[\sum_{i=1}^{n}\mathbf{x}_i E^T(\hat{\mathbf{t}}_i|\mathbf{x}_i)\right]\left[\sum_{i=1}^{n}E(\hat{\mathbf{t}}_i\hat{\mathbf{t}}_i^T|\mathbf{x}_i) + \beta^{-1}\mathbf{A}\right]^{-1} \\ &\hat{\alpha}_i = \frac{m}{\|\mathbf{p}_i\|^2} \\ &\hat{\beta} = \frac{nm}{\sum_{i=1}^{n}\{\mathbf{x}_i^T\mathbf{x}_i - 2E^T(\hat{\mathbf{t}}_i|\mathbf{x}_i)\hat{\mathbf{P}}^T\mathbf{x}_i + Tr[E(\hat{\mathbf{t}}_i\hat{\mathbf{t}}_i^T|\mathbf{x}_i)\hat{\mathbf{P}}^T\hat{\mathbf{P}}]\}} \end{aligned} \tag{11}$$

where $\mathbf{A} = \mathrm{diag}\,(\alpha_i, i = 1,2,\ldots,d)$ is a diagonal matrix, $\mathrm{Tr}(\cdot)$ is an operator for trace value calculation. Detailed derivation of the EM algorithm is provided in Appendix A. Therefore, by updating E-step and M-step until all of the parameters satisfy a suitable convergence criterion, such as 10^{-4}, the optimal values of **P** and β can be determined.

Mixture Bayesian regularization of PPCA

When the process dataset comes from several different operation modes, the single Bayesian PCA-based modeling method is not sufficient. Hence, this subsection extends the Bayesian regularization treatment of PPCA to its mixture form, thus a mixture Bayesian regularization method of PPCA is developed. Suppose a total of C operation conditions are incorporated, the distribution of the process variable can be calculated as:

$$p(\mathbf{x}) = \sum_{c=1}^{C} p(\mathbf{x}|c)p(c) \tag{12}$$

where $p(c)$ is the mixing proportion of each operation mode, under the constraint that $\sum_{c=1}^{C} p(c) = 1$, $p(\mathbf{x}|c)$ is the conditional distribution function of each operation mode, which follows Gaussian distribution as $N(0, \mathbf{P}_c\mathbf{P}_c^T + \beta_c^{-1}\mathbf{I})$, where $\mathbf{P}_c$ and β_c are the corresponding parameters in the cth local Bayesian PCA model. For Bayesian regularization, a hyperparameter matrix is defined as follows:

$$\alpha = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1d} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2d} \\ \vdots & \vdots & \alpha_{ci} & \vdots \\ \alpha_{C1} & \alpha_{C2} & \cdots & \alpha_{Cd} \end{bmatrix} \tag{13}$$

where $c = 1,2,\ldots,C$, $i = 1,2,\ldots,d = m-1$. Then the aim of the mixture Bayesian PCA model development can be formulated as follows: given a process dataset $\mathbf{X} \in R^{n\times m}$, estimate the parameter sets $\Theta_c = \{\mathbf{P}_c,\beta_c,\alpha_c\}(c = 1,2,\ldots,C)$ for each local Bayesian PCA model.

First, the log posterior distribution function is given as:

$$\ln p(\mathbf{P}|\mathbf{X}) = \ln\prod_{j=1}^{n} p(\mathbf{x}_j|\mathbf{P}) - \frac{1}{2}\sum_{i=1}^{d}\alpha_{ci}p(c)\|\mathbf{p}_i\|^2 + \text{const}$$
$$= \sum_{j=1}^{n}\ln\left[\sum_{c=1}^{C}p(\mathbf{x}_j|c,\mathbf{P})p(c)\right] - \frac{1}{2}\sum_{i=1}^{d}\alpha_{ci}p(c)\|\mathbf{p}_i\|^2 + \text{const} \quad (14)$$

Then, a new EM algorithm can be constructed for estimation of the parameter set Θ_c in the mixture Bayesian PCA model. In the E-step, we are given the parameters Θ_{old} obtained in the previous step, our aim is to estimate the expected sufficient statistics of the latent variable posterior distribution and the mode posterior probability $p(c|\mathbf{x},\Theta_{\text{old}})$. Through the Bayesian rule, the posterior probabilities of the mode and the latent variable can be calculated as follows:

$$p(c|\mathbf{x},\Theta_{\text{old}}) = \frac{p(\mathbf{x}|c,\Theta_{\text{old}})p(c|\Theta_{\text{old}})}{p(\mathbf{x}|\Theta_{\text{old}})} \quad (15)$$

$$p(\mathbf{t}|\mathbf{x},c,\Theta_{\text{old}}) = \frac{p(\mathbf{x}|\mathbf{t},c,\Theta_{\text{old}})p(\mathbf{t}|c,\Theta_{\text{old}})}{p(\mathbf{x}|c,\Theta_{\text{old}})} \quad (16)$$

where $p(c|\Theta_{\text{old}})$ is the proportion value calculated in the previous step, and $p(\mathbf{x}_i|c, \Theta_{\text{old}})$ follows the multivariate Gaussian distribution $N(0,\mathbf{P}_c\mathbf{P}_c^T + \beta_c^{-1}\mathbf{I})$. The estimated mean and variance values of the latent variable are given as:

$$E(\hat{\mathbf{t}}|\mathbf{x},c,\Theta_{\text{old}}) = \mathbf{M}_c^{-1}\mathbf{P}_c^T\mathbf{x}$$
$$E(\hat{\mathbf{t}}\hat{\mathbf{t}}^T|\mathbf{x},c,\Theta_{\text{old}}) = \beta_c^{-1}\mathbf{M}_c^{-1} + E(\hat{\mathbf{t}}|\mathbf{x},c,\Theta_{\text{old}})E^T(\hat{\mathbf{t}}|\mathbf{x},c,\Theta_{\text{old}}) \quad (17)$$

where $\mathbf{M}_c = \mathbf{P}_c^T\mathbf{P}_c + \beta_c^{-1}\mathbf{I}$. The M-step involves estimating the optimal parameter values through the maximum likelihood method, which are given as:

$$p(c|\Theta_{\text{new}}) = \frac{1}{n}\sum_{j=1}^{n}p(c|\mathbf{x}_j,\Theta_{\text{old}}) \quad (18)$$

$$\mathbf{P}_c^{\text{new}} = \left(\sum_{j=1}^{n}p(c|\mathbf{x}_j,\Theta_{\text{old}})\mathbf{x}_jE(\mathbf{t}_j^T|\mathbf{x}_j,c,\Theta_{\text{old}})\right)\times$$
$$\left(\sum_{j=1}^{n}p(c|\mathbf{x}_j,\Theta_{\text{old}})E(\mathbf{t}_j\mathbf{t}_j^T|\mathbf{x}_j,c,\Theta_{\text{old}}) + p(c|\Theta_{\text{new}})\beta_c^{-1}\mathbf{A}_c\right)^{-1} \quad (19)$$

$$\alpha_{ci}^{\text{new}} = \frac{m}{\|\mathbf{p}_{ci}^{\text{new}}\|^2} \quad (20)$$

$$\beta_c^{\text{new}} = \frac{nm}{\sum_{j=1}^{n}p(c|\mathbf{x}_j,\Theta_{\text{old}})\{\mathbf{x}_j^T\mathbf{x}_j - 2E^T(\mathbf{t}_j|\mathbf{x}_j)\mathbf{P}_c^{\text{new}T}\mathbf{x}_j + Tr[E(\mathbf{t}_j\mathbf{t}_j^T|\mathbf{x}_j)\mathbf{P}_c^{\text{new}T}\mathbf{P}_c^{\text{new}}]\}} \quad (21)$$

where $\mathbf{A}_c = \text{diag}(\alpha_{ci}, i = 1,2,\ldots,d)$ is a diagonal matrix, $c = 1,2,\ldots,C$, $\mathbf{p}_{ci}^{\text{new}}$ is the ith column of the new estimated loading matrix $\mathbf{P}_c^{\text{new}}$. Detailed derivation of the EM algorithm for mixture BPCA model is provided in Appendix B. Updating the E-step and the M-step recursively, we can finally get an optimal value of the parameter set $\Theta_c^{\text{opt}} = \{\hat{\mathbf{P}}_c, \hat{\beta}_c, \hat{\alpha}_c\}(c = 1,2,\ldots,C)$ for each local Bayesian PCA model. Thus, the mixture Bayesian regularization model of PPCA has been constructed. Different from the traditional mixture PPCA model, the effective number of principal components in each local BPCA model plane can be determined automatically by the new method.

Multimode Process Monitoring Based on MBPCA

On the basis of the developed mixture BPCA model in Section "Mixture Bayesian Regularization of PPCA," a specific monitoring scheme for multimode processes can be constructed. The following two subsections give detailed illustrations of the proposed monitoring and mode localization approaches.

Monitoring scheme development

Suppose a new data sample $\mathbf{x}_{\text{new}}$ has been collected from the process, we first construct two traditional monitoring statistic (T^2 and SPE) in each of the C local BPCA model spaces, which can be calculated as follows:

$$\mathbf{t}_{c,\text{new}} = \hat{\mathbf{Q}}_c\mathbf{x}_{\text{new}}$$
$$T_{c,\text{new}}^2 = \mathbf{t}_{c,\text{new}}^T(\hat{\mathbf{Q}}_c(\hat{\mathbf{P}}_c\hat{\mathbf{P}}_c^T + \hat{\beta}_c^{-1}\mathbf{I})\hat{\mathbf{Q}}_c^T)^{-1}\mathbf{t}_{c,\text{new}} \quad (22)$$

$$\mathbf{e}_{c,\text{new}} = (\mathbf{I} - \hat{\mathbf{P}}_c\hat{\mathbf{Q}}_c)(\mathbf{x}_{\text{new}} - \mu_k)$$
$$\text{SPE}_{c,\text{new}} = \mathbf{e}_{c,\text{new}}^T(\hat{\beta}_c^{-1}\mathbf{I})^{-1}\mathbf{e}_{c,\text{new}} \quad (23)$$

where $c = 1,2,\ldots,C$, $\hat{\mathbf{Q}}_c = \hat{\mathbf{P}}_c^T(\hat{\mathbf{P}}_c\hat{\mathbf{P}}_c^T + \hat{\beta}_c^{-1}\mathbf{I})^{-1}$. In the mixture PPCA monitoring method, a total of C monitoring charts can be built, the one with the largest posterior probability value is selected for monitoring. Through the Bayesian rule, the posterior probability of each operation mode can be calculated as:

$$p(c|\mathbf{x}_{\text{new}},\Theta^{\text{opt}}) = \frac{p(\mathbf{x}_{\text{new}}|c,\Theta^{\text{opt}})p(c|\Theta^{\text{opt}})}{p(\mathbf{x}_{\text{new}}|\Theta^{\text{opt}})} \quad (24)$$

.

Hence, the final T^2 and SPE monitoring statistics of the mixture PPCA method can be determined as:

$$T_{\text{final,new}}^2 = T_{c,\text{new}}^2\{\max(p(c|\mathbf{x}_{\text{new}},\Theta^{\text{opt}})), c = 1,2,\cdots,C\} \quad (25)$$

$$\text{SPE}_{\text{final,new}} = \text{SPE}_{c,\text{new}}\{\max(p(c|\mathbf{x}_{\text{new}},\Theta^{\text{opt}})), c = 1,2,\cdots,C\}. \quad (26)$$

Then, the process can be monitored by two normalized statistics, the control limits of which are selected as 1.[12] In our opinion, however, some useful information may lose through the hard assignment of the mixture PPCA monitoring method. Besides, if two or more posterior probability values are comparative, it is difficult to make a final monitoring decision. In such a special case, it can be inferred that the monitoring performance may be deteriorated. In contrast, a soft assignment strategy is proposed to enhance the monitoring performance, which is based on Bayesian transformation and probabilistic combination.

Since different local BPCA model spaces have their own monitoring statistics and control limits, it is difficult to combine them directly. Here, the Bayesian method is used to turn the monitoring results into fault probabilities, which are given as[16]:

$$P_{T^2}^c(F|\mathbf{x}_{\text{new}}) = \frac{P_{T^2}^c(\mathbf{x}_{\text{new}}|F)P_{T^2}^c(F)}{P_{T^2}^c(\mathbf{x}_{\text{new}})}$$
$$P_{T^2}^c(\mathbf{x}_{\text{new}}) = P_{T^2}^c(\mathbf{x}_{\text{new}}|N)P_{T^2}^c(N) + P_{T^2}^c(\mathbf{x}_{\text{new}}|F)P_{T^2}^c(F) \tag{27}$$

$$P_{\text{SPE}}^c(F|\mathbf{x}_{\text{new}}) = \frac{P_{\text{SPE}}^c(\mathbf{x}_{\text{new}}|F)P_{\text{SPE}}^c(F)}{P_{\text{SPE}}^c(\mathbf{x}_{\text{new}})}$$
$$P_{\text{SPE}}^c(\mathbf{x}_{\text{new}}) = P_{\text{SPE}}^c(\mathbf{x}_{\text{new}}|N)P_{\text{SPE}}^c(N) + P_{\text{SPE}}^c(\mathbf{x}_{\text{new}}|F)P_{\text{SPE}}^c(F) \tag{28}$$

where $P_{T^2}^c(N)$, $P_{T^2}^c(F)$ and $P_{\text{SPE}}^c(N)$, $P_{\text{SPE}}^c(N)$ represent the prior probabilities of the normal and faulty process conditions, which can be defined as:

$$P_{T^2}^c(N) = P_{\text{SPE}}^c(N) = 1-\gamma$$
$$P_{T^2}^c(F) = P_{\text{SPE}}^c(F) = \gamma \tag{29}$$

where γ is the significant level. To obtain the fault probabilities of the new data sample, we should also know the conditional probabilities. In the present article, they are defined as follows:

$$P_{T^2}^c(\mathbf{x}_{\text{new}}|N) = \exp\left\{-\frac{T_{c,\text{new}}^2}{T_{c,\text{lim}}^2}\right\} \quad P_{T^2}^c(\mathbf{x}_{\text{new}}|F) = \exp\left\{-\frac{T_{c,\text{lim}}^2}{T_{c,\text{new}}^2}\right\} \tag{30}$$

$$P_{\text{SPE}}^c(\mathbf{x}_{\text{new}}|N) = \exp\left\{-\frac{SPE_{c,\text{new}}}{SPE_{c,\text{lim}}}\right\}$$
$$P_{\text{SPE}}^c(\mathbf{x}_{\text{new}}|F) = \exp\left\{-\frac{SPE_{c,\text{lim}}}{SPE_{c,\text{new}}}\right\} \tag{31}$$

.

After the monitoring results of the new data sample in each local BPCA model space has been obtained, we can easily calculate the final T^2 and SPE monitoring statistics by combining them through their posterior probabilities, which are given as follows:

$$T_{\text{final,new}}^2 = \sum_{c=1}^{C} p(c|\mathbf{x}_{\text{new}}, \Theta^{\text{opt}})P_{T^2}^c(F|\mathbf{x}_{\text{new}}) \tag{32}$$

$$\text{SPE}_{\text{final,new}} = \sum_{c=1}^{C} p(c|\mathbf{x}_{\text{new}}, \Theta^{\text{opt}})P_{\text{SPE}}^c(F|\mathbf{x}_{\text{new}}) \tag{33}$$

.

On the basis of these two final monitoring statistics, we can easily judge the process behavior by examining if these two statistics have violated their corresponding control limits γ. If their values exceed γ, some fault could be detected, otherwise, the process is normal, and the monitoring procedures should be kept on.

Mode localization

As a part of the process monitoring task, mode localization is also very important. Knowing the exact operation mode of the monitored data sample, the engineer can easily locate the production of the process, and the problem will also be easily found if there is any. In this subsection, a new mode localization method is proposed based on the mixture BPCA monitoring framework. Having calculated the posterior probabilities of the monitored data sample corresponds to different operation modes, we can simply locate the mode information through these posterior probabilities, thus the mode with the largest posterior probability value should be determined as the current operation mode. However, this method may not function well when an unknown operation mode happened in the process. On the basis of the posterior probability localization method, the data sample from the new operation mode will be assigned to one of the known operation mode, which is not the real case. In another word, mode localization based on the posterior probability is not reliable.

To emphasize the reliability for mode localization, we intend to use the joint probability analysis method. Similar to the posterior probability, successful localization will be obtained if the data sample was collected under known operation modes. However, different from the posterior probability method, the joint probability analysis can also successfully identify the change of the process. That is, if the process is changed to a new operation mode, the joint probabilities of the monitored data sample with all known operation modes will approach to zero. To get the joint probability values of the new data sample $\mathbf{x}_{\text{new}}$ with different operation modes, the conditional probabilities should be calculated first, which follows a multivariate Gaussian distribution. The conditional probability function are given as:

$$p(\mathbf{x}|c) = \int p(\mathbf{x}|\mathbf{t}, c)p(\mathbf{t}|c)p(c)d\mathbf{t} \tag{34}$$

.

For the new data sample $\mathbf{x}_{\text{new}}$, its conditional probability can be calculated as $p(\mathbf{x}_{\text{new}}|c) = p(\mathbf{x}|c)_{\mathbf{x}_{\text{new}}}$. Then, the joint probabilities values of $\mathbf{x}_{\text{new}}$ can be given as follows:

$$\text{JP}_{c,\text{new}} = p(\mathbf{x}_{\text{new}}|c)p(c|\mathbf{x}_{\text{new}}, \Theta^{\text{opt}}) \tag{35}$$

where $c = 1,2,\ldots,C$. Through analysis of Eq. 35, it can be inferred that if the process is changed to a new operation mode, the value of the conditional probability of $\mathbf{x}_{\text{new}}(p(\mathbf{x}_{\text{new}}|c))$ will be very small. Although the mode posterior probability value $p(c|\mathbf{x}_{\text{new}},\Theta^{\text{opt}})$ will not decrease to zero, since they are restricted to $\sum_{c=1}^{C} p(c|\mathbf{x}_{\text{new}}, \Theta^{\text{opt}}) = 1$, by multiplying the mode posterior probability value with the conditional probability value, the joint probability value of $\mathbf{x}_{\text{new}}$ will quickly decrease to zero. Hence, if all of the joint probability values are approached to zero, a new operation mode can be judged.

Case Studies

In this section, two case studies are provided to evaluate the efficiency of the proposed method. The first one is a numerical example with six variables, and the other one is a real industrial polypropylene production process application case.

A numerical example

This numerical example consists of six variables, which are driven by 3 Gaussian latent variables, the relationships between them are given by:

$$\begin{aligned} \mathbf{x}_1 &= \mathbf{P}_1\mathbf{t}_1 + \mathbf{e}_1, \quad \mathbf{t}_1 \in N(-1,1), \quad \mathbf{e}_1 \in N(0,0.1) \\ \mathbf{x}_2 &= \mathbf{P}_2\mathbf{t}_2 + \mathbf{e}_2, \quad \mathbf{t}_2 \in N(0,1), \quad \mathbf{e}_2 \in N(0,0.1) \\ \mathbf{x}_3 &= \mathbf{P}_3\mathbf{t}_3 + \mathbf{e}_3, \quad \mathbf{t}_3 \in N(1,1), \quad \mathbf{e}_3 \in N(0,0.1) \end{aligned} \tag{36}$$

where $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3 \in R^6$, $\mathbf{t}_1$, $\mathbf{t}_2$, $\mathbf{t}_3 \in R^3$ are Gaussian distributed latent variables with different mean values, and $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3 \in R^6$ are Gaussian process noises with an equal variance 0.1. The three loading matrices $\mathbf{P}_1$, $\mathbf{P}_2$, and $\mathbf{P}_3$ are randomly selected as 6×3 matrices. Therefore, three operation modes can be simulated through Eq. 36. To build the mixture BPCA and the mixture PPCA models, 200 data samples of each cluster have been generated, thus a total of 600 data samples are used for model construction. To simulate the process fault, two cases are assumed as follows:

Case 1: A step bias by adding 0.5 to the second variable of the first operation mode is introduced starting from sample 101 to 200.

Case 2: A ramp change by adding $0.01(i-100)$ to the first variable of the second operation mode is introduced starting from sample number $i = 101$ to $i = 200$.

First, the two mixture models are constructed with full possible latent variables direction. Denote the loading matrices of mixture PPCA as $\mathbf{O}_1$, $\mathbf{O}_2$, and $\mathbf{O}_3$, the hinton diagrams of the loading matrices of both models are given in Figure 1. In this figure, the red rectangle represents a positive value of the element in the corresponding matrix, and the green one represents a negative value. Seen from the left three sub-figures, it is very clear that last two latent variable directions have been switched off by the mixture BPCA model. In contrast, the importance of the latent variable direction in mixture PPCA loading matrices cannot be easily differentiated, since the values of their elements are comparative. Therefore, through the Bayesian regularization method, the number of the effective latent variables can be determined automatically. Also, the importance of each effective latent variable direction can be determined by the value of the hyperparameter α. However, due to the inherent defect of the PPCA method, the number of effective latent variables should be determined by other methods, such as cross validation.

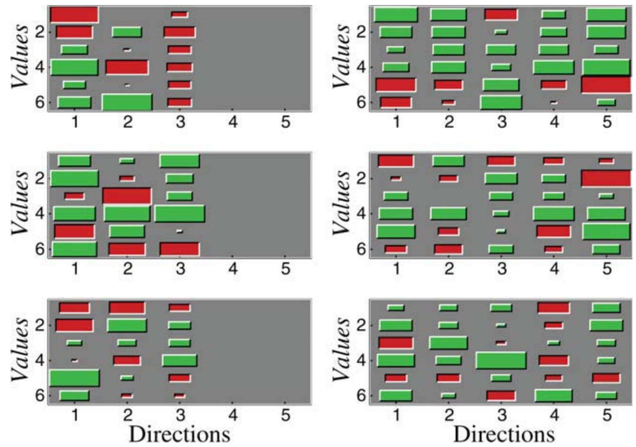

Figure 1. Hinton diagrams of loading matrices of both mixture PPCA and mixture BPCA.

Left: Mixture BPCA and Right: mixture PPCA. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

To test the monitoring performance of the proposed method, the two generated fault cases are used. The results of the first fault by both mixture PPCA and mixture BPCA methods are shown in Figure 2, in which the control limits of both monitoring statistics have been selected as 99%. It can be seen that this fault can hardly be detected by the mixture PPCA model, since both two monitoring statistics are under their control limits. In contrast, the fault can be successfully detected by the mixture BPCA model after it has

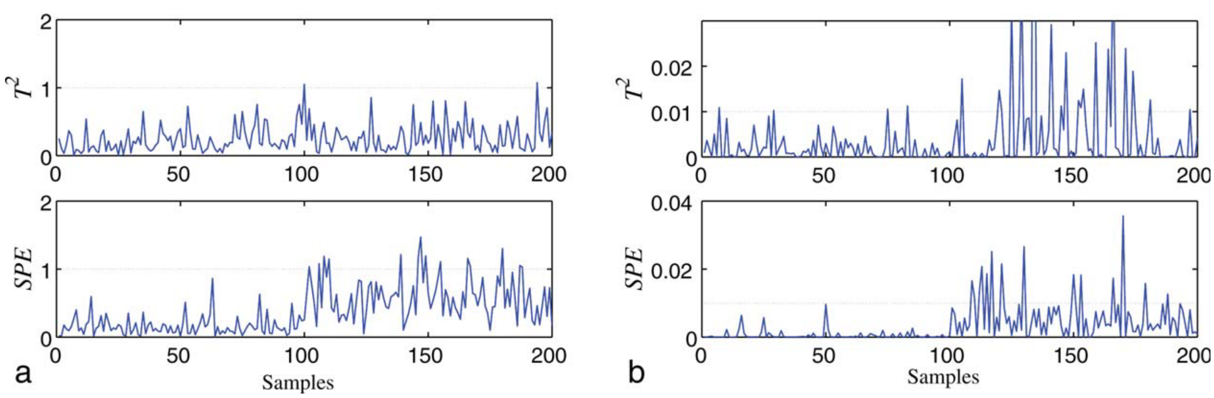

Figure 2. Monitoring results of fault 1.

(a) Mixture PPCA and (b) mixture BPCA. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

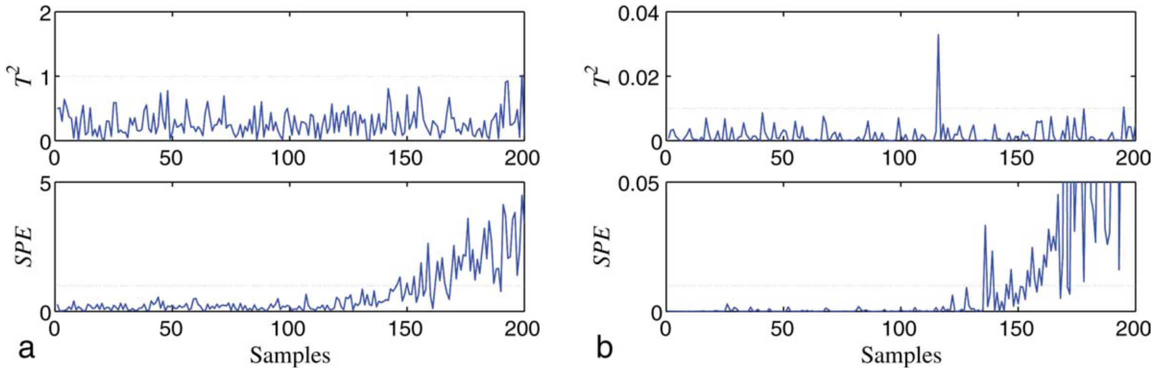

Figure 3. Monitoring results of fault 2.

(a) Mixture PPCA and (b) mixture BPCA. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

been introduced into the process. Similarly, the monitoring performance of the second fault can also be improved by the proposed method, which is given in Figure 3. Although the SPE statistic of both methods cannot detect this fault, the T^2 statistic of mixture BPCA can detect the fault earlier than that of mixture BPCA. To examine the mode information of these two fault cases, their mode localization results are provided in Figure 4. Judging from Figure 4a, we can infer that it first ran under mode one, and then some fault was introduced. The results presented in Figure 4b exhibits that the second fault case first ran under mode two, and then the fault happened.

To simulate mode change and new operation mode cases, two more datasets have also been generated, which are listed as follows:

Case 3: Hundred data samples are collected under the first operation mode, then the process is changed to the second operation mode, a same number of data sample are collected.

Case 4: A new operation mode $\mathbf{x}_4 = \mathbf{P}_4\mathbf{t}_4 + \mathbf{e}_4$, $\mathbf{t}_4 \in N(-3,1)$, $\mathbf{e}_4 \in N(0,0.1)$ is introduced.

The mode localization results of both two cases are given in Figure 5. It is straightforward that the mode change can be detected by the proposed method, since both operation modes have been modeled. For the new operation mode, the mode localization method can also give correct result, which can be found in Figure 5b. Because of the small probability values of all the three operation modes, the new case should be deemed as a new operation mode.

Polypropylene production process application study

As an important material, polypropylene has been widely used in many fields, such as chemical industry, light industry, medical industry, etc. With the increased market demands of different product types, producing different characteristics of polypropylene has become a key issue of this industry process. A typical polypropylene production device always contains a catalytic body system, which comprises of $TiCl_4$, triethylaluminum (TEAL), and diphenyldimethoxysilane (DONOR). To produce different brands of productions, three reactors are connected in series. The flowchart of this process is given in Figure 6. As seen in the flowchart, this process consists of four major units: the catalytic body system and three reactors. To record the characteristic of this process, over 40 variables are measured online. However, in the present study, we have selected 14 important variables for process monitoring purpose, which are highly correlated with the final product quality. These 14 monitored variables are listed in Table 1.

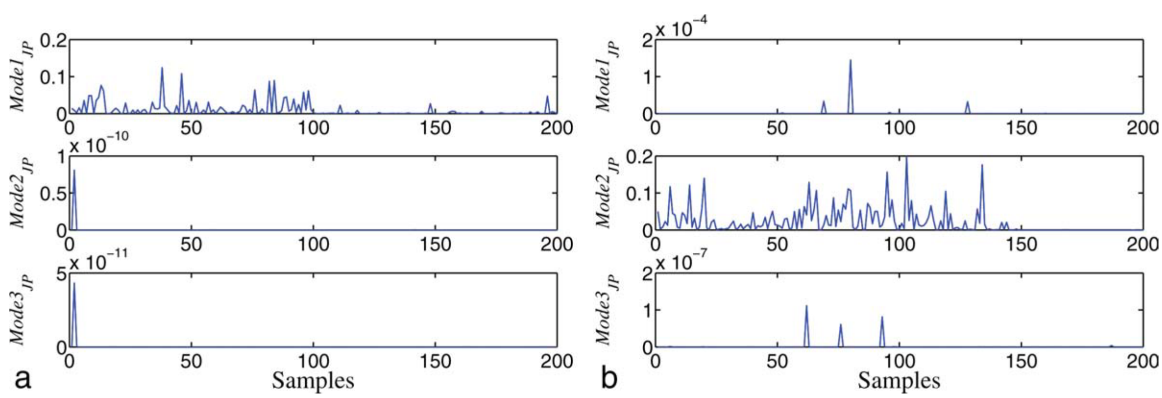

Figure 4. Mode localization results.

(a) The first fault and (b) the second fault. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

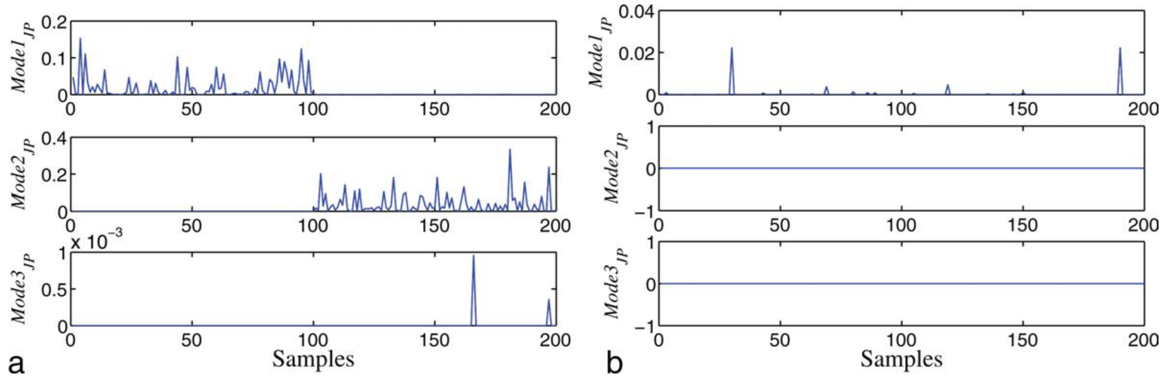

Figure 5. Mode localization results of mode change and new mode cases.

(a) Mode change and (b) new mode. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

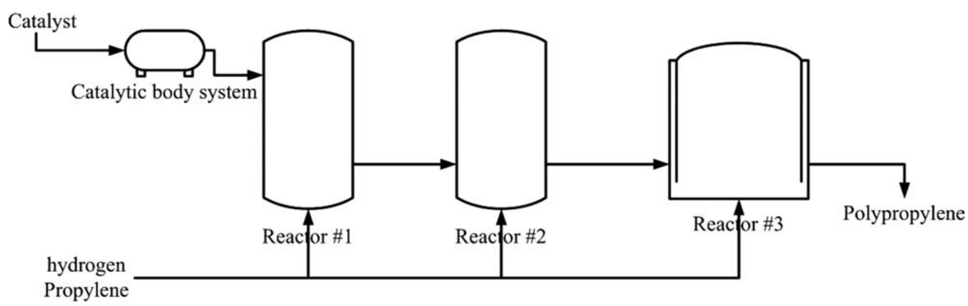

Figure 6. Flowchart of the polypropylene production process.

To develop the monitoring model, three different types of process datasets have been collected, each of which contains 200 data samples. Figure 7 exhibits the data characteristic of these three dataset, through which we can easily find that these three datasets belong to three different operation modes. To evaluate the monitoring performance of the proposed method, two additional fault datasets have also been collected, both which consist of 200 data samples. These two fault cases are described as follows:

Case 1: The process initially operated under the first operation mode, 100 data samples have been collected. Then, a step change of the TEAL flow has been caused, another 100 data samples are recorded.

Case 2: The process initially operated under the third operation mode, 100 data samples have been collected. Then a step change of the TEAL flow has been caused, another 100 data samples are recorded.

After the process datasets have been screened and preprocessed, we are ready to construct the monitoring models. Running the modeling development process 50 times, the automatic selected dimensionalities of three different subspaces are plotted in Figure 8. The mean values of these three

Table 1. Monitoring Variables in Polypropylene Production Process

No.	Measured Variables
1	Hydrogen concentration of the first reactor
2	Hydrogen concentration of the second reactor
3	Density of the first reactor
4	Density of the second reactor
5	TEAL flow
6	DONOR flow
7	Atmer-163 flow
8	Propylene feed of the first reactor
9	Propylene feed of the second reactor
10	Power for the first reactor
11	Power for the second reactor
12	Lever of the second reactor
13	Temperature of the first reactor
14	Temperature of the second reactor

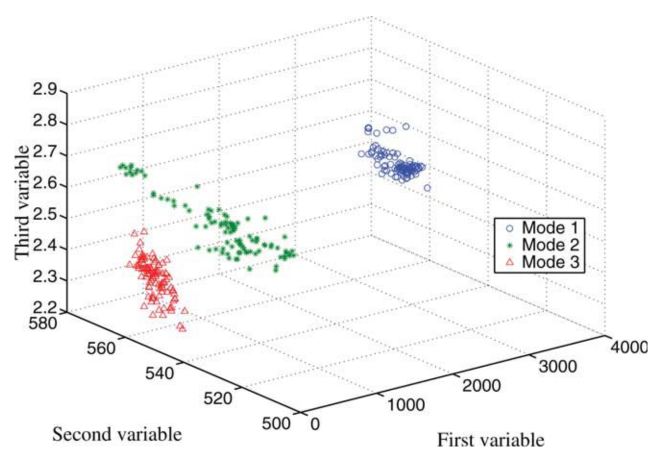

Figure 7. Three-dimensional data characteristic of three different datasets in polypropylene production process.

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

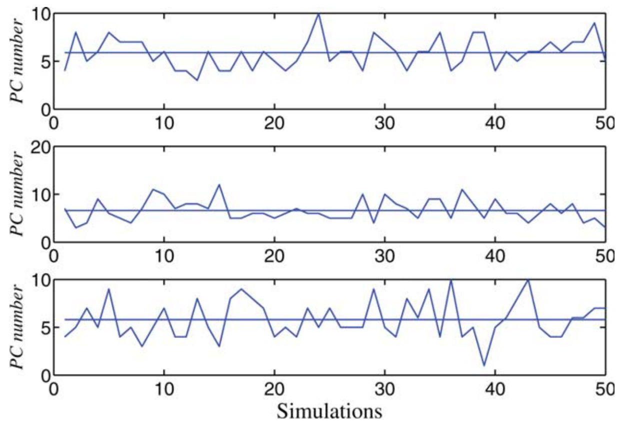

Figure 8. Automatic determined number of effective latent variables in each subspace.

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

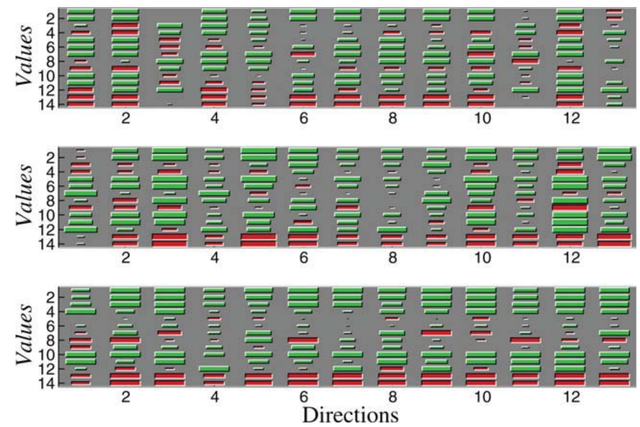

Figure 10. Hinton diagrams of the loading matrix of mixture PPCA.

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

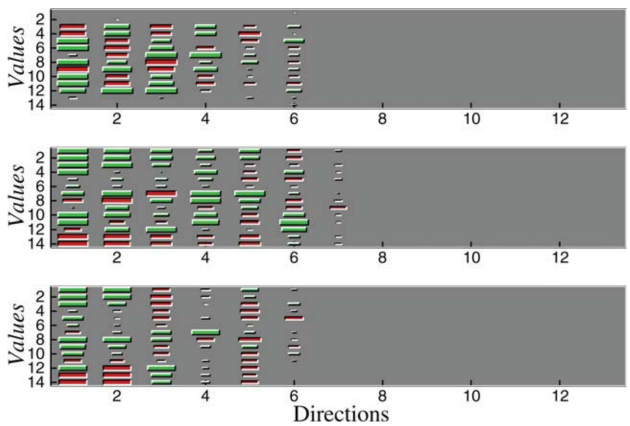

Figure 9. Hinton diagrams of the loading matrix of mixture BPCA.

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

selected dimensionalities are 5.88, 6.91, and 5.94. Thus, the three dimensionalities can be determined as 6, 7, and 6 for monitoring model construction. Similarly, the hinton diagrams of both loading matrices of the mixture PPCA and mixture BPCA models are given in Figures 9 and 10, respectively. As been expected, the dimensionality of the latent variable space can be easily determined by the mixture BPCA model, which can be clearly obtained from Figure 9. However, again, the latent variable dimensionality cannot be automatically determined by the mixture PPCA model. Depending on the results presented in Figure 10, it is very difficult to select effective principal components. However, to be fair, the same numbers of principal components are selected in each sub model space for the mixture PPCA model development. For process monitoring, T^2 and SPE statistic of both methods have been constructed, the control limits of all monitoring statistics are selected as 99%.

Depending on multiple running results of both monitoring methods, we have found that the monitoring performance of both fault cases have been greatly improved by the mixture BPCA-based method. One realization of the first fault case obtained by both two methods are given in Figure 11, through which one can see that both two monitoring statistics of mixture BPCA can successfully detect the fault. However, only the SPE statistic of mixture PPCA can continuously detect the fault, the monitoring performance of the T^2

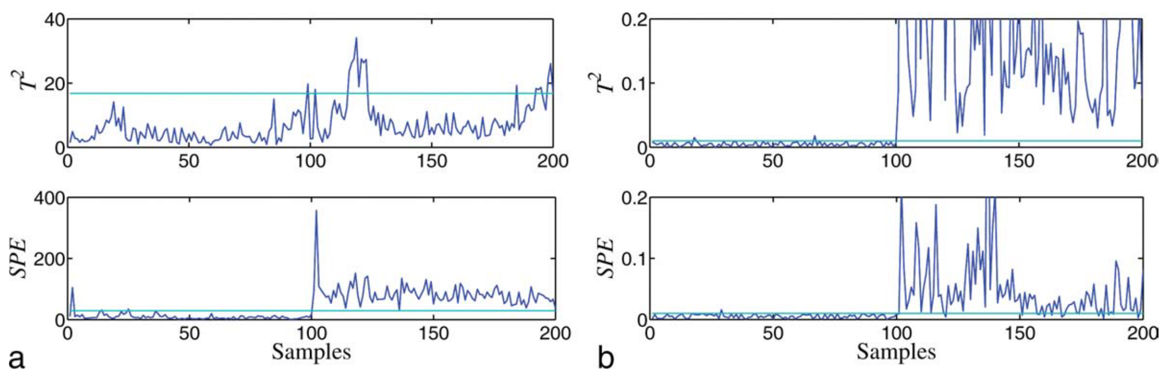

Figure 11. Monitoring results of the first fault case.

(a) Mixture PPCA and (b) mixture BPCA. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

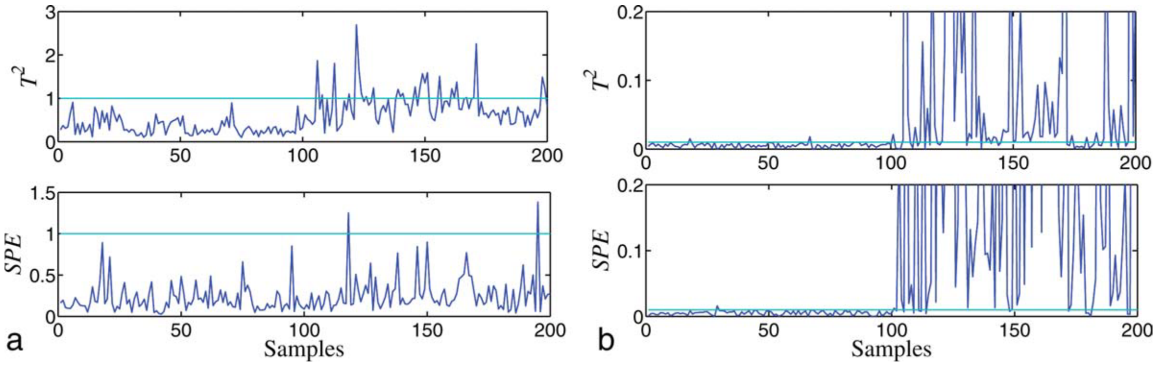

Figure 12. Monitoring results of the second fault case.

(a) Mixture PPCA and (b) mixture BPCA. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

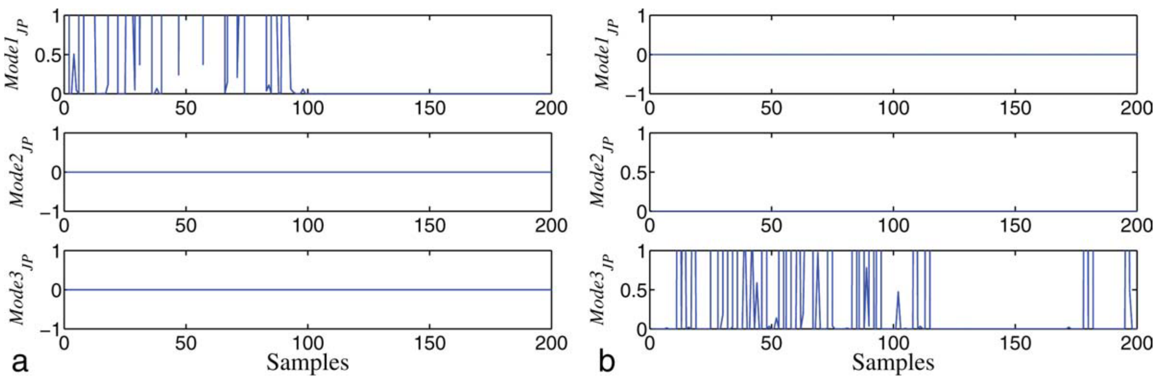

Figure 13. Mode localization results.

(a) The first fault case and (b) the second fault case. [Color figure can be viewed in the online issue, which is available at wileyonline library.com.]

statistic is very poor. Similarly, the monitoring performance improvement of the mixture BPCA method for the second fault case can also be obtained, which is clearly presented in Figure 12. Compared to mixture PPCA, which can hardly detect this fault, both two monitoring statistics of mixture BPCA can detect it immediately after the fault happened. The mode localization results of both two fault cases are shown in Figure 13, depending on which one can easily conclude that these two fault cases happened under the first and the third operation modes, respectively.

In addition, the noise level of this process is examined. A total of 300 data samples of each operation mode have been collected for model development and noise variance estimation. To test the efficiency of the mixture BPCA method under limited data sample cases, we intend to change the number of modeling data samples from 30 to 300. Therefore, a total of 28 noise estimation results have been generated through the mixture BPCA modeling. These estimated results of the noise variance are shown together in Figure 14, through which we can find that the modeling efficiency of the mixture BPCA method is rarely degraded when the number of data samples is changed. Therefore, when the number of process data samples is limited, which could possibly happen in multimode processes, the mixture BPCA method can still provide efficient modeling and monitoring results.

Conclusions

In the present article, a mixture Bayesian regularization method of PPCA has been elaborately demonstrated for

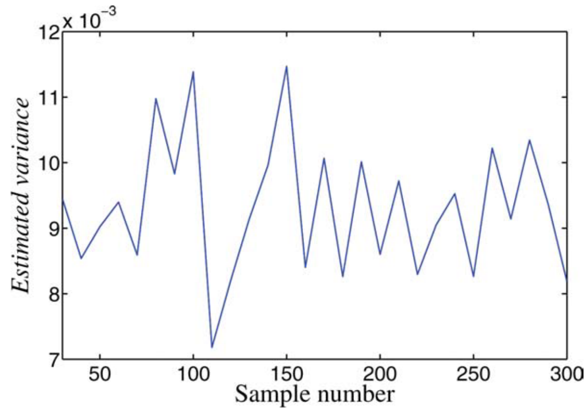

Figure 14. Estimated noise variance under different numbers of data samples.

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

multimode process monitoring purpose. Compared with the traditional PPCA method, the main advantage of the new method is that the effective dimensionality of the latent variable in each local space can be determined automatically. Also, it can easily determine the importance of each retained principal component, which is controlled by a hyperparameter. In addition, a new mode localization approach has been proposed, which is based on the joint probability analysis method. On the basis of the illustrated results of two case studies, the performances of monitoring and mode localization have both been greatly improved by the proposed method. However, it is worth to notice that this method cannot determine if the process is operated under a new operation mode or some other changes have happened. Without appropriate process knowledge or expert experience, one can hardly obtain the exact condition of the process. How to incorporate some necessary process information into the developed monitoring scheme is one of our future research directions. Besides, how to extend the proposed method to its nonlinear and non-Gaussian forms is also included in our future research work.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant Nos. 60774067, 60736021) and the National 863 High Technology Research and Development Program of China (2009AA04Z154).

Literature Cited

1. Qin SJ. Statistical process monitoring: basics and beyond. *J Chemometrics.* 2003;17:480–502.
2. Wang X, Kruger U, Lennox B. Recursive partial least squares algorithms for monitoring complex industrial processes. *Contr Eng Prac.* 2003;11:613–632.
3. Kruger U, Dimitriadis G. Diagnosis of process faults in chemical systems using a local partial least squares approach. *AIChE J.* 2008;54:2581–2596.
4. Zhao CH, Wang FL, Mao ZH, Lu NY, Jia MX. Quality prediction based on phase-specific average trajectory for batch processes. *AIChE J.* 2008;54:693–705.
5. Zhang YW, Qin SJ. Improved nonlinear fault detection technique and statistical analysis. *AIChE J.* 2008;54:3207–3220.
6. AlGhazzawi A, Lennox B. Monitoring a complex refining process using multivariate statistics. *Control Eng Pract.* 2008;16:294–307.
7. Tipping ME, Bishop CM. Probabilistic principal component analysis. *J Royal Stat Soc.* 1999;61:611–622.
8. Kim DS, Lee IB. Process monitoring based on probabilistic PCA. *Chem Intell Lab Syst.* 2003;67:109–123.
9. Tipping ME, Bishop CM. Mixtures of probabilistic principal component analysis. *Neural Comput.* 1999;11:443–482.
10. Chen JH, Liu JL. Mixture principal component analysis models for process monitoring. *Ind Eng Chem Res.* 1999;38:1478–1488.
11. Chen JH, Liu JL. Using mixture principal component analysis networks to extract fuzzy rules from data. *Ind Eng Chem Res.* 2000; 39:2355–2367.
12. Choi SW, Park JH, Lee IB. Process monitoring using a Gaussian mixture model via principal component analysis and discriminant analysis. *Comput Chem Eng.* 2004;28:1377–1387.
13. Choi SW, Martim EB, Morris AJ. Fault detection based on a maximum-likelihood principal component analysis (PCA) mixture. *Ind Eng Chem Res.* 2005;44:2316–2327.
14. Bishop CM, Bayesian PCA. *Advanced in Neural Information Proceeding Systems.* 1999;11:382–388.
15. Nounou MN, Bakshi BR, Goel PK, Shen XT. Bayesian principal component analysis. *J Chemometrics.* 2002;16:576–595.
16. Bishop CM. *Pattern Recognition and Machine Learning.* Springle, 2006; 559–603.

Appendix A: Derivation of the EM Algorithm for the Bayesian PCA Method

In the E-step of the EM algorithm, the expected sufficient statistics of the latent variable can be calculates as follows:

$$E(\hat{\mathbf{t}}|\mathbf{x}) = E(\hat{\mathbf{t}}) + \mathrm{cov}(\hat{\mathbf{t}},\mathbf{x})[\mathrm{var}(\mathbf{x})]^{-1}[\mathbf{x} - E(\mathbf{x})] \quad \text{(A1)}$$

$$E(\hat{\mathbf{t}}\hat{\mathbf{t}}^T|\mathbf{x}) = \mathrm{var}(\hat{\mathbf{t}}|\mathbf{x}) + E(\hat{\mathbf{t}}|\mathbf{x})E^T(\hat{\mathbf{t}}|\mathbf{x})$$
$$\mathrm{var}(\hat{\mathbf{t}}|\mathbf{x}) = \mathrm{var}(\hat{\mathbf{t}}) - \mathrm{cov}(\hat{\mathbf{t}},\mathbf{x})\mathrm{var}^{-1}(\mathbf{x})\mathrm{cov}(\mathbf{x},\hat{\mathbf{t}}) \quad \text{(A2)}$$

where $E(\hat{\mathbf{t}}) = 0$, $\mathrm{var}(\mathbf{x}) = \mathbf{PP}^T + \beta^{-1}\mathbf{I}$, $E(\mathbf{x}) = 0$, $\mathrm{vat}(\hat{\mathbf{t}}) = \mathbf{I}$, and

$$\mathrm{cov}(\hat{\mathbf{t}},\mathbf{x}) = E(\hat{\mathbf{t}}\mathbf{x}^T) - E(\hat{\mathbf{t}})E(\mathbf{x}^T) = \mathbf{P}^T \quad \text{(A3)}$$

$$\mathrm{cov}(\mathbf{x},\hat{\mathbf{t}}) = E(\mathbf{x}\hat{\mathbf{t}}^T) - E(\mathbf{x})E(\hat{\mathbf{t}}^T) = \mathbf{P}. \quad \text{(A4)}$$

Then, Eqs. A1 and A2 come to

$$E(\hat{\mathbf{t}}|\mathbf{x}) = \mathbf{P}^T(\mathbf{PP}^T + \beta^{-1}\mathbf{I})^{-1}\mathbf{x} = (\mathbf{P}^T\mathbf{P} + \beta^{-1}\mathbf{I})^{-1}\mathbf{P}^T\mathbf{x} \quad \text{(A5)}$$

$$E(\hat{\mathbf{t}}\hat{\mathbf{t}}^T|\mathbf{x}) = \mathbf{I} - \mathbf{P}^T(\mathbf{PP}^T + \beta^{-1}\mathbf{I})^{-1}\mathbf{P} + E(\hat{\mathbf{t}}|\mathbf{x})E^T(\hat{\mathbf{t}}|\mathbf{x})$$
$$= \beta^{-1}(\mathbf{P}^T\mathbf{P} + \beta^{-1}\mathbf{I})^{-1} + E(\hat{\mathbf{t}}|\mathbf{x})E^T(\hat{\mathbf{t}}|\mathbf{x}). \quad \text{(A6)}$$

If we define $\mathbf{M} = \mathbf{P}^T\mathbf{P} + \beta^{-1}\mathbf{I}$, Eq. 10 can be obtained.

Notice Eqs. 3 and 9, maximizing the log posterior distribution function with respect to $\mathbf{P}$ and β, and set them to zero, thus

$$\frac{\partial[\ln p(\mathbf{P}|\mathbf{X})]}{\partial \mathbf{P}} = 0 \quad \text{(A7)}$$

$$\frac{\partial[\ln p(\mathbf{P}|\mathbf{X})]}{\partial \beta} = 0. \quad \text{(A8)}$$

Denote $\Lambda = \beta^{-1}\mathbf{I}$, Eqs. A7 and A8 can be calculated as:

$$\sum_{i=1}^{n}[\Lambda^{-1}(\mathbf{x}_i - \mathbf{P}\hat{\mathbf{t}}_i)\hat{\mathbf{t}}_i^T] - \mathbf{PA} = 0 \quad \text{(A9)}$$

$$\Lambda^T\mathbf{PA} - \sum_{i=1}^{n}[(\mathbf{x}_i - \mathbf{P}\hat{\mathbf{t}}_i)\hat{\mathbf{t}}_i^T] = 0. \quad \text{(A10)}$$

Simplifying Eqs. A9 and A10, we can easily obtain

$$\hat{\mathbf{P}} = [\sum_{i=1}^{n}\mathbf{x}_i\hat{\mathbf{t}}_i]\left[\sum_{i=1}^{n}\hat{\mathbf{t}}_i\hat{\mathbf{t}}_i^T + \beta^{-1}\mathbf{A}\right]^{-1} \quad \text{(A11)}$$

$$\hat{\beta} = \frac{nm}{\sum_{i=1}^{n}\{\mathbf{x}_i^T\mathbf{x}_i - 2\hat{\mathbf{t}}_i)\mathbf{P}^T\mathbf{x}_i + Tr[(\hat{\mathbf{t}}_i\hat{\mathbf{t}}_i^T)\mathbf{P}^T\mathbf{P}]\}}. \quad \text{(A12)}$$

Appendix B: Derivation of the EM Algorithm for the Mixture Bayesian PCA Method

Similar to the single BPCA model, the E-step of the mixture BPCA model can be developed as follows. Thus, the expected sufficient statistics of the latent variable can be calculates as:

$$E(\hat{\mathbf{t}}|\mathbf{x},c,\Theta_{\text{old}}) = E(\hat{\mathbf{t}}|c,\Theta_{\text{old}}) + \text{cov}(\hat{\mathbf{t}},\mathbf{x}|c,\Theta_{\text{old}})[\text{var}(\mathbf{x}|c,\Theta_{\text{old}})]^{-1}[\mathbf{x} - E(\mathbf{x}|c,\Theta_{\text{old}})] \quad \text{(B1)}$$

$$E(\hat{\mathbf{t}}\hat{\mathbf{t}}^T|\mathbf{x},c,\Theta_{\text{old}}) = \text{var}(\hat{\mathbf{t}}|\mathbf{x},c,\Theta_{\text{old}}) + E(\hat{\mathbf{t}}|\mathbf{x},c,\Theta_{\text{old}})E^T(\hat{\mathbf{t}}|\mathbf{x},c,\Theta_{\text{old}})\text{var}(\hat{\mathbf{t}}|\mathbf{x},c,\Theta_{\text{old}}) = \text{var}(\hat{\mathbf{t}}|c,\Theta_{\text{old}}) - \text{cov}(\hat{\mathbf{t}},\mathbf{x}|c,\Theta_{\text{old}})[\text{var}(\mathbf{x}|c,\Theta_{\text{old}})]^{-1}\text{cov}(\mathbf{x},\hat{\mathbf{t}}|c,\Theta_{\text{old}}) \quad \text{(B2)}$$

where $c = 1,2,\ldots,C$, $E(\hat{\mathbf{t}}|c,\Theta_{\text{old}}) = 0$, $\text{var}(\mathbf{x}|c,\Theta_{\text{old}}) = \mathbf{P}_c\mathbf{P}_c^T + \beta_c^{-1}\mathbf{I}$, $E(\mathbf{x}|c,\Theta_{\text{old}}) = 0$, $\text{var}(\hat{\mathbf{t}}|c,\Theta_{\text{old}}) = \mathbf{I}$, and

$$\text{cov}(\hat{\mathbf{t}},\mathbf{x}|c,\Theta_{\text{old}}) = E(\hat{\mathbf{t}}\mathbf{x}^T|c,\Theta_{\text{old}}) - E(\hat{\mathbf{t}}|c,\Theta_{\text{old}})E(\mathbf{x}^T|c,\Theta_{\text{old}}) = \mathbf{P}_c^T \quad \text{(B3)}$$

$$\text{cov}(\mathbf{x},\hat{\mathbf{t}}|c,\Theta_{\text{old}}) = E(\mathbf{x}\hat{\mathbf{t}}^T|c,\Theta_{\text{old}}) - E(\mathbf{x}|c,\Theta_{\text{old}})E(\hat{\mathbf{t}}^T|c,\Theta_{\text{old}}) = \mathbf{P}_c \quad \text{(B4)}$$

.

Then, Eqs. B1 and B2 come to

$$E(\hat{\mathbf{t}}|\mathbf{x},c,\Theta_{\text{old}}) = \mathbf{P}_c^T(\mathbf{P}_c\mathbf{P}_c^T + \beta_c^{-1}\mathbf{I})^{-1}\mathbf{x} = (\mathbf{P}_c^T\mathbf{P}_c + \beta_c^{-1}\mathbf{I})^{-1}\mathbf{P}_c^T\mathbf{x} \quad \text{(B5)}$$

$$E(\hat{\mathbf{t}}\hat{\mathbf{t}}^T|\mathbf{x},c,\Theta_{\text{old}}) = \mathbf{I} - \mathbf{P}_c^T(\mathbf{P}_c\mathbf{P}_c^T + \beta_c^{-1}\mathbf{I})^{-1}\mathbf{P}_c + E(\hat{\mathbf{t}}|\mathbf{x},c,\Theta_{\text{old}})E^T(\hat{\mathbf{t}}|\mathbf{x},c,\Theta_{\text{old}}) = \beta_c^{-1}(\mathbf{P}_c^T\mathbf{P}_c + \beta_c^{-1}\mathbf{I})^{-1} + E(\hat{\mathbf{t}}|\mathbf{x},c,\Theta_{\text{old}})E^T(\hat{\mathbf{t}}|\mathbf{x},c,\Theta_{\text{old}}). \quad \text{(B6)}$$

Different from the EM algorithm for the single BPCA model, the proportion value of each mode $p(c|\Theta_{\text{new}})$ should also be updated through the M-step in the mixture BPCA model structure, which are calculated as follows:

$$p(c|\Theta_{\text{new}}) = \frac{1}{n}\sum_{i=1}^{n} p(c|\mathbf{x}_i,\Theta_{\text{old}}) \qquad p(c|\mathbf{x},\Theta_{\text{old}}) = \frac{p(\mathbf{x}|c,\Theta_{\text{old}})p(c|\Theta_{\text{old}})}{p(\mathbf{x}|\Theta_{\text{old}})}. \quad \text{(B7)}$$

The updated value $\mathbf{P}_c^{\text{new}}$, β_c^{new} can be determined by maximizing the log posterior probability function. Setting the derivative of ln $p(\mathbf{P}|\mathbf{X})$ respect to $\mathbf{P}_c$ and β_c as zero

$$\frac{\partial[\ln p(\mathbf{P}|\mathbf{X})]}{\partial \mathbf{P}_c} = 0 \quad \text{(B8)}$$

$$\frac{\partial[\ln p(\mathbf{P}|\mathbf{X})]}{\partial \beta_c} = 0. \quad \text{(B9)}$$

Refer to the parameter updating process in Appendix A, we can easily obtain

$$\mathbf{P}_c^{\text{new}} = \left(\sum_{i=1}^{n} p(c|\mathbf{x}_i,\Theta_{\text{old}})\mathbf{x}_i E(\hat{\mathbf{t}}_i^T|\mathbf{x}_i,c,\Theta_{\text{old}})\right) \times \left(\sum_{i=1}^{n} p(c|\mathbf{x}_i,\Theta_{\text{old}})E(\hat{\mathbf{t}}_i\hat{\mathbf{t}}_i^T|\mathbf{x}_i,c,\Theta_{\text{old}}) + p(c|\Theta_{\text{new}})\beta_c^{-1}\mathbf{A}_c\right)^{-1} \quad \text{(B10)}$$

$$\beta_c^{\text{new}} = \frac{nm}{\sum_{i=1}^{n} p(c|\mathbf{x}_i,\Theta_{\text{old}})\{\mathbf{x}_i^T\mathbf{x}_i - 2E^T(\hat{\mathbf{t}}_i|\mathbf{x}_i)\mathbf{P}_c^{\text{new}T}\mathbf{x}_i + Tr[E(\hat{\mathbf{t}}_i\hat{\mathbf{t}}_i^T|\mathbf{x}_i)\mathbf{P}_c^{\text{new}T}\mathbf{P}_c^{\text{new}}]\}}. \quad \text{(B11)}$$

Manuscript received Sept. 15, 2009, and revision received Jan. 21, 2010.