Batch-to-Batch Steady State Identification Based on Variable Correlation and Mahalanobis Distance

ARTICLE *in* INDUSTRIAL & ENGINEERING CHEMISTRY RESEARCH · OCTOBER 2009

Impact Factor: 2.59 · DOI: 10.1021/ie901107h

Yuan Yao
National Tsing Hua University

Chunhui Zhao
Zhejiang University

Furong Gao
The Hong Kong University of Science and T…

Batch-to-Batch Steady State Identification Based on Variable Correlation and Mahalanobis Distance

Yuan Yao, Chunhui Zhao, and Furong Gao*

Department of Chemical Engineering and Biomolecular Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong SAR, P. R. China

Online steady state identification (SSID) is an important task to ensure the quality consistence of final products in batch processes. Additionally, it is also critical for satisfactory control of many batch processes. The existing approach for batch process SSID is based on the multiway principal component analysis (MPCA) method, which requires history data from dozens of batches for process modeling. Consequently, this limits its online applications. In this work, principal component analysis (PCA) models are built for each batch from which the variable correlation information is extracted. The changes in variable correlation structures are then quantified with a PCA similarity factor. At the same time, the Mahalanobis distances between batch trajectories are also calculated to indicate the changes in variable trajectory magnitudes. These two types of information are then used for online SSID in batch processes. This method is more suitable for online applications and can solve the problems of uneven operation durations. Additionally, it can be easily extended to deal with non-Gaussian information and multiphase batch process characteristics. Application examples show the effectiveness of the proposed method.

1. Introduction

In industrial manufacturing, batch processes are widely applied. Different from continuous processes, batch processes can be indexed along either time direction within each cycle or batch direction from cycle to cycle.[1,2] With respect to the batch index, most batch process operations can be divided into two stages: batch-to-batch start-up and steady-state operation. During the runs in batch-to-batch start-up periods, the incoming materials usually have not mixed well, and the material properties and the machine conditions have not been stabilized either. Consequently, the batch operations in start-up are unsteady and cannot guarantee the satisfied products with consistent qualities. Reliable products are only manufactured in steady-state batch operations. Usually, the durations of the start-up periods are unknown and varied from one process to another. The defective products could not be rejected until a series of laboratory analyses are made, which costs quite a lot of labor, material, and financial resources. Therefore, an efficient method is desired for the online identification of the operation states of batch processes which can indicate the product qualities without lab analysis.

Meanwhile, the efficient online steady state identification (SSID) is also critical for satisfactory batch process control.[3] In process control, the steady-state models are widely utilized. Because of process nonstationarity, it is necessary to update the model parameters frequently, in order to make the model accurately reflect the nature of the process. Such parameter adjustments should only be made with the process data collected under approximately steady state conditions instead of in the batch-to-batch start-up periods.

Initially, the definition of steady state was made for continuous processes,[4] which requires the mean of time series data to be constant, regardless of whether the noises and disturbances are stationary or not. Since batch processes have the characteristics of their own, the steady state cannot be defined in the same way.[3] Batch processes operate in a cyclical manner. In each cycle, the process variable measurements form a set of variable trajectories over the batch duration. Usually, there are cross-correlations between process variables. Meanwhile, the variable trajectories are dynamic (autocorrelated) with certain magnitudes. Therefore, the batch process steady state can be defined in the following ways: at steady state, the variable correlation structures remain stable from batch to batch; meanwhile, the magnitudes of variable trajectories are similar between batches.

There are different types of existing methods developed by the researchers and the industry for SSID.[4] Among them, the method based on a ratio of variances[5] plays an important role. With this method, the data variance in a most recent window is calculated in two different ways. When the process is at the steady state, these two calculated variances have similar values which make the ratio between them close to one. Otherwise, the ratio will be much larger. Since the ratio is dimensionless and independent of the measurement level, this method can be applied to different processes. An alternative method was proposed by Cao and Rhinehart.[4] In contrast to the conventional method, they use three exponentially weighted moving (EWM) filters to estimate the sample average and variance, so that the past samples do not have to be stored. Their method was extended to multivariate cases by Brown and Rhinehart.[6] Subsequently, Ruin et al.[7] combined principal component analysis (PCA) with SSID, and performed SSID on the principal components (PCs) instead of the raw process data. To the best of our knowledge, a recently published paper by Aguado et al.[3] first introduced the related research on SSID in batch processes. In their method, a multiway principal component analysis (MPCA) model is utilized to summarize the batch trajectory information. Then, SSID is performed on PCs and residuals to detect the batch process steady state. However, in order to build a batch process model, MPCA usually requires history data from dozens of batches. It limits the online application of this method.

In this work, the problem of batch process online SSID is solved by dealing with the changes in both variable correlation structures and trajectory magnitudes. The proposed method only uses process data from each single batch in PCA modeling.

* To whom correspondence should be addressed. Tel: (852)-2358-7139. Fax: (852)-2358-0054. E-mail: kefgao@ust.hk.

10.1021/ie901107h CCC: $40.75
Published on Web 10/20/2009

Therefore, large amounts of history data are not needed, which indicates that this method is more suitable for online steady state detection. Moreover, it can be easily extended to deal with non-Gaussian information and multiphase batch process features.

This paper is organized as follows. In section 2, the preliminaries are reviewed, including the PCA/MPCA technique and the existing SSID methods related to this work. The proposed method, including the motivations, basic idea, and detailed procedure, is introduced in section 3. Then, application results are shown in section 4, which verify the effectiveness of the proposed method. Finally, the conclusions are drawn in section 5 to summarize the paper.

2. Preliminaries

2.1. Fundamentals of PCA and MPCA. 2.1.1. PCA and Dynamic PCA. In statistics, correlation indicates the strength and direction of the linear relationship between variables, which can be analyzed by PCA. Mathematically, PCA is an orthogonal linear transformation performed on a data matrix like $X(n \times m)$, where n is the number of samples and m is the number of variables. As shown in eq 1, it decomposes X as

$$X = TP^{\mathrm{T}} = \sum_{j=1}^{m} \mathbf{t}_j \mathbf{p}_j^{\mathrm{T}} = \mathbf{t}_1 \mathbf{p}_1^{\mathrm{T}} + \mathbf{t}_2 \mathbf{p}_2^{\mathrm{T}} + \cdots + \mathbf{t}_m \mathbf{p}_m^{\mathrm{T}} \quad (1)$$

where $\mathbf{t}_j(n \times 1)$ is the principal component (PC) vector which is orthogonal to each other, $\mathbf{p}_j(m \times 1)$ is the orthonormal loading vector which projects data into score space, T is the score matrix, and P is the loading matrix reflecting variable correlation structure. Algebraically, $\|\mathbf{t}_j\|$ is equal to the jth largest eigenvalue of the covariance matrix $\Sigma = X^{\mathrm{T}}X$, and $\mathbf{p}_j$ is the corresponding eigenvector. Therefore, most of systematical variation information is extracted by retaining first several orthogonal PCs as follows:

$$X = \sum_{j=1}^{A} \mathbf{t}_j \mathbf{p}_j^{\mathrm{T}} + \sum_{k=A+1}^{m} \mathbf{t}_k \mathbf{p}_k^{\mathrm{T}} = \hat{X} + E \quad (2)$$

where A is the number of retained PCs, $\hat{X}$ is the reconstructed data matrix, and E is the residual matrix containing noise information.

Before performing PCA on a data set, normalization is a necessary step to eliminate the effects of variable units and measuring ranges. The most common way of normalization includes removing means and equalizing variances. The formula is shown below:

$$\tilde{x}_{i,j} = \frac{x_{i,j} - \bar{x}_j}{s_j} \qquad (i = 1, \cdots, n; j = 1, \cdots, m) \quad (3)$$

where i is the sample index, j is the variable index, $\bar{x}_j$ is the mean value of variable x_j, and s_j is the standard deviation of variable x_j. More details of PCA can be found in ref 8.

Dynamic PCA (DPCA)[9] extends PCA to analyze process dynamics. In DPCA, first, an expanded data matrix is formed, which contains both current process measurements and lagged measurements in past several steps. Then, PCA is performed on it. By doing so, both variable cross-correlation information and the process dynamics indicated by the relationship between current and past variables can be simultaneously extracted.

2.1.2. MPCA. As introduced previously, PCA can only deal with two-dimensional data matrices. It cannot be directly applied to batch process modeling, since the data collected from a batch process are usually represented by a three-dimensional data matrix $\underline{X}(I \times J \times K)$, where I is the number of total batches, J is the number of process variables, and K is the number of total sampling time intervals in a batch.

The basic idea of MPCA[10,11] follows: First, $\underline{X}$ is unfolded to a two-dimensional data matrix $X(I \times KJ)$, by keeping the dimension of batches and merging variable and time dimensions. Then, PCA can be used to reveal variable correlation structure and systematic variation information by analyzing X. Since each batch operation is treated as one single sample, MPCA usually requires dozens of historical batches in modeling to obtain enough statistical information.

2.2. Related Works on SSID. 2.2.1. SSID Based on a Ratio of Variances. Let y_1, ..., y_n represent n successive observations. The ordinary formula for variance estimation is

$$s^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n - 1} \quad (4)$$

where $\bar{y}$ is the mean of the n observations; s^2 measures variance independently of the order of the observations. If a trend is present among the observations, the s^2 will include the effect of the trend.

Another estimation of variance can be derived from the mean square successive difference:

$$\delta^2 = \frac{\sum_{i=1}^{n-1} (y_{i+1} - y_i)^2}{n - 1} \quad (5)$$

$\delta^2/2$ is an estimation of variance which minimizes the trend effect.[12]

Von Neumann[13] suggested that the ratio of the mean square successive difference to the variance $\eta = \delta^2/s^2$ is suitable as a basis to judge whether a trend exists or not. On condition that the data are in a moving window and there is no trend in this window, the value of $R = 2/\eta$ is expected to be near 1; on the other hand, if the data follow a curve, $R = 2/\eta$ is statistically greater than 1. The critical values for trend detection can be found in ref 5.

This method was modified by Cao and Rhinehart[4] in 1995 to avoid the use of moving window. In their method, the sample average and variance are estimated using exponentially weighted moving (EWM) filters:

$$y_{f.i} = \lambda_1 y_i + (1 - \lambda_1) y_{f.i-1} \quad (6)$$

$$v_{f.i}{}^2 = \lambda_2 (y_i - y_{f.i-1})^2 + (1 - \lambda_2) v_{f.i-1}{}^2 \quad (7)$$

$$\delta_{f.i}{}^2 = \lambda_3 (y_i - y_{i-1})^2 + (1 - \lambda_3) \delta_{f.i-1}{}^2 \quad (8)$$

where y_i is the ith observation of the process variable, $y_{f.i}$ is the filtered value (EWM average) of y_i, $v_{f.i}{}^2$ relates to the filtered value of the variance, and $\delta_{f.i}{}^2$ is the filtered value of the mean square successive difference. λ_1, λ_2. and λ_3 are the filter parameters.

Therefore, the ratio of variances can be calculated as follows:

$$R = \frac{(2 - \lambda_1) v_{f.i}{}^2}{\delta_{f.i}{}^2} \quad (9)$$

Similar to $2/\eta$, when the process is steady, the value of R is close to 1. If a trend exists, both the numerator and the denominator will be affected. Because the numerator term will increase more and persist longer, R will be greater than 1, which indicates the nonsteady state. The critical values of R statistic can be looked up in the distribution table.[14]

Compared to the conventional method, Cao and Rhinehart's method does not need a moving window, which avoids the storage of the past data and the selection of the window length. However, their method also requires three filter parameters be specified. It is difficult to exactly tell which method is more convenient as both methods are reasonable and valid in SSID.

2.2.2. Batch Process SSID Using MPCA. To identify the steady state in batch process operation, Aguado et al.[3] combined MPCA with SSID technique. Their method can be divided into two major steps. At the first step, MPCA is utilized to extract batch variable trajectory information with a relatively small number of uncorrelated PCs. Meanwhile, the residuals of the model are summarized into a statistic called distance to model (DmodX)[15] which indicates the model fitness. Next, at the second step, the ratio of variances is used as an index to show the state of each PC or DmodX along the batch-to-batch direction and to tell whether the batch process is in the start-up period or in the steady state period.

3. Online SSID for Batch Processes

3.1. Motivations. For batch processes, the MPCA-based SSID method works well when the MPCA model is available. However, when it is applied online to detect the end of batch process start-up, sufficient history data are not always available for MPCA modeling. It limits the online application potential of this method, especially when the target batch process has a fast start-up.

Another shortcoming of this method is that MPCA modeling assumes all batches have the same operation durations. However, batch processes frequently have uneven-length operation time, especially when batch processes are in the start-up periods. During start-up, the process properties keep changing, which often leads to significant difference of operation durations among batches. Although many different trajectory synchronization methods have been proposed, each method still has certain limitations.[16] Therefore, a method for batch process modeling and SSID without trajectory synchronization is desired.

In general, to achieve a better online SSID for batch processes, we need to develop a method which can model the process with limited (even minimum number of) history data and is able to deal with uneven lengths of operation durations. At the same time, this method should be able to compress the high-dimensional batch process data to a small number of indicators. These indicators are desired to reflect the changes in process nature related closely to the process states and are much easier to be monitored and analyzed than the raw measurements.

3.2. Detection of Correlation Structure Changes. 3.2.1. Basic Idea. In the start-up period of a batch process, the material and instrument properties may have not been stable. This may cause the variable correlation structure to change, including cross-correlations and autocorrelations (dynamics). On the contrary, the variable correlation structures should be similar among batches, when the batch process start-up has finished. Therefore, the change in variable correlations is a good indicator of unsteady state.

3.2.2. Extraction of Correlation Information. The first step in SSID for batch processes, based on correlation structures, is to extract the correlation information from the measurements in each batch. For easy understanding, first, let us take the assumption that each batch has the same length of operation

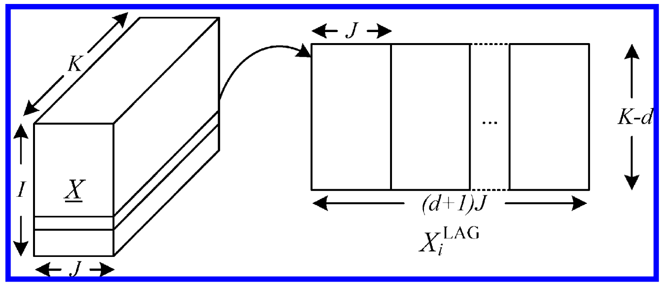


Figure 1. Transformation of batch process data matrix.

durations. In a three-dimensional batch process data matrix $\underline{X}(I \times J \times K)$, each horizontal slice $X_i(K \times J)$ $(i = 1,2,...,I)$ is a two-dimensional data matrix corresponding to the data from the ith batch, which can be further transformed into a lagged form $X_i^{\mathrm{LAG}}((K-d) \times (d+1)J)$ $(i = 1,2,...,I)$ as shown in eq 10 and Figure 1:

$$X_i^{\mathrm{LAG}} = [\mathbf{X}_i(0) \quad \mathbf{X}_i(1) \quad \ldots \quad \mathbf{X}_i(d)] = \begin{bmatrix} \mathbf{x}_i^{\mathrm{T}}(d+1) & \mathbf{x}_i^{\mathrm{T}}(d) & \ldots & \mathbf{x}_i^{\mathrm{T}}(1) \\ \mathbf{x}_i^{\mathrm{T}}(d+2) & \mathbf{x}_i^{\mathrm{T}}(d+1) & \ldots & \mathbf{x}_i^{\mathrm{T}}(2) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_i^{\mathrm{T}}(K) & \mathbf{x}_i^{\mathrm{T}}(K-1) & \ldots & \mathbf{x}_i^{\mathrm{T}}(K-d) \end{bmatrix} \tag{10}$$

$$\mathbf{x}_i^{\mathrm{T}}(k) = [x_{i,1}(k) \quad x_{i,2}(k) \quad \cdots \quad x_{i,J}(k)] \tag{11}$$

where d is the total length of lagged steps, i is the batch index, J is the total number of variables, $\mathbf{X}_i(m)$ is the measurement data matrix at m steps before, $\mathbf{x}_i^{\mathrm{T}}(k)$ is the measurement data vector at the kth sampling interval in the ith batch, and $x_{i,j}(k)$ is the measurement value of the jth variable at the kth sampling interval in the ith batch. These lagged batch data matrices are then normalized as eq 3 to eliminate the effects of units and measuring ranges.

When the batch process has uneven operation durations from batch to batch, the form of X_i^{LAG} does not change much. Without losing generality, suppose there are K_i number of sampling intervals in the ith batch. $X_i^{\mathrm{LAG}}((K_i - d) \times (d+1)J)$ can be arranged as eq 12. It is easy to understand that eq 10 is a special case of eq 12.

$$X_i^{\mathrm{LAG}} = [\mathbf{X}_i(0) \quad \mathbf{X}_i(1) \quad \ldots \quad \mathbf{X}_i(d)] = \begin{bmatrix} \mathbf{x}_i^{\mathrm{T}}(d+1) & \mathbf{x}_i^{\mathrm{T}}(d) & \ldots & \mathbf{x}_i^{\mathrm{T}}(1) \\ \mathbf{x}_i^{\mathrm{T}}(d+2) & \mathbf{x}_i^{\mathrm{T}}(d+1) & \ldots & \mathbf{x}_i^{\mathrm{T}}(2) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_i^{\mathrm{T}}(K_i) & \mathbf{x}_i^{\mathrm{T}}(K_i-1) & \ldots & \mathbf{x}_i^{\mathrm{T}}(K_i-d) \end{bmatrix} \tag{12}$$

Then, to reveal the variable correlation structure in each batch, PCA is adopted to decompose the normalized matrices. Suppose that X_i^{LAG} has been normalized,

$$X_i^{\mathrm{LAG}} = T_i P_i^{\mathrm{T}} \tag{13}$$

As introduced in section 2.1.1, the loading matrix P_i $(i = 1,2,...,I)$ models the variable correlations in the ith batch. At the same time, the eigenvalue information corresponding to this PCA model is stored in a singular-value diagonal matrix S_i, where

$$S_i = \mathrm{diag}(\lambda_1^i, \lambda_2^i, ..., \lambda_J^i) \tag{14}$$

and λ_j^i is the jth biggest eigenvalue in the PCA model corresponding to the ith batch. Since the model is built for information extraction instead of monitoring, there is no necessity to select the number of retained PCs. Small weights will be assigned to the unimportant PCs in the model comparisons, as introduced later in section 3.2.3.

The idea of the above procedure is equivalent to performing DPCA on the data of each batch. Therefore, the parameter d can be determined using the same method adopted in DPCA modeling.[9] With a proper d, both variable cross-correlation and autocorrelation information in batch i is extracted by P_i. A simplified choice is setting the value of d to be equal to 0. Thus, X_i^{LAG} is the same as X_i, and P_i only reflects the variable cross-correlations, while the dynamics information is ignored. In many situations, only using the cross-correlation information may be sufficient to achieve satisfied SSID results.

3.2.3. Comparison of Correlation Structures. After PCA decompositions, the batch model P_{i} contains variable correlation information. However, each P_i is a matrix instead of a scalar value. The SSID method based on the ratio of variances cannot be directly utilized to analyze P_i. To solve this problem, the batch PCA models are compared to each other. Each comparison result is represented with a single quantitative value. These quantitative values can be organized in sequence, which can be analyzed by the SSID method based on the ratio of variances. The details are as following.

Several types of criteria have been proposed to compare different PCA models.[17−20] In this work, a PCA similarity factor proposed by Yao and Gao[20] is adopted to evaluate the similarities between P_i of each batch, which is a summary of the angles between corresponding PC pairs in two models. The formula of the similarity between two PCA models is as below:

$$S_{\mathrm{PCA}}^{\mathrm{paired}}(P_1, S_1, P_2, S_2) = \frac{\sum_{j=1}^{J} (\lambda_j^1 \lambda_j^2) \cos^2 \theta_{jj}}{\sum_{j=1}^{J} \lambda_j^1 \lambda_j^2} \quad (15)$$

where P_i is the loading matrix of the ith PCA model; S_i is the singular-value diagonal matrix of the ith PCA model; λ_j^i is the jth largest eigenvalue in the ith PCA model, indicating the importance of each PC; and θ_{jj} is the angle between the directions of the jth PCs in two PCA models. The value of $\cos \theta_{jj}$ can be calculated as

$$\cos \theta_{jj} = P_1(:, j).{*}P_2(:, j)/(\|P_1(:, j)\|\|P_2(:, j)\|) = P_1(:, j).{*}P_2(:, j) \quad (16)$$

where $P_i(:, j)$ is the jth loading vector of the ith PCA model.

In online SSID for batch processes, we calculate the similarities between a referenced model and the batch PCA models in sequence. Practically, the PCA model of the first batch can be used as the referenced one. By doing so, a series of similarity values between the first batch PCA model and those of the following batches are computed as

$$\mathrm{SIM}_i = S_{\mathrm{PCA}}^{\mathrm{paired}}(P_1, S_1, P_i, S_i) \quad (i = 2, 3, ...) \quad (17)$$

where P_i and S_i are the loading matrix and the singular-value diagonal matrix of the batch PCA model of the ith batch. Particularly, $\mathrm{SIM}_1 = 1$.

After a batch process start to operate, the variable correlations may keep changing until the start-up period finishes. Usually, such changes are monotone, which means that the similarities between the first batch and the following batches will become smaller and smaller during start-up. Then, after the process comes into the steady state, these similarity values become similar to each other and do not vary much from batch to batch. Therefore, the sequence of SIM_i can be regarded as an indicator revealing the changes in the variable correlation structures and relating to the process state changes. Considering the situations that the correlation structures have nonmonotone changes in start-up, the SIM_i can still be used. In those situations, the values of SIM_i have significantly larger variations in the start-up than the variations at the steady state.

Please note that since the total number (J) of process variables does not change with batches, the dimensions of matrix P_i and S_i are always ($J \times J$), even if the lengths of batch durations are unequal. Therefore, the uneven duration problem will not affect the calculations of $S\mathrm{IM}_i$.

3.2.4. SSID for Correlation Structure Changes. With the calculations of the similarities between the first batch and the following batches, a vector **SIM** is available, where **SIM** = [SIM_1, SIM_2, SIM_3, ...]. During online applications, the length of the vector **SIM** is expanded whenever a new batch operation finishes.

The ordinary SSID methods can be used in trend detection of **SIM**, just like their applications on a single variable in the continuous process. Either the conventional SSID method based on a ratio of variances[5] or Cao and Rhinehart's method[4] can be used. The common features of these methods are also inherited. When applying the conventional SSID method, there is a trade-off in the determination of the window length. A wider window leads to the better insensitivity to noise, while a narrower window causes less delay. Similar situations exist in the application of Cao and Rhinehart's method. The choices of the filter parameters λ_1, λ_2, and λ_3 should balance the requirements of the detection accuracy and the efficiency. Generally, smaller values of the parameters better distinguish nonsteady states and steady states, while larger values mean a more rapid tracking. The problem of parameter selection is not going to be discussed in this work. Readers who have deep interests in this problem can consult the reference[21] for more detailed information.

The trend detection results based on **SIM** imply the batch process operation states. Only when **SIM** is at steady states, does the batch process have the possibility to be at steady state. On the contrary, if there is a trend in **SIM**, the batch process is not steady.

3.3. Detection of Trajectory Magnitude Changes. 3.3.1. Basic Idea. To conclude whether a batch process is at steady state or not, it is not enough to only check the changes of variable correlations. For some batch processes, the changes of correlation structures are not obvious in start-up period, while there are significant drifts in the magnitudes of variable trajectories. Usually, such situations are caused by the slow-response variables with large time constants. These variables may take several operation cycles to approach their set points (normal trajectories) slowly, which means the variable trajectories will keep changing from cycle to cycle until the batch process begins to operate at the steady state. Therefore, the trajectory change is another indicator of the unsteady state in batch operations.

3.3.2. Comparison of Trajectory Magnitudes. The Mahalanobis distance,[22] which can be used to compare two data sets that may have similar spatial orientation but are located far apart, can be adopted to assess the differences in the magnitudes of variable trajectories. This factor is useful when two data sets have similar PCA model structures but the values of the process variables are different.

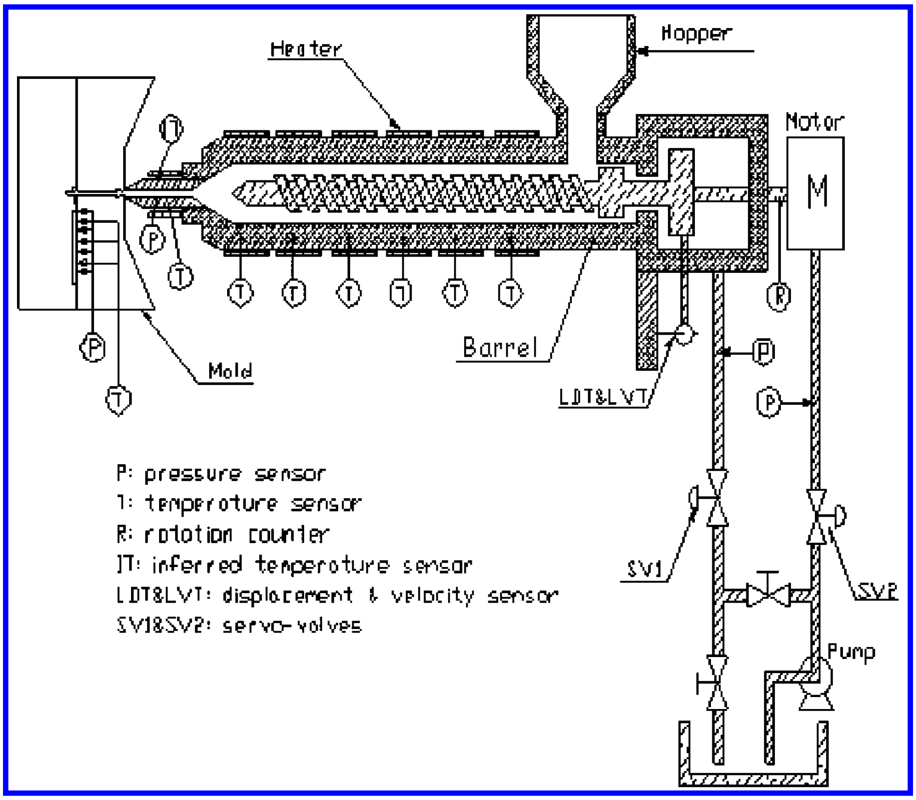


Figure 2. Simplified illustration of an injection molding machine.

The Mahalanobis distance between two random vectors **x** and **y** of the same distribution with the covariance matrix Σ is defined as

$$\Phi = \sqrt{(\mathbf{x} - \mathbf{y})^{\mathrm{T}}\Sigma^{+}(\mathbf{x} - \mathbf{y})} \tag{18}$$

where Σ^{+} is the pseudoinverse of Σ.

Similarly, the Mahalanobis distance between the variable trajectories in a referenced batch and a new batch can be represented as

$$\Phi = \sqrt{(\bar{\mathbf{x}}_{\mathrm{R}} - \bar{\mathbf{x}}_{\mathrm{N}})^{\mathrm{T}}\Sigma_{\mathrm{R}}^{+}(\bar{\mathbf{x}}_{\mathrm{R}} - \bar{\mathbf{x}}_{\mathrm{N}})} \tag{19}$$

where Σ_{R}^{+} is the pseudoinverse of the covariance matrix of the referenced batch data, $\bar{\mathbf{x}}_{\mathrm{R}}$ and $\bar{\mathbf{x}}_{\mathrm{N}}$ are the sample means of the referenced batch and the new batch, respectively, which can be calculated as

$$\bar{\mathbf{x}}_{\mathrm{R}} = \frac{\sum_{k=1}^{K_{\mathrm{R}}} \mathbf{x}_{\mathrm{R}}(k)}{K_{\mathrm{R}}} \tag{20}$$

$$\bar{\mathbf{x}}_{\mathrm{N}} = \frac{\sum_{k=1}^{K_{\mathrm{N}}} \mathbf{x}_{\mathrm{N}}(k)}{K_{\mathrm{N}}} \tag{21}$$

$\mathbf{x}_{\mathrm{R}}(k)$ and $\mathbf{x}_{\mathrm{N}}(k)$ are the column vectors of the measurements data at the kth sampling interval in the referenced batch and the new batch, respectively, and K_{R} and K_{N} are the total numbers of sampling intervals in the two batches, respectively.

In online SSID for batch processes, the first batch can be regarded as the reference one. Then, the Mahalanobis distances between the following batches and the first batch are computed as follows:

$$\Phi_i = \sqrt{(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_i)^{\mathrm{T}}\Sigma_1^{+}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_i)} \tag{22}$$

when $i = 1$, $\Phi_i = 0$.

For the batch processes with significant trajectory magnitude changes in start-up, the values of Φ_i have large variations from batch to batch until coming into the steady state. Particularly, if such changes are monotone, the distances between the first batch and the following batches will enlarge along batch direction, leading to larger and larger Φ_i in start-up period. Such large variations or monotone trends will vanish only after the process comes into the steady state, which means the sequence of Φ_i is suitable to be used as a complementarity of **SIM** in the online identification of batch process steady state.

Since the Mahalanobis distance is calculated on the basis of the average values of the variable trajectories in each batch, the uneven length problem in operation durations is solved easily. At the same time, the measurement noises are filtered out, which makes the SSID based on the distance more robust.

3.3.3. SSID for Trajectory Magnitude Changes. Similar to the SSID based on **SIM**, the sequence of Φ_i can be arranged as a vector $\mathbf{\Phi} = [\Phi_1, \Phi_2, \Phi_3, ...]$ whose length is expanded whenever a new operation cycle finishes. Either the conventional

Table 1. Description of the Process Variables

no.	variable description	unit
1	nozzle temperature	°C
2	stroke	mm
3	screw velocity	mm/sec
4	injection cylinder pressure	bar
5	plastication pressure	bar
6	SV1 opening	%
7	SV2 opening	%

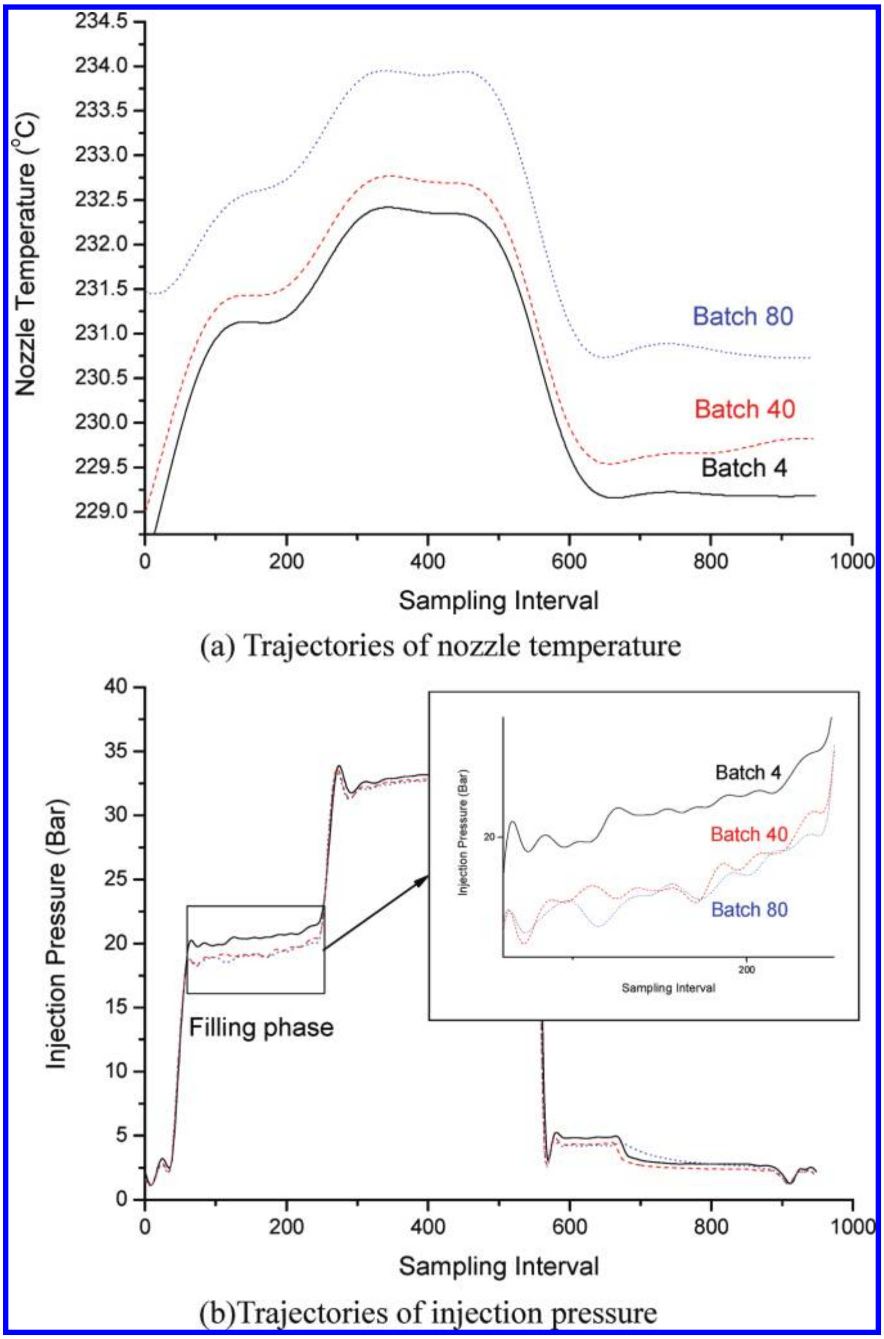


Figure 3. Variable trajectories in injection molding.

SSID method based on a ratio of variances or Cao and Rhinehart's method can be performed to detect the trend in **Φ** which indicates the process states.

3.4. Procedure of Online SSID for Batch Processes. The overall procedure of batch process online SSID is summarized as following.

(1) Select parameters for PCA modeling and SSID, including the lagged steps d and the length of moving window or the filter parameters.

(2) Whenever the data matrix X_i of a new batch is available, calculate Φ_i as eq 22.

(3) Transform the data matrix X_i to the lagged batch data matrix X_i^{LAG}.

(4) Build batch PCA model based on X_i^{LAG}. Calculate P_i and S_i.

(5) Calculate SIM_i as eq 17.

(6) Perform SSID on **Φ**.

(7) If a trend is detected in **Φ**, go to step 2. Otherwise, go to the next step.

(8) Perform SSID on **SIM**.

(9) If a trend is detected in **SIM**, go to step 2. Otherwise, the process is at steady state.

3.5. Method Extensions. 3.5.1. Dealing with non-Gaussian Information. An implied assumption of performing PCA is that all variables are normally distributed. When there is significant non-Gaussian information contained in process data, PCA may not extract it well. In those situations, another data analysis method called independent component analysis (ICA)[23] can be utilized in online SSID for batch processes instead of PCA.

Different from PCA, ICA regards the process measurements as the mixtures of several independent components. It searches for components that are statistically independent

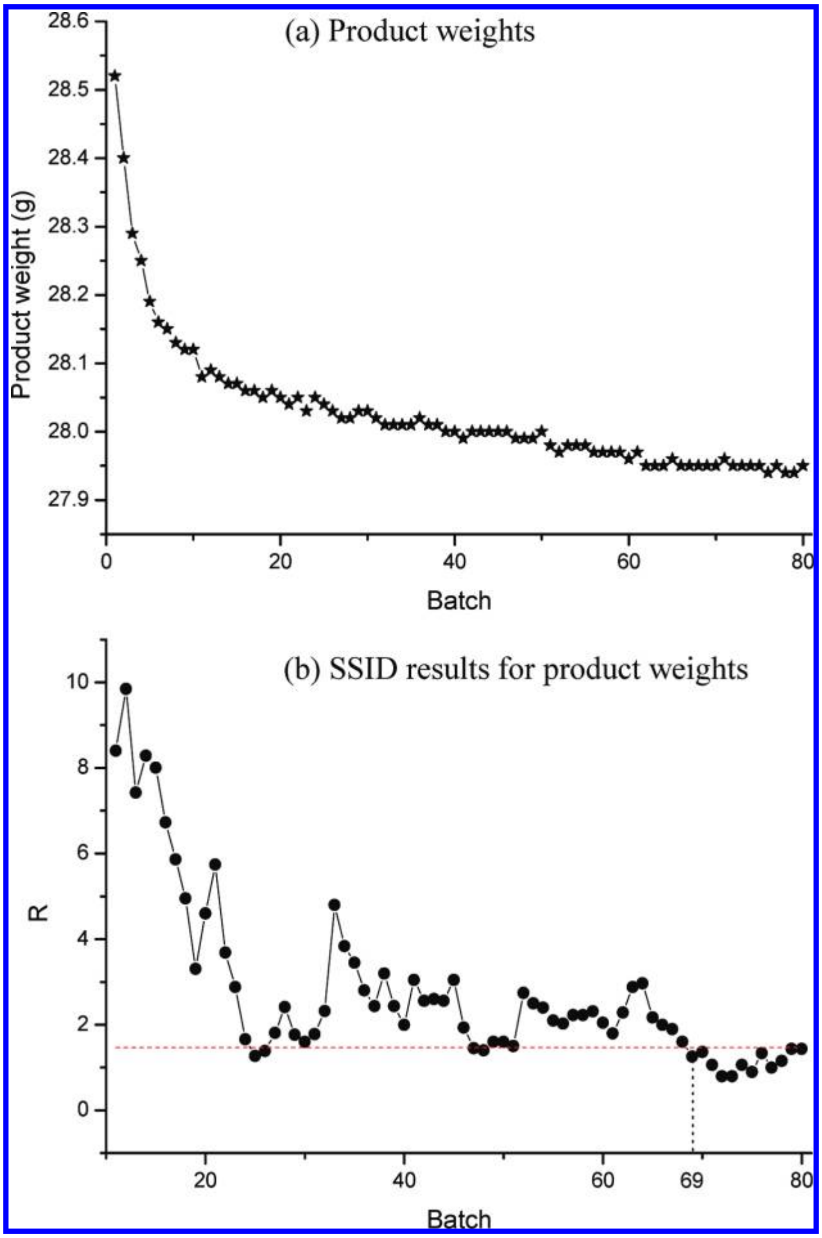


Figure 4. Trend detection for product weights.

to each other and non-Gaussian. The formula of ICA is like the following:

$$X = AS + E \tag{23}$$

where X is the data matrix, A is the mixing matrix, S is the independent component matrix, and E is the residual matrix.

Recently, Ge and Song[24] proposed an ICA similarity factor which can be calculated from the main angles between two ICA subspaces as following:

$$\mathrm{SIM}^{\mathrm{ICA}} = \frac{1}{r}\sum_{i=1}^{r}\cos^2\varphi_i \tag{24}$$

$$\cos\varphi_i = \max_{\mathbf{a}\in\mathbf{A}_{i-1}}\max_{\mathbf{b}\in\mathbf{B}_{i-1}}\left(\frac{\mathbf{a}^{\mathrm{T}}\mathbf{b}}{\|\mathbf{a}\|\|\mathbf{b}\|}\right) = \frac{\mathbf{a}_i^{\mathrm{T}}\mathbf{b}_i}{\|\mathbf{a}_i\|\|\mathbf{b}_i\|} \tag{25}$$

where φ_i is the ith main angle between two ICA subspaces, $\mathbf{a}_i$ and $\mathbf{b}_i$ are the vectors corresponding to φ_i, $\mathbf{A}_0$ and $\mathbf{B}_0$ are two subspaces corresponding to two different ICA models, and $\mathbf{A}_i$ and $\mathbf{B}_i$ are the new subspaces formed by removing $\mathbf{a}_i$ and $\mathbf{b}_i$ from $\mathbf{A}_{i\text{-}1}$ and $\mathbf{B}_{i\text{-}1}$. For more details about ICA and ICA similarity factor, please refer to ref 24.

For online SSID for batch processes, when the process variables follow significant non-Gaussian distribution, the procedure in subsection 3.4 can be revised by integrating ICA into the trend detection. After transforming the data matrix X_i to the lagged batch data matrix X_i^{LAG} as described in step 2, an ICA model is built as eq 23 in step 3. Then, a $\mathrm{SIM}_i^{\mathrm{ICA}}$ statistic, the similarity between the ICA models of the ith batch and the first batch, can be calculated and arranged into a vector **SIMICA**. Then, go to the steps 5−6 listed in subsection 3.4. After step 6, SSID is performed to detect the trend in **SIMICA**. If there is a trend detected in **SIMICA**, go to step 1. Otherwise, the process is identified to be at steady state.

3.5.2. Dealing with Multiphase Characteristic. Multiple operation phases are important characteristics of many batch processes. In multiphase batch processes, different phases have different variable correlations. If such characteristic is

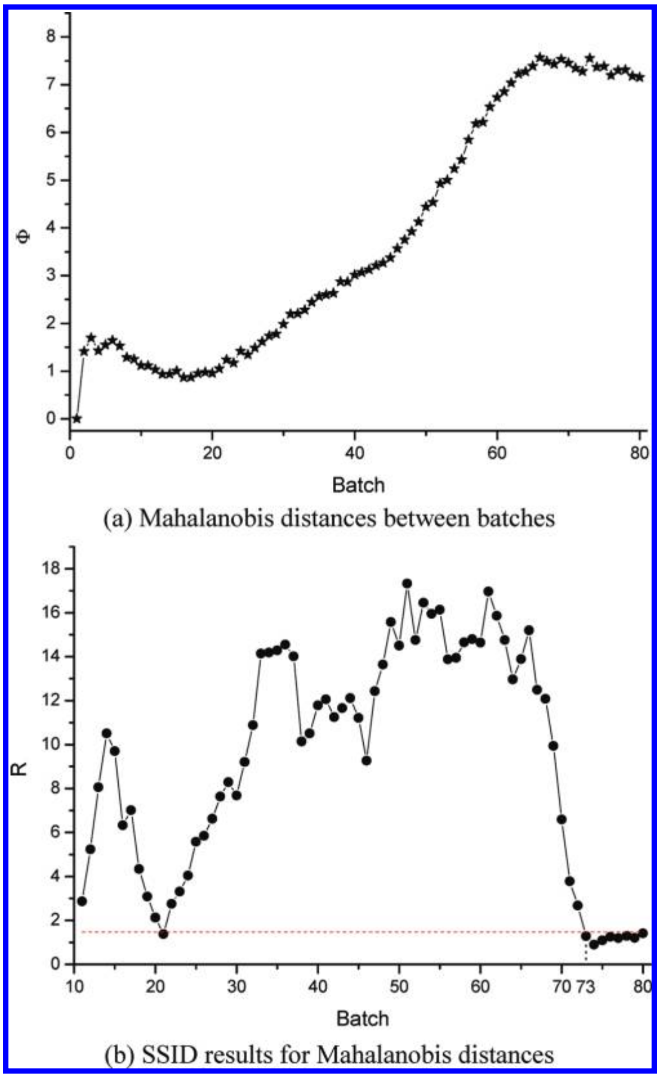


Figure 5. Trend detection for Mahalanobis distances in injection molding process.

taken into consideration during online SSID for batch processes, the process model can be more accurate and hopefully the detection results will be better.

The phase division methods based on variable correlation structure changes can be utilized to divide a batch process into different phases.[20,25,26] Then the procedure in section 3.4 can be carried out in each phase separately. Only when all phases are at steady state, the batch process is identified to be at steady state.

4. Application Results

4.1. SSID for Start-up in Injection Molding Process. 4.1.1. Injection Molding Process. Injection molding is an important polymer processing technology, which transforms polymer materials into various shapes and types of products. Figure 2 shows a simplified diagram of a typical reciprocating-screw injection molding machine with measurement instrumentations.

As a typical batch process, injection molding operates in cycles. In each cycle, operation starts with mold close. Then, during filling, the screw moves forward and pushes polymer melt into the mold cavity. After the mold is fully filled, additional melt is packed into the mold at a high pressure, to compensate for the material shrinkage caused by cooling and

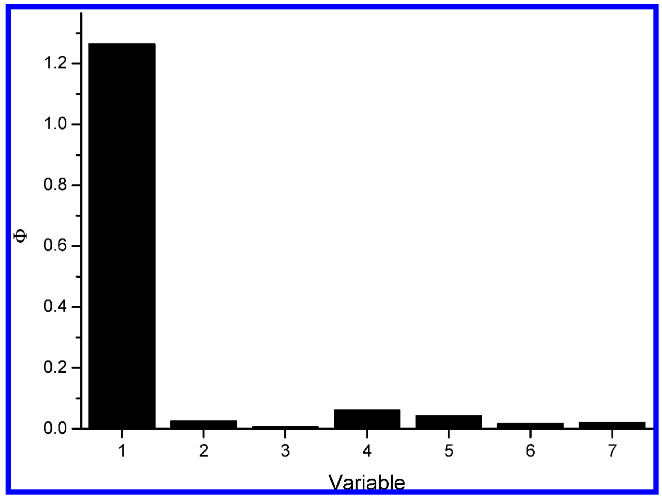


Figure 6. Mahalanobis distances between each pair of variables in an unsteady batch and a steady batch.

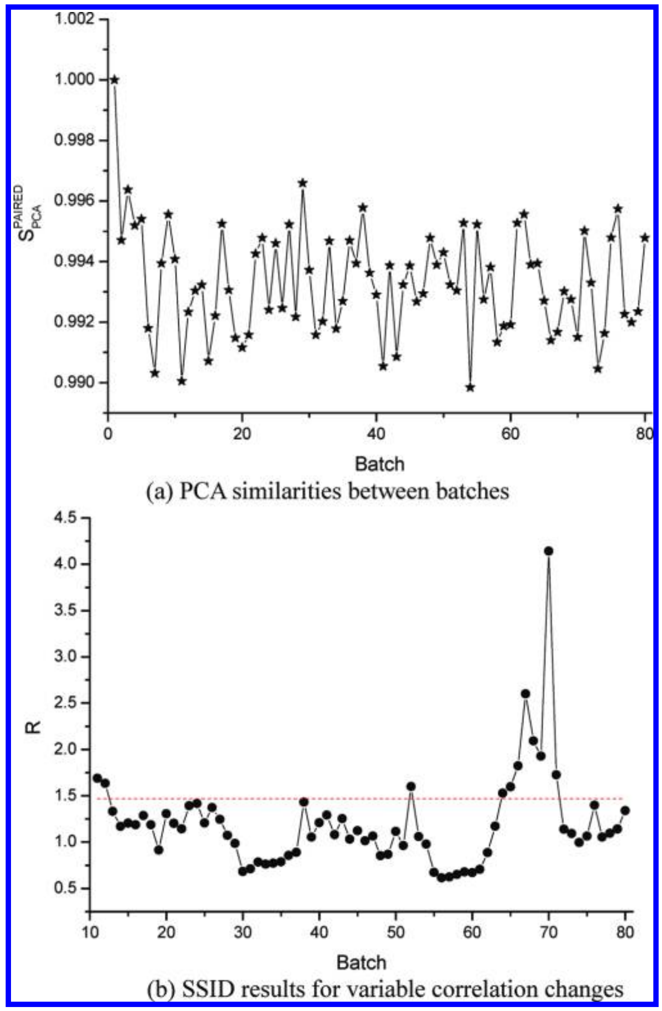
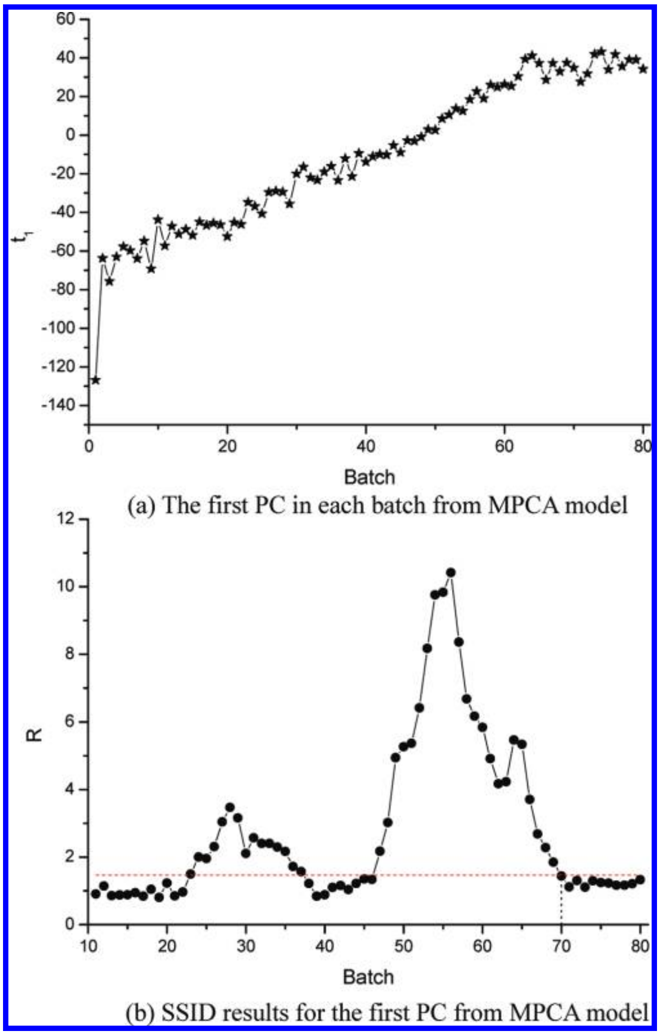


Figure 7. Trend detection for changes of variable correlations in injection molding process.

solidification of the material. Such high pressure is kept for a while until the gate is frozen and the material in the mold cannot flow back into the injection unit any longer. Then, the material in the mold is cooled and solidified, so that it can be rigid enough to be ejected from the mold without damage. Meanwhile, in the early period of cooling, the screw rotates to plasticate the polymer in the barrel and to convey the melt to the front of barrel, which is a preparation for the next cycle.



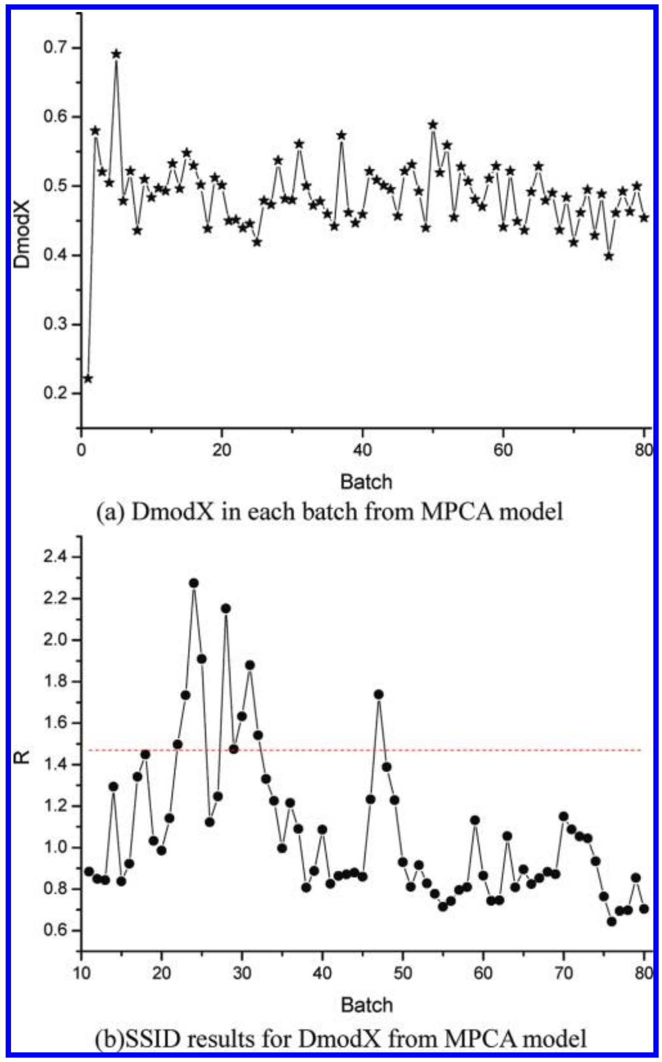
(a) The first PC in each batch from MPCA model

(b) SSID results for the first PC from MPCA model

Figure 8. Trend detection for the first PC from MPCA model.

4.1.2. Injection Molding Start-up. On the basis of the process knowledge, we know that injection molding is a proper batch process for the application and verification of the proposed online SSID method. In injection molding start-up, the barrel is first heated. After the barrel temperature is close to the requirement of normal operation, the process starts to operate. However, the process still needs quite a few batches to accomplish start-up. In the beginning, the materials in barrel are not well mixed, and the melt temperature also takes time to rise. Since temperature has a large time constant, the drifts in temperature trajectories are the dominant process features in injection molding start-up. Additionally, the material properties are not stable in this period. The material viscosity and density decrease with rising temperature, which further affects the magnitudes of pressure variables. For example, due to the changes in viscosity, it requires lower injection pressure to maintain the same injection speed in filling. Similar situations also occur in other operation phases. This kind of drifts in temperature and pressure variables will last until the start-up finishes and the process enters the steady state. Meanwhile, since the material density decreases, the final product weights also decrease from batch to batch in start-up.

The material used in the experiment is high-density polyethylene (HDPE). As shown in Table 1, seven process variables are selected for modeling, while product weights are measured as the quality variable. The operating conditions are set as following: injection velocity is 20 mm/sec; packing-holding time is fixed to be 3 s. In total, 80 batch runs are conducted, with



(a) DmodX in each batch from MPCA model

(b)SSID results for DmodX from MPCA model

Figure 9. Trend detection for the residuals from MPCA model.

uneven cycle durations. The filtered trajectories of nozzle temperature (melt temperature in the nozzle) and injection cylinder pressure are plotted in Figure 3, which confirm the analysis in the previous paragraph and show obvious start-up features.

4.1.3. SSID for Injection Molding Process. As discussed in the previous subsection, the trend of product weights reflects whether the injection molding process is at steady state or in start-up period. Therefore, all weights are plotted in Figure 4a, and the SSID method is performed to analyze the trend of weights from batch to batch. In this work, the conventional SSID method based on the ratio of variances is utilized for illustration, while the window length is selected to be 10. The SSID results with the 95% control limit are shown in Figure 4b. According to the weights analysis, the SSID results show that the process becomes steady after batch 69.

Although the product weights can reflect the state of the injection molding process, online quality measurements are usually unavailable. Therefore, it is preferable to use online measurements of process variables for SSID. The proposed online SSID method for batch processes is conducted following the steps described in subsection 3.4.

The Mahalanobis distances between the first batch and the following batches are plotted in Figure 5a, while Figure 5b shows the SSID results based on the distance information. The window length is selected as 10 for SSID. With Mahalanobis

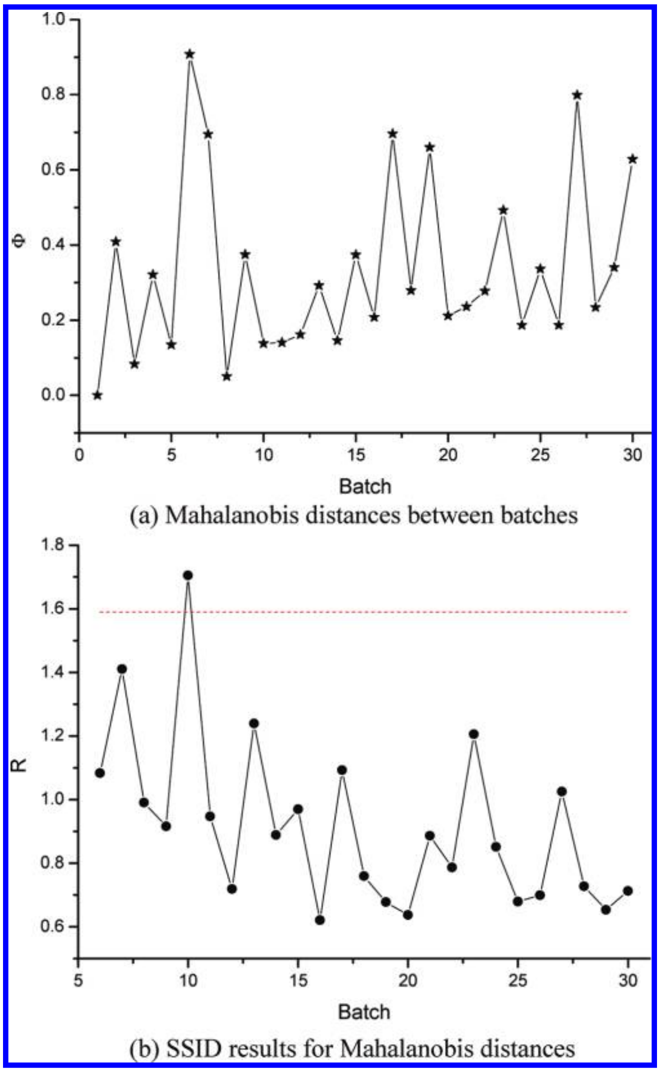


Figure 10. Trend detection for Mahalanobis distances in the simulated process.

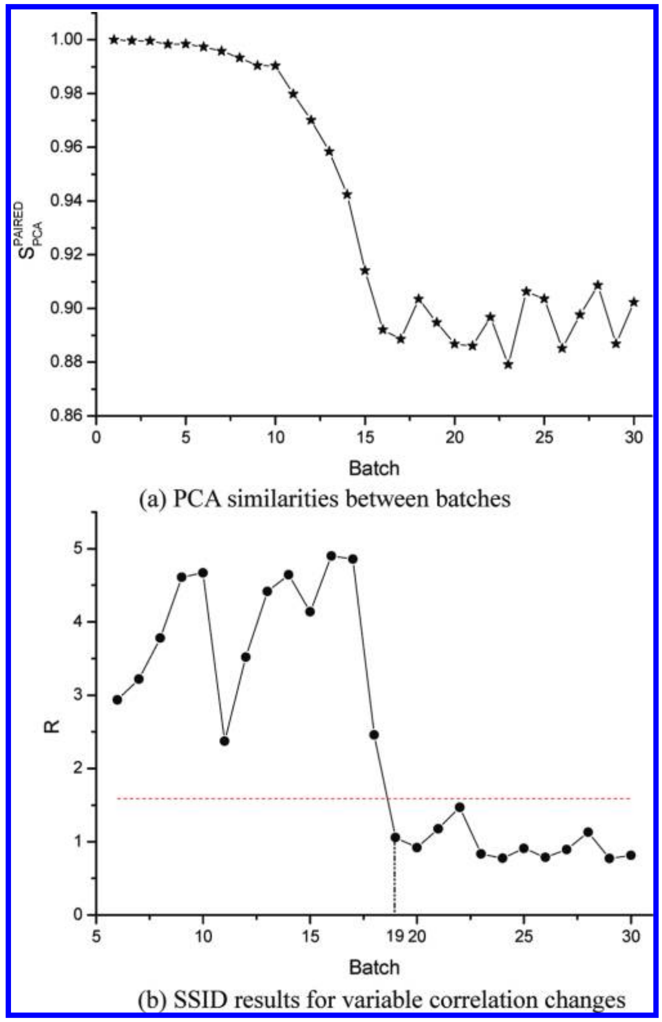


Figure 11. Trend detection for changes of variable correlations in the simulated process.

distances, the start-up period is detected to finish in batch 73. It is a little bit slower than the detection based on weights. However, all measurement data used in this detection can be easily obtained online. Further analysis can be conducted by looking at the Mahalanobis distance between each pair of variables measured in two different batches. Figure 6 reveals the differences between a batch in start-up (batch 3) and a batch at steady state (batch 70). The variables are numbered according to the numbers in Table 1. It is quite obvious that the values of the temperature variable change most significantly during start-up period, while the drifts in pressure variables are relatively small. The least affected variable is injection velocity, because it is closed-loop controlled. Such analysis results are consistent with the prior process knowledge. More knowledge about process characteristics can be dug out, if the detection and the analysis are performed after phase division, as discussed in subsection 3.5.2.

In the detection of the trend in variable correlation structures, the lagged batch data matrices X_i^{LAG} is organized with parameter $d = 0$, which means only the variable cross-correlations are taken into consideration in this application. The window length for SSID is selected as 10. Figure 7a shows the values of the PCA similarities between the first batch and the following batches, where no significant drift is found. Figure 7b confirms this judgment; despite that there are some spikes in the plot. This means that, during injection molding start-up, the batch-to-batch differences are mainly caused by the drifts in the temperature variables, while there is no significant change in the variable cross-correlations.

For comparison, the MPCA-based SSID method is also applied to injection molding data analysis. The window length for the SSID is selected to be the same as that used in the proposed method. Before building a MPCA model, the data of each batch are cut to the same length. It is observed that the first PC contains most trend information. Therefore, the trajectory of the first PC from batch to batch is plotted in Figure 8a and the SSID results are shown in Figure 8b. It can be found in this application that the steady state is detected at batch 70, which is a slightly more efficient detection than the results of the proposed method. However, there are more missed alarms. The trajectory of DmodX and the corresponding SSID results are shown in Figure 9. No significant trend is detected. Since the DmodX measures the model fitness, it indicates no significant change in variable correlations, which is consistent with the conclusion drawn from the proposed method.

4.2. SSID for a Simulated Batch Process. 4.2.1. Process Description. The second illustration example is a simulated batch process with two features. First, in start-up, this process has drifts in variable correlation structure, while the variable trajectory magnitudes do not change a lot. Therefore, it is proper to verify the use of the detection efficiency of the proposed method on the trends of variable correlation changes. Second, this process starts up quickly in 15 batches. In such a situation, there is not sufficient history data for MPCA modeling, which means the MPCA-based batch process SSID method is not suitable to be applied online to detect the states of such processes. The online detection ability of the proposed method can be tested.

There are three process variables in this simulation. Before the 16th batch, the process is in start-up with a model as follows:

$$\begin{aligned} x_1(i,k) &= 0.1 + 0.9*x_1(i,k-1) + \varepsilon_1 \\ x_2(i,k) &= 0.7 + 0.3*x_2(i,k-1) + \varepsilon_2 \\ x_3(i,k) &= (1.5 - \alpha*i)*x_1(i,k) + \\ &\quad (-0.5 + \alpha*i)*x_2(i,k) + \varepsilon_3 \ \alpha = 0.1 \end{aligned} \tag{26}$$

where α describes the drifts in variable cross-correlations, and ε_1, ε_2, ε_3 are white noises. From the 16th batch, the process comes into the steady-state operation which is formulated as

$$\begin{aligned} x_1(i,k) &= 0.1 + 0.9*x_1(i,k-1) + \varepsilon_1 \\ x_2(i,k) &= 0.7 + 0.3*x_2(i,k-1) + \varepsilon_2 \\ x_3(i,k) &= -0.1*x_1(i,k) + 1.1*x_2(i,k) + \varepsilon_3 \end{aligned} \tag{27}$$

The Mahalanobis distances between the first batch and the following ones are plotted in Figure 10a. The SSID with a window length of 5 tells us that there is no trend in variable trajectory magnitudes, as shown in Figure 10b. On the other hand, in Figure 11a, the PCA similarity values shows the trend in correlation structures clearly. By performing SSID, Figure 11b identifies the steady state from the 19th batch with a delay of three batches. This is a reasonable and efficient detection. In this detection, the parameter d is selected as 0 in the arrangement of the lagged batch data matrices X_i^{LAG}. The window length is also chosen to be 5 in SSID.

5. Conclusions

Online SSID for batch processes is an important task to identify the end of start-up period which is the beginning of manufacturing the reliable products. It also ensures the satisfactory batch process control. In this work, an SSID method is developed to better online identify the operation states of batch processes. This method extracts both correlation structure information and variable trajectory magnitude information, using the paired PCA similarity factor and the Mahalanobis distance, respectively. Both the two types of information are utilized in SSID, which is a full-scale monitoring of process states. Another advantage of this method is that it models each batch individually. Even though there is not sufficient history data, this method can still be conducted for online SSID. The experiment and simulation examples illustrate the effectiveness of the proposed method.

Acknowledgment

This work was supported in part by Hong Kong Research Grant Council under project number: CERG 613107, and China National 973 program under project number: 2009CB320603.

Literature Cited

(1) Lu, N.; Yao, Y.; Gao, F.; Wang, F. Two-dimensional dynamic PCA for batch process monitoring. *AIChE J.* **2005**, *51*, 3300–3304.

(2) Bonvin, D.; Srinivasan, B.; Hunkeler, D. Control and optimization of batch processes: Improvement of process operation in the production of specialty chemicals. *IEEE Control Syst. Mag.* **2006**, *26*, 34–45.

(3) Aguado, D.; Ferrer, A.; Seco, A.; Ferrer, J. Using unfold-PCA for batch-to-batch start-up process understanding and steady-state identification in a sequencing batch reactor. *J. Chemom.* **2008**, *22*, 81–90.

(4) Cao, S.; Rhinehart, R. An efficient method for on-line identification of steady state. *J. Process Control* **1995**, *5*, 363–374.

(5) Crow, E.; Davis, F., *Statistics Manual*; Dover Publications: Mineola, NY, 1960.

(6) Brown, P.; Rhinehart, R. Demonstration of a method for automated steady-state identification in multivariable systems. *Hydrocarbon Process.* **2000**, *79*, 79–83.

(7) Ruiz, G.; Castellano, M.; González, W.; Roca, E.; Lema, J. Algorithm for steady state detection of multivariate process: Application to anaerobic wastewater digestion process. Proceedings of the 2nd International IWA Conference on Automation in Water Quality Monitoring (AutMoNet), Viena, Austria, 2004.

(8) Jackson, J. *A User's Guide to Principal Components*; Wiley: New York, 1991.

(9) Ku, W.; Storer, R.; Georgakis, C. Disturbance detection and isolation by dynamic principal component analysis. *Chemom. Intell. Lab. Syst.* **1995**, *30*, 179–196.

(10) Nomikos, P.; MacGregor, J. Monitoring batch processes using multiway principal component analysis. *AIChE J.* **1994**, *40*, 1361–1375.

(11) Nomikos, P.; MacGregor, J. Multivariate SPC charts for monitoring batch processes. *Technometrics* **1995**, 41–59.

(12) Von Neumann, J.; Kent, R.; Bellinson, H.; Hart, B. The mean square successive difference. *Ann. Math. Stat.* **1941**, 153–162.

(13) Von Neumann, J. Distribution of the ratio of the mean square successive difference to the variance. *Ann. Math. Stat.* **1941**, 367–395.

(14) Cao, S.; Rhinehart, R. Critical values for a steady state identifier. *J. Process Control* **1997**, *7*, 149–152.

(15) Eriksson, L.; Academy, U. *Multi- and Megavariate Data Analysis*; Umetrics Academy: Umea, Sweden, 2001.

(16) Lu, N.; Gao, F.; Yang, Y.; Wang, F. PCA-based modeling and on-line monitoring strategy for uneven-length batch processes. *Ind. Eng. Chem. Res.* **2004**, *43*, 3343–3352.

(17) Krzanowski, W. Between-groups comparison of principal components. *J. Am. Stat. Assoc.* **1979**, 703–707.

(18) Singhal, A.; Seborg, D. Pattern matching in historical batch data using PCA. *IEEE Control Syst. Mag.* **2002**, *22*, 53–63.

(19) Zhao, S.; Zhang, J.; Xu, Y. Monitoring of processes with multiple operating modes through multiple principle component analysis models. *Ind. Eng. Chem. Res.* **2004**, *43*, 7025–7035.

(20) Yao, Y.; Gao, F. Phase and transition based batch process modeling and online monitoring. *J. Process Control* **2009**, *19*, 816–826.

(21) Bhat, S.; Saraf, D. Steady-state identification, gross error detection, and data reconciliation for industrial process units. *Ind. Eng. Chem. Res.* **2004**, *43*, 4323–4336.

(22) Mardia, K.; Kent, J.; Bibby, J., *Multivariate Analysis*; Academic Press: New York, 1979.

(23) Herault, J.; Jutten, C. Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture. *Signal Process.* **1991**, *24*, 1–10.

(24) Ge, Z.; Song, Z. Process Monitoring Based on Independent Component Analysis−Principal Component Analysis (ICA−PCA) and Similarity Factors. *Ind. Eng. Chem. Res.* **2007**, *46*, 2054–2063.

(25) Lu, N.; Gao, F.; Wang, F. Sub-PCA modeling and on-line monitoring strategy for batch processes. *AIChE J.* **2004**, *50*, 255–259.

(26) Lu, N.; Yang, Y.; Wang, F.; Gao, F. A stage-based monitoring method for batch processes with limited reference data. Proceedings of the 7th International Symposium on Dynamics and Control of Process Systems (Dycops-7); Boston, MA, 2004.

Received for review July 9, 2009
Revised manuscript received September 7, 2009
Accepted October 2, 2009

IE901107H