Stage-Based Process Analysis and Quality Prediction for Batch Processes

Ningyun Lu and Furong Gao*

Department of Chemical Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong SAR

A process analysis and quality prediction scheme is proposed based on a stage-based PLS modeling for batch processes. Without any requirement of prior process knowledge, the scheme first divides a batch process into stages of different process characteristics. Subsequently, a strategy is developed to identify stages that have critical influences on concerned qualities, defined as critical-to-quality stages. Within these critical-to-quality stages, an algorithm is then further developed to identify the variables that have significant contributions to the quality variations. Finally, based on the identified nature of quality and stage relationships, a set of algorithms is developed for online quality prediction. The applications of the proposed scheme to injection molding show that the proposed analysis and quality prediction are not only effective but are also able to enhance process understanding and identify specific variables and periods for quality improvement.

1. Introduction

Batch processes are widely used today for producing higher value-added products to meet rapidly changing market. Semiconductor processing, injection molding, fermentation, and most bio-processes are all batch processes in nature. Competition and demand for consistent and high-quality product have spurred the development of quality-related researches for batch processes.

Online quality control of batch processes is, however, difficult due to complicated nonlinear process behaviors and the lack of online quality measurements. Significant efforts have been made for the development of methods for quality prediction, among which, multivariate statistical modeling is widely used as it is derived directly from historical data with little prior process knowledge, and it has superior ability in handling high-dimensional and correlated process data.

Multi-way principal component analysis (MPCA) and multi-way partial least square (MPLS) models, pioneered by Nomikos and MacGregor,[1,2] are the most popular multivariate statistical methods for batch processes with good applications in the industry. MPLS models use process variables over the entire batch course as the inputs; they require future process measurements in the current batch for online monitoring and quality prediction. The monitoring and quality prediction accuracy is, therefore, dependent on the prediction accuracy of the future data.[3] Furthermore, many multistage batch processes have time-specific effects of process variables on the final qualities. MPLS method is inefficient in revealing time-specific relationships for such processes.

Extensions of PLS/MPLS modeling for interpretation and prediction improvement have been reported, for example, by selecting the key process variables[4−6] or focusing on the critical-to-quality time periods.[6,8] Wold and co-workers used the orthogonal signal correction (OSC) algorithm to remove process variables that have weak relations to the product quality from PLS inputs.[4,5] This idea was also echoed by Chu et al.[6] for their MPLS modeling, where bootstrapping technology was adopted for input variable selection. Yu and MacGregor,[7] however, argued against the removal of "unimportant" process variables in OSC+PLS method. A strategy of combining the PLS modeling with canonical correlation analysis (CCA)[7] was proposed to explore effectively the relationship between process variables and the product qualities. Unfortunately, it can be only used as a post-analysis tool.

On the issue of focusing on the time-specific effects of process variables on the final product quality, the pathway multi-block PLS algorithm has been developed by Duchesne and MacGregor[8] by introducing intermediate quality measurements to isolate the local effects of process variables to the final product quality. For most industrial processes, online measurements of intermediate quality are rarely available. They did not provide any means of dividing a batch process into blocks required in the method. The bootstrapping improved MPLS of Chu et al.[6] can isolate the local effects of process variables on the final quality based on so-called VIP (variable importance in the projection of MPLS). The computation of bootstrapping-based variable selection is however very heavy, and this method is limited to processes with variables having significant time-specific effects.

Considering that multiplicity of stage is an inherent nature of many batch processes. Each stage has its own underlying characteristics, and a batch process can exhibit significantly different behaviors over different stages. It is therefore natural to develop stage-based statistical modeling methods to reflect the inherent stage nature to improve the performances of process monitoring and quality control. A stage-based sub-PCA modeling method[9] has been developed by the authors, and it has been shown that the stage PCA modeling can overcome many difficulties of MPCA-based monitoring for batch processes.

This paper extends the stage-based modeling to online quality prediction and analysis. This is accomplished

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

10.1021/ie048852l CCC: $30.25
Published on Web 04/02/2005

in the following. For a batch process with online measurements of process variables and offline measurements of the end product qualities, a stage-based PLS modeling method is proposed. The PLS modeling can automatically divide the batch process into stages without the requirement of prior process knowledge. Variables of different stages may have different effects on the final product qualities. A stage-based process analysis strategy is then developed. The concept of critical-to-quality stage is introduced; the method for identifying key process variables with significant contributions to quality variations for each critical-to-quality stages is also developed. Finally, a set of algorithms is proposed for online quality prediction. The end product qualities are divided into two types: qualities determined by only one specific stage and qualities determined by more than one stages. For the first type, an online quality prediction model will be developed in the corresponding critical-to-quality stage; for the second type, stacked modeling[10,11] is used to combine the effects of different stages for quality prediction.

The proposed stage-based process analysis and quality prediction can have the following advantages:

(i) It allows process analysis to be conducted on each stage. Stage here is classified according to the change of process characteristics.

(ii) It allows someone to explore the relationship of qualities over different stages, to determine critical-to-quality stages, and to identify key variables in each stage.

(iii) It allows earlier quality prediction in critical-to-quality stage; it avoids the prediction of future process data required for MPLS-based methods.

(iv) It has a simpler model structure, as a conventional PLS model is directly used in each stage, and it does not require unfolding of three-way batch data during online prediction.

The proposed analysis and quality prediction scheme is verified on a typical batch process, injection molding. A prediction comparison is also conducted for the proposed method, the classical MPLS model,[2] and the MPLS model using bootstrapping-based generalized variable selection[6] (abbreviated as B-MPLS model in this paper). The remainder of the paper is organized as follows. The proposed stage-based PLS modeling method is introduced in section 2, followed by the stage-based process analysis in section 3 and the stage-based quality prediction in Section 4. Illustrative experimental results on injection molding are shown in Section 5. Finally, conclusions are drawn in Section 6.

2. Stage-Based PLS Modeling

The proposed stage-based PLS modeling is an extension of stage-based sub-PCA modeling method proposed previously by the authors. The sub-PCA modeling was based on the recognition that (1) a batch process may be divided into several stages reflected by its changing process correlation nature; and (2) despite that the process may be time varying, the correlation of its variables will be largely similar within the same stage. Changes in the correlation may be used to indicate changes of process stages. In the stage-based sub-PCA modeling method, stage division is made according to changes in the correlation of the process variables. In this work, stage division will be made according to changes in the correlation between process and quality variables. This will allow us to determine stage-specific effects of process variables on the final product qualities. Similar to sub-PCA modeling, a clustering algorithm will be adopted for stage division.

Consider an even-length batch process with J_x process variables measured over time intervals k ($k = 1, 2, ..., K$) and J_y quality variables at the end of each batch; a data matrix of dimensions $J_x \times K$ for process variables and a vector of dimension $1 \times J_y$ for quality variables are generated from each batch. A set of I number of history batch data hence result in a three-way process data matrix, $\underline{X}(I \times J_x \times K)$, and a two-way quality data matrix, $Y(I \times J_y)$. For industrial processes with varying batch length, several methods[12−14] are available to deal with uneven-length batch data, among which the indicator variable method can be the first choice if there exists a prior knowledge to determine a proper indicator variable for the process.[14] $\underline{X}$ is normalized as in MPCA/MPLS modeling, that is, centered and scaled across the batches, to remove process nonlinearity along the trajectories and to focus on batch-to-batch variation. Two-way quality data matrix Y should also be normalized to have zero means and unit variances.

Similar as in the stage-based sub-PCA modeling, $\underline{X}$ is first cut along the time axis, resulting in K time−slice matrices, $\tilde{X}_k(I \times J_x)(k = 1, 2, ..., K)$. $\tilde{X}_k$ gives batch-to-batch variations of process variables at time interval k. PLS algorithm is applied to $\{\tilde{X}_k, Y\}$ at each time interval, generating K number of time−slice regression models:

$$\tilde{X}_k = \tilde{T}_k(\tilde{P}_k)^T + E$$

$$Y = \tilde{U}_k(\tilde{Q}_k)^T + F \tag{1}$$

The above time−slice PLS model can be written in a compact form as:

$$Y = \tilde{X}_k \cdot \tilde{\Theta}_k + F^*$$

$$\hat{Y}_k = \tilde{X}_k \cdot \tilde{\Theta}_k \tag{2}$$

According to the aforementioned recognition, process data should have similar statistical features when the underlying process correlation is the same. The regression parameter matrix, $\tilde{\Theta}_k(J_x \times J_y)$, containing the correlation information between process and quality variables at time interval k, should remain similar in the same stage and show significant difference over different stages. Changes in process correlation can be reflected by the changes in the regression parameter matrices $\tilde{\Theta}_k(k = 1, ..., K)$. A *k-means* clustering algorithm, as detailed in Appendix, is adopted for stage division. Process stages can be determined by the clustering results associated with process operation time.

Stage division results based on clustering of time−slice PLS regression parameter matrices could be different from that of stage-based sub-PCA modeling, where clustering is conducted on the time−slice PCA loading matrices. It is likely that a batch process may be characterized by several quality variables, and different quality variables may have different correlation to different variables over different periods (or stages). The proposed stage division method requires no prior process knowledge; it divides automatically a batch process into appropriate number of stages. This allows analysis of different local effects of process variables on

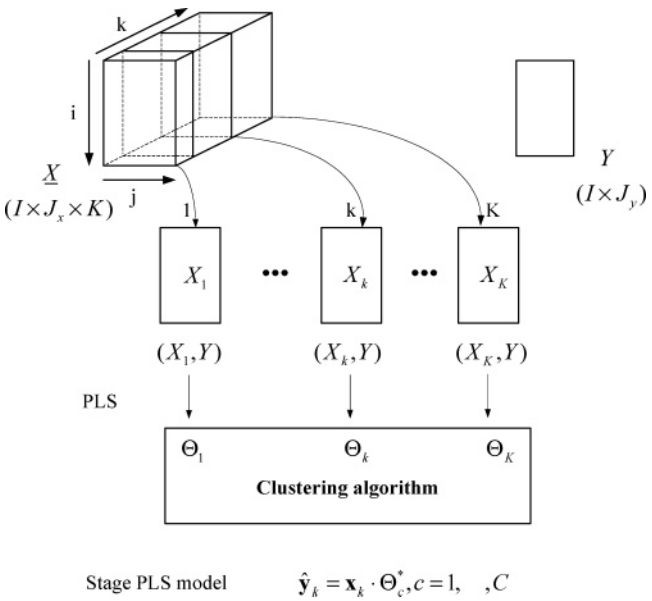

Figure 1. Illustration of stage-based PLS modeling.

the product qualities. The stage division also allows the conventional PLS model to be applied directly to batch processes.

Suppose that K time−slice regression parameter matrices are clustered into C groups, indicating C number of stages of different correlation structure between process and quality variables, a representative regression parameter matrix for stage c can be computed by

$$\Theta_c^* = \frac{1}{n_{\text{stage\_}c}} \sum_k \tilde{\Theta}_k (c = 1, 2, ..., C; k = 1,2, ..., n_{stage\_c}) \tag{3}$$

where $n_{\text{stage\_}c}$ is the data length in stage c. Stage PLS model for quality prediction can be formulated as

$$\hat{\mathbf{y}}_k = \mathbf{x}_k \cdot \Theta_c^*, \quad c = 1, 2, ..., C, \quad k = 1, ..., n_{\text{stage\_}c} \tag{4}$$

where $\mathbf{x}_k$ is composed of process measurements at time k, belonging to stage c. An illustration of the proposed stage-based PLS modeling scheme is given in Figure 1.

3. Stage-Based Quality-Related Process Analysis Method

In batch processes, it is common that the product qualities are determined by some key variables of some critical stages, and different qualities may be determined by different stages. For example, in injection molding, product surface defects, such as jetting and record grooves, depend only on injection stage; while product dimension are mainly determined by packing−holding stage. We define a stage that has most important contribution to a quality as a critical-to-quality stage. A batch process may have several critical-to-quality stages if a product has several quality measures. For different quality measures, critical-to-quality stages may be different or the same but determined by different variables, depending on the process characteristics. It is therefore important to identify critical-to-quality stages and find out in each stage the key process variables that contribute most to the quality variations. The stage-based PLS modeling method as formulated in section 2, in fact, can provide such a platform for the stage-based quality-related analysis.

We propose to determine critical-to-quality stages by measuring the goodness-of-fit of the stage PLS models. Stages with prediction models that can explain significant quality variations are defined as critical-to-quality stages. The multiple coefficient of determination, R^2, in linear regression analysis[15] can be adopted to evaluate the fitness of each stage PLS model and to present the variables' contributions of these stages to the final quality variations.

The index R^2 can be calculated for each quality variable y_{j_y} ($j_y = 1, ..., J_y$) at every sampling interval k:

$$R^2_{j_y,k} = 1 - \frac{\sum_{i=1}^{I} (y_{i,j_y,k} - \hat{y}_{i,j_y,k})^2}{\sum_{i=1}^{I} (y_{i,j_y} - \bar{y}_{j_y})^2} \quad (j_y = 1, ..., J_y) \tag{5}$$

where i, j_y, and k are indices of batch run, quality variables, and sampling time, respectively; y_{i,j_y} is the real measurement; $\bar{y}_{j_y}$ is the average across batches; and $y_{i,j_y,k}$ is the quality prediction at time k, calculated by calling the corresponding stage PLS model. Obviously, $R^2_{j_y,k}$ ranges from 0 to 1.

For each quality variable y_{j_y}, larger $R^2_{j_y,k}$ indicates better fitness of the corresponding stage PLS model, that is, the stage PLS model is more accurate and reliable for the prediction of quality variable y_{j_y}. $R^2_{j_y}$ indices over the batch duration can therefore be used to determine critical-to-quality stages. Like in linear regression, F-test can be used to test the significance of R^2:[16]

$$F = \frac{n - m - 1}{m} \left(\frac{R^2}{1 - R^2} \right)$$

The critical values of R^2 can be conversely calculated by the above equation, where the critical values of F-statistic with significance factors ($\alpha = 0.01$ and/or 0.05) can be found up in the statistical tables of F-distribution. If the average $R^2_{j_y,k}$ at stage c is larger than the critical values, stage c is defined as the critical stage for quality variable y_{j_y}.

The purposes of stage-based process analysis are two-fold: to find the critical stages for each quality variables and to analyze the stage effects of process variables on the quality variation, where as the above R^2 index is used to find critical-to-quality stages, a new index will be introduced for finding key process variables in each critical-to-quality stages.

With stage PLS model of eq 4, the prediction of quality variable y_{j_y} for the reference data sets can be written as:

$$\hat{\mathbf{y}}_{j_y,k} = [\mathbf{x}_{1,k}, \mathbf{x}_{2,k}, ..., \mathbf{x}_{J_x,k}] \cdot \begin{bmatrix} \Theta_c^*(1, j_y) \\ \vdots \\ \Theta_c^*(j_x, j_y) \end{bmatrix}$$

$$= \sum_{j_x=1}^{J_x} \mathbf{x}_{j_x,k} \times \Theta_c^*(j_x, j_y) \tag{6}$$

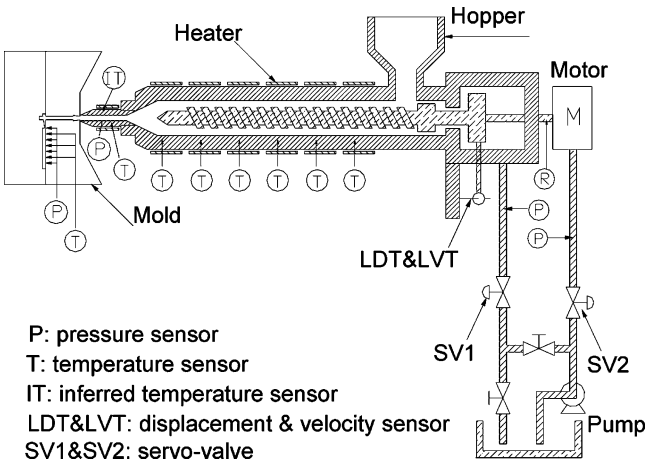

Figure 2. Simplified schematic of injection molding machine and its major measurements.

Table 1. Selected Process and Quality Variables of Injection Molding Process

no.	description	unit
	Process Variables	
1	cavity temperature	°C
2	nozzle pressure	bar
3	stroke	mm
4	injection velocity	mm/s
5	hydraulic pressure	bar
6	plastication pressure	bar
7	back pressure	bar
8	cavity pressure	bar
9	SV1	%
10	SV2	%
11	barrel temperature	°C
12	mold temperature	°C
	Quality Variables	
1	weight	g
2	length	mm
3	jetting	–
4	record grooves	–

Table 2. Operation Conditions for Training Data Sets by DOE

packing pressure (bar)	barrel temperature			mold temperature (°C)
	180 °C	200 °C	220 °C	
150	A	–	G	15
300	–	M	–	15
450	B	–	J	15
150	C	–	I	35
300	–	N	–	35
450	D	–	J	35
150	E	–	K	55
300	–	O	–	55
450	F	–	L	55

$\hat{y}_{j_y}$ is a weighted sum of process measurements. The contribution by process variable x_{j_x} can be computed by

$$\hat{\mathbf{y}}_{j_y,k}|\mathbf{x}_{j_x,k} = [0, ..., \mathbf{x}_{j_x,k}, ..., 0]\cdot\Theta_c^*(j_x, j_y) \tag{7}$$

The contribution rate of process variable x_{j_x} to $\hat{y}_{j_y}$ can then be defined as

$$C_{\mathbf{y}_{j_y},\mathbf{x}_{j_x},k} = 1 - \frac{||\mathbf{y}_{j_y} - \hat{\mathbf{y}}_{j_y,k}|\mathbf{x}_{j_x,k}||}{||\mathbf{y}_{j_y}||} \tag{8}$$

The larger $C_{\mathbf{y}_{j_y},\mathbf{x}_{j_x},k}$ is, the more significant does process variable x_{j_x} contribute to the variation of quality variable y_{j_y}. This can therefore be used to determine key process variables in each critical-to-quality stage.

Regression parameter matrix of each stage PLS model may further be explored for the details of the relationship between process variables and the final product qualities. The above proposed stage-based process analysis can be useful for the quality improvement as it provides what process variables in which stages should be better controlled for quality improvement.

Table 3. Process Settings for Injection Molding Process

parameters	settings
material	high-density polyethylene (HDPE)
injection velocity	8~40 mm/s
injection stroke	38.5 mm
packing-holding time	6 s
plastication back pressure	5 bar
screw rotation speed	80 rpm
cooling time	15 s

Table 4. Operation Conditions for Validation Data Sets

no.	packing pressure (bar)	barrel temperature (°C)	mold temperature (°C)	injection velocity (mm/s)
1	150	180	15	8
2	275	190	30	20
3	275	210	40	20
4	300	200	35	24
5	300	200	35	20
6	150	180	15	40
7	325	190	40	20
8	325	210	30	20
9	450	220	35	24

4. Online Quality Prediction Algorithms

Quality variables in a batch process can be divided into two types: qualities determined by only one specific stage and qualities determined by more than one stages. Two online quality prediction algorithms are hence developed in this section, correspondingly.

For the first case, the PLS models of the critical-to-quality stages can be simply used without any modification for online quality prediction, mathematically represented by

$$\hat{y}_{j_y,k} = \begin{cases} \mathbf{x}_k\cdot\Theta_c^*(:, j_y) & \text{if stage } c \text{ is critical to } y_{j_y} \\ \text{null} & \text{if stage } c \text{ is not critical to } y_{j_y} \end{cases}$$

$$j_y = 1, 2, ..., J_y; \quad c = 1, 2, ..., C; \quad k = 1, 2, ..., n_{\text{stage\_}c} \tag{9}$$

If, according to R^2 indices, the quality variable y_{j_y} is mainly determined by stage c (i.e., stage c is the critical-to-y_{j_y} stage), the prediction model will be activated to estimate y_{j_y} at each sampling intervals of that stage with no prediction of y_{j_y} in any other stages.

The predicted quality may vary slightly with time within the same stage; the variation may be caused by measurement noises and modeling errors as the stage PLS model is an averaged time−slice PLS model of that stage. This variation is small, as seen in section 5. Similar to the final predicted qualities in MPLS model-based methods, the averaged value of $\hat{y}_{j_y,k}$ in stage c, $\hat{y}^*_{j_y,c}$, is defined as the end-of-stage prediction.

For the second case, the variations in quality are determined by more than one stage. Two or more critical-to-quality stages have cumulative but different effects on the product quality. This case allows a larger

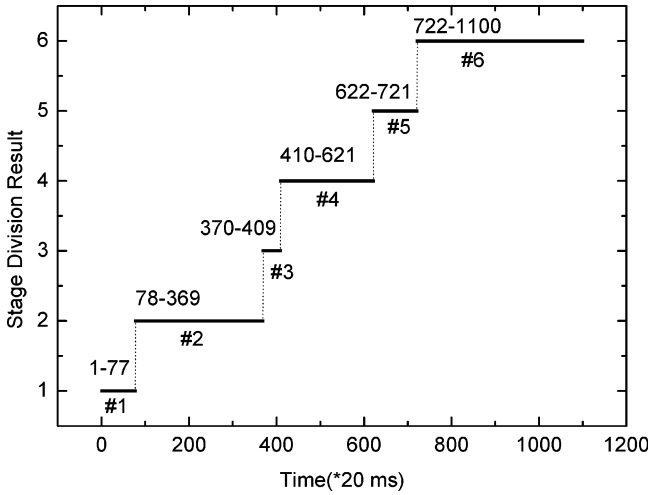

Figure 3. Stage division result for injection molding process.

degree for online quality compensation; predicted deviation in the earlier critical-to-quality stages may be compensated in the current or future critical-to-quality stages. A strategy combining multi-stage PLS models will have to be developed; stacked modeling methods [10,11] can be adopted to weight each stage PLS model for the combination.

Without losing generality, assuming that quality variable y_{j_y} has two critical stages, stage c_1 and c_2. These two stages may explain different parts of quality variations. Accurate quality prediction becomes available in the last critical stage. Online quality prediction in stage c_2 can be formulated as

$$\hat{y}_{j_y,k} = w_1 \cdot y^*_{j_y,c_1} + w_2 \cdot \mathbf{x}_k \cdot \Theta^*_{c_2}(:, j_y) \quad k = 1, 2, ..., n_{c_2} \quad (10)$$

where w is stage weight, $\hat{y}^*_{j_y,c_1}$ is the quality prediction at the end of stage c_1, $\mathbf{x}_k$ is the current process measurement, $\Theta^*_{c_2}$ is a regression parameter matrix of stage c_2, and n_{c_2} is the length of stage c_2. Weighted sum of end-of-stage predicted values is defined as final-stage prediction, formulated as

$$\hat{y}_{j_y} = \sum_c w_c y^*_{j_y,c} \quad (11)$$

Stage weight w_c can be obtained by methods such as the least square regression (LSR), by minimizing $\sum_{i=1}^{I}(y_{i,j_y} - \sum_c w_c \hat{y}^*_{i,j_y,c})^2$ for the reference data set, where i, j_y, and c are indices of batch runs, quality variables, and critical-to-y_{j_y} stages. If the end-of-stage predictions in the different critical-to-quality stages are correlated, ridge regression, principal component regression (PCR), or partial least square (PLS) can be used to calculate the weight vector in stacked regression instead of LSR method.[11]

5. Illustration on Injection Molding

5.1. Injection Molding. Injection molding is an ideal process for the application and verification of the proposed stage-based modeling methods for monitoring and quality prediction.[14] Figure 2 shows a simplified illustration of a typical reciprocating-screw injection molding machine with instrumentations. Key process conditions such as temperatures, pressures, displacement, and velocity can be measured online by their corresponding transducers, while quality variables are available only at the end of each batch run. The selected process and quality variables are listed in Table 1, among which the dimension qualities (weight and length) can be directly measured by instruments; while the surface defects (jetting and record grooves) are quantified by a process operator expert before modeling.

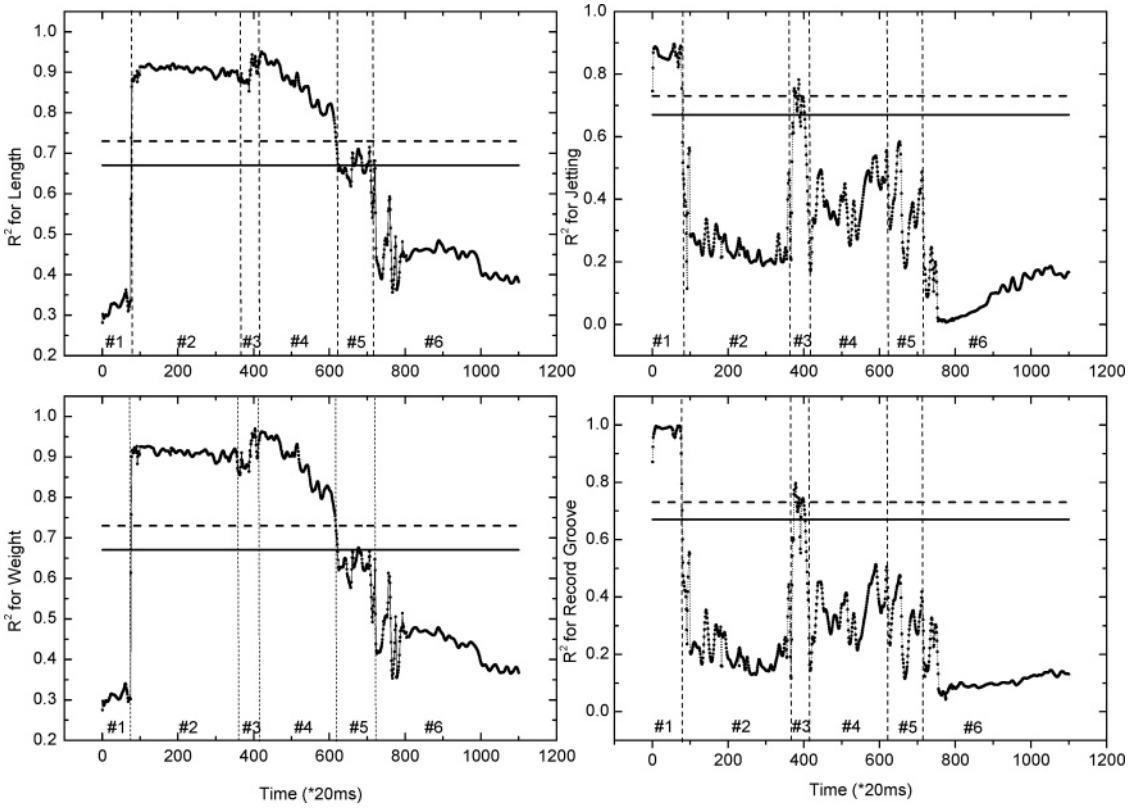

Figure 4. R^2 plots of four quality variables in injection molding process.

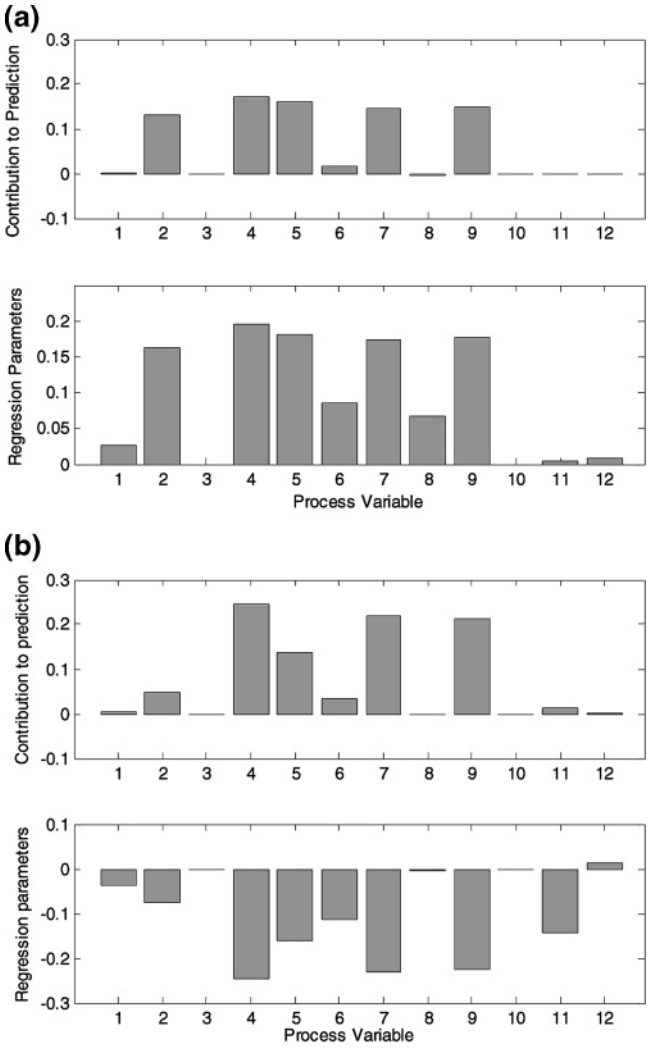

Figure 5. Key process variables in critical-to-surface stage (stage 1). (a) For jetting. (b) For record grooves.

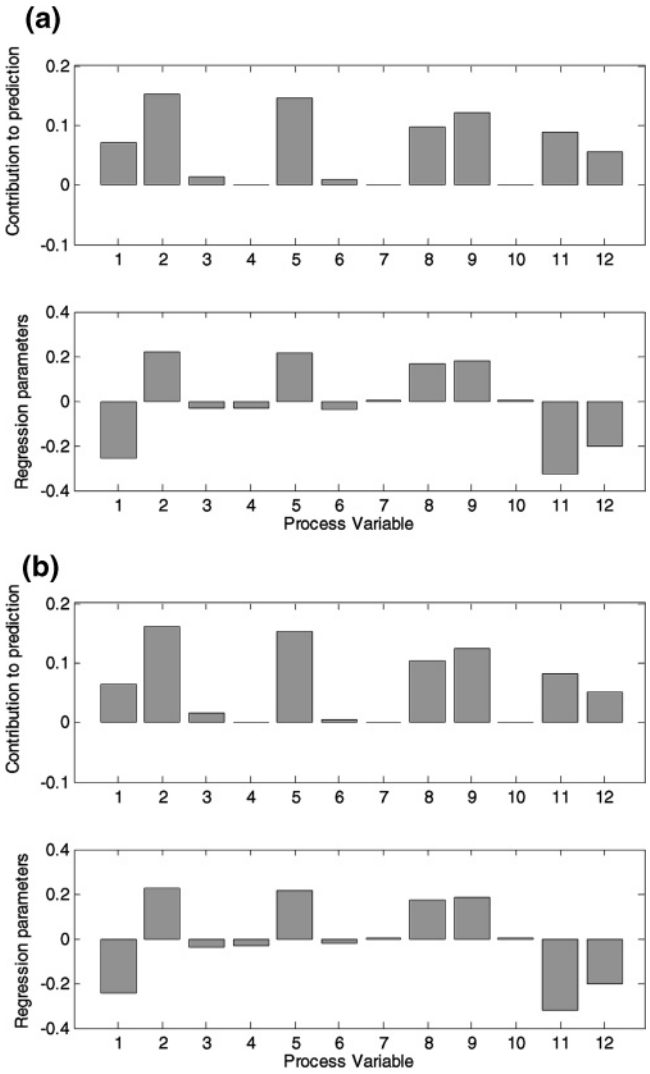

Figure 6. Key process variables in critical-to-dimension stage (stage 2). (a) For length. (b) For weight.

Experimental design as shown in Table 2 is applied to excite the process, generating training data sets for modeling and analysis. The range of DOE factors are determined according to the machine and mold specifications.[17] The first 12 conditions in the design, treatments A−L form a balanced DOE with packing pressure taking three levels (low = 150 bar, middle = 300 bar, and high = 450 bar) and barrel temperature (low = 180 °C and high = 220 °C) and mold temperature (low = 15 °C and high = 55 °C) of two levels. Treatments M, N, and O are added with barrel temperature (middle = 200 °C) and mold temperature (middle = 35 °C) both taking the centered value and packing pressure of three levels. As jetting and record grooves are associated with injection velocity, four additional experiments with different velocity settings (low = 8 mm/s and high = 40 mm/s) are conducted in conditions A, C, and D. All other process settings are shown in Table 3.

All together, 33 training batch runs were collected for modeling and post-batch analysis. Nine experiments with the settings shown in Table 3 were conducted as validation data sets to evaluate the proposed method. The details of each of those nine experiments are listed in Table 4.

5.2. Illustration of Stage-Based Process Analysis. (a) Stage Division. As different injection velocities result in different filling time in the injection stage,[14] reference batch runs therefore have varying data length. To arrange these batch data into matrices, batches with different injection velocities are aligned to have the same filling length, taking injection stroke as an indicator variable, as the values of injection stroke decreases monotonically from 60 mm to 21.5 mm.The total displacement is fixed at 38.5 mm at the end of injection stage. There are finally 78 samples in injection stage after alignment. The dimensions of the final data matrices $\underline{X}$ and Y are 33 × 12 × 1100 and 33 × 4, respectively.

According to the modeling procedure, $\underline{X}$ is first cut into 1100 time−slice matrices $\tilde{X}_k$ (k = 1, ..., 1100). PLS is applied to $\{\tilde{X}_k, Y\}$, generating 1100 time−slice regression parameter matrices $\tilde{\Theta}_k$. The clustering algorithm is divided the process into six stages, as shown in Figure 3. This stage division result is slightly different from that of the stage-based sub-PCA modeling method,[14] where the process is divided into four stages, in consistent with the real operation stages: injection, packing−holding, plastication, and cooling. From Figure 3, the first two stages and the last stage are in agreement with the actual operation stages of injection, packing−holding and cooling; but the plastication operation is partitioned into two sub-stages, stages 4 and 5. These two stages indeed have different effects on the final

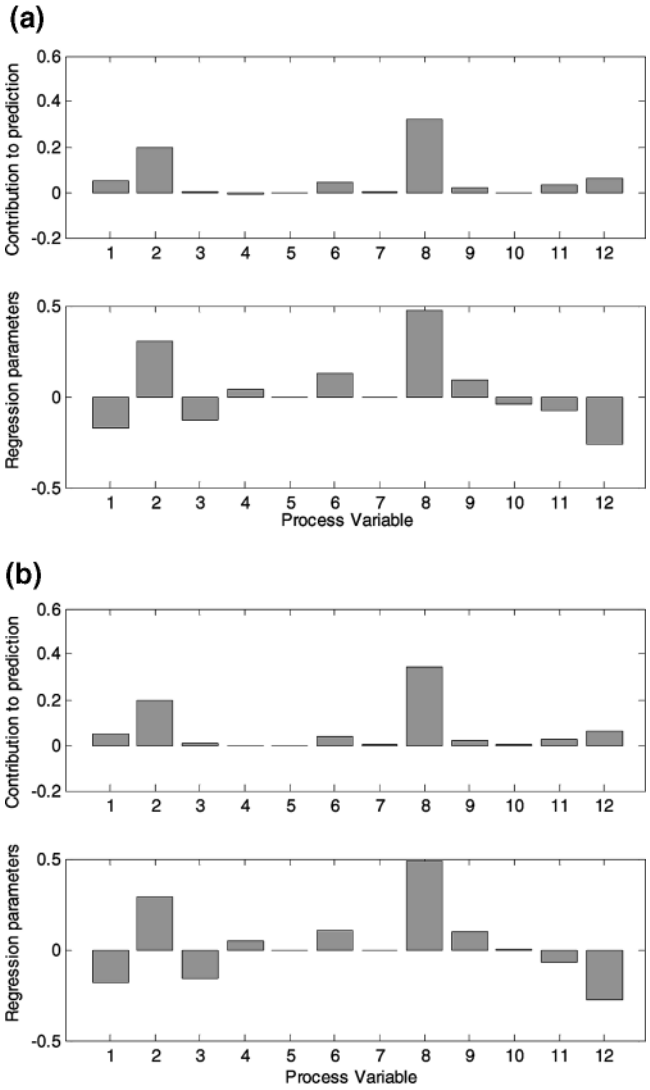

Figure 7. Key process variables in critical-to-dimension stage (stage 4). (a) For length. (b) For weight.

product quality and have different correlation between process and quality variables, as explained later. A short period between packing−holding and plastication stages forms a separate stage, that is, stage 3 in Figure 3. This stage corresponds to the operation of "suck-back" to retract screw for a certain distance to prevent the melt-drooling, which has little impact on quality. This stage will be neglected in the following stage-based process analysis.

Figure 4 shows the R^2 indices for the four quality variables over the batch duration. Obviously, the surface qualities, jetting and record grooves, have large and stable R^2 in the first stage (i.e., in injection stage) while the dimension qualities, length and weight, have significant relation to stages 2 and 4. The first stage is therefore defined as critical-to-surface stage, and stages 2 and 4 are critical-to-dimension stages. Detailed analysis is given in the following subsections.

(b) Process Analysis in Critical-to-Surface Stage. Figure 5 shows the contributions of process variables to the prediction of jetting and record grooves and the corresponding stage regression parameters that indicate the correlation pattern between process and quality variables in the first stage, respectively.

From Figure 5a, the variation of nozzle pressure (no. 2), injection velocity (no. 4), hydraulic pressure (no. 5), back pressure (no. 7), and manipulated variable SV1 (no. 9) can accurately reflect the variation of jetting in the products. These process variables have positive correlation to jetting, which means that high pressures and large injection velocity in the injection stage cause jetting in the final products. The same group of process variables also determines the variation of record grooves as shown in Figure 5b. The correlation between these process variables and record grooves, however, are negative, indicating that low pressures and slow injection velocity are responsible for record grooves. The negative correlation between jetting and record grooves, in fact, agrees well to the process characteristics.[17]

Table 5. Prediction Performance Comparison for both Training and Test Data Sets

	mean squared error			
methods	length	length	jetting	RG
Training Data Sets				
MPLS	**0.0055**	**0.0099**	0.0394	0.0147
the proposed	**0.0048**	**0.0147**	**0.0205**	**0.0019**
B-MPLS	0.0213	0.0497	**0.0046**	**0.0004**
Test Data Sets				
MPLS	**0.0190**	**0.0272**	0.4632	0.3065
the proposed	**0.0254**	**0.0194**	**0.1007**	**0.0244**
B-MPLS	10.45	0.08451	**0.0712**	**0.0156**

(c) Process Analysis in Critical-to-Dimension Stages. In stage 2 (i.e., the packing−holding stage), dominant process variables include pressure variables (nos. 2, 5, and 8), temperature variables (no.1, 11, and 12), and manipulated variable SV1, as shown in Figures 6. The stage regression parameters can explain how these variables affect the product dimension qualities. From Figure 6, pressure variables have positive relation with the variations of length and weight; while temperature variables are negatively related with the dimension qualities. This implies that larger pressures and lower temperatures result in larger and heavier products, which again agrees well with the physics of polymer processing.

Stage 4 (i.e., the earlier phase of plastication stage) is also recognized by the algorithm as a critical-to-dimension stage according to the R^2 plot. Dividing the plastication operation into two sub-stages is reasonable in our experiments because the process shows different characteristics before and after mold gate is frozen. From Figure 7, cavity pressure (no. 8) is the most dominant variable reflecting variations in the dimension qualities. This is because cavity pressure is mainly determined by the volume of material filled in the cavity and has close relation with nozzle pressure (no. 2) and mold temperature (no. 12) before the mold gate is frozen in the earlier phase of plastication. The variations of the above three process variables therefore can accurately reflect the dimension variation of filled materials in the mold. The decreasing explained rate in stage 4 as shown in the R^2 plot may be caused by the solidification in the filled polymer in mold and gradually frozen in the gate.

Without the use of prior process knowledge, the above analysis results in the following observations for the injection molding, which agree well with the process characteristics:

(i) This injection molding process can be divided into six stages; different stages have different effects on the final product qualities.

(ii) The surface qualities are determined by the injection stage (stage 1) where nozzle pressure (no. 2),

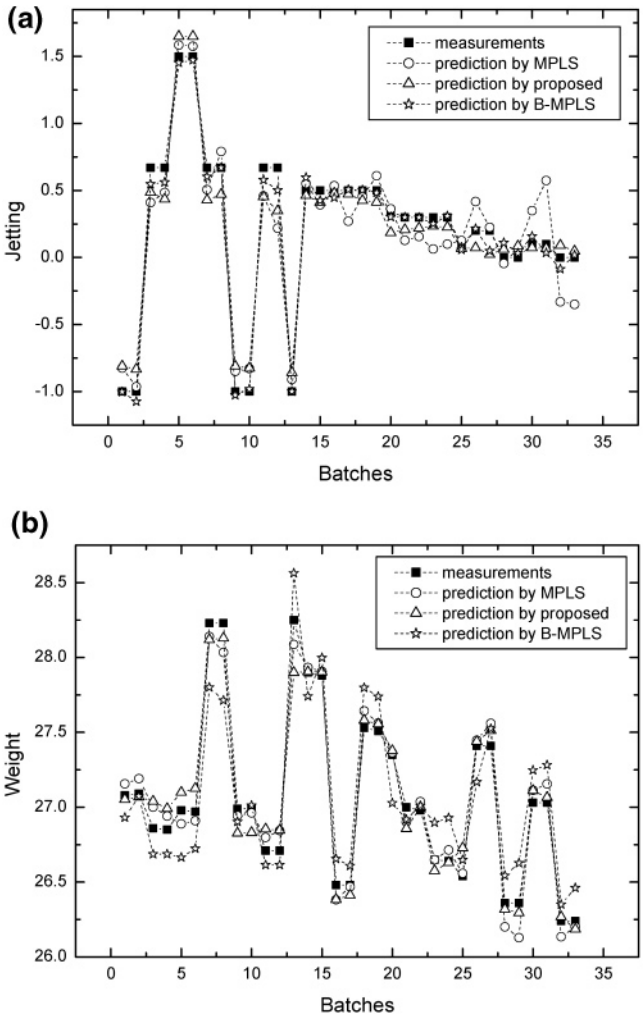

Figure 8. Prediction comparison for training batch runs.

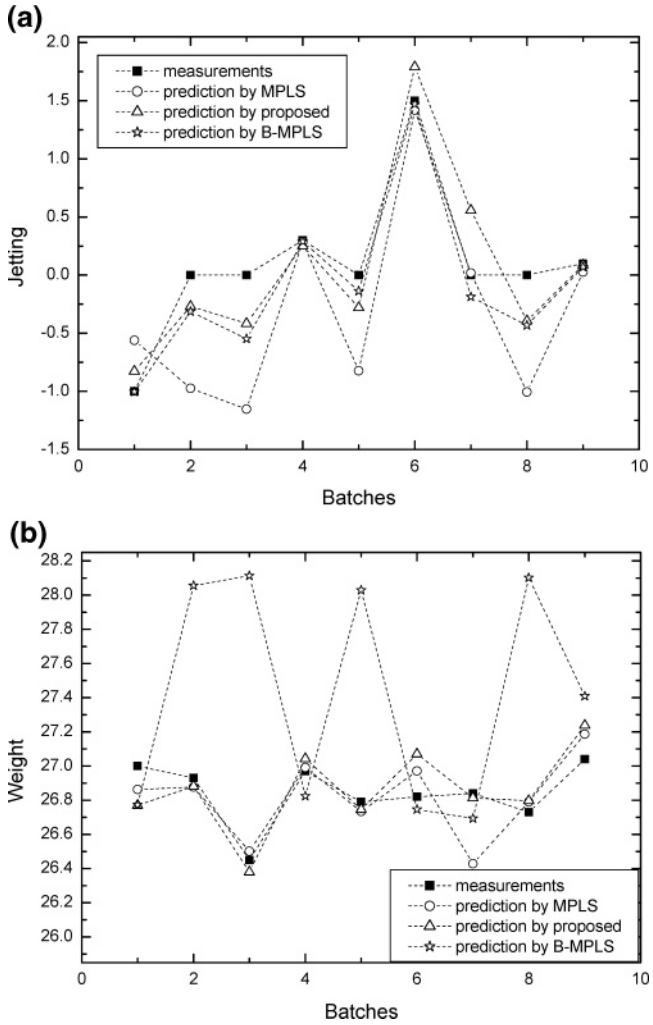

Figure 9. Prediction comparison for test batch runs.

injection velocity (no. 4), hydraulic pressure (no. 5), back pressure (no. 7), and manipulated variable SV1 (no. 9) are key process variables that have large contributions to the variation of surface attributes.

(iii) The dimension qualities are mainly determined by the packing−holding stage (stage 2). Larger pressures (nos. 2, 5, and 7) and lower temperatures (nos. 1, 11, and 12) lead to heavier product. Cavity pressure (no. 8) and nozzle pressure (no. 2) at the earlier phase of plastication stage (stage 4) can also have significant effects on dimension variations.

(iv) The remaining period of plastication (stage 5) and the cooling stage (stage 6) have little effects on the final product qualities of our concern here.

(v) For on-line quality prediction, surface qualities can be predicted by process variables in the first stage; while the dimension qualities can be predicted in the fourth stage.

5.3. Illustration of Stage-Based Quality Prediction. The above stage-based quality-related process analysis can divide the four quality variables into two types. Surface qualities are determined only by the first stage; while the dimension qualities have close relations with stages 2 and 4. Online predictions of surface attributes are given in the first stage; predictions of weight and length are available from stage 4.

A comparison of prediction performance is conducted for MPLS model, the proposed stage PLS model, and the bootstrapping-based MPLS model (B-MPLS), as shown in Figures 8 and 9 and Table 5. Figures 8 and 9 show the quality predictions by the three methods for the training and validation data sets, respectively. Table 5 summarizes the prediction performances. From Figures 8 and 9 and Table 5, conclusions can be drawn that the MPLS model is applicable for prediction of quality variables that are cumulatively determined by several stages; bootstrapping-based improved MPLS model is applicable for prediction of quality variables that have significant time-specific correlation to process variables trajectories; the proposed stage-based PLS model can give reasonable predictions for both types of quality variables.

Online predictions of the product weight for test batch 4 are compared in Figure 10, using the MPLS model and the proposed stage PLS model, respectively. Offline prediction results shown in Figure 9 suggest that the proposed stage PLS model has a comparable accuracy to the MPLS model; while online predictions of Figure 10 show the superiority of the proposed method. The online prediction of the weight by the proposed method is not only more accurate but also can be made much earlier at the beginning of stage 4 rather than at the end of batch.

The above experimental results show that the proposed stage-based process analysis and quality prediction methods can indeed be successfully applied to obtain stage information, to explore stage-specific effects

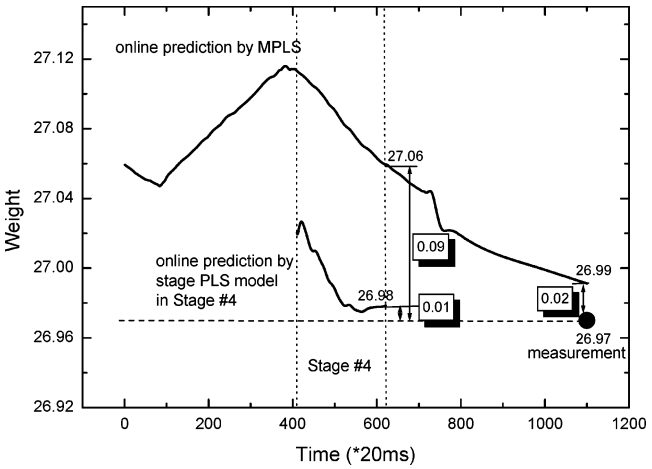

Figure 10. Online weight prediction for test batch 4 by MPLS and the proposed method.

of process variables, to find the key process variables, and to give earlier and accurate online quality prediction.

6. Conclusions

A scheme of process analysis and on-line quality prediction for multistage batch processes has been proposed. On the basis of the stage-based PLS modeling, a batch process can be divided into stages, and stages can be identified for their different effects on the end product qualities. A stage-based quality-related process analysis has been proposed and illustrated to find critical-to-quality stages. Quality prediction is also made stage-based to result in earlier and accurate online prediction. The proposed analysis and modeling can not only enhance process understanding but also lead to identification of specific process variables and periods for quality improvement. Application to injection molding process shows good feasibility of the proposed scheme.

Acknowledgment

This project was supported in part by the Hong Kong Research Grant Council under Project No. 601104.

Appendix: A *k-means* Clustering Algorithm

Inputs: $\{\tilde{\Theta}_1, \tilde{\Theta}_2, ..., \tilde{\Theta}_K\}$, patterns to be partitioned.

Outputs: Cluster centers, $\{W_1, W_2, ..., W_C\}$; membership of $\tilde{\Theta}_k$, $m(k)$: $\tilde{\Theta}_k \rightarrow \{1, 2, ..., C\}$; and the number of clusters, C.

Index variables: iteration index, i, and pattern index, k.

The Euclidean distance is used to identify the dissimilarity between two patterns $\tilde{\Theta}_{k_1}$ and $\tilde{\Theta}_{k_2}$, $\text{dist}(\tilde{\Theta}_{k_1}, \tilde{\Theta}_{k_2}) = \sum_{j_y=1}^{j_y} |\tilde{\Theta}_{k_1}(:, j_y) - \tilde{\Theta}_{k_2}(:, j_y)|^2$.

The clustering algorithm for classifying the K number of PLS regression parameter matrix $\tilde{\Theta}_k$ is given as follows:

1. Choose C^0 ($i = 0$) cluster centers W_c^0 ($c = 1, 2, ..., C^0$) from K patterns along time series. Practically, the initial cluster centers can be assumed to be uniformly distributed in the pattern set.

2. Merge clusters of which the distance between their centers, $\text{dist}(W_{c_1}^{i-1}, W_{c_1}^{i-1})$, is below a predetermined threshold ϵ.

3. Calculate the distances between each $\tilde{\Theta}_k$ to all the centers, $\text{dist}(\tilde{\Theta}_k, W_c^{i-1})$, and assign $\tilde{\Theta}_k$ to the closest center $W_{c^*}^{i-1}$, and its membership is assigned to be $m(k) = c^*$.

4. Eliminate the centers that catch few patterns after a set number of iteration $i > I\_num$ to avoid singular clusters.

5. Update the number of cluster centers to be C^i; recompute the new cluster centers, W_c^i ($c = 1, 2, ..., C^i$), using the current cluster membership, $m(k)$.

6. Go back to step 2 if a convergence criterion is not met. Typical convergence criteria are minimal changes in the cluster centers and/or minimal rate of decrease in squared errors.

Literature Cited

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

(2) Nomikos, P.; MacGregor, J. F. Multi-way partial least squares in monitoring batch processes. *Chemom. Intell. Lab. Syst.* **1995**, *30*, 97.

(3) Cho, H.-W.;Kim, K.-J. A method for predicting future observations in the monitoring of a batch process. *J. Qual. Technol.* **2003**, *35*, 59.

(4) Wold, S.; Antti, H.; Lindgren, F.; Öhman, J. Orthogonam signal correction of near-infrared spectra. *Chemom. Intell. Lab. Syst.* **1998**, *44*, 175.

(5) Trygg, J.; Wold, S. Orthogonal projection to latent structures (O−PLS). *J. Chemom.* **2002**, *16*, 119.

(6) Chu, Y.-H.; Lee, Y.-H.; Han, C. Improved quality estimation and knowledge extraction in a batch process by bootstrapping-based generalized variable selection. *Ind. Eng. Chem. Res.* **2004**, *43*, 2680.

(7) Yu, H.; MacGregor, J. F. Post processing methods (PLS−CCA): simple alternatives to preprocessing methods (OSC−PLS). *Chemom. Intell. Lab. Syst.* **2004**, *73*, 199.

(8) Duchesne C.; MacGregor C. D. Multivariate analysis and optimization of process variable trajectories for batch processes. *Chemom. Intell. Lab. Syst.* **2000**, *51*, 125.

(9) Lu, N.; Gao, F.; Wang, F. A sub-PCA modeling and on-line monitoring strategy for batch processes. *AIChE J.* **2004**, *50*, 255.

(10) Breiman, L. Stacked regression. *Machine Learn.* **1996**, *24*, 49.

(11) Zhang, J.; Martin, E. B.; Morris, A. J. Kiparssider C. Inferential estimation of polymer quality using stacked neural networks. *Comput. Chem. Eng.* **1997**, *21*, s1025.

(12) Nomikos, P.; MacGregor, J. F. Multivariate SPC charts for monitoring batch processes. *Technometrics* **1995**, *37*, 41.

(13) Kassidas, A.; MacGregor, J. F.; Taylor, P. A. Synchronization of batch trajectories using dynamic time warping. *AIChE J.* **1998**, *44*, 864.

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

(15) Johnson, R.; Wichern, D. *Applied Multivariate Statistical Analysis*; Prentice Hall: Upper Saddle River, NJ, 2002.

(16) Golberg, M. A.; Cho, H. A. *Introduction to Regression Analysis*; WIT Press: Boston, 2004.

(17) Yang, Y. Injection molding: from process to quality control. Ph.D. Thesis, The Hong Kong University of Science & Technology, 2004.

Received for review November 29, 2004
Revised manuscript received February 27, 2005
Accepted March 7, 2005

IE048852L