IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received October 24, 2017, accepted November 16, 2017, date of publication December 4, 2017, date of current version February 14, 2018.

Digital Object Identifier 10.1109/ACCESS.2017.2778095

Time Sequential Phase Partition and Modeling Method for Fault Detection of Batch Processes

A. XIAOFENG YE[1], B. PEILIANG WANG[1,2], AND C. ZEYU YANG[3]
[1]Institute of Electron Devices and Application, Hangzhou Dianzi University, Hangzhou 310016, China
[2]Institute of Information and Control Technology, Huzhou University, Huzhou 313000, China
[3]Institute of Industrial Process Control, Zhejiang University, Hangzhou 310027, China

Corresponding author: B. Peiliang Wang (wpl@zjhu.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 61573137.

ABSTRACT Different operation phases in batch processes cover distinguishing behaviors, so establishing statistical models for each identical phase become an effective way for batch monitoring. In this paper, a new adaptive phase partition and online fault detection method is proposed, which can track the phase's transition by time sequence and has less reliance on parameters' selection. The discussion and analysis of this proposed method follows. In this proposed method, the information contained in every sample time will be evaluated, and the change tendency of feature is demonstrated on a batch prospect. Then, two control bounds are designed for the feature tendency, the stable, and the transitional phases that have a different feature level and play certainly roles in process operation, will be identified automatically. For online monitoring, the new fault detection strategy is composed of modeling the PCA and PLS statistical methods for each identified phase, three statistics are established to ensure the data-decomposing reliable. The proposed method is applied to the industrial penicillin fermentation process, and the experimental result shows better performance in phase partition and fault detection.

INDEX TERMS Multi-phase monitoring, adaptive phase partition method, information evaluation, feature tendency, fault detection strategy.

I. INTRODUCTION

The scale of product and process flow has sharply risen with the development of modern industry. The quality and safety of production attract increasing attention. Batch processes, as a significant production method, are widely used in the fields of chemicals, semiconductors, food, etc. Its multi operation character contributes to produce a product step by step safely and allows us to achieve high-quality production at a low cost.

The monitoring and fault detection model of industrial batch process is generally established from historical production data, apart from keeping the product program on an efficient, safe and stable operational status, the monitoring model also helps to find and eliminate abnormal situations in time by analyzing the statistical information among measured data. Multi-way principal component analysis (MPCA) [1]–[3] and Multi-way partial least squares (MPLS) [4]–[6], both are multivariate statistical process control (MSPC) methods, have commonly been used in data analysis and process monitoring. Duchesne and Mcgregor [6] first applied MPLS method to monitor the quality of batch production. The proposed method takes all the history data as model's input to extract the relative features between the process variables and quality-related variables. However, the MPLS model fails to consider the high correlation contained in measured data and the influence of the phase reaction in local regions, both of them have great impact on the final production quality.

In fact, the industrial batch process has many inherent characteristics, such as multi-phases and local reactions. Multi-phases based process monitoring method has been widely reported [7]–[9]. Zhao *et al.* [8] proposed that the correlation of variables in the batch process does not vary with time, but closely relate to the process's operation and mechanism. For example, the injection process can be divided into three phases according to the product program. Among these phases, the measured variables show different data characteristics in different operation states and modes, while the correlation of measured variables keeps stable in the same mode. From a statistical view, the mean, variance, and correlation of the measured variables show the phases' diffidence. Therefore, the multiphase method is widely used in quality-related fault monitoring [9]–[11]. Duchesne and Mcgregor [6] proposed the multi-block PLS trajectory method for fault detection. This method extracts the features that related to the multi-phases and quality with the information of middle process variables. However, the intermediate quality information is rare in actual process, which limits its actual application. Lu and Gao [12] extracted the principal components of time slice along the time duration to indicate the transition of phase's characteristics. They found that the phase's characteristics keep similar in the same phase and differ between phases. Finally, the k_mean clustering algorithm is employed to phase's identification, and the result performs well. Zhao *et al.* [13] proposed the soft partition method for phase identification to overcome the shortcomings of k_mean algorithm, such as hard division for each sample data. The basic idea of phase partition based on the cluster-based method [12]–[16] is that process features follow with phases' switch and keep similar in the same phase. These methods can effectively capture the dynamic characteristics without filling the unknown and missing data. However, there are still some deficiencies in these methods. For example, the cluster-based algorithm doesn't consider the time sequence of phase duration, and it easily leads to the time crossover of sample data. In addition, the neglect of the characteristics from one phase to another also descends the accuracy of the representative model. Moreover, the clustering methods are often limited by the choice of parameters, such as the initial center, the initial number of clusters, and the minimum phase's time duration.

In order to overcome the shortcomings of cluster-based method in phase identification, it is beneficial to analyze the inherent nature of process at each sample time in which we could capture the time-variant phase's information [11], [17]–[19]. Therefore, an indicator that has the ability to reveal the change of the phase's features in a running process is required. In fact, a number of phase partition methods based on feature indicator have been developed in recent years [20]–[24]. Sun *et al.* [20] track the cumulative percent variance of the principal component to get a phase division effect. However, it fails to capture the variation in different batches, and be insensitive to phase switch. Chang *et al*. [21] have proposed a two-step phase partition method by tracking the variation of principle components' number and direction. After the two-step program finished, the stable and transition phase will be identified according to the tendency of mean load matrix's similarity. Zhao *et al*. [22] have proposed an SSPP algorithm which considers the time sequence for phase partition, the phases and transition patterns are in fact identified by capturing their influences on monitoring performance. In addition, many effective indicators had been designed for phase partition [23]–[25], such as the dissimilarity index, where the information of neighbor time slice is integrated for phase partition.

Commonly, all of the phase partition methods mentioned above not only analyze the information contained in sample times but also develop a process monitoring model for each phase. However, none of them consider the changing information from whole batch's prospect, the phase partition results are often constrained by the local phase duration view, in other words, they lacks a unified and global index to compare the information in different time slices that are not in neighbor sample times. The hidden Markov model (HMM) chain [26], [27] is an effective pattern recognition method, its rich mathematical structure enables it to learn a system state's transition progress from training data, which is suitable to explain and assess the change of information from a global prospect [28]–[30].

In this paper, a new adaptive phase partition method and online fault detection method is proposed. It can track the phase's transition by time sequence from a batch's prospect. Here, we assume the batch duration is the same for different batches. To begin with, the information contained in every sample time will be evaluated, and the change tendency of feature is demonstrated on a batch prospect. The tendency of feature revolution not only implies phase's duration length, but also provides the information of state transition. The stable and the transitional phases that have different feature level and play certainly roles in process operation, will be identified automatically, meanwhile, the short time duration between phases and the transition process are also well described. Discussion and analysis are followed. When the partition program finished, the batch duration will be divided into several time intervals. For online monitoring, the PLS-based monitoring method aims to extract the relative information between process variables and quality variables, but fails to provide interpretation to the residual part of process variables, which brings the confusion to fault location and reduces the monitoring model's accuracy. The new fault detection strategy is composed of modeling the PCA and PLS statistical methods in each identified phase, three statistics are established to ensure the data decomposition reliable. The proposed method is applied to a typical batch process, industrial penicillin fermentation process. The experiment proves that the proposed method has a better performance in monitoring and fault detection.

The remained of paper is organized as follows. The phase partition method, with less reliance on parameter selection, is proposed to separate the process duration into several phases automatically and discussion is conducted in section II. In section III, a monitoring frame is established for batch processes, online and offline model is discussed. A typical industry batch process is used to confirm the efficiency of multi-phase identification and fault detection, the analysis follows in section IV. Finally, we present our conclusions in section V.

II. METHODOLOGY

A. PCA AND PLS METHOD

Partial least squares (PLS) [3], [4] is one of the important data-decomposed methods in multivariate statistical analysis. It focuses on finding the most relative and explainable directions to extract low-dimensional features between matrixes X and Y. These explainable directions pay attention to predictive ability from input space to output space. Multi-way partial least squares method (MPLS) is an extensive method of PLS. It firstly transforms the 3-dimensional matrix into 2-dimensional matrix with the variable or batch unfolded method, and then PLS is employed to extract low-dimension feature on those 2-dimensional matrixes subsequently. These two unfolded methods mentioned above not only preserve the nonlinear time-vary character of batch data, but also keep the average trajectory cross batches. Generally, the variable unfolded method does not need to predict the unknown data, and it also does not require the same batch duration. The batch unfolded method refers to the equal duration of batch's trajectory, and the unknown data must be filled. The MPLS data-decomposed form for 2-dimension matrixes X and Y follows:

$$X = TP^T + E = \sum_{i=1}^{k} t_i p_i + E$$
$$Y = UQ^T + F = \sum_{i=1}^{k} u_i q_i^T + F \tag{1}$$

Where the matrixes T and P are the score and load matrix of X. U and Q are the score and load matrixes of Y. k is the number of latent variables retained. E and F are the residuals matrixes of X and Y, respectively.

The PLS method is a predictive regression method, which is responsible for the data space decomposition from one to another. For online process monitoring, the traditional PLS-based monitoring method pays more attention to the relevant information between process variables and quality variables than the fault information's location that may appears in a less relevant correlation. In other words, there may be a large variable information exist in the X's residual matrix E. In order to ensure the integrity of the monitoring, the PCA method is further carried out to extract the residual information in matrix E, and the extracted residual form is as follows:

$$E = T_x P_x + \tilde{E}$$
$$T_x = EP_x \tag{2}$$

The matrix X can be further decomposed into the following form:

$$X = TP^T + T_x P_x + \tilde{E} \tag{3}$$

When the MPLS and PCA method are devoted to monitoring model, the statistics and corresponding control limits are calculated from off-line training data, as TABLE 1 shows:

TABLE 1. Statistics and corresponding control limits.

Statistics	Control limits
$T_i^2 = \sum_{i=1}^{k} (\frac{t_{ij}}{\sigma_j})^2$	$T^2 = \frac{k(K^2-1)}{K(K-k)} F_{k,n-k,\alpha}$
$Q_i = \sum_{j=1}^{n} (X_{i,j} - \tilde{X}_{i,j})^2$	$Q = \sum_{j=1}^{J} e_{ij}^2 \sim g\chi^2_{h,\alpha}$

Among them, t_{ij} is the j_{th} principal component vector in the PLS model corresponding to the i_{th} sample. σ_j is the standard deviation of the j principal component, k is the number of principal components, K is the number of training samples, and $X_{i,j}$ and $\tilde{X}_{i,j}$ are the measured value and the reconstructed value of the i_{th} sample. $g = v_k/2m_k$, $h = 2m_k/v_k$, m_k, v_k, respectively, represent the mean and variance of all Q_i statistics at the k_{th} time interval, and Q is the calculated control limits.

For on-line monitoring, the new collected data $x_{new}(1 \times J)$ is standardized by the training data, the low-dimensional principal component score vector t_{new} and residual vector e_{new} are obtained by the following form.

$$t_{new} = x_{new} w/(p^T w)$$
$$e_{new} = x_{new} - t_{new} p^T \tag{4}$$

Among them, w is the weight coefficient of X in the PLS model, which reflects the influence of elements in x_{new}.

Furthermore, the detection statistics T_{new} and Q_{new} are calculated based on t_{new} and e_{new}. If $T_{new} > T^2$ or $Q_{new} > Q$, the process may be out of control, which indicates that there may be a fault and the specific situation needs further analysis and verification.

B. HIDDEN MARKOV MODEL CHAIN

The hidden Markov model (HMM) [25], [26] chain is a probabilistic model of time series. It describes a transit process for an unobservable random state sequence to infer an observable data's generation probability. The HMM model can be symbolized as $\lambda = (A, B, \pi)$. The state probability transition matrix A and the state probability vector π jointly confirm a Markov chain, and both of them also generate an unobservable hidden state sequence. The observable probability matrix B and state sequence commonly determine an observable data's generation probability. FIGURE 1 shows a process of hidden state sequence transit with in HMM model.

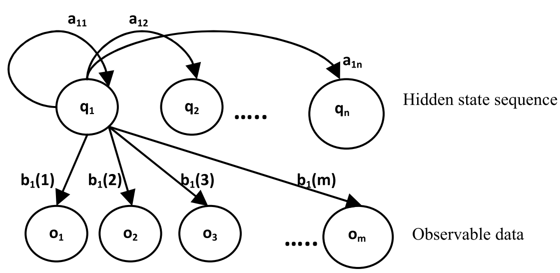


FIGURE 1. A process of hidden state sequence transit in HMM model.

The HMM model's parameters are defined as follows:

1) State probability transition matrix $A = [a_{ij}]_{N \times N}$. A is the transit probability for hidden state q_j transfer to q_i when the sample time t steps to $t+1$, where $a_{ij} = p(S_{t+1} = q_i | S_t = q_j)$, $i = 1, 2.., N, j = 1, 2, .., N$
2) Observation probability matrix $B = [b_{kj}]_{M \times N}$, $b_j(k) = p(o_t = v_k | S_t = q_j)$, $k = 1, 2, .., M$, $j = 1, 2, .., N$,

where $b_j(k)$ is the probability of observable data v_k that generated by state q_j on t_{th} sample time.

3) State probability vector π, π is the initial probability of the state q_i, where $\pi_i = p(S_1 = q_i)$, $i = 1, 2, .., N$.

A new parameter training method of the HMM model was proposed in paper [31], [32]. Setting the initial state number as N, the dimension of the initial probability distribution vector and the output observation data as M, which is determined by experience and cross validation, the state transition probability matrix A has a discrete distribution, and the output observable matrix B has a continuous distribution, both of them are learned by the forward algorithm [31]. The observable probability matrix B is generated by a Gauss Mixture model on each hidden state, $b_k(o) = \sum_{j=1}^{M} c_{kj} G(o, u_{kj}, \sigma_{kj})$, where o is the observable data vector and c_{kj}, u_{kj}, σ_{kj} correspond to the mixture coefficient, mean vector and covariance matrix of the j_{th} Gauss component on k_{th} hidden state, respectively. The number of random state sequences in the HMM model will affect the performance of the system, therefore, our paper uses Bayesian information criterion [10] to determine the state sequence's number as 5 and the number of Gauss component in Gauss Mixture model as 5.

The information contained in the observable data $O = \{o_1, o_2, .., o_M\}$ are assessed by the HMM model as follows:

$$I_{est} = P(o_1, o_2, o_3, ..o_N) = log(\prod_{j=1}^{N} b_j(k)) \tag{5}$$

Where $b_j(k) = P(o_t = v_k | S_t = q_j)$, k $= 1, 2, .., M$ and $j = 1, 2, .., N$. I_{est} is generated by the inner Markov chain in HMM model, which is evaluated by the transit of inner hidden state. Its value represents the relevant level for observable data $O = \{o_1, o_2, \ldots, o_k, \ldots, o_M\}$, therefore, when the o_j is closer with other vector in data set O, the larger I_{est} is, in other words, more closer in the data space.

C. PHASE PARTITION METHOD

Industrial batch process has many operation phases, and different phases present different data distribution. For online monitoring, a single PLS-based model trained by whole batches data proves poor monitoring performance. It is necessary to develop multi-phase models for batch process to handle the features contained in smaller and local phases.

In the current work, the assumption is that the batch duration is the same for different batches. The three-dimensional historical batch data $X(I \times J \times K)$ will be transformed into a 2-dimensional matrix by the batch or variable unfolded method. Where, I refer to the batch's number, J is the dimension of process variable, K is the batch duration. The batch method keeps the main dynamic characteristics along time and batches, and the variable method retains the nonlinear time-varying trajectory. The advantages of two methods mentioned above can be integrated in actual practice. $X(I \times J \times K)$ is firstly arranged as $X(J \times IK)$ by the variable unfolded method, and its class quantities are eliminated. Then, $X(J \times IK)$ is transferred as time slices $X(I \times JK)$ with batch unfolded method. The proposed method views $X_i(I \times J)$, $i = 1, 2, .., K$, as a basic time slice unit to capture the time-varying feature along the time direction. The time slice based on batch unfolded method implies the phase's transition information, which allows us to track the tendency of process's feature revolution from batch and time prospect. The basic phase partition program is shown in FIGURE 2.

In this section, an adaptive phase partition method based on time slice is delivered. It mainly takes the principle components of each time slice unit into consider. The phase partition progress we proposed is consists of two parts, generation of features tendency and segment of features tendency. **1)** Generation of features tendency. The PCA method is often used to extract the principle components (PCs) in time slice unit. It has the ability to find the largest explanatory orthogonal basis for data distribution, which provides a feature extraction scheme for every sample time. While the PCA method lacks the ability to compare and assess the information in two different time slices. Therefore, an indicative index is required to be designed to effectively show the feature change in batch process. In this paper, The HMM model, due to its powerful mathematical structure, is employed to fit system status. When all time slices' PCs had been extracted by PCA, the features tendency of batch process will get with the help of HMM's overall evaluation from batch view. **2)** Segment of features tendency. The features tendency shows the relevant level of changing features with time sequence, and the distribution character of features tendency is adopted for the phase partition. Finally, the batch duration will be divided into three type phases: the stable phases, the transition phases, and the shorter phase duration intervals of batch process are well described. What's more, the shorter phase duration intervals that have fast feature shifts are commonly identified, and a merging strategy is carried out to handle these special regions.

This phase partition method relies on the information contained in time direction, and the phase result follows with time sequence. The proposed phase partition method pays less attention to parameter selection and helps to reduce prior experience on a special condition in practical application.

1) DATA PREPROCESSING

First, the history process data $X(I \times J \times K)$ is standardized to $\bar{X}$(IK × J), whose mean is 0 and variance is 1 according to the variable unfolded method. Second, the data matrix $\bar{X}(IK \times J)$ is transformed into $\widehat{X}(I \times KJ)$ by the unfolded forms are shown in FIGURE 3; the basic time slice matrix is $\widehat{X}_k(I \times J)$, where $k = 1, 2, 3, .., K$

2) TIME SLICE's INFORMATION EXTRACTION

The principal component's information is extracted using the PCA method in all K time slices, and the number of principal components (PCs) is determined from the variance

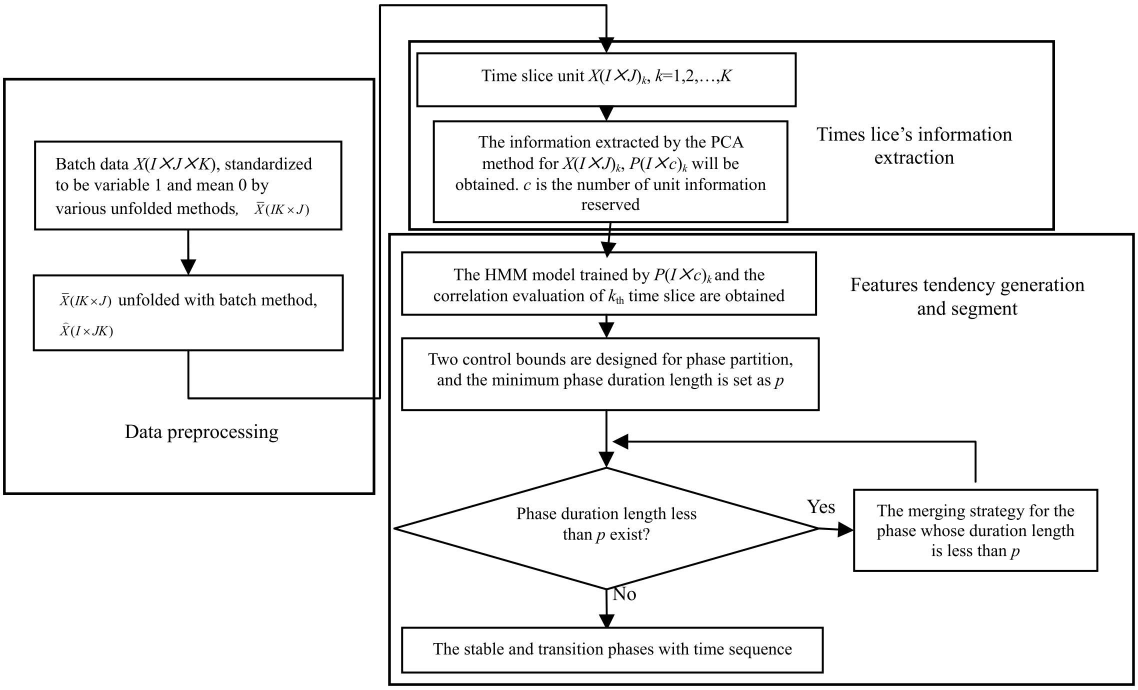


FIGURE 2. The program for phase partition.

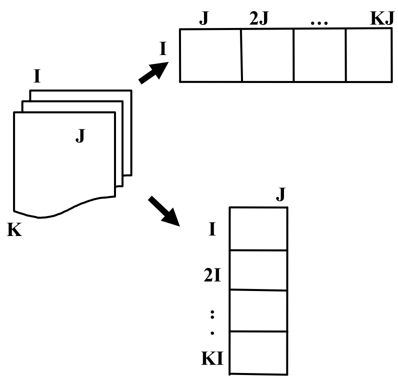


FIGURE 3. Three-dimensional matrixes with the batch and variable unfolded methods.

accumulation rate (>90%), recording the maximum number of PCs in all K time slice as c. To guarantee the comparison of the changing information in each time slice, the degree of principal information held in all K time slices is unified by the number c. The PCA information extraction for the k_{th} time slice matrix is as follows.

$$\begin{aligned} \widehat{X}_k &= T_{kc}P_{kc} + E \\ T_{kc} &= \widehat{X}_k P_{kc} \end{aligned} \tag{6}$$

Among them, T_{kc} and P_{kc} are the score and load matrix of time slice $\widehat{X}_k(I\times J)$ and c is the maximum number of principal components in all K time slices.

3) GENERATION OF FEATURES TENDENCY

After the principal components of each time slice has been extracted by the PCA method, all K load matrix P_{kc} are obtained. The K load matrixes represent the data space distribution of time slices. We use the HMM model to fit these data information, and the quantitative the correlation analysis evaluated by HMM model is shown in FIGURE 4 and the detailed assessment program follows:

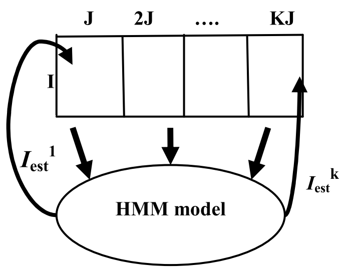


FIGURE 4. Time slice correlation analysis based on the HMM model.

a) Using all K load matrix P_{kc} to train the HMM model.
b) The trained HMM model is used to evaluate the correlation of all K load matrix P_{kc}, and the evaluation index $I_{\text{est}}^{\text{i}}$ is attained by equalization (5), where $i = 1, 2, .., K$.

This evaluation index is based on the overall significant system fitness where comprehensively evaluate between the current i_{th} sample time and the other sample time that excludes the i_{th}.

4) SEGMENT OF FEATURES TENDENCY

The evaluation index $I_{\text{est}}^{\text{i}}$ shows a stable variation degree in some time intervals, some switch rapidly and last short, which corresponds to the longer duration operation phase and the transition period in the industrial process mode. Therefore, the variation degree of evaluation in the sample time interval can be viewed as the basis for phase division, which is reasonable for multi-phase identification and helpful to improve the detection accuracy and sensitivity of the monitoring model.

a) Calculating the mean and variance of all K time slice evaluation index $I_{\mathrm{est}}^{\mathrm{i}}$, as $I_{\mathrm{esti\_mean}}$ and $I_{\mathrm{esti\_std}}$.

b) Recording the i_{th} time slice evaluation index as $I_{\mathrm{esti}}^{\mathrm{i}}$ and the control upper and lower bounds of $I_{\mathrm{est}}^{\mathrm{i}}$ value as

$$I_{esti\_h} = I_{esti\_mean} + \alpha I_{esti\_std}$$

and

$$I_{esti\_l} = I_{esti\_mean} - \alpha I_{esti\_std}$$

Where, α is the control bound factor that determines the result of phase identity. The larger α is, the less the identical phase's number is. The smaller α is, the easier to exert shorter time duration and more phases. Therefore, the selection of α should be following actual process data. The value of P_{index} (i) shows the control area of the i_{th} time slice hold.

$$P_{\mathrm{index}}(i) = \begin{cases} -1 & I_{esti}^{i} < I_{esti\_l} \\ 0 & I_{esti\_l} < I_{esti}^{i} < I_{esti\_h} \\ 1 & I_{esti}^{i} > I_{esti\_h} \end{cases} \tag{7}$$

c) Continuous time intervals with different lengths are obtained by merging the sample time points with the same P_{index} (i) values, and p is set as the minimum time duration length of a phase. Define a time interval whose length is shorter than p as relatively short time duration of a phase. The single sample time or short time intervals $L_{\mathrm{min}}^{\mathrm{j}}$ are obtained by the serial positioning of all time intervals. The short time intervals $L_{\mathrm{min}}^{\mathrm{j}}$ means the short phase duration of swift features, and they may appear in a stable operation or a transition operation of the batch process. We carry out a precise scheme to handle such conditions. Thus, the merge direction of short time interval $L_{\mathrm{min}}^{\mathrm{j}}$ is determined by using the following form.

$$abs(mean(I_{esti}^{j}) - I_{esti}^{prev}) - abs(mean(I_{esti}^{j}) - I_{esti}^{next}) = \begin{cases} > 0 & j \in next \\ < 0 & j \in prev \end{cases} \tag{8}$$

Among them, $I_{\mathrm{esti}}^{\mathrm{prev}}$ and $I_{\mathrm{esti}}^{\mathrm{next}}$ are the average values of I_{esti} in the continuous time interval before $L_{\mathrm{min}}^{\mathrm{j}}$ and after $L_{\mathrm{min}}^{\mathrm{j}}$, where their phase duration length is larger than p. The $abs(*)$ and $mean(*)$ are the operation of absolute and mean in math, respectively. $I_{\mathrm{esti}}^{j} \in L_{\mathrm{min}}^{j}$, $\mathrm{j} = 1, 2, .., h, h < K$.

D. DISCUSSION AND ANALYSIS

With the segment of feature tendency, the batch duration is divided into multi-phases. The advantage to use correlation evaluation as an indicator for phase partition is that it offers a significant correlation evaluation in a global perspective. Furthermore, the changing degree of time slice information evaluation in the stable and transitional phases is devoted to identify multi-phases in the whole batch duration.

1) THE UNIFIED INFORMATION CONTAINED IN TIME SLICES

Time slice is viewed as the basic analytical unit, because different information contained in sample time point of batch process. To ensure the comparability of information between time slices, the PCA method is applied to extract the information of time slice whose principal components number is determined by the cumulative variance rate ($> 90\%$) method, and the largest number of principal components in all the time slices is considered as a unified information reserve. This makes the information of a time slice comparable and ensures the consistency of HMM model information evaluation.

2) PHASE IDENTIFICATION AND MERGING STRATEGY

When the phase partition method is employed to identify the stable and transition phases in batch process, all sample time will be collected into several time intervals with diffident length. $P_{\mathrm{index}}(i)$ shows the index of $I_{\mathrm{esti}}^{\mathrm{i}}$ in the i_{th} time slice, which is relative to control limits $I_{\mathrm{esti\_h}}$ and $I_{\mathrm{esti\_l}}$. The stable phase often performs a long time interval and same P_{index} value, while in the transition phase whose features shift fast as well as the $P_{\mathrm{index}}(i)$ value, the time interval with same $P_{\mathrm{index}}(i)$ value has a short duration and out of rule. We can partition phases roughly according to the duration length of continuous P_{index} value. The sample time point with the same $P_{\mathrm{index}}(i)$ will be collected in a continuous interval, then, many time intervals with different lengths are obtained. Define a time interval whose length is shorter than p as relatively short time duration of a phase. Therefore, the time intervals can be divided into three kinds roughly, stable intervals, transition intervals, single time points and short duration intervals. Single time points and short duration intervals means shorter phase durations with swift features, and they may appear in a stable operation or a transition operation of a batch process. We carry out a precise scheme to handle such conditions. The merge direction of a short time interval will be determined by using the following form. If they stay between transition intervals, it is reasonable to attribute them to transition intervals with the transition operation's characteristics. If their neighbors are stable operations, they may be discrete time intervals or single sample times in the normal process condition; it is better to classify them as partials of stable intervals, instead of abandoning them. If the single time point and short duration intervals hold as narrow stable intervals and transition intervals, the mean of $I_{\mathrm{esti}}^{\mathrm{i}}$ of the three intervals above will be compared to make sure of the certain merging direction of the short duration intervals. Finally, the whole reaction time of the process is divided into the stability phases and transition phases adaptively.

III. ONLINE MONITORING STRATEGY

To obtain the process monitoring model, batches process variable data $X(I \times J \times K)$ and quality data $Y(I \times M \times K)$ are arranged as $X(IK \times J)$ and $Y(IK \times M)$ by the variable unfolded method, where their mean and variance are standardized to 0 and 1. Second, $X(IK \times J)$ and $Y(IK \times M)$ are rearranged

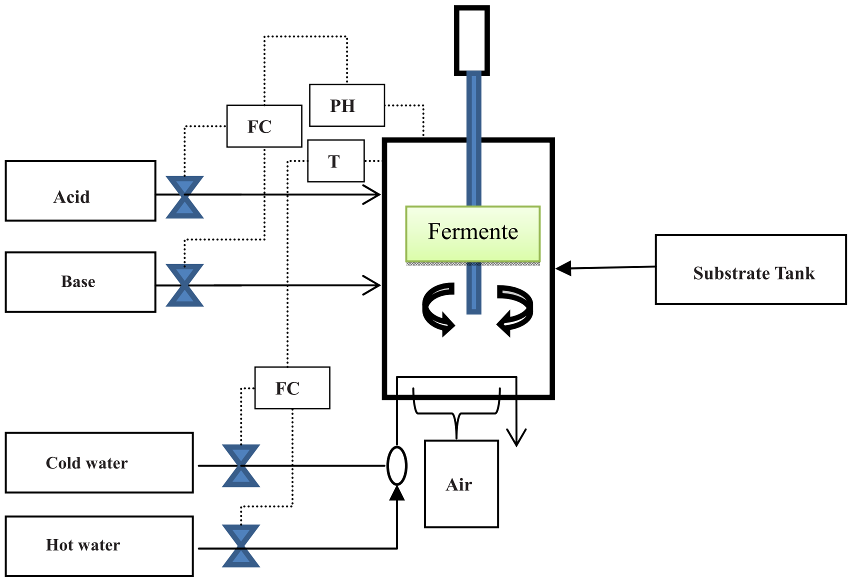


FIGURE 5. The fermentation process of penicillin production.

as $X(I \times JK)$ and $Y(I \times MK)$ by the batch unfolded method. The time slice $X_i(I \times J)$ is viewed as the basic unit for the phase partition, where $i = 1, 2, .., K$. After n phases have been identified, the PLS and PCA models are used to set up the quality-related monitoring model in each identified phase. Finally, as the TABLE 1 depicted, the statistical control limit T^2 is established for T_y^2 and T_x^2, and the statistical control limit Q is established for the residual matrix $\tilde{E}$.

When the new process data $X_{\text{new}}(1 \times J)$ is sampled online, it is firstly normalized by mean and variance obtained from historical training data, and the standardized data $x_{\text{new}}(1 \times J)$ is obtained. Then, the statistics t_y, t_x and e are calculated in the corresponding phase's model with $x_{\text{new}}(1 \times J)$.

$$\begin{aligned} t_y &= x_{new} w/(p^T w) \\ \tilde{x} &= x_{new} - tP \\ t_x &= \tilde{x} P_x \\ e &= \tilde{x} - t_x P_x \end{aligned} \tag{9}$$

Online fault detection and monitoring is conducted by continuously comparing the statistics t_y, t_x with T^2 or and e with Q. If any statistic is beyond the statistical control limit, the process may be out of control, there is a fault. Otherwise, the process is in the normal state.

IV. ILLUSTRATION AND DISCUSSION

A. PROCESS DESCRIPTION

The proposed phase partition and process monitoring method is applied to the actual multiphase process in this section. This phase partition method is effective in separating the stable operation phases for batch process. The transition process between neighbor's phases can also be expressed clearly. It overcomes the specific problems, such as time crossover and phase disorder, and the partitioned result follows with time sequence, which enhances the phase's interpretability. After the phases have been identified, online fault detection strategy is composed of modeling the PCA and PLS statistical methods in each identified phase, three statistics are established to assure the data-decomposing reliable. The effect of parameter α on monitoring performance is also analyzed.

Penicillin is one of the most widely used antibiotics in clinical medicine. It is a typical semi-batch process, which covers nonlinear, dynamic, and multiphase characteristics. The penicillin fermentation process is a two-microbe metabolic process. In the first phase, the growth and reproduction of the bacteria are carried out under certain conditions. When the concentration of the bacteria reaches the required level, penicillin is produced as a metabolite. In the penicillin fermentation process, in order to ensure the final penicillin production, the concentration of the bacteria must be maintained at a certain level. Therefore, sugar, nitrogen and other nutrients need to be added continually. FIGURE 5 shows the program for the penicillin fermentation production process. The temperature and pH are under closed-loop control, and the feeding process is under open-loop control.

Ali Cinar led his team in 2002 to develop a penicillin production simulation model Pensim 2.0. As a standard test platform, the model has been widely used in the field of batch process monitoring and fault diagnosis. This platform is a standard platform for researchers to process monitoring for home and abroad, with which you can simulate the microbial concentration, pH, carbon concentration, CO_2 concentration, penicillin concentration, oxygen concentration and heat generation under different operating conditions in the penicillin production process.

The penicillin fermentation production process can be divided into three phases: the cell growth period, the initial fermentation period and the fermentation period. Its batch

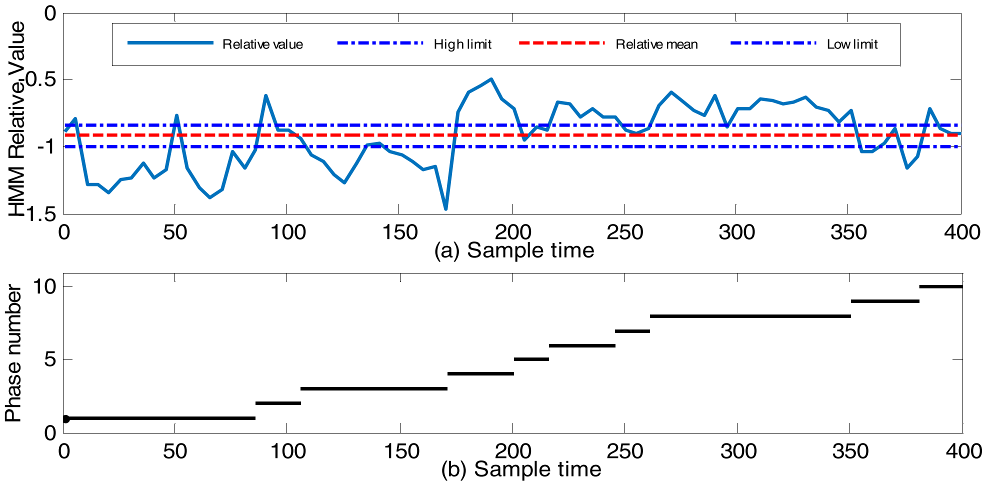


FIGURE 6. **(a) The correlation results of each time slice evaluated by HMM model (solid line). The control bounds are based on the evaluation result distribution (point line), and the dashed line is the average evaluation level; (b) the phase division result according to the phase division method when the control factor $\alpha = 0.25$.**

reaction time is 400 hours, and the sample time is 1 hour. In this experiment, 10 process variables and 2 quality variables are selected for quality-fault monitoring, as shown in TABLE 2. In order to make the training data reliable and adequate, this paper has produced 80 batches normal data as the reference database for model training. Among them, 70 batches are model training data, and 10 batches are test data.

TABLE 2. **Process and quality variables in the monitoring of penicillin.**

No.	Process variables	No.	Quality variables
x_1	Mixing rate(r/min)	y_1	Biomass conc(g/H)
x_2	Aeration rate(L/h)	y_2	Penicillin conc(g/L)
x_3	Substrate feed rate(L/H)		
x_4	Feed temperature(K)		
x_5	Dissolved oxygen rate(%)		
x_6	pH		
x_7	Alkali flow rate(g/H)		
x_8	Acid flow rate(g/H)		
x_9	Temperature(K)		
x_{10}	Cold water flow rate(L/H)		

The fault types of the penicillin fermentation process can be introduced by setting the amplitude and fault time for three variables. They are the aeration rate, agitator power, and substrate feed rate. The introduction of the specific form is shown in TABLE 3. All training batches are assume to have equal duration, which results in the three-way array data matrix $X(I \times 10 \times 400)$ and the quality data matrix $Y(I \times 2 \times 400)$, where I denotes the number of batches for normal and fault cases. 70 normal batches data are used for model development, and the other 10 batches are used for model test. The form of the fault cases is in TABLE 3.

B. PHASE PARTITION DEVELOPMENT

Firstly, the principal component's information of all time slice matrices is extracted by PCA method, and the number of

TABLE 3. **Fault types of the penicillin fermentation process.**

Various.	Fault.
Aeration rate (L/h)	1.Magnitude ramp+5%,when the time interval is 100 and the terminal is400
	2. Magnitude ascend +2% as step form, when the time interval is 100 and the terminal is 400
Agitator power(W)	3. Magnitude ramp+2%, when the time interval is 100 and the terminal is 400
	4. Magnitude ascend +5% as step form, when the time interval is 100 and the terminal is 400
Substrate feed rate(L/h)	5. Magnitude ramp+5%, when the time interval is 100 and the terminal is 400
	6. Magnitude ascend +2% as step form, when the time interval is 100 and the terminal is 400

principal components is determined by the variance accumulation rate (>90%). In order to guarantee the comparison of the changing information in all time slices, the maximum PCs number of time slices in this partition procedure is 6. According to the phase partition method in section 2, setting the control factor $\alpha = 0.25$ and the minimum phase duration time $p = 10$. FIGURE 6 shows the numerical evaluation determined by the HMM model and the phase partition results with the control factor $\alpha = 0.25$. It can be seen from FIGURE 6 (a) that when the time is50, the value of the evaluation curves rises and I_{esti} crosses the lower bound, the P_{index} value is different from the time point before 50, moreover, the length of the time interval around 50 with the same P_{index} value is less than the minimum phase duration p, which should be viewed as a relatively short time interval for subsequent merger to determine where this time period belongs to. It is known from FIGURE 6 (b) that the shorter time interval around 50 is determined as same as the continuous stable time interval before 50. In addition, the P_{index} value has 7 times change in time interval [300, 400] in FIGURE 6 (a), 3 phases are identified according to the minimum time duration and the characteristics of I_{esti}'s mean in the neighbor time interval in FIGURE 6 (b).

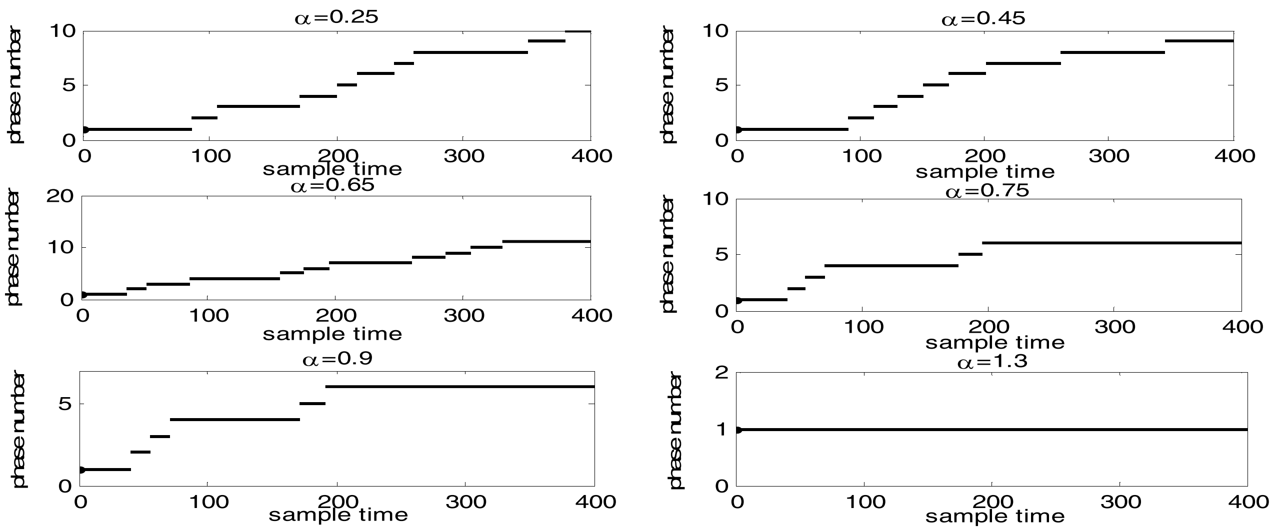


FIGURE 7. The selection of the different control boundary factors α for the phase division method and its phase partition results.

FIGURE 7 shows the phase partition results when different control factors α are selected. It can be seen that the smaller α is, the more phases we divide. Fewer long time duration phases mean swifter change degree of the process feature. As the α value increases, the phase that we had identified gradually decreases, the number of the stable and long period phases increases, and more sample times are considered to be similar with each other and attributed to the same stable phase. The transition patterns are missing since they are accommodated into neighboring phases, and the transition patterns are the time intervals with short time durations or the single points between the stable neighbor phases. For example, when α is set to 0.25, 10 phases are separated from the batch process, including 3 stable phases and 7 transition phases. When the α value increases to 0.75, the number of transition phases reduces to 4. Especially, the phase partition method that we proposed loses its ability when the value of α is larger than 1.3.it is noted that the partition phase is agreeable with the actual penicillin fermentation process when α value falls in the interval [0.75, 0.9]. In addition, the transition process between the stable phases is also shown to improve the monitoring process.

From the result, it is clear that the proposed phase partition method can divide the whole process into several continuous time duration intervals as time series. What's more, the transition and stable phases can be expressed, and the transition between phases has shorter time interval that more dynamic process characteristics than the stable ones. The division results using the proposed method are more direct and easy to understand, and no extra post-processing is required.

C. PHASE MONITORING MODEL

Based on the proposed phase partition method, different multi-statistical space models are developed for each identified phase and transition phase by the variable unfolded method, and batch data is assigned to the same time segment. To show the feasibility and ability of process monitoring, FIGURE 8 (a) shows proposed method for a normal batch with α is 0.75. Compared with FIGURE 8 (b), we can find that the detection limit follows as the phase partition in FIGURE 8 (a), which shows the continuity of phase's characteristics. When α is 0.75, the process is divided into 6 phases, including 3 stable and long duration time phases. In FIGURE 8 (a), the detection limit T_{y}^2 does not show the multi phase character because the numerical statistical limits are close in the first few phases. The detection limit Q_{x} is more indicative to the phase division. Considering the detect results of the normal batch, the results of the proposed method have more advantages than the PLS-based method, and the statistics of the normal batches are overall distributes below the statistical limit in FIGURE 8 (a). While, the Q^2 statistic of the PLS method in FIGURE 8(b) leads to the error detection at the beginning of the process, and this phenomenon last for a long time, which can provide error guidelines for the operation and product safety.

The cause of fault 4 is shown in TABLE 2. The Fault 4 occurs at the 100_{th} time point and lasts until the end of the process. FIGURE 9 (a) shows the detect result of the proposed monitoring method for fault 4 when control factor α is 0.75, the control limit T_{y}^2 has detected the fault happen at the 100_{th} time point and lasts until the end of the process, which shows better fault detection capabilities. FIGURE 9 (b) shows the detect result for fault 4 according to the PLS-based method. The control limit T^2 fails to detect the fault at the 100_{th} time point, what's worse, its fault missing and detection rate is low after the fault occurs, and fails to indicate the fault start time and duration of fault 4.The control limit Q^2 appears error detection at the beginning of the process, which reduces the reliability of the monitoring. In addition, the Q^2 statistic has an obvious numerical jump when the fault occurs and lasts to

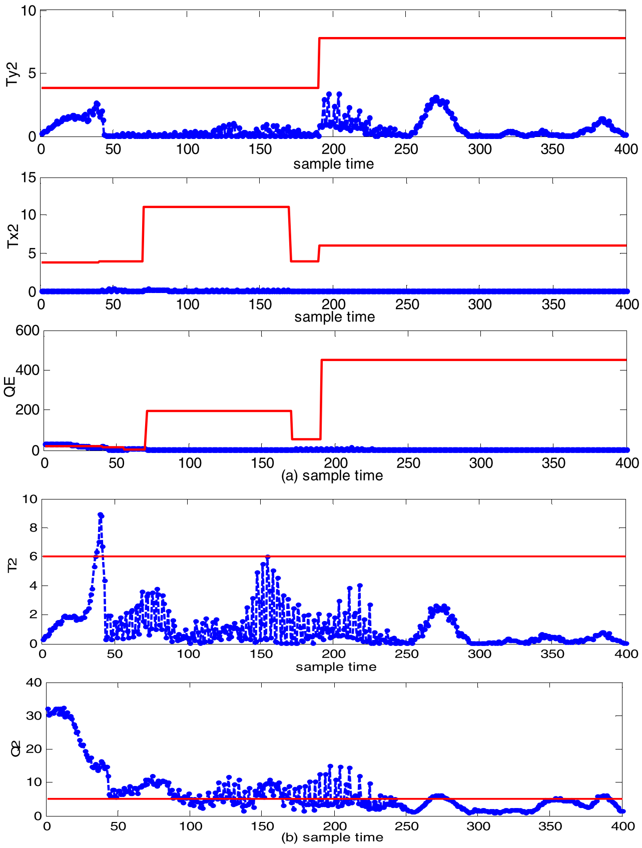


FIGURE 8. (a) $\alpha = 0.75$, the detection results of the proposed method for a normal batch. (b) The detection result of the traditional PLS method for a normal batch. The solid line corresponds to the detection limit at 95% confidence, and the point line is the calculated statistic.

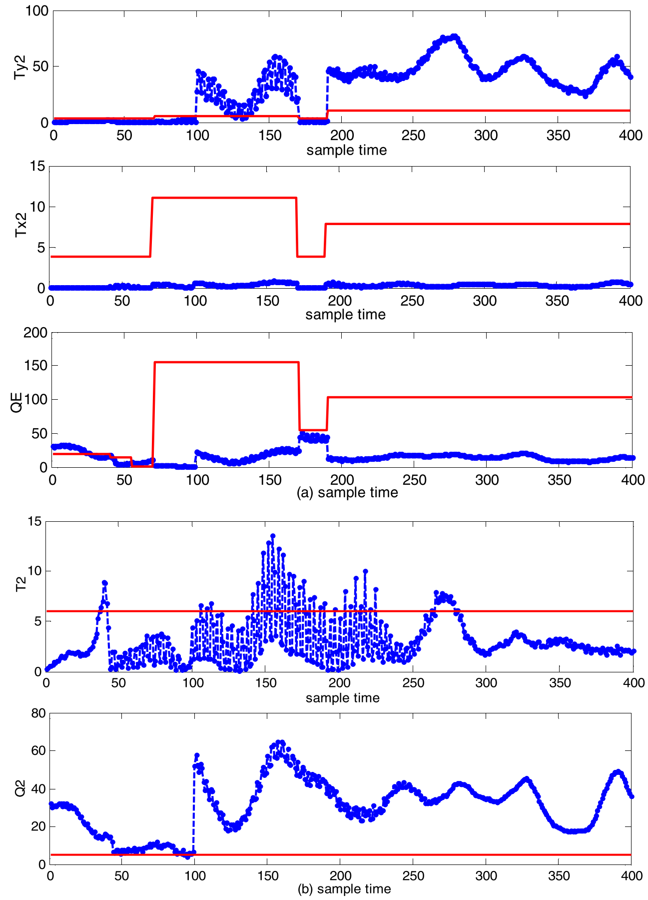


FIGURE 9. (a) $\alpha = 0.75$, the detection results of the proposed method for fault 4. (b) The detection results of the traditional PLS method for fault 4. The solid line corresponds to the detection limit at 95% confidence, and the point line is the calculated statistic.

the end of the process. From the detect results we can come to the conclusion that the proposed method has better detection performance than the traditional PLS-based method in process monitoring because the global monitoring model needs to handle the high correlation contained in training data, while the proposed method partition the data into several phase models, which enhance and strengthen the influence of local reaction processes that contribute to monitoring performance.

D. DISCUSSION OF THE CONTROL FACTOR α

Considering the influence of α value on the detection efficiency and fault recognition capability, FIGURE 10 and 11 show the fault detection rate (*FDR*) and the missing detection rate (*MDR*) detection results of the proposed multi-phase PLS method with different α values. Among them, the *FDR* is defined as the rate of the fault detection in the time interval [100, 400], and *MDR* is the rate of the normal sample error to judge as the fault one before the fault occurs. As the TABLE 3 figures out, fault 1 occurs at the 100_{th} time point, corresponding to the process variable x_1, has an increasing magnitude of 5%. FIGURE 10 (a) shows that the proposed multi-phase PLS method has a higher *FDR*, the *FDR* is more than 0.92 at different α value, which reveals a good detection performance. The control limit T_y^2 has the highest detection rate at α value interval [0.7, 0.9], which is defined as the optimal α interval for T_y^2, and the phase partition results are shown in FIGURE 7. The *FDR* of control limit Q_x^2 reaches its maximum point at α value interval [0.3, 0.6], but the *FDR* is less than T_y^2's. In FIGURE 10 (b), the *MDR* of control limit T_x^2 is the lowest and has the best monitoring ability. Due to the α interval of the maximum *FDR* for Q_x^2 has a larger *MDR* when compares to the optimal α interval of T_y^2, the value of control boundary factor α is appropriate in the numerical range [0.7, 0.9].

FIGURE 11 shows the *FDR* and *MDR* detection results of the proposed multi-phase PLS method for fault 3 with different α values. The *FDR* of detection limit T_x^2 and Q_x^2 are below 0.5, which shows poor detection ability, and the *FDR* of detection limit T_y^2 keeps increasing in the α interval [0.8, 1] and remains above 0.8 in FIGURE 11 (a).As we can see in FIGURE 11 (b), the T_x^2 statistic has the minimum *MDR* and remains unchanged at 0 with different α values. The minimum *MDR* of detection limit Q_x^2 exists in the α interval [0, 0.5], and the maximum *FDR* of α interval for T_y^2 corresponding to the maximum *MDR* is approximately 0.14. Compared with detection limit T_x^2 and Q_x^2, the *FDR* of the

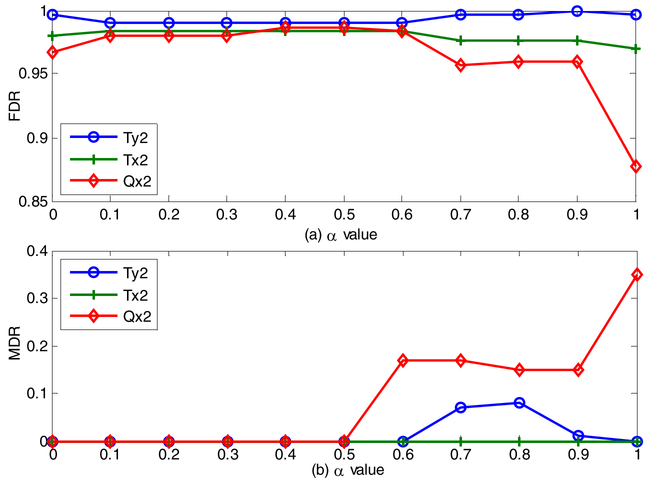


FIGURE 10. **(a) *FDR* of fault 1 with different α values; (b) *MDR* of fault 1 with different α values.**

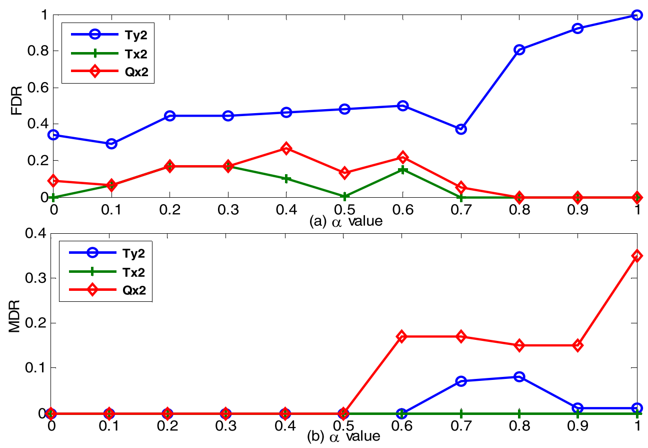


FIGURE 11. **(a) *FDR* of fault 3 with different α values; (b) *MDR* of fault 3 with different α values.**

statistical limit T_y^2's distribution has a great advantage in detection performance. When the α is 0.9, the detection rate of T_y^2 is approximately 0.85, while the maximum detection rate of Q_x and T_x^2 are 0.3 and 0.25, although the *MDR* of T_y^2 statistic is slightly larger than that two when α is 0.9. Therefore, combining the actual detection results, the α value should be selected in interval [0.85, 1].

V. CONCLUSION

In view of the multi-phases characteristics in batch processes, a new phase partition and online fault detection method is proposed to improve the precision and sensitivity of process monitoring. The time slice contains the potential features information about the batch phases, which is viewed as unit cell to partition batch process into several phase duration. The stable and the transitional phases, reveal different statistical character, will be identified automatically. It not only strengthens interpretation for process, but also reduces the complex of model. For online monitoring, online fault detection strategy is composed of modeling the PCA and PLS statistical methods for each identified phase, which aims to assure the data-decomposing reliable. The proposed method is applied to the industrial penicillin fermentation process. The experimental result covers better performance in phase partition and fault detection.

REFERENCES

[1] Z. Ge, Z. Song, and F. Gao, "Review of recent research on data-based process monitoring," *Ind. Eng. Chem. Res.*, vol. 52, no. 10, pp. 3543–3562, 2013.

[2] Z. Lv, J. Qingchao, and Y. Xuefeng, "Batch process monitoring based on multi subspace multi-way principal component analysis and time-series Bayesian inference," *Ind. Eng. Chem. Res.*, vol. 53, pp. 6457–6466, 2014.

[3] B. Wang, Q. Jiang, and X. Yan, "Fault detection and identification using a Kullback-Leibler divergence based multi-block principal component analysis and Bayesian inference," *Korean J. Chem. Eng.*, vol. 31, no. 6, pp. 930–943, 2014.

[4] A. Höskuldsson, "PLS regression methods," *J. Chemometrics*, vol. 2, no. 3, pp. 211–228, 1988.

[5] Z. Ge, Z. Song, S. X. Ding, and B. Huang, "Data mining and analytics in the process industry: The role of machine learning," *IEEE Access*, vol. 5, pp. 20590–20616, 2017, doi: 10.1109/ACCESS.2017.2756872.

[6] C. Duchesne and J. F. MacGregor, "Multivariate analysis and optimization of process variable trajectories for batch processes," *Chemometrics Intell. Lab. Syst.*, vol. 51, no. 1, pp. 125–137, 2000.

[7] X. Chen, X. Gao, Y. Zhang, and Y. Qi, "Enhanced batch process monitoring and quality prediction based on multi-phase multi-way partial least squares," in *Proc. Comput. Appl. Chem.*, 2011, pp. 32–36.

[8] C. Zhao, F. Wang, Y. Yao, and F. Gao, "Phase-based statistical modeling, online monitoring and quality prediction for batch processes," *Acta Automat. Sinica*, vol. 36, no. 3, pp. 366–374, 2010.

[9] L. Zhao, C. Zhao, and F. R. Gao, "Phase transition analysis based quality prediction for multi-phase batch processes," *Chin. J. Chem. Eng.*, vol. 20, no. 6, pp. 1191–1197, 2012.

[10] C. Undey and A. Cinar, "Statistical monitoring of multistage, multiphase batch processes," *IEEE Control Syst.*, vol. 22, no. 5, pp. 40–52, Oct. 2002.

[11] Y. Yao, W. Dong, L. Zhao, and F. Gao, "Multivariate statistical monitoring of multiphase batch processes with between-phase transitions and uneven operation durations," *Can. J. Chem. Eng.*, vol. 90, no. 6, pp. 1383–1392, 2012.

[12] N. Lu and F. Gao, "Stage-based process analysis and quality prediction for batch processes," *Ind. Eng. Chem. Res.*, vol. 44, no. 10, pp. 3547–3555, 2005.

[13] C. Zhao, F. Wang, N. Lu, and M. Jia, "Stage-based soft-transition multiple PCA modeling and on-line monitoring strategy for batch processes," *J. Process Control*, vol. 17, no. 9, pp. 728–741, Oct. 2007.

[14] X. T. Doan, R. Srinivasan, P. M. Bapat, and P. P. Wangikar, "Detection of phase shifts in batch fermentation via statistical analysis of the online measurements: A case study with rifamycin B fermentation," *J. Biotech.*, vol. 132, no. 2, pp. 156–166, 2007.

[15] S. K. Maiti, R. K. Srivastava, M. Bhushan, and P. P. Wangikar, "Real time phase detection based online monitoring of batch fermentation processes," *Process Biochem.*, vol. 44, no. 8, pp. 799–811, 2009.

[16] F. Shen, Z. Ge, and Z. Song, "Multivariate trajectory-based local monitoring method for multiphase batch processes," *Ind. Eng. Chem. Res.*, vol. 54, no. 4, pp. 1313–1325, 2015.

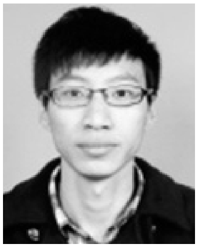
[17] W. Li, C. Zhao, and F. Gao, "Sequential time slice alignment based unequal-length phase identification and modeling for fault detection of irregular batches," *Ind. Eng. Chem. Res.*, vol. 54, no. 41, pp. 10020–10030, 2015.

[18] Y.-Q. Chang, Y.-S. Lu, F.-L. Wang, S. Wang, and S.-M. Feng, "Sub-stage PCA modelling and monitoring method for uneven-length batch processes," *Can. J. Chem. Eng.*, vol. 90, no. 1, pp. 144–152, 2012.

[19] N. Lu, F. Gao, Y. Yang, and F. Wang, "PCA-based modeling and on-line monitoring strategy for uneven-length batch processes," *Ind. Eng. Chem. Res.*, vol. 43, no. 13, pp. 3343–3352, 2004.

[20] W. Sun, Y. Meng, A. Palazoglu, J. Zhao, H. Zhang, and J. Zhang, "A method for multiphase batch process monitoring based on auto phase identification," *J. Process Control.*, vol. 21, no. 4, pp. 627–638, 2011.

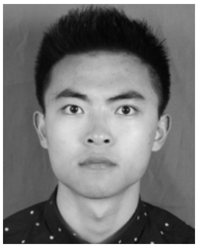
[21] Y. Chang, S. Wang, S. Tan, F. Wang, and Z. Mao, "MPCA based phase identification method and its application to process monitoring," in *Proc. IEEE Conf. Decision Control*, Dec. 2012, pp. 1245–1252.
[22] C. Zhao and Y. Sun, "Step-wise sequential phase partition (SSPP) algorithm based statistical modeling and online process monitoring," *Chemometrics Intell. Lab. Syst.*, vol. 125, pp. 109–120, Jun. 2013.
[23] W. Dong, Y. Yao, and F. Gao, "Phase analysis and identification method for multiphase batch processes with partitioning multi-way principal component analysis (MPCA) model," *Chin. J. Chem. Eng.*, vol. 20, pp. 1121–1127, Dec. 2012.
[24] L. Y. Jiang, B. J. Xu, J. H. Xi, and G. X. Fu, "Application of phase division based on dissimilarity index in batch process monitoring," *Adv. Mater. Res.*, vol. 566, pp. 134–139, Sep. 2012.
[25] C. Zhao, F. Wang, F. Gao, N. Lu, and M. Jia, "Adaptive monitoring method for batch processes based on phase dissimilarity updating with limited modeling data," *Ind. Eng. Chem. Res.*, vol. 46, no. 14, pp. 4943–4953, 2007.
[26] J. D. Bryan and S. E. Levinson, "Autoregressive hidden Markov model and the speech signal," *Proc. Comput. Sci.*, vol. 61, pp. 328–333, 2015.
[27] S. Calinon, F. D'halluin, E. L. Sauser, D. G. Caldwell, and A. G. Billard, "Learning and reproduction of gestures by imitation," *IEEE Robot. Autom. Mag.*, vol. 17, no. 2, pp. 44–54, Jun. 2010.
[28] H. Cao, D. Jiang, J. Pei, E. Chen, and H. Li, "Towards context-aware search by learning a very large variable length hidden Markov model from search logs," in *Proc. Int. Conf. World Wide Web*, Madrid, Spain, Apr. 2009, pp. 191–200.
[29] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Trans. Signal Process.*, vol. 46, no. 4, pp. 886–902, Apr. 1998.
[30] M. Gales and S. Young, "The application of hidden Markov models in speech recognition," *J. Found. Trends Signal Process.*, vol. 1, pp. 195–304, Jan. 2008.
[31] L. Bahl, P. Brown, P. de Souza, and R. Mercer, "Maximum mutual information estimation of hidden Markov model parameters for speech recognition," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, vol. 11. Apr. 1986, pp. 49–52.
[32] W. L. Kendall, G. C. White, J. E. Hines, C. A. Langtimm, and J. Yoshizaki, "Estimating parameters of hidden Markov models based on marked individuals: Use of robust design data," *Ecology*, vol. 93, pp. 913–920, Apr. 2012.



A. XIAOFENG YE received the B.Sc. degree in electronic and information engineering from Hangzhou Dianzi University, Hangzhou, China, in 2015, where he is currently pursuing the master's degree with the Institute of Electron Devices and Application. His main research interests include process monitor modeling and fault detection.

B. PEILIANG WANG received the B.Sc. degree in industrial electrical automation and the M.S. degree in control theory and control engineering from Zhejiang University, Hangzhou, China, in 1986 and 2005.

From 2008 to 2009, he served as a Visiting Scholar with Zhejiang University and the University of Duisburg-Essen (2015), Germany. He is currently a Professor with the School of Engineering, Huzhou University, China, and a part-time Master Tutor with Hangzhou Dianzi University and the Zhejiang University of Technology, China. He is currently a member of the committee on the fault diagnosis and safety of technical process of the Chinese Association of Automation, the Executive Director and the Deputy Director of the Teaching Committee of the Zhejiang Association of Automation, and the Chairman of the Huzhou Association of Automation. His research interests include pattern recognition and intelligent control, process monitoring and diagnosis, and industrial automation.



C. ZEYU YANG received the B.Sc. degree in electronic and information engineering from Wenzhou University, Wenzhou, China, in 2014, and the master's degree from the Institute of Electron Devices and Application, Hangzhou Dianzi University, Hangzhou, China, in 2017. He is currently pursuing the Ph.D. degree with the Institute of Industrial Process Control, Zhejiang University. His main research interests include process control, fault detection, and identification.