Monitoring of Processes with Multiple Operating Modes through Multiple Principle Component Analysis Models

Shi Jian Zhao,[†] Jie Zhang,*[,‡] and Yong Mao Xu[†]

Department of Automation, Tsinghua University, Beijing 100084, China, and School of Chemical Engineering and Advanced Materials, University of Newcastle, Newcastle upon Tyne NE1 7RU, U.K.

Because many of the current multivariate statistical process monitoring (MSPM) techniques are based on the assumption that the process has one nominal operating region, the application of these MSPM approaches to an industrial process with multiple operating modes would always trigger continuous warnings even when the process itself is operating under another normal steady-state operating conditions. Adopting principal angles to measure the similarities of any two models, this paper proposes a multiple principal component analysis model based process monitoring methodology. Some popular multivariate statistical measurements such as squared prediction error and its control limit can be incorporated straightforwardly to facilitate process monitoring. The efficiency of the proposed technique is demonstrated through application to the monitoring of the Tennessee−Eastman challenge process and an industrial fluidized catalytic cracking unit. The proposed scheme can significantly reduce the amount of false alarms while tracking the process adjustment.

1. Introduction

Statistical process control forms the basis of performance monitoring and the detection of process malfunctions. With immense amounts of data collected into modern monitoring and control systems, there has been increasing interest in pursuing methods that are capable of grasping the essentials in these highly correlated data. Owing to the adoption of some powerful multivariate projection techniques such as principal component analysis (PCA) and projection to latent structures (PLS), multivariate statistical process monitoring (MSPM) has been applied to diverse processes and achieved great success in the past decade.[1,2]

The current MSPM approaches, however, have problems when applied to a process with multiple operating modes where the process characteristics vary significantly when operating at different modes, accompanied by the typical phenomenon of numerous false alarms. It has been observed in practice that the application of the conventional MSPM technique to a process with multiple operating modes can trigger continuous warnings even when the process itself is operating under another steady-state nominal operating conditions.

This is not accidental because many of the current MSPM techniques are based on the assumption that the process has one nominal operating region, at most with some slight normal deviations because of the so-called slow time-varying properties that can be handled successfully by recursive PCA (RPCA).[3] However, the RPCA approach lacks the ability of coping with processes with multiple operating modes because of the various changes in product specifications, feed flow rate, compositions, and set points. In most cases, RPCA fails to distinguish the conscious process operating condition changes from the possible process malfunctions. This is particularly true when the feed is changed manually and the process varies quite significantly.

Recently, some approaches have been proposed to address the issues associated with multiple operating modes under certain relatively restrictive assumptions. For example, on the basis of the partial common principal component model, which requires the first A retained principal components to be common to all groups,[4] Lane et al.[5] adopted a common subspace model, which demands that all of the retained eigenvectors of different groups span the same subspace, to monitor a multiproduct semibatch process. On the other hand, assuming that the numbers of the retained eigenvectors are the same in all of the clusters, Hwang and Han[6] proposed a monitoring methodology using hierarchical clustering and a super-PCA model.

To enhance the monitoring performance and reduce false alarms, a multiple PCA model based process monitoring method is proposed in this paper. Multiple PCA models are developed for different operating modes, and the principal angle is used to measure the similarities between any two PCA models that usually have different numbers of retained principal components. In addition, this definition is extended to measure the similarities between more than two groups of PCA models simultaneously with different retained numbers of principal components. Within this proposed technique, some popular multivariate statistics such as squared prediction error (SPE) and its control limit can thus be determined directly to facilitate process monitoring.

This paper is organized as follows. In the next section, the definition and calculation scheme for the principal angle are introduced. This metric is then utilized to measure the similarities between two PCA models. A multiple PCA model based MSPM technique is proposed in section 3. The efficiency of the proposed technique is demonstrated by application to the Tennessee−Eastman challenge process (TECP) and an industrial fluid-

* To whom correspondence should be addressed. Tel.: +44-191-2227240. Fax: +44-191-2225292. E-mail: jie.zhang@newcastle.ac.uk.

[†] Tsinghua University.

[‡] University of Newcastle.

10.1021/ie0497893 CCC: $27.50
Published on Web 09/30/2004

ized catalytic cracking (FCC) unit in section 4, and section 5 concludes this paper.

2. Similarity between PCA Models

2.1. Principal Angles. It is well-known in linear algebra that the angle between two nonzero vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathscr{R}^m$ is defined as θ, which satisfies

$$\cos\theta = \frac{|\mathbf{u}^{\mathrm{T}}\mathbf{v}|}{||\mathbf{u}||_2||\mathbf{v}||_2}$$

and $\theta \in [0, \pi/2]$. As a natural extension, the concept of principal (or canonical) angles is originally proposed to measure the distance or angle between two subspaces of higher dimensional linear vector spaces and finds many important applications in statistics and numerical analysis.[7]

Let $\mathbf{F}$ and $\mathbf{G}$ be given subspaces of real space $\mathscr{R}^m$, and assume for convenience that $p = \dim(\mathbf{F}) \geq \dim(\mathbf{G}) = q \geq 1$. The q smallest principal angles $\theta_k \in [0, \pi/2]$, $k = 1, 2, ..., q$, between $\mathbf{F}$ and $\mathbf{G}$ are recursively defined as

$$\theta_1 = \cos^{-1}\left[\max_{\mathbf{u}\in\mathbf{F}} \max_{\mathbf{v}\in\mathbf{G}} (\mathbf{u}^{\mathrm{T}}\mathbf{v})\right] = \cos^{-1}[\mathbf{u}_1{}^{\mathrm{T}}\mathbf{v}_1] \quad (1)$$

subject to $||\mathbf{u}||_2 = ||\mathbf{v}||_2 = 1$, and

$$\theta_k = \cos^{-1}\left[\max_{\mathbf{u}\in\mathbf{F}} \max_{\mathbf{v}\in\mathbf{G}} (\mathbf{u}^{\mathrm{T}}\mathbf{v})\right] = \cos^{-1}[\mathbf{u}_k{}^{\mathrm{T}}\mathbf{v}_k] \quad (2)$$

subject to $||\mathbf{u}||_2 = ||\mathbf{v}||_2 = 1$, $\mathbf{u}_i{}^{\mathrm{T}}\mathbf{u} = 0$, and $\mathbf{v}_i{}^{\mathrm{T}}\mathbf{v} = 0$, where $i = 1, 2, ..., k - 1$ for $k = 2, ..., q$.

The vectors $(\mathbf{u}_1, ..., \mathbf{u}_q)$ and $(\mathbf{v}_1, ..., \mathbf{v}_q)$ are called the principal vectors of the pair of spaces. It should be noted that the principal angles satisfy

$$0 \leq \theta_1 \leq ... \leq \theta_q \leq \frac{\pi}{2} \quad (3)$$

Notice that $\theta_q = 0$ if and only if $\mathbf{F} \supseteq \mathbf{G}$ and $\theta_1 = \pi/2$ if and only if $\mathbf{F} \perp \mathbf{G}$.

While eqs 1 and 2 serve to define θ_k, they are not trivial to use in practice. Numerically stable methods to compute the canonical structure through singular value decomposition (SVD) have been proposed by Björck and Golub.[8] If the columns of $\mathbf{Q_F} \in \mathscr{R}^{m\times p}$ and $\mathbf{Q_G} \in \mathscr{R}^{m\times q}$ generate orthonormal bases for $\mathbf{F}$ and $\mathbf{G}$, respectively, let the SVD of $\mathbf{Q_F}^{\mathrm{T}}\mathbf{Q_G}$ be

$$\mathbf{Q_F}^{\mathrm{T}}\mathbf{Q_G} = \mathbf{YCZ}^{\mathrm{T}} \quad (4)$$

where $\mathbf{C} = \mathrm{diag}(c_1,...,c_q)$ and $c_1 \geq c_2 \geq ... \geq c_q$. There exists a concise conclusion that $[\mathbf{u}_1, ..., \mathbf{u}_p] = \mathbf{Q_F Y}$, $[\mathbf{v}_1, ..., \mathbf{v}_q] = \mathbf{Q_G Z}$, and $\cos\theta_k = c_k$, $k = 1, ..., q$.

The definition of eqs 1 and 2 is interesting because of its applicability for two subspaces with possibly different dimensions. It is noteworthy that there are alternative measurements for the similarity between subspaces such as approaches presented by Ramsay et al.[9] and Knyazev and Argentati.[10] It should be pointed out that the method by Ramsay et al.[9] is identical with the principal angles when applied to column spaces of matrices.[10] In addition, it is more appealing in some cases if the similarities between one subspace and other more than one subspaces can be measured directly. This extension is deferred to the appendix for brevity.

2.2. Similarity Metric for PCA Models. Consider data matrix $\mathbf{X} = [\mathbf{x}_1, ..., \mathbf{x}_m] \in \mathscr{R}^{n\times m}$, where n stands for the number of samples and m the number of measured process variables and quality variables. Hereinafter, we assume that the same m variables are measured for different data groups. PCA aims at finding projection axes (loadings) $\mathbf{P} = [\mathbf{p}_1, ..., \mathbf{p}_A] \in \mathscr{R}^{m\times A}$, $A \leq m$, which represent the directions with maximum variations and thus provide a simpler and more parsimonious representation of the original data set:[11]

$$\mathbf{X} = \mathbf{TP}^{\mathrm{T}} + \mathbf{E} \quad (5)$$

where $\mathbf{T} \in \mathscr{R}^{n\times A}$ is the principal component scores representing projections of the original data set onto principal component loading $\mathbf{P}$, while $\mathbf{E} \in \mathscr{R}^{n\times m}$ is the residual and $\mathbf{p}_k$ is the eigenvector corresponding to the kth largest eigenvalue of $\mathbf{X}^{\mathrm{T}}\mathbf{X}$. The retained number of principal components, A, is usually determined through cross-validation.[12] Newly collected multivariate observation $\mathbf{x} \in \mathscr{R}^{m\times 1}$ can then be fitted as

$$\hat{\mathbf{x}} = \mathbf{PP}^{\mathrm{T}}\mathbf{x} \quad (6)$$

Here $\mathbf{P}^{\mathrm{T}}\mathbf{P} = \mathbf{I}$, and hence $\mathbf{P}$ forms an orthonormal basis of the retained principal component subspace. Hereinafter, the norms "subspace" and "model" are interchangeable unless explicitly stated. For any two PCA models with usually unequal retained numbers of principal components, the definition of principal angles in eqs 1 and 2 can then be utilized to measure the similarities or distance between them. The corresponding calculation can be performed using the aforementioned SVD approach.[8] In the case of more than two PCA models presented, apart from comparing one model with another, it is also viable to calculate the similarity between one specific PCA model and other PCA models simultaneously according to theorem 1 in the appendix.

One implicit but crucial assumption for the definition of principal angles must be explicitly clarified here. It is obvious that in the aforementioned definitions every basis of the subspace is treated in a peer-to-peer fashion. This is natural when the models are applied in prediction but seems not always appropriate in other cases such as comparison of the extent to which the original data spaces generating the PCA models differ. For example, assume two PCA models with the principal loadings $\mathbf{P}_1$ and $\mathbf{P}_2$, which contain the same p columns but in different orders. It is apparent that, from the definition of eqs 1 and 2, $\theta_1 = \theta_2 = ... = \theta_p = 0$, which reflects the fact that the subspaces spanned by $\mathbf{P}_1$ and $\mathbf{P}_2$ are the same. Nevertheless, it is also obvious that these two PCA models are distinct because they correspond to different "data clouds" in the original space.

There exist two possible solutions from two distinct points of view. On the one hand, to interpret the similarities or differences between these PCA models, we can consider the pairs of principal vectors of the counterpart subspaces because these vectors are defined with respect to the original space and hence may be interpreted by reference to the coefficients in each vector.[13] On the other hand, it is also reasonable and, in some senses, more appealing to account for the relative importance of the axes in every subspace.

One straightforward remedy is to assign some possible unequal weights for the axes in a certain subspace. Recognizing that the original definition is usually no longer equivalent to the presentation in theorem 1 in

the appendix, the more extensible formulation in theorem 1 in the appendix is adopted as the starting point. For PCA models, the ratio of the explained variance by a specific axis is frequently utilized to reflect the relative importance of different loadings. Denoting $\lambda_{i,j}$, $j = 1,...,$ d_i, the eigenvalue corresponding to the jth axis (column) of the ith PCA model, and replacing the loading $\mathbf{P}_i$ by $\mathbf{P}_i\mathbf{\Lambda}_i^{1/2}$ in eq 11 yield

$$\bar{\mathbf{H}} = \mathbf{\Lambda}_i^{1/2}\,\mathbf{P}_i^{\mathrm{T}}\left(\sum_{j\neq i}\mathbf{P}_j\mathbf{\Lambda}_j\mathbf{P}_j^{\mathrm{T}}\right)\mathbf{P}_i\mathbf{\Lambda}_i^{1/2} \tag{7}$$

where $\mathbf{\Lambda}_i = \mathrm{diag}(\tilde{\lambda}_{i,1}, ..., \tilde{\lambda}_{i,d_i})$, $i = 1, ..., g$, and $\tilde{\lambda}_{i,j} = \lambda_{i,j}/m$, $j = 1, ..., d_i$. The conclusions in theorem 1 in the appendix can then be applied directly, and the similarities between PCA models can be interpreted straightforwardly.

3. A MSPM Scheme Based on Multiple PCA Models

For some processes with multiple normal operating modes, it would be appropriate to develop multiple PCA models in order to provide finer resolution of the underlying process. When multiple PCA models are employed, some specific problems must be taken into account. These generally include (a) how to construct the individual PCA models, (b) how to integrate the results from different models, and (c) when and how to incorporate new models. The proposed multiple PCA model based MSPM scheme is detailed in the following.

3.1. Data Preprocessing. Because data from real industrial processes is usually corrupted by outliers and noise, two corresponding treatments are adopted sequentially for all modeling data.

Quite often, the measurements are generally corrupted with outliers and gross errors. To address this issue, a moving median filter[14] is employed to adapt both the possible set-point alterations and slow process variations. In this technique, the median of a window containing an odd number of observations is found by sliding the window over every measurement. With appropriate selection of the filter window length, it is possible to prevent this approach from tracking the change of the process unrealistically or, conversely, failing to respond to changes at all. In the proposed MSPM scheme, historical data for modeling are preprocessed using this method. The selection of the filter window length is based on the engineering information of the monitored processes, and automatic selection guidelines are currently under investigation.

Furthermore, the collected process measurements are inherently contaminated by measurement noise. This undesirable influence should be removed to extract the relevant information from the measurements while possible misleading or biased results in subsequent stages will arise otherwise. Denoising using wavelet shrinkage is widely adopted and proven effective in many applications.[15−17]

After these pretreatments, data from different groups, where segmentation of the original data will be discussed in the next section, are then autoscaled to have zero mean and unit variance.

3.2. Model Construction and Comparison. During the modeling stage, based on certain a priori information of the monitored process, the preprocessed historical process operating data are preliminarily segmented into a number of groups, $\mathbf{X}_1^{(0)}$, ..., $\mathbf{X}_{n_0}^{(0)}$, which generally have unequal lengths. Each group is usually expected to correspond to a different operating mode resulting from aspects such as varying feedstock, manufacturing planning, or manual set-point alterations. A PCA model is developed for each group of data.

It is apparent that the key problem within this context is how to assess the similarity of different PCA models. The calculation scheme introduced in section 2.2 provides one possible solution by comparing all of the first smallest principal angles one by one and then concluding (method 1). This method characterizes the similarity elaborately, while if only the overall similarity of two PCA models is required, there exists a more concise conclusion. In fact, according to lemma 1 in the appendix, the overall similarity of PCA models $\mathbf{P}_i$ and $\mathbf{P}_j$, with $p = \dim(\mathbf{P}_i) \geq \dim(\mathbf{P}_j) = q \geq 1$, can be defined as

$$\theta_0^{\,2} = \sum_{k=1}^{q}\cos^2\theta_k = \sum_{k=1}^{q}\lambda_k(\mathbf{P}_j^{\mathrm{T}}\mathbf{P}_i\mathbf{P}_i^{\mathrm{T}}\mathbf{P}_j) = \mathrm{trace}(\mathbf{P}_j^{\mathrm{T}}\mathbf{P}_i\mathbf{P}_i^{\mathrm{T}}\mathbf{P}_j) \tag{8}$$

where operator $\lambda_k(\mathbf{A})$ stands for the kth largest eigenvalue of matrix $\mathbf{A}$. Thus, the sum of all of the eigenvalues of $\mathbf{P}_j^{\mathrm{T}}\mathbf{P}_i\mathbf{P}_i^{\mathrm{T}}\mathbf{P}_j$, or, equivalently, the trace of $\mathbf{P}_j^{\mathrm{T}}\mathbf{P}_i\mathbf{P}_i^{\mathrm{T}}\mathbf{P}_j$, equals the sum of squares of the cosines of all of the principal angles, and θ_0^2 characterizes the overall similarities of two PCA models (method 2). It is apparent that θ_0^2 lies between 0 and q, while $\theta_0^2 = 0$ if and only if $\mathbf{P}_i \perp \mathbf{P}_j$ and $\theta_0^2 = q$ if and only if $\mathbf{P}_i \supseteq \mathbf{P}_j$.

In some cases only the first principal angle θ_1 is utilized for simplicity to measure the similarities between any two PCA models[18] (method 3). This stems from the assumption that, for different PCA models of the same process, a few most important principal component loadings are usually very similar, and this fact is represented by the relative magnitude of θ_1.

For the above three methods, both methods 2 and 3 provide a single quantity to measure the similarity of two PCA models and can be applied directly in practice at the cost of the possibility of inspecting this similarity more elaborately. On the other hand, method 1 provides the details of the similarity of the models together with the demand for interpreting the obtained results based on the knowledge about the process. The final choice depends on the specific requirements of the underlying objective and the practitioners themselves.

The model construction issue can then be addressed as follows. Denoting $\mathbf{X}_1^{(0)}$, ..., $\mathbf{X}_{n_0}^{(0)}$ as the initial data groups and setting $i = 0$, construct PCA models through the following iterative procedure:

(1) For each group $\mathbf{X}_k^{(i)}$, PCA model $\mathbf{P}_k^{(i)}$ is *independently* obtained through the traditional PCA method, and the number of retained principal components can be determined through cross-validation[12] with the testing data set as a part of $\mathbf{X}_k^{(i)}$, $k = 1, ..., n_i$.

(2) Calculate the similarities $\mathbf{S}_{j,k}^{(i)}$, $j = 1, ..., n_i$, $k = 1, ..., j - 1$, between PCA models $\mathbf{P}_j^{(i)}$ and $\mathbf{P}_k^{(i)}$ using one of the above methods.

(3) If any $\mathbf{S}_{j,k}^{(i)} < \theta_t$ (for methods 1 and 3) or $\mathbf{S}_{j,k}^{(i)} > \theta_t$ (for method 2) where θ_t is the preselected threshold of similarities, merge the corresponding data groups $\mathbf{X}_j^{(i)}$

and $\mathbf{X}_k^{(i)}$, denote these newly generated data groups as $\mathbf{X}_1^{(i+1)}$, ..., $\mathbf{X}_{n_i+1}^{(i+1)}$, let $i = i + 1$, and go to step 1; otherwise, stop.

These finally retained PCA models are therefore guaranteed to be away from each other, measured by principal angles. The selection of θ_t is application-dependent, and it determines the capability of the employed number of models. For method 2, in general, the bigger the threshold θ_t is, the less different the employed models can be and hence the more models that may be contained and vice versa. A similar threshold can also be constructed for the overall similarity defined in the appendix. In practice, if it is sure that the known different modes can be well segmented, the threshold θ_t can be determined based on the similarities of the existing models at hand. In this case, conditions of step 3 in the prescribed model developing procedure will then be satisfied automatically and therefore unnecessary.

In addition, although under most circumstances this approach constructs a separate model for every operating mode, it must be clarified that this is not always the case. In fact, even though every data group $\mathbf{X}_k^{(0)}$ corresponds to a specific operating mode initially, some data groups could be merged in step 3 of the proposed approach. In other words, it is possible that several operating modes have been represented by one single model if they are very similar.

3.3. Process Monitoring and Fault Detection. Within this proposed scheme, some widely used multivariate statistic such as SPE, which is defined as $||\mathbf{x} - \hat{\mathbf{x}}||^2 = \sum_{i=1}^{m}(x_i - \hat{x}_i)^2$ for a new observation $\mathbf{x} = [x_1, ..., x_m]^T$, can be straightforwardly calculated from a specific PCA model to facilitate process monitoring. Its corresponding control limit can be determined through traditional methods.[19]

After all of these PCA models are constructed, they can now be applied to perform online process monitoring. Every newly collected observation is first scaled by the corresponding mean and standard deviation of each PCA model and then approximated by it. The one yielding the minimum SPE is finally accepted and utilized to determine whether a warning should be triggered. In addition, it should be noted that these PCA models may have been altered by the procedure of model updating or new model incorporation illustrated in the following section. As a consequence, control limits of the control charts generally vary according to the adopted PCA models.

This model fusion strategy is reasonable based on the idea that one new sample should belong to the model that fits it best or, in other words, yields the minimum SPE. It is apparent that this problem is objective-specific and there probably exist other fusion approaches from corresponding different points of view.

3.4. Model Updating. Industrial processes often experience time-varying changes because of, for example, catalytic decaying and instrument drifts. It is thus desirable to update the model recursively when newly collected data are available. Various approaches such as RPCA[3] and PLS[20,21] are proposed and applied successfully in diverse fields in the case when only one single model is employed.

With multiple PCA models, the newly collected data should be processed sequentially as follows:

(1) Determine the model (called here the *adopted model*) it belongs to according to the strategy introduced in section 3.3.

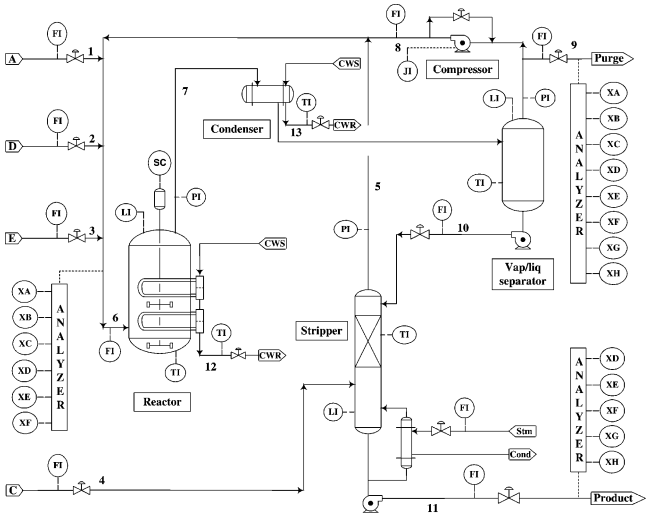

Figure 1. Schematic diagram of the TECP.

(2) If these new data do not trigger a warning, they will then be incorporated into the *adopted model*, which can thus be updated recursively analogously to the single-model case.[3]

However, one should be careful that the model updating should not violate the model similarity constraints. On the other hand, unlike the single-model case,[3] the data that trigger a warning are not definitely regarded as the consequence of process malfunctions but will probably be further utilized to construct a new model, which will be discussed next.

3.5. New Model Incorporation. Some newly online collected data will probably exceed the upper control limit (UCL) of the corresponding PCA model and trigger a warning, which can be attributed to either process abnormality or the advent of a new normal operating mode that has not been included in current models. For the first case, auxiliary analysis tools such as a contribution plot or a principal scores scatter plot can be employed for further fault diagnosis. If the presence of process abnormality is eliminated, then a new model will be incorporated provided that adequate samples have been collected and its distance to every existing one exceeds the predefined threshold. The model incorporation can be performed at run time if necessary and will not cause a break in the monitoring.

The data for constructing the new model can be prepared either automatically based on certain strategy, for example, gathering all of the recent samples that are in the normal operation region, or manually at appropriate times by the operators. The latter is preferred because it is still rather challenging at present to distinguish the normal unseen data effectively from the abnormal data automatically.

4. Application Results and Discussion

4.1. Case Study 1: The TECP. 4.1.1. Process Description. The TECP presented by Downs and Vogel in 1993[22] is a model of a complex industrial chemical process. It has been widely used as a benchmark case study for different purposes in the past and is used here to assess the efficiency of the proposed strategy. As shown in Figure 1, the process has five major unit operations: a reactor, a product condenser, a vapor−liquid separator, a recycle compressor, and a product stripper. It involves two simultaneous gas−liquid exo-

Table 1. Six Operating Modes of TECP

mode	desired G/H mass ratio	desired production (kg/h)
1	50/50	14 076
2	10/90	14 077
3	90/10	11 111
4	50/50	maximum
5	10/90	maximum
6	90/10	maximum

Table 2. Number of Retained Principal Components and CPV for Five PCA Models Corresponding to Modes 2−6

	mode 2	mode 3	mode 4	mode 5	mode 6
retained PCs	11	7	10	11	8
CPV	0.7149	0.7187	0.7044	0.7267	0.7109

thermic reactions that produce two products G and H from four reactants A and C−E:

$$A(g) + C(g) + D(g) \rightarrow G(l)$$

$$A(g) + C(g) + E(g) \rightarrow H(l)$$

Also present are an inert B and a byproduct F involving two additional reactions. The process consists of 12 manipulated variables and 41 measurements available for monitoring or control, of which 22 are continuous while the rest, 19, are gas chromatographic measurements at the sampling interval of 0.1 or 0.25 h.

The original process is itself strictly open-loop unstable, and in the absence of feedback, small perturbations will cause a large pressure transient and eventually lead to a shutdown within 1 h.[23] A certain control strategy thus must be introduced, and several candidates are available.[23−27] Here the decentralized control presented by Ricker is employed for its capability of less variability in the product rate and quality and of operating on-spec for long periods without feedback from composition measurements.[27] For the sake of practical consideration, the 22 continuous outputs among the 41 measurements are used for monitoring and the sampling interval is 0.01 h. This process possesses six modes of process operation at three different G/H mass ratios (stream 11), which are listed in Table 1 and are normally dictated by market demand or capacity limitations.[22] The corresponding optimal steady-state operating conditions can be found in ref 28. Both the used steady-state data and simulation programs are available at Ricker's home page.[29]

4.1.2. Monitoring of Mode Alternations and Process Transition. In this case study, the process first runs for 10 h at each mode, except mode 1, under normal operation, and 1000 samples are collected for each mode. These data are then utilized to construct five PCA models. The number of retained principal components is determined by discarding the remaining minor principal components when the cumulative percentage of variance (CPV) explained is greater than 0.70, and the results are listed in Table 2.

Their similarities are checked using the root-mean-square-similarity (RMSS) measurement:

$$\mathrm{RMSS} = \sqrt{\frac{\sum_{k=1}^{q} \cos^2 \theta_k}{q}} \qquad (9)$$

which scales the definition of eq 8 to make the result lie between 0 and 1 and can be approximately regarded as the overall similarities between two PCA models. The results are illustrated in Table 3. Because these operating modes are well-defined, the threshold is determined as 0.94 (≈cos 20°) based on their similarities. In addition, both a RPCA[3] model and a global PCA model based on all of the collected data are also developed for the purpose of comparison. These models are then applied to process monitoring.

Table 3. Similarities among the Five PCA Models Corresponding to Modes 2−6

	mode 2	mode 3	mode 4	mode 5	mode 6
mode 2	1.0000	0.8514	0.9011	0.9346	0.8656
mode 3	0.8514	1.0000	0.8884	0.8467	0.8853
mode 4	0.9011	0.8884	1.0000	0.9117	0.8358
mode 5	0.9346	0.8467	0.9117	1.0000	0.8749
mode 6	0.8656	0.8853	0.8358	0.8749	1.0000

After running for about 72 h under normal conditions in mode 4, the process operation switches to mode 1. In this case, the three approaches produce similar results and all trigger continuous alarms as illustrated in Figure 2, where model switching takes place at sampling point 200.

Because so many alarms are presented continuously, it is reasonable to assume that operators recognize the reason for this and try to develop a new model. The data at steady state in mode 1 are then collected, and a new PCA model is constructed. It retains 11 principal components, and its similarities with the existing PCA models are listed in Table 4. The overall similarities are calculated using eq 13 in the appendix, and it represents the similarity between the newly developed model and all of the previously developed models. When it is recognized that its similarities with the existing PCA models are less than the aforementioned predefined threshold 0.94, this newly developed model is incorporated, and these six models are then utilized together to monitor the process. For a fair comparison, a new global PCA model is developed as well with the new data at steady state in mode 1 incorporated.

This process switches to mode 3 and then transits from steady state under normal operation of mode 3 to mode 1. Figures 3−5 show respectively the monitoring performance of RPCA, global PCA, and the proposed approach. In these figures, the solid lines represent the 99% control limits, the dash-dotted lines represent the 95% control limits, and the process switches from mode 3 to mode 1 at sampling point 200. Because the RPCA algorithm cannot recover from the previous mode alteration, a new RPCA model is reconstructed by using entirely the steady-state data in mode 3. Unfortunately, a result analogous to Figure 2 occurs, and continuous alarms are incurred right after the mode alteration, as shown in Figure 3. This is in sharp contrast with the results illustrated in Figure 5 by the six PCA models that recognize the mode alterations correctly. In fact, only a few alarms are triggered, which can be attributed to both the measurement noise and the transition itself. In addition, a benefit of the proposed approach is that the control charts also distinguish the current operation modes of the process correctly. For the global PCA model, as shown in Figure 4, although it behaves well for mode 1, it incurs many false warnings when the process is operated in mode 3. Besides, it also tends to generate relatively small T^2 values, which is also demonstrated in the following experiments.

These results indicate that the proposed approach is very efficient in monitoring the process with multiple

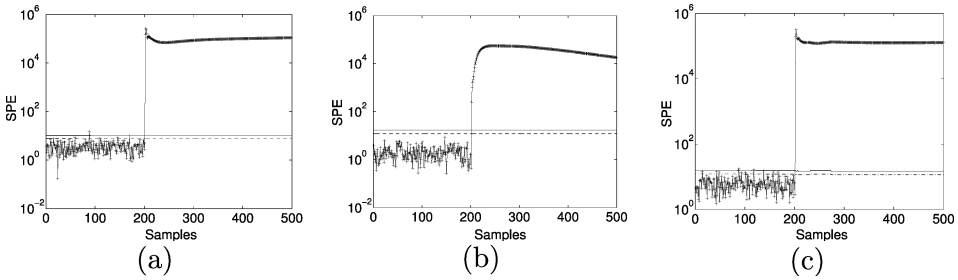

Figure 2. SPE control chart by (a) RPCA, (b) global PCA, and (c) the proposed approach with 95% (dash-dotted line) and 99% (solid line) control limits.

Table 4. Similarities between the Newly Developed Model and the Existing Ones

	mode 2	mode 3	mode 4	mode 5	mode 6	overall
new model	0.9380	0.8768	0.9033	0.9184	0.8698	4.8511

operating modes and is reliable enough to adapt possible mode alterations provided the corresponding mode has been included in one of the PCA models. On the other hand, although it is effective to address slow and normal process changes, RPCA is liable to failure when the process switches to another operating mode. This deficiency cannot be overcome by intuitionistic thought of employing modeling data under as many different modes as possible. This will certainly impose much of the burden on the model construction, which is undesirable in practice, whereas the more important reason is that regardless of how many observations are employed initially, if the process is operated under a specific mode for a long time, data under this mode will gradually dominate the RPCA model by virtue of adaptive model updating. This eventually causes RPCA to be ineffective in responding to the mode alterations.

For the global PCA model if data corresponding to different operating modes exhibit certain dissimilarities due to different means or variances, the developed model tends to model some operating modes better at the cost of poor characterization of the other modes. Operating modes with relatively less samples and variances will generally be more poorly explained, whereas this deficiency cannot be overcome by merely retaining more principal components. For the previous example, operating mode 3 is not well characterized, which can also be demonstrated from Tables 3 and 4 by its similarities with the other operating modes, with the consequence of numerous false warnings even when the process is operated normally. On the other hand, the global PCA approach also lacks the capability of adapting possible slow changes in the industrial processes.[3]

4.1.3. Fault Detection. To assess the efficiency of the proposed approach for fault detection, some predefined disturbances (refer to Table 8 of ref 22 for details) are introduced to the process in a specific mode. Owing to the previous discussion on RPCA, when the process switches to a new mode, a new PCA model is developed using the corresponding nominal data with the original one discarded and then employed for further fault detection. Here for brevity only one example is illustrated.

In this case, the process operated under normal conditions in mode 5 and slow drift of reaction kinetics (IDV 13) started at sample 100. This is generally rather challenging to detect, whereas the proposed algorithm catches this deviation after approximately 37 min of its first advent. The corresponding SPE and T^2 charts are illustrated in Figure 6. RPCA presents comparable results as expected and is shown in Figure 7. However, it should be emphasized here that the RPCA model needs to be first set up when the process is switched to a new operating mode before it can be used for monitoring. The RPCA method therefore has difficulties in handling the situation where a fault is present soon after operating mode switching. As shown in Figure 8, the global PCA model also detects this process malfunction after about 44 min, which is slower than the proposed approach for about 7 min. These results indicate that RPCA can be approximately regarded as a special case of the proposed approach for processes with one operating mode only or, equivalently, the proposed approach as an extension of RPCA to the case of processes with multiple operating modes, whereas for the global PCA method, it ensembles the data from all different operating modes together and will therefore be less sensitive or provide coarser resolution of the process than the proposed approach, which develops local models according to specific operating modes. Thus, the global PCA method can be less sensitive to faults occurring in certain process operating modes.

4.2. Case Study 2: Industrial FCC Unit. The proposed approach is also applied to an industrial FCC unit where there are totally 74 online measured variables. These variables (tags) are categorized into four groups including the monitoring of reactor−regenerator, carbon deposition, fractionator, and stabilization process, while only the first group is presented here for the sake of simplicity in illustration.

4.2.1. Plant Description. The FCC unit plays a very important role in the oil refinery, and it converts heavy fractions to gasoline, C_3−C_4 cuts, and petrochemicals. It consists of three major parts including the reactor−regenerator, fractionator, and stabilization process. Among them, the considered reactor−regenerator is the upstream part and is also the most important section.

The plant shown in Figure 9 distinguishes itself from popular FCC units by possessing two-stage regenerators. The feedstock is injected at the bottom of the riser. At this point, the catalyst from the second-stage regenerator and feedstock vaporization with feedstock droplets get mixed, and reaction starts up. This provokes a high gas velocity, which entrains catalyst particles upward, and reactions take place in the riser. The catalyst containing coke at the riser exit is separated from the gases in cyclones. The catalyst is then flushed with a steam flow in the stripper to minimize the hydrocarbon entrainment to the first regenerator. The coke on the catalyst is burned off by air in the first and second regenerators consecutively. Fresh catalyst is added to the second regenerator at predefined regular intervals. The coke combustion provides the thermal energy necessary to vaporize the feedstock and to

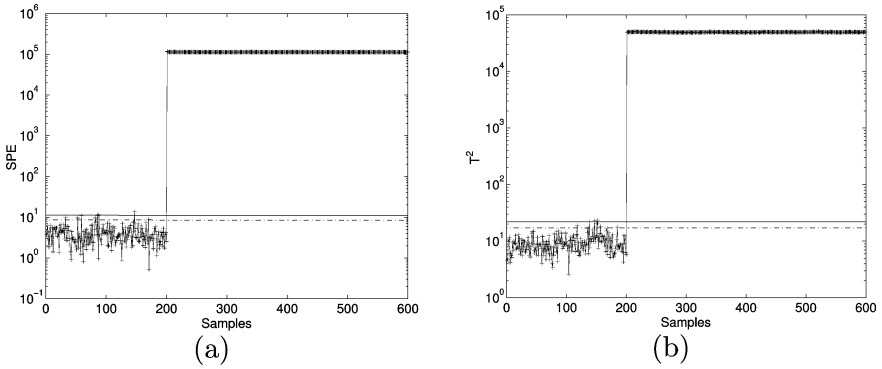

Figure 3. SPE (a) and T^2 (b) control charts from RPCA.

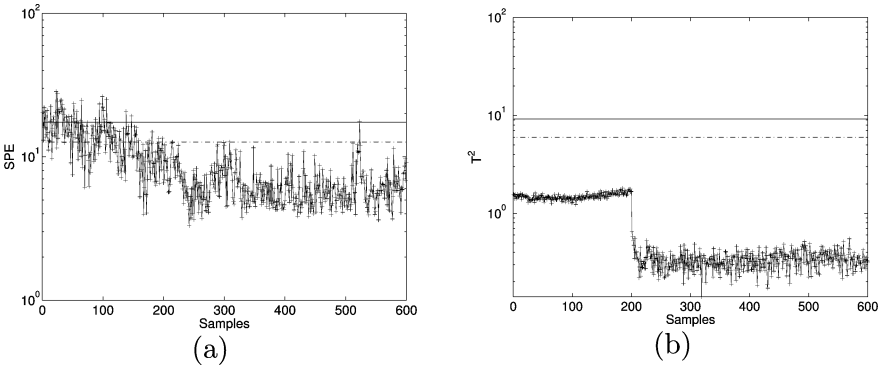

Figure 4. SPE (a) and T^2 (b) control charts from global PCA.

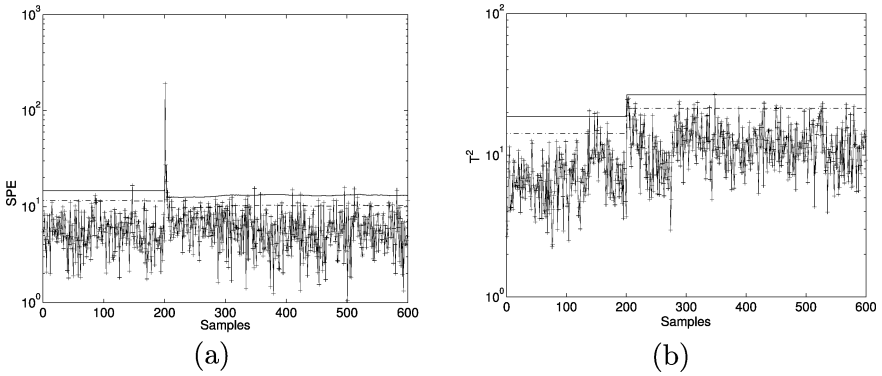

Figure 5. SPE (a) and T^2 (b) control charts from the proposed approach.

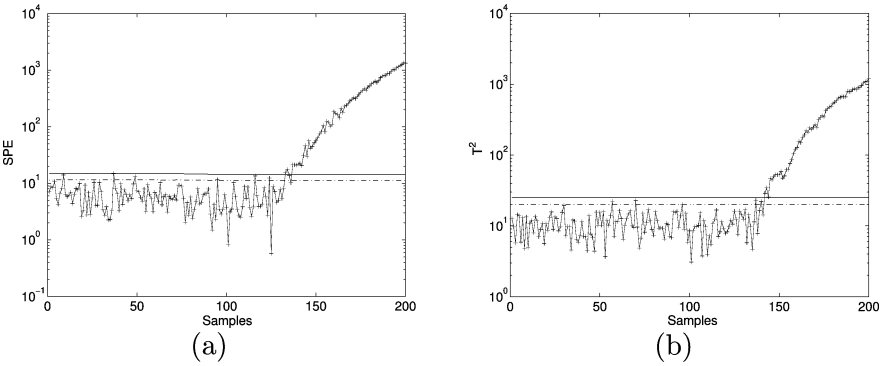

Figure 6. SPE (a) and T^2 (b) control charts from the proposed approach with 95% (dash-dotted line) and 99% (solid line) control limits.

compensate for the reaction endotherm, and this energy is transported by the hot catalyst.

4.2.2. Construction of the Nominal Model. There are totally 26 highly correlated variables in the reactor−regenerator part, and the sampling interval is 1 min. It must be noted, however, that these variables will not definitely correspond to some specific measurement of the process under consideration and, in fact, may be the combination of several associated ones. This stems from the motivation that a certain combination reflects the operation in essence better than the individuals themselves. Take, for example, although the feedstock is in-

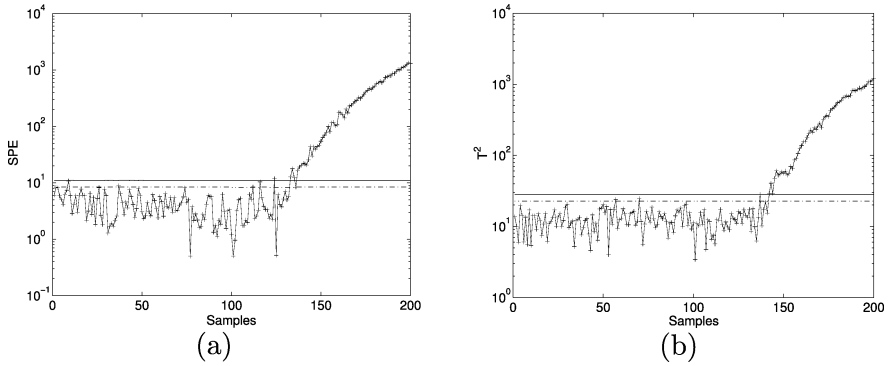

Figure 7. SPE (a) and T^2 (b) control charts from RPCA with 95% (dash-dotted line) and 99% (solid line) control limits.

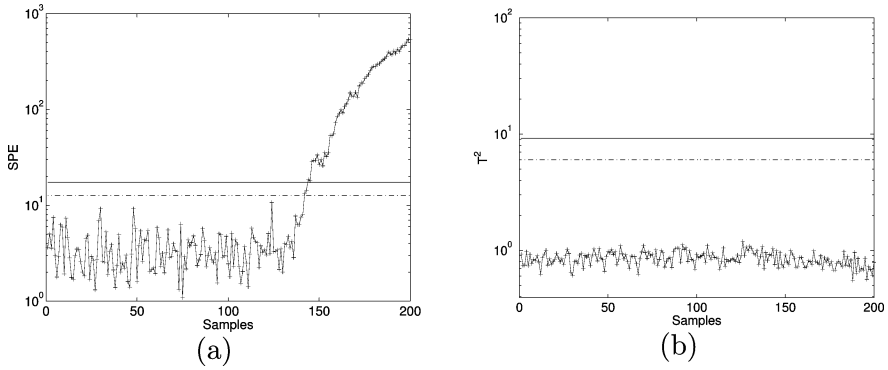

Figure 8. SPE (a) and T^2 (b) control charts from the global PCA with 95% (dash-dotted line) and 99% (solid line) control limits.

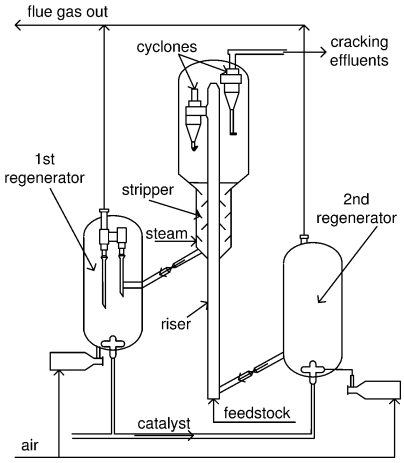

Figure 9. Schematic diagram of the reactor−regenerator of the FCC process.

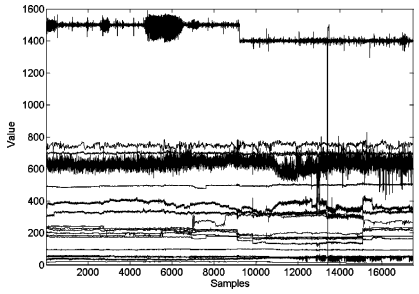

Figure 10. Original data for three PCA models.

Table 5. Model Information

model no.	no. of samples	retained PC no.
I	6000	9
II	3000	10
III	4300	12

Table 6. Similarities of the Models

	I	II	III
I	1.0000	0.6751	0.8146
II	0.6751	1.0000	0.8513
III	0.8146	0.8513	1.0000

jected at the bottom of the riser through four different pipelines and measured at corresponding sites, the finally adopted variable is their sum. From our practice, it is observed that employing the combination of the variables will greatly decrease the amount of false warnings.

Up to 17 500 samples of historical process operation data are available (see Figure 10), and they are then segmented into three data groups manually that contain 6000, 3000, and 4300 samples, respectively, with some transitional data discarded. It is apparent that some observations are contaminated by measurement noises, while data from different groups manifest strong dissimilarities.

These data are first preprocessed using the aforementioned methods, and three PCA models are then constructed independently. Cross-validation is employed to determine the retained numbers of principal components, and the results are illustrated in Table 5. In this case, all of the models retain more than nine principal components because of the fact that there were no dominant sources of variability, which mostly influence the process behavior. On the other hand, this result also indicates that it is necessary to retain different numbers of principal components for different PCA models.

Their similarities are then checked using RMSS defined in eq 9, and the results are listed in Table 6. Because these data groups are chosen specifically, the threshold is chosen as 0.86 (≈cos 30°) based on similarities of the existing models. This clustering behavior is also demonstrated through plotting of the first two

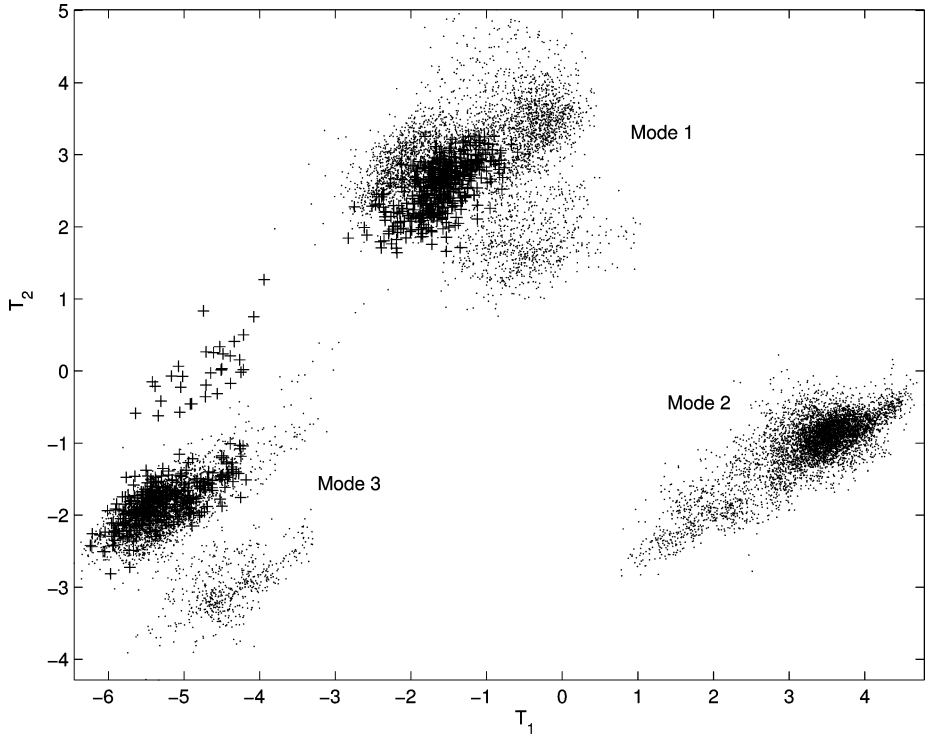

Figure 11. $\mathbf{T}_1$ vs $\mathbf{T}_2$ of the PCA model developed from the original data, where + denotes the online collected data while the others are the modeling data.

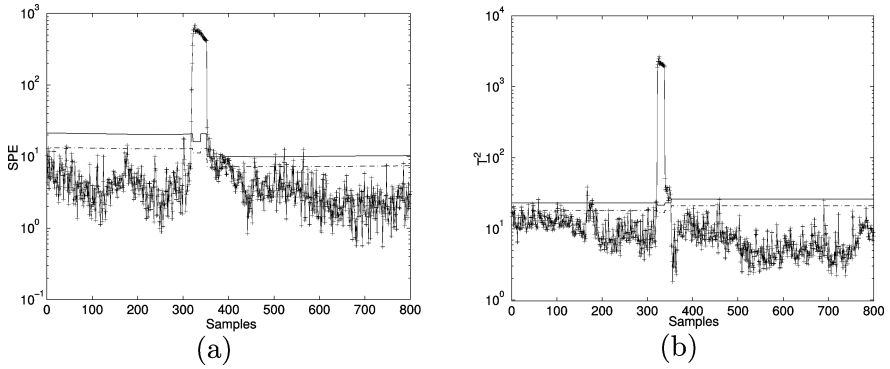

Figure 12. Real-time SPE (a) and T^2 (b) monitoring charts by the proposed algorithm with 95% (dash-dotted line) and 99% (solid line) control limits.

scores $\mathbf{T}_1$ and $\mathbf{T}_2$, as shown in Figure 11, of the PCA model developed from collecting all of the modeling data together.

In addition, a RPCA model and a global one using all of the 17 500 samples are also developed for the purpose of comparison. Both the SPE and T^2 charts are employed for the proposed approach. The UCLs for the SPE and T^2 charts are both chosen as 0.95 and 0.99. Note that each individual PCA model by the proposed algorithm has its own UCL independently.

4.2.3. Real-Time Monitoring. Here the operating condition of this process is altered around sampling point 300, and if the scores of these online data are plotted together with the modeling data, as shown in Figure 11, this trend can also be easily identified. This clustering characteristic can be demonstrated using other approaches as well.[30] The corresponding SPE and T^2 control charts for all methods are given in Figures 12−14, respectively.

Note that both UCL of the SPE and T^2 charts by RPCA and the proposed approach vary to some extent. For RPCA, it varies slightly because of model updating, while for the proposed method, it varies mainly because of real-time adoption of the specific different models. It is shown that, for the monitoring under consideration, the RPCA method tracks the changes of the process well but is sensitive to the variations and deteriorates sharply in the presence of mode alterations. The global PCA model triggers numerous false warnings even if the process itself is running normally. On the other hand, the proposed scheme can reduce the amount of false alarms to a great extent while at the same time track the process adjustment effectively. In fact, it triggers a few alarms in the transition stage only, whereas it rapidly adapts to the adjustment within 50 samples.

5. Conclusions

Many of the current MSPM techniques, including RPCA, are based on the assumption that the process has one nominal operating region and they are liable to producing continuous false warnings when applied to processes with multiple operating modes. A multiple PCA model based process monitoring method is pro-

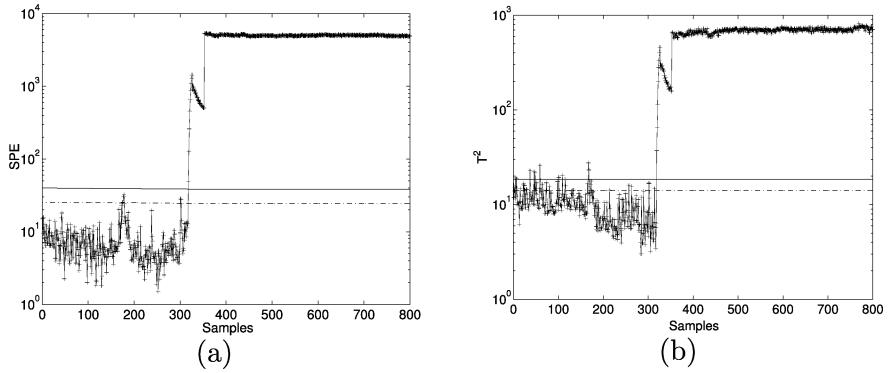

Figure 13. Real-time SPE (a) and T^2 (b) monitoring charts by RPCA with 95% (dash-dotted line) and 99% (solid line) control limits.

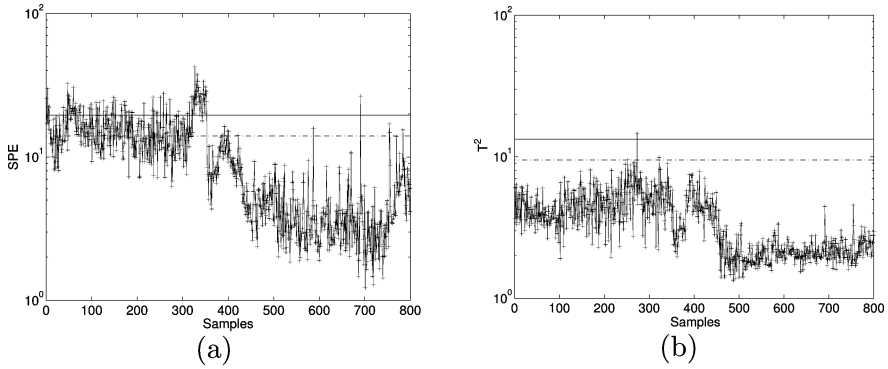

Figure 14. Real-time SPE (a) and T^2 (b) monitoring charts by the global PCA with 95% (dash-dotted line) and 99% (solid line) control limits.

posed in this paper for monitoring processes with multiple normal operating modes. Some popular multivariate statistics such as SPE and its control limit can then be incorporated straightforwardly. Its efficiency is demonstrated through application to both the TECP with six predefined operating modes and an industrial FCC unit. Because most processes in practice experience different operating modes, the proposed approach is expected to have broad applicability in the realistic monitoring of industrial processes.

Acknowledgment

The financial support of both the UK EPSRC (Grant GR/R10875) and the National High Technology Research and Development Program of China (863 Program) (No. 2001AA413320) is greatly appreciated. The authors thank Professor N. Lawrence Ricker at University of Washington for providing the TECP applications and for many useful instructions.

Appendix: Extension to the Case of More Than Two Subspaces

This extension is based on an equivalent definition. Actually, the following can be concluded from eq 4.

Lemma 1. Let $\theta_k = \cos^{-1}\sigma_k$, $k \le q$, representing the kth smallest principal angle between subspaces $\mathbf{F}$ and $\mathbf{G}$ in $\mathcal{R}^m$, and the columns of $\mathbf{Q_F} \in \mathcal{R}^{m\times p}$ and $\mathbf{Q_G} \in \mathcal{R}^{m\times q}$ generate orthonormal bases for $\mathbf{F}$ and $\mathbf{G}$, respectively; then σ_k^2 is equal to the kth largest eigenvalue of $\mathbf{Q_F}^T\mathbf{Q_G}\mathbf{Q_G}^T\mathbf{Q_F}$, while $\mathbf{u}_k = \mathbf{Q_F}\mathbf{y}_k$ and $\mathbf{v}_k = \mathbf{Q_G}\mathbf{z}_k$ where $\mathbf{y}_k$ and $\mathbf{z}_k$ are, up to equivalence, the eigenvectors corresponding to the kth largest eigenvalue of $\mathbf{Q_F}^T\mathbf{Q_G}\mathbf{Q_G}^T\mathbf{Q_F}$ and $\mathbf{Q_G}^T\mathbf{Q_F}\mathbf{Q_F}^T\mathbf{Q_G}$, respectively.

Proof. Because eq 4 holds for the first k smallest principal angles, lemma 1 is then straightforward based on the relationship between the singular values of any matrix $\mathbf{A}$ and the corresponding eigenvalues of $\mathbf{AA}^T$. In addition, it is apparent that this definition is symmetrical with respect to the two subspaces; therefore, σ_k^2 is also equal to the kth largest eigenvalue of $\mathbf{Q_G}^T\mathbf{Q_F}\mathbf{Q_F}^T\mathbf{Q_G}$, and meanwhile $\mathbf{u}_k = \mathbf{Q_F}\mathbf{y}_k$ and $\mathbf{v}_k = \mathbf{Q_G}\mathbf{z}_k$ where $\mathbf{y}_k$ and $\mathbf{z}_k$ are, upon equivalence, the eigenvectors corresponding to the kth largest eigenvalue of $\mathbf{Q_F}^T\mathbf{Q_G}\mathbf{Q_G}^T\mathbf{Q_F}$ and $\mathbf{Q_G}^T\mathbf{Q_F}\mathbf{Q_F}^T\mathbf{Q_G}$, respectively.

Supposing an arbitrary unit vector in the subspace $\mathbf{F}$ with coordinates $\mathbf{y}$, its coordinates with reference to the original space are therefore the components of $\mathbf{u} = \mathbf{Q_F}\mathbf{y}$ and its projection onto the subspace $\mathbf{Q_G}$ is $\mathbf{v} = \mathbf{Q_G}\mathbf{Q_G}^T\mathbf{u}$. According to the orthogonality principal, the angle $0 \le \delta \le \pi/2$ between vectors $\mathbf{u}$ and $\mathbf{v}$ satisfies

$$\cos^2\delta = \mathbf{v}^T\mathbf{v} = \mathbf{u}^T\mathbf{Q_G}\mathbf{Q_G}^T\mathbf{Q_G}\mathbf{Q_G}^T\mathbf{u} = \mathbf{u}^T\mathbf{Q_G}\mathbf{Q_G}^T\mathbf{u}$$

or equivalently, $\cos^2\delta = \mathbf{y}^T\mathbf{Q_F}^T\mathbf{Q_G}\mathbf{Q_G}^T\mathbf{Q_F}\mathbf{y}$. The quantity $\cos^2\delta$ attains its maximum when $\mathbf{y}$ is the eigenvector corresponding to the largest eigenvalue of $\mathbf{Q_F}^T\mathbf{Q_G}\mathbf{Q_G}^T\mathbf{Q_F}$. This indicates that finding the smallest principal angle can be interpreted as the process of searching for vectors in two subspaces that most nearly parallel each other. It is obvious that this interpretation applies to the other principal angles as well.

It is much more appealing that this presentation can be extended to the case with more than two subspaces. That is, a new metric based on the smallest principal angles between more than two subspaces can be presented to measure the overall similarities of one specific subspace to all other subspaces simultaneously. In fact, it can be formulated as follows:

Definition 1. For any $g \ge 2$ subspaces $\mathbf{S}_i \in \mathcal{R}^{m\times d_i}$, $i = 1, ..., g$, where $d_i = \dim(\mathbf{S}_i) \ge 1$ and without loss of generality, assuming that $d_1 \ge d_2 \ge ... \ge d_g$, denote their

corresponding orthonormal bases as $\mathbf{Q}_{\mathbf{S}_i} \in \mathcal{R}^{m \times d_i}$ with $\mathbf{Q}_{\mathbf{S}_i}{}^{\mathrm{T}}\mathbf{Q}_{\mathbf{S}_i} = \mathbf{I}_{d_i}$. Let $\mathbf{u}_i \in \mathcal{R}^{m \times 1}$ be an arbitrary unit vector in the subspace $\mathbf{S}_i$ and $\delta_{i,j}$, $j \neq i$, be the angle between $\mathbf{u}_i$ and $\mathbf{u}_j$; then the first smallest principal angle between $\mathbf{S}_i$ and subspaces $\mathbf{S}_j$, $j \neq i$, is attained when $\mathbf{u}_i$ maximizes

$$V_i = \sum_{j \neq i} \cos^2 \delta_{i,j} \tag{10}$$

It is apparent that this definition is equivalent to eq 1 when $g = 2$ and eq 1 is therefore a special case of eq 10. Analogously, this definition can be regarded as the process of pursuing a unit vector in the subspace $\mathbf{S}_i$, which is most nearly parallel to the vectors in the subspaces $\mathbf{S}_j$, $j \neq i$. On the basis of this definition, the following theorem holds.

Theorem 1. $\mathbf{u}_i = \mathbf{Q}_{\mathbf{S}_i}\mathbf{y}$ maximizes V_i in eq 10 when $\mathbf{y}$ is the eigenvector corresponding to the largest eigenvalue of

$$\mathbf{H} = \mathbf{Q}_{\mathbf{S}_i}{}^{\mathrm{T}}\left(\sum_{j \neq i} \mathbf{Q}_{\mathbf{S}_j}\mathbf{Q}_{\mathbf{S}_j}{}^{\mathrm{T}}\right)\mathbf{Q}_{\mathbf{S}_i} \tag{11}$$

Proof. Supposing an arbitrary unit vector in the subspace $\mathbf{S}_i$ with coordinates $\mathbf{y}$, its coordinates with reference to the original space are therefore the components of $\mathbf{u} = \mathbf{Q}_{\mathbf{S}_i}\mathbf{y}$ and its projection onto the subspace $\mathbf{Q}_{\mathbf{S}_j}$, $j \neq i$, is $\mathbf{u}_j = \mathbf{Q}_{\mathbf{S}_j}\mathbf{Q}_{\mathbf{S}_j}{}^{\mathrm{T}}\mathbf{u}$. The minimum angle $0 \leq \delta_{i,j} \leq \pi/2$ between these vectors $\mathbf{u}$ and $\mathbf{u}_j$ satisfies

$$\cos^2 \delta_{i,j} = \mathbf{u}_j{}^{\mathrm{T}}\mathbf{u}_j = \mathbf{u}^{\mathrm{T}}\mathbf{Q}_{\mathbf{S}_j}\mathbf{Q}_{\mathbf{S}_j}{}^{\mathrm{T}}\mathbf{Q}_{\mathbf{S}_j}\mathbf{Q}_{\mathbf{S}_j}{}^{\mathrm{T}}\mathbf{u} = \mathbf{u}^{\mathrm{T}}\mathbf{Q}_{\mathbf{S}_j}\mathbf{Q}_{\mathbf{S}_j}{}^{\mathrm{T}}\mathbf{u}$$

or equivalently

$$\cos^2 \delta_{i,j} = \mathbf{y}^{\mathrm{T}}\mathbf{Q}_{\mathbf{S}_i}{}^{\mathrm{T}}\mathbf{Q}_{\mathbf{S}_j}\mathbf{Q}_{\mathbf{S}_j}{}^{\mathrm{T}}\mathbf{Q}_{\mathbf{S}_i}\mathbf{y} \tag{12}$$

Therefore

$$V_i = \sum_{j \neq i} \cos^2 \delta_{i,j} = \mathbf{y}^{\mathrm{T}}\mathbf{Q}_{\mathbf{S}_i}{}^{\mathrm{T}}\left(\sum_{j \neq i} \mathbf{Q}_{\mathbf{S}_j}\mathbf{Q}_{\mathbf{S}_j}{}^{\mathrm{T}}\right)\mathbf{Q}_{\mathbf{S}_i}\mathbf{y} = \mathbf{y}^{\mathrm{T}}\mathbf{H}\mathbf{y} \tag{13}$$

It is apparent that when $\mathbf{y}$ is the eigenvector corresponding to the largest eigenvalue of $\mathbf{H}$, V_i is maximized, and this proves theorem 1.

The obtained maximum, V_i, can thus be considered as a measure of closeness of the subspace $\mathbf{S}_i$ to all of the other subspaces $\mathbf{S}_j$ and $\mathbf{u}_i$ provides the average component in $\mathbf{S}_i$ that agrees most closely with all of the subspaces $\mathbf{S}_j$. Furthermore, this conclusion can also be extended to define all of the first $d = \min(d_1, \ldots, d_g)$ smallest principal angles analogously. However, because the concerned subspaces have different dimensions, the interpretation of the results will differ in practice. The detailed discussion will be covered in the next section with respect to different contexts.

Literature Cited

(1) Martin, E. B.; Morris, A. J.; Zhang, J. Process performance monitoring using multivariate statistical process control. *IEE Proc. Control Theory Appl.* **1996**, *143*, 132−144.

(2) Qin, S. J. Statistical process monitoring: basics and beyond. *J. Chemom.* **2003**, *17*, 480−502.

(3) Li, W.; Yue, H. H.; Valle-Cervantes, S.; Qin, S. J. Recursive PCA for adaptive process monitoring. *J. Process Control* **2000**, *10*, 471−486.

(4) Flury, B. K. Two Genealizations of the Common Principal Component Model. *Biometrika* **1987**, *74* (1), 59−69.

(5) Lane, S.; Martin, E. B.; Kooijmans, R.; Morris, A. J. Performance monitoring of a multi-product semi-batch process. *J. Process Control* **2001**, *11*, 1−11.

(6) Hwang, D.-H.; Han, C. Real-time monitoring for a process with multiple operating modes. *Control Eng. Pract.* **1999**, *7*, 891−902.

(7) Golub, G. H.; Van Loan, C. F. *Matrix Computations*, 3rd ed.; The Johns Hopkins Press: Baltimore, MA, 1996.

(8) Björck, Å.; Golub, G. H. Numerical Methods for Computing Angles Between Linear Subspaces. *Math. Comput.* **1973**, *27* (123), July, 579−594.

(9) Ramsay, J. O.; Berge, J. T.; Styan, G. P. H. Matrix Correlation. *Psychometrika* **1984**, *49* (3), 403−423.

(10) Knyazev, A. V.; Argentati, M. E. Principal Angles between Subspaces in an A-Based Scalar Product: Algorithms and Perturbation Estimates. *SIAM J. Sci. Comput.* **2002**, *23* (6), 2008−2040.

(11) Jackson, J. E. *A User's Guide to Principal Components*; John Wiley & Sons: New York, 1991.

(12) Wold, S. Cross-Validatory Estimation of the Number of Components in Factor and Principal Components Models. *Technometrics* **1978**, *20* (4), 397−405.

(13) Krzanowski, W. J. Between-group comparison of Principal Components. *J. Am. Stat. Assoc.* **1979**, *74* (367), 703−707.

(14) Tukey, J. W. *Exploratory Data Analysis*; Addison-Wesley: Reading, MA, 1977.

(15) Donoho, D. L. *Denoising by soft thresholding*; Technical Report; Department of Statistics, Stanford University: Stanford, CA, 1992.

(16) Donoho, D. L. Nonlinear Wavelet Methods for Recovery of Signals, Densities, and Spectra from Indirect and Noisy Data. *Proceedings of Symposia in Applied Mathematics*; American Mathematical Society: Providence, RI, 1993; Vol. 00.

(17) Strang, G.; Nguyen, T. *Wavelets and filter banks*; Wellesley: Wellesley-Cambridge, 1996.

(18) Zhao, S. J.; Xu, Y. M.; Zhang, J. A Multiple PCA Model based Technique for the Monitoring of Processes with Multiple Operating Modes. In *European Symposium on Computer-Aided Process Engineering−14. Computer-Aided Chemical Engineering*; Barbosa-Póvoa, A., Matos, H., Eds.; Elsevier BV: Lisbon, Portugal, 2004; Vol. 18, pp 865−870.

(19) Nomikos, P.; MacGregor, J. F. Multivariate SPC Charts for Monitoring Batch Processes. *Technometrics* **1995**, *37* (1), 41−59.

(20) Qin, S. J. Recursive PLS algorithms for adaptive data modeling. *Comput. Chem. Eng.* **1998**, *22* (4/5), 503−514.

(21) Dong, D.; McAvoy, T. J. Nonlinear principal component analysis-based on principal curves and neural networks. *Comput. Chem. Eng.* **1997**, *20* (1), 65−78.

(22) Downs, J. J.; Vogel, E. F. A Plant-wide Industrial Process Control Problem. *Comput. Chem. Eng.* **1993**, *17* (3), 245−255.

(23) Ricker, N. L.; Lee, J. H. Nonlinear Model Predictive Control of the Tennessee Eastman Challenge Process. *Comput. Chem. Eng.* **1995**, *19* (9), 961−981.

(24) McAvoy, T. J.; Ye, N. Base Control for the Tennessee Eastman Problem. *Comput. Chem. Eng.* **1994**, *18* (5), 383−413.

(25) Ricker, N. L. Model predictive control of a continuous, nonlinear, two-phase reactor. *J. Process Control* **1993**, *3* (2), 109−123.

(26) Lyman, P. R.; Georgakis, C. Plant-wide control of the Tennessee Eastman Problem. *Comput. Chem. Eng.* **1995**, *19* (3), 321−331.

(27) Ricker, N. L. Decentralized control of the Tennessee Eastman Chanllenge Process. *J. Process Control* **1996**, *6* (4), 205−221.

(28) Ricker, N. L. Optimal Steady-state Operation of the Tennessee Eastman Chanlenge Process. *Comput. Chem. Eng.* **1995**, *19* (9), 949−959.

(29) Ricker, N. L. Tennessee Eastman Challenge Archive. Available at http://depts.washington.edu/control/LARRY/TE/download.html, 1999.

(30) Johannesmeyer, M. C.; Singhal, A.; Seborg, D. E. Pattern Matching in Historical Data. *AIChE J.* **2002**, *48* (9), 2022−2038.

Received for review March 18, 2004
Revised manuscript received August 13, 2004
Accepted August 13, 2004

IE0497893