Between-Phase Calibration Modeling and Transition Analysis for Phase-Based Quality Interpretation and Prediction

Chunhui Zhao, Furong Gao, and Youxian Sun
Laboratory of Industrial Control Technology, Dept. of Control Science and Engineering, Zhejiang University, Hangzhou, 310027, China

DOI 10.1002/aic.13790
Published online March 21, 2012 in Wiley Online Library (wileyonlinelibrary.com).

The quality-concerned between-phase transition analysis is performed and an improved calibration modeling strategy is designed for quality prediction and interpretation in multiphase batch processes. From the between-phase viewpoint, the quality-related phase behaviors are decomposed and two subspaces are separated. In common subspace, the underlying quality-relevant variation stays invariable between the neighboring phases, showing the common contribution to quality. The other part changes with the alternation of phases and has the different influences on quality interpretation, termed specific subspace here. Based on subspace separation, between-phase transition regions are distinguished from steady phases. Different models are developed in steady phases and transition regions respectively for online quality prediction. Offline quality analyses are also conducted in two subspaces to explore the time cumulative effects. The proposed method gives an interesting insight into the phase behaviors and between-phase transitions for quality prediction. The feasibility and performance of the proposed method are illustrated with a typical multiphase batch process. © 2012 American Institute of Chemical Engineers *AIChE J,* 59: 108–119, 2013
Keywords: multiphase batch processes, between-phase transition, subspace separation, quality interpretation and prediction

Introduction

Data-based multivariate calibration methods have been widely used to establish a quantitative relationship between process measurement (**X**) and quality property (**Y**). Accurate qualitative and quantitative calibration analysis may help avoiding cumbersome and costly chemical measurements. In practice, calibration modeling and analysis can often be accomplished with familiar, conventional statistical techniques,[1–11] such as multiple linear regression, principal component regression, canonical correlation analysis (CCA), and partial least squares (PLS). The subject of calibration modeling and quality interpretation arouses new issues and demand specific solutions when it refers to multiphase batch processes where various phases generally operate orderly under the domination of different physical phenomena, revealing different effects on the final quality. How to characterize them efficiently is a challenging problem.

Multiway partial least squares (MPLS) model[12] uses process variables over the entire batch course as the input, which reveals well the time correlations throughout the cycle and, thus, shows efficiency for analysis of cumulative effects on quality. However, for multiphase batches, it is generally deemed that if the data are handled in a single matrix, the effect of one segment tends to be hidden more or less by the influence of another. The resulting tribulation is that the hidden effect could be useful in quality-concerned process analysis and control. It is commonly accepted that in multiphase cases, more underlying information can be explored by dividing the data into meaningful blocks either by the types of variables or by the part of the process they originate from and building multiple specific models instead of single modeling of all data. The effect of each block can be seen clearly and, thus, more comprehensive process understanding and quality analysis can be expected.

Considering that the phase multiplicity is an inherent nature of many batch processes, various strategies have been reported and can be put into quality prediction for online or offline applications.[13–25] One is to model the variable correlation within each phase under the influence of other phases by multiblock PLS (MBPLS).[13–18] Compared with MPLS, the advantage of MBPLS is mainly to allow easier interpretation of both the roles of each smaller meaningful block and the integrated contribution of all blocks. Qin et al.[14] have made a comprehensive review of multiblock algorithms and reported the super scores of MBPLS were identical to the scores of regular MPLS and, thus, achieved the same quality prediction performance. Further, Reinikainen and Hoskuldsson et al.[19] reported a priority PLS regression analysis method and its successful application to a multistep industrial process, which gave multiple phases descending priority following operation time sequence. The idea was to compute the predictor information first only on the basis of first phase. When no more significant latent variables (LV)s could be found from the first phase, the next predictor information was extracted from the data of the second phase so that the quality information that is not modeled by the previous phases will be left to be explained by the following phases. Zhao et al.[20] presented a two-level multiphase statistical analysis strategy to probe the phase-wise local and cumulative effects on quality interpretation and prediction by calibration modeling in both predictor and quality spaces. It revealed within each phase which part of process variation was responsible for quality variations and which part of

Correspondence concerning this article should be addressed to C. Zhao at chhzhao@zju.edu.cn.

quality variation was dominated. For online quality prediction, the subPLS modeling algorithm was first developed by Lu and Gao,[21] where all measurements within the same phase could use the same regression model. It was based on such a presupposition that despite the time-varying batch operation trajectory, the correlations between process and quality variables should remain similar within the same phase. Further development[22–25] has also been reported to further extend this research line.

In this work, a between-phase calibration modeling strategy is proposed for transition analysis and quality prediction and interpretation. For multiphase batch processes, how the effects of process variation on quality evolve with the phase changes is an interesting issue. The key point or most concerned is, thus, how to capture them along with phase alternation and process evolvement. This also requires the attention to the between-phase transitions and their difference from the steady phases. From this motivation, a multiset regression modeling method, termed multiset regression analysis (MsRA),[26] which was proposed to relate the inherent quality-related predictor variation across multiple data sets, can be used here as the basic modeling method. Following the theoretical developments and its property analysis in the previous work,[26] this study addresses the potential of using the said algorithm to solve one practical problem, the quality prediction and interpretation in multiphase batch processes with between-phase transitions. It should be noted that here MsRA only focuses on two neighboring phases for between-phase transition analysis and quality prediction. From the quality-concerned viewpoint, each of the neighboring phases can be separated into two subspaces, termed between-phase common and specific subspaces, respectively. In the between-phase common subspace, the quality-concerned process variability stays invariable from one phase to another, thus, revealing the same contributions to quality properties. That is, they explain the same part of quality variability repetitively. In the left between-phase specific subspace, the process variability is more specific, revealing different effects on quality interpretation between the two neighboring phases. That is, they explain different parts of quality characteristics in a complementary way. For one extreme case, if the quality-concerned systematic variations in the two neighboring phases can be all characterized in the common subspace, they will have quite the same quality contributions even though they may be described by different regression weights. The purposes of between-phase calibration analysis are two-fold: to explore the transition regions between two neighboring phases and give more meaningful insight into the phase behaviors with respect to quality prediction. Here, it should be noted that for quality analysis, the between-phase transition is different from that in process monitoring, as they check the changes of different underlying process variations. In comparison with the traditional calibration methods, the separation of the common and specific predictor variations and the dual care paid to both are the major differences.

Methodology

Transition problem in quality prediction

For multiphase batches, in general, different phases have different characteristics and reveal different effects on quality. From the quality-concerned viewpoint, the process may be divided into several modeling phases as indicated by the changes of the quality-relevant inherent characteristics. In process monitoring, the between-phase transition problem has been noticed.[27] It is in general deemed that the process actually covers steady status and dynamic transition status from a starting phase to the target phase. Valid process monitoring during transitions are attracting increasing attention.[27–29] Commonly, the transition pattern shows the gradual changes between two neighboring phases. The underlying characteristics of the transition patterns at the beginning resemble the characteristics of the starting phase, whereas those later transition patterns resemble the characteristics of the target phase. Based on the above recognition, some corresponding monitoring strategies have been developed.[27–29]

However, the between-phase transition problem is different and requires special attention with respect to quality prediction. Instead of only considering the process variables, dual attention is paid to both process variable correlations and those between process variables and quality. In this study, it is deemed that the contribution to quality may not completely change from one phase to another. That is, although the regression relationship may change as indicated by the prediction models, one part of the underlying quality-related predictor variation will stay invariable or consistent between the two neighboring phases. This reveals the same power of quality prediction and interpretation, as the prediction is made by directly regressing quality on the extracted regression scores (that is, the predictor variation). The difference in between-phase predictor variation reflects the different quality interpretability. Therefore, how the predictor variation information changes from one phase to another actually reveals the between-phase changeover of quality interpretability. Instead of isolating the effects of single phase on qualities, it is necessary to gain a detailed quality-concerned insight into the underlying phase characteristics from the between-phase viewpoint, which can provide important information of the changing quality interpretability from one phase to another. The key is how to separate the two different types of predictor variations in each phase.

Based on the above analysis, the phase behavior should be modeled in a quite different way from the conventional calibration methods. The MsRA[26] algorithm includes two versions, MsRA-score and MsRA-weight, which extract the similar regression scores and weights, respectively. Actually, the quality prediction and interpretation is directly related with the regression scores where the same scores will make the same contribution to quality. The regression weights are just for the calculation of scores, where even the similar weights can be figured out, the regression scores may be quite different, revealing quite different quality interpretability. In the study, as the quality interpretability is the concerned, the regression scores should be used as the analysis subject and MsRA-score algorithm can be used here as the basic modeling method for between-phase analysis. With this method, the original phase measurement space can be separated into two different subspaces, each subspace containing different underlying predictor information and revealing different effects on qualities. One subspace is called the common subspace, which is composed of the similar predictor variations between two neighboring phases. The other is called the between-phase specific predictor subspace, which is supported by the unique predictor variation in each phase. The two subspaces give the similar and different quality interpretability between two neighboring phases, respectively. The separation of the common and specific phase information and the attention given to both provide a good platform for the identification of transition regions.

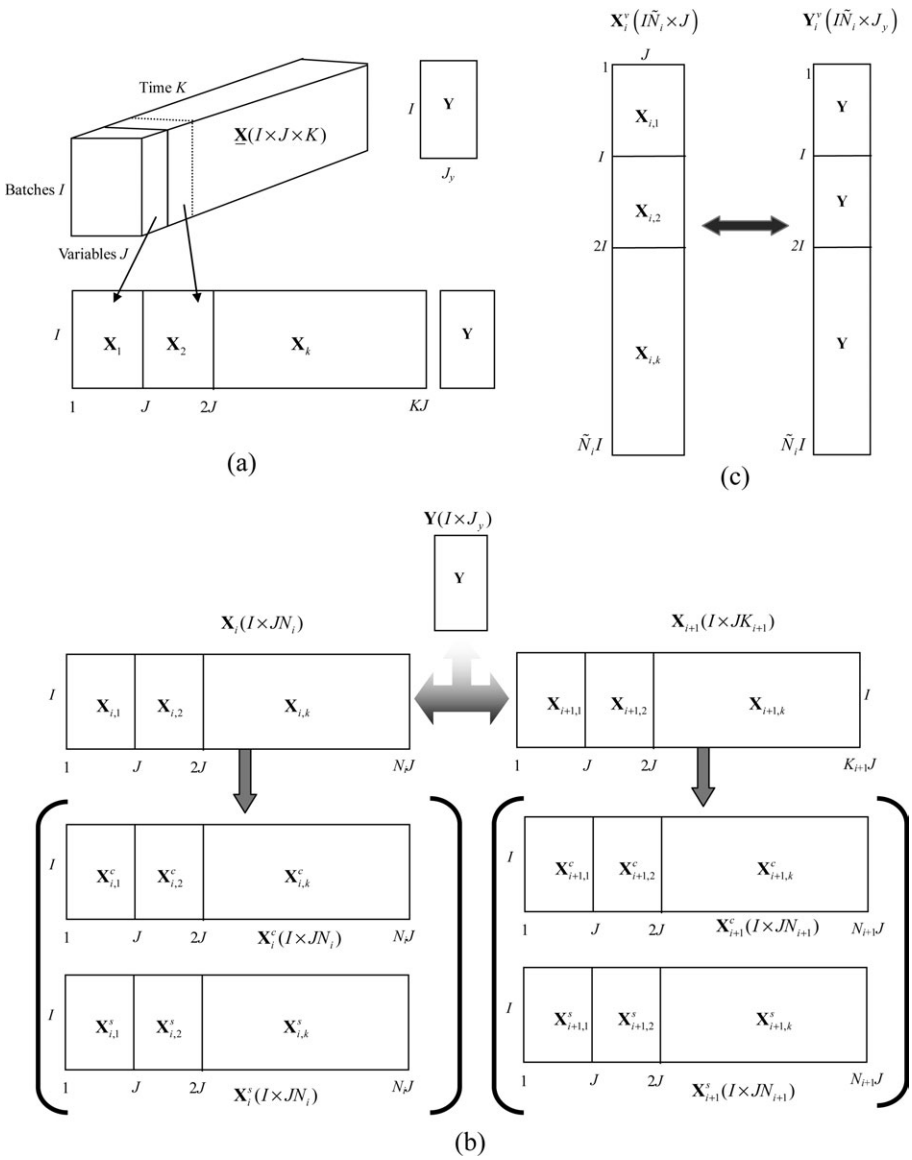

Figure 1. Illustration of the between-phase regression modeling scheme (a) Batch-wise unfolding and data normalization, (b) between-phase subspace separation for the *i*th and *i*+1th phases, and (c) regression modeling for online quality prediction with respect to the *i*th steady phase.

Between-phase subspace separation

In each batch run, assume that J process variables are measured online at $k = 1, 2, \ldots, K$ time instances throughout the batch and J_y quality variables are obtained offline. Then, process observations collected from similar I batches can be organized as a three-way array $\underline{\mathbf{X}}(I \times J \times K)$ and a corresponding quality matrix $\mathbf{Y}(I \times J_y)$ as shown in Figure 1a. At each time, the means of each column are subtracted to approximately eliminate the main nonlinearity. Each variable is scaled to unit variance to handle different measurement units, thus, giving each equal weight. In this work, the batches are of equal length without special declaration so that the specific process time can be used as an indicator for data normalization.

Phase information should be identified before between-phase analysis. In multiphase modeling, one important issue is how to get the phase marks (i.e., phase division) before designing the phase model. Various strategies[21,23,30–34] have been reported from different viewpoints and based on different principles, providing a rich database for phase division, and can be put into practical process monitoring. In this work, assuming that no prior process knowledge is available, C phases can be readily identified along time direction by clustering algorithm as what was done in previous work,[21] where the PLS weights are used to evaluate the similarity of time-slices, revealing the changes of process-quality relationships. Certainly, the phase clustering information can also be modified using prior expertise. Then, the process data in each phase are arranged as $\mathbf{X}_i(I \times JN_i)$ $(i = 1, 2, \ldots, C)$, where I is the number of batches and N_i is the number of time samples in each phase. They are corresponding to the same quality data, $\mathbf{Y}(I \times J_y)$.

The between-phase similarity in quality prediction and interpretation is deemed to be driven by some common regression scores, which are directly related to quality. Using MsRA-score algorithm,[26] the common regression scores are extracted focusing on two adjacent phases, which can describe the process variables by linear combinations. Here,

it should be noted that for each phase, when it is analyzed with different adjacent phases, the subspace separation results may be different, revealing different between-phase relationships with respect to quality analysis.

As shown in Figure 1b, for any two neighboring phases, $\mathbf{X}_i(I \times JN_i)$ and $\mathbf{X}_{i+1}$ ($I \times \mathrm{JN}_{i+1}$), they may have a different number of variables but the same samples. Under the dual supervision of betweens-set relationship and quality variables ($\mathbf{Y}(I \times J_y)$), the between-phase common regression scores ($\mathbf{T}_i$ ($I \times R_i^c$), where R_i^c is the number of retained common scores) are obtained in each of the two neighboring phases by the two-step MsRA algorithm[26]

$$\begin{aligned}\bar{\mathbf{R}}_i &= \bar{\mathbf{A}}_i\left(\bar{\mathbf{P}}_i^{\mathrm{T}}\bar{\mathbf{A}}_i\right)^{-1}\\ \bar{\mathbf{T}}_i &= \mathbf{X}_i\bar{\mathbf{R}}_i\\ \mathbf{T}_i &= \bar{\mathbf{T}}_i\mathbf{A}_i = \mathbf{X}_i\bar{\mathbf{R}}_i\mathbf{A}_i = \mathbf{X}_i\mathbf{R}_i\end{aligned} \tag{1}$$

where $\overline{\mathbf{A}}_i$ is composed of the weight vectors $\mathbf{a}_i$ calculated in the first step, and $\overline{\mathbf{P}}_i$ is the regression loading matrix for $\mathbf{X}_i$ by the first-step MsRA. The weight coefficients ($\overline{\mathbf{R}}_i$) to compute scores ($\overline{\mathbf{T}}_i$) directly from original $\mathbf{X}_i$ is expressed based on the similar calculation in PLS. $\overline{\mathbf{T}}_i$ is the scores from the first-step calculation. $\mathbf{A}_i$ is composed of weight vectors $\mathbf{a}_i$ obtained in the second step. The final weight vectors ($\mathbf{R}_i$) are obtained by the two-step calculation to compute scores ($\mathbf{T}_i$) directly from the original $\mathbf{X}_i$.

For two neighboring phases, $\mathbf{X}_i$ ($I \times JN_i$) and $\mathbf{X}_{i+1}$ ($I \times JN_{i+1}$), the common scores are extracted to achieve the maximal correlations. Then, the representative global/common scores can be calculated by averaging the common scores from the two phases

$$\mathbf{T}_{\mathrm{g}} = \frac{1}{2}(\mathbf{T}_i + \mathbf{T}_{i+1}) \tag{2}$$

Thus, each original observed data space ($\mathbf{X}_i$) is separated into two different parts, $\mathbf{X}_i^c$ and $\mathbf{X}_i^s$

$$\begin{aligned}\mathbf{P}_i^{c\mathrm{T}} &= \left(\mathbf{T}_{\mathrm{g}}^{\mathrm{T}}\mathbf{T}_{\mathrm{g}}\right)^{-1}\mathbf{T}_{\mathrm{g}}^{\mathrm{T}}\mathbf{X}_i\\ \mathbf{X}_i^c &= \mathbf{T}_{\mathrm{g}}\mathbf{P}_i^{c\mathrm{T}} = \mathbf{T}_{\mathrm{g}}\left(\mathbf{T}_{\mathrm{g}}^{\mathrm{T}}\mathbf{T}_{\mathrm{g}}\right)^{-1}\mathbf{T}_{\mathrm{g}}^{\mathrm{T}}\mathbf{X}_i\\ \mathbf{X}_i^s &= \mathbf{X}_i - \mathbf{X}_i^c = \left(\mathbf{I} - \mathbf{T}_{\mathrm{g}}\left(\mathbf{T}_{\mathrm{g}}^{\mathrm{T}}\mathbf{T}_{\mathrm{g}}\right)^{-1}\mathbf{T}_{\mathrm{g}}^{\mathrm{T}}\right)\mathbf{X}_i\end{aligned} \tag{3}$$

where $\mathbf{P}_i^{c\mathrm{T}}$ are the ($R_i^c \times JN_i$)-dimensional loadings or linear combination coefficients corresponding to the common prediction scores. Actually, $\mathbf{G}_{\mathbf{T}_{\mathrm{g}}} = \mathbf{T}_{\mathrm{g}}(\mathbf{T}_{\mathrm{g}}{}^{\mathrm{T}}\mathbf{T}_{\mathrm{g}})^{-1}\mathbf{T}_{\mathrm{g}}{}^{\mathrm{T}}$ is the orthogonal projector onto the column space of $\mathbf{T}_{\mathrm{g}}$, and $H_{\mathbf{T}_{\mathrm{g}}} = \mathbf{I} - \mathbf{G}_{\mathbf{T}_{\mathrm{g}}} = \mathbf{I} - \mathbf{T}_{\mathrm{g}}(\mathbf{T}_{\mathrm{g}}{}^{\mathrm{T}}\mathbf{T}_{\mathrm{g}})^{-}\mathbf{T}_{\mathrm{g}}{}^{\mathrm{T}}$ is the antiprojector with respect to the column space of $\mathbf{T}_{\mathrm{g}}$. Therefore, from another viewpoint, the two subspaces can also be obtained by projecting $\mathbf{X}_i$ onto two different projectors as $\mathbf{X}_i\,\mathbf{G}_{\mathbf{T}_{\mathrm{g}}}$ and $\mathbf{X}_i\mathbf{H}_{\mathbf{T}_{\mathrm{g}}}$. It is clear that the two predictor subspaces are orthogonal to each other as $\mathbf{X}_i^{c\mathrm{T}}\mathbf{X}_i^s = \mathbf{X}_i{}^{\mathrm{T}}\,\mathbf{G}_{\mathbf{T}_{\mathrm{g}}}\,\mathbf{X}_i = \mathbf{0}$.

Between-phase transition identification

In the between-phase common subspace, the variables enclose the same systematic predictor variations, as all of them are the linear combinations of common scores where the combination coefficients are indicated by $\mathbf{P}_i^c$ ($JN_i \times R_i^c$) in each phase. The reconstructed predictor variation $\mathbf{X}_i^c$ ($I \times JN_i$) is explained by the common scores and their linear combination relationship $\mathbf{P}_i^c$ ($JN_i \times R_i^c$). For each time-slice process data, the corresponding linear combinations is a $J \times R_i^c$ matrix ($\mathbf{P}_{i,k}^c$ ($J \times R_i^c$)) separated from $\mathbf{P}_i^c$ ($JN_i \times R_i^c$), which is composed of linear combination coefficients associated with R_i^c common scores for J process variables at each time. It is deemed that the linear combinations should be similar except with normal oscillations along time direction within the same phase. The linear combinations would be different from one phase to its neighboring one, representing different reconstruction relationships of common scores and different ways common scores act on predictor information. Therefore, by checking the changes of time-slice $P_{i,k}^c$ ($J \times R_i^c$) along time direction, the between-phase changeover can be indicated to capture the between-phase transitions.

The similarity of $P_{i,k}^c$ ($J \times R_i^c$) along time direction should be evaluated. Here, based on the phase division result, the phase center can be calculated by averaging all $P_{i,k}^c$ ($J \times R_i^c$) within the same phase

$$\mathbf{P}_i^{*c} = \frac{1}{N_i}\sum_{k\in i}\mathbf{P}_{i,k}^c \tag{4}$$

Then, the similarity between each time-slice and the phase center is calculated based on the angle between each pair of loadings

$$Simi\left(\mathbf{P}_i^{*c},\mathbf{P}_{i,k}^c\right) = \frac{1}{R_i^c}\sum_{j=1}^{R_i^c}\left(\lambda_j\frac{\left\langle\mathbf{P}_{i,j,k}^c,\mathbf{P}_{i,j}^{*c}\right\rangle}{\left\|\mathbf{P}_{i,j,k}^c\right\|\left\|\mathbf{P}_{i,j}^{*c}\right\|}\right)^2 = \frac{1}{R_i^c}\sum_{j=1}^{R_i^c}\left(\lambda_j\cos\theta_j\right)^2 \tag{5}$$

where, $\|\ \|$ is the module operator, and $\langle\ \rangle$ calculates the inner product. cos θ_{j} means the angle between the phase center and time-slice loading corresponding to each score as indicated by subscript j. The weights (λ_j) attached to each loading direction is the correlation coefficient between the extracted common scores. Clearly, the similarity index ranges within [0, 1], in which, the larger one means more significant similarity.

Another useful analysis subject to help understanding the between-phase transition is time-slice $\mathbf{X}_{i,k}^c$ ($I \times J$) which are separated from the phase-representative $\mathbf{X}_i^c$ ($I \times JN_{i+1}$). The reconstructed predictor variation information for quality prediction in the common subspace can be quantitatively evaluated at each time by

$$R^2\mathrm{X}_k = \frac{\sum_{j=1}^{J}\sum_{m=1}^{I}\hat{x}_{m,k,j}^{c\,2}}{\sum_{j=1}^{J}\sum_{m=1}^{I}x_{m,k,j}{}^2} \tag{6}$$

where, $\hat{x}_{m,k,j}^c$ is the reconstructed jth process variable for mth batch and kth samples by common scores. $x_{m,k,j}$ is the corresponding measurement.

It is regarded in general that $R^2\,\mathrm{X}_k^c$ may be similar and only with normal oscillation within the same phase. It should be noted that sometimes, $R^2\,\mathrm{X}_k^c$ may be similar for two neighboring phases, indicating the similar ratio of common predictor variation to all predictor variation. When the process evolves to the transition regions, $R^2\,\mathrm{X}_k^c$ will give the corresponding indication. The time-varying profile of $R^2\,\mathrm{X}_k^c$ can be used to help that of *Simi* index for identifying the transition patterns between two neighboring phases. Here, to more clearly reveal their changing trend, their profiles can be smoothed by Kalman filter[35,36] to remove the influence of those singular points and normal oscillations. By plotting the time-varying profile, gradually increasing or decreasing trends are distinguished from those steady statuses and identified as the transition patterns. Comparatively, the time when some successive *Simi*

values show small changes is identified as the turning point for steady phase. In this way, the steady phase and the transition region will be separated from each other.

Online quality prediction model

After the analysis of between-phase transition, the phase information can be updated by adjusting the phase boundary, called steady phase here. The time-slices in each steady phase (i) and the following transition region are denoted as $\underline{\tilde{\mathbf{X}}}_i$ ($I \times J \times \tilde{N}_i$) and $\underline{\tilde{\mathbf{X}}}_{i-t}$ ($I \times J \times \tilde{N}_{i-t}$), respectively, where $\tilde{N}_i$ is the number of time-slices in the steady phase (i) and $\tilde{N}_{i-t}$ is the number in transition region following the ith steady phase. In general, $\tilde{N}_i$ should be smaller than the original number (N_i) in each phase, as some samples are separated and classified into the transition region. All samples within the same steady phase are arranged by variable-unfolding, $\mathbf{X}_i^v$ ($I\tilde{N}_i \times J$). Correspondingly, the quality data are duplicated within the same steady phase to get a consistent observation dimension, $\mathbf{Y}_i^v$ ($I\tilde{N}_i \times J_y$), as shown in Figure 1c. PLS-CCA algorithm,[6] where CCA is used as a postprocessing on the PLS results, is used to design the steady phase-based regression model:

$$\begin{aligned} \mathbf{T}_i^v &= \mathbf{X}_i^v \mathbf{R}_i^v \\ \mathbf{Q}_i^{v\mathrm{T}} &= \left(\mathbf{T}_i^{v\mathrm{T}} \mathbf{T}_i^v\right)^{-1} \mathbf{T}_i^{v\mathrm{T}} \mathbf{Y}_i^v \\ \Theta_i^v &= \mathbf{R}_i^v \mathbf{Q}_i^{v\mathrm{T}} \end{aligned} \tag{7}$$

where the steady phase-based PLS-CCA weights model, $\mathbf{R}_i^v$ ($J \times A_i^v$) is intrinsically doubly controlled by PLS and CCA weights; A_i^v is the dimension of the PLS-CCA model, that is, the number of retained PLS-CCA scores. $\mathbf{Q}_i^v$ are the phase loadings for qualities. Θ_i^v is the regression coefficients matrix for online quality prediction.

For transition regions, to reveal their more time-varying characteristics, the time-slice PLS-CCA model will be developed and used for online quality prediction

$$\begin{aligned} \mathbf{T}_{i-t,k} &= \mathbf{X}_{i-t,k} \mathbf{R}_{i-t,k} \\ \mathbf{Q}_{i-t,k}{}^{\mathrm{T}} &= \left(\mathbf{T}_{i-t,k}^{\mathrm{T}} \mathbf{T}_{i-t,k}\right)^{-1} \mathbf{T}_{i-t,k}{}^{\mathrm{T}} \mathbf{Y} \\ \Theta_{i-t,k} &= \mathbf{R}_{i-t,k} \mathbf{Q}_{i-t,k}^{\mathrm{T}} \end{aligned} \tag{8}$$

where the terms have the similar denotation with those shown in Eq. 7.

Therefore, for the new measurement vector $\mathbf{x}_k^{\mathrm{T}}$ ($1 \times J$) at each time in each steady phase (i) or the transition region ($i-t$), a realtime quality prediction can be obtained online

$$\hat{y}_k = \mathbf{x}_k^{\mathrm{T}} \Theta_k = \begin{cases} \Theta_k = \Theta_i^v & k \in \text{the steady phase} \\ \Theta_k = \Theta_{i-t,k} & k \in \text{the transition region} \end{cases} \tag{9}$$

The online quality predictions may be time-varying within the same steady phase, which may be caused by the measurement noises and modeling errors, just to name a few. The oscillation of predicted quality variation in transition region can also be caused by the changes of prediction relationship and predictor information besides the aforementioned oscillatory factors.

Offline quality interpretation

Despite of the realtime strength, online quality prediction, however, only uses the predictor information isolated at each time. It fails in revealing the effects of cumulative process behaviors and between-phase relationship on quality interpretation. Some questions naturally arise: How do the process patterns cumulatively act on quality in each steady phase and transition region? Moreover, under the influences of between-phase relationship, will the phase behaviors play differently? On the basis of subspace separation, this would be analyzed during offline between-phase quality analysis.

First, all time-slices within the same steady phase and the following transition region share the same common scores ($\mathbf{T}_\mathrm{g}$) which is calculated by Eq. 2. The quality information explained by the common scores can be readily calculated

$$\begin{aligned} \tilde{\mathbf{Q}}_i^{c\mathrm{T}} &= \left(\mathbf{T}_\mathrm{g}^{\mathrm{T}} \mathbf{T}_\mathrm{g}\right)^{-1} \mathbf{T}_\mathrm{g}^{\mathrm{T}} \mathbf{Y} \\ \hat{\mathbf{Y}}_i^c &= \mathbf{T}_\mathrm{g} \tilde{\mathbf{Q}}_i^{c\mathrm{T}} = \mathbf{T}_\mathrm{g} \left(\mathbf{T}_\mathrm{g}^{\mathrm{T}} \mathbf{T}_\mathrm{g}\right)^{-1} \mathbf{T}_\mathrm{g}^{\mathrm{T}} \mathbf{Y} \end{aligned} \tag{10}$$

where, $\tilde{\mathbf{Q}}_i^c$ is the loading matrix for quality with respect to the common scores. $\hat{\mathbf{Y}}_i^c$ is the predicted quality variation in the common subspace.

In the phase-specific subspace, the process data contain no information associated with the common scores and, thus, reveal different and specific quality-related variation to each phase. All time-slices in the specific subspace within the same steady phase are arranged by batch-unfolding, $\tilde{\mathbf{X}}_i^s$ ($I \times \tilde{N}_i\, J$). Similarly, the time-slices in the specific subspace within the following transition region are arranged as $\tilde{\mathbf{X}}_{i-t}^s$ ($I \times \tilde{N}_{i-t}^s\, J$). The PLS–CCA decomposition[6] is performed in each specific subspace with respect to the steady phase and transition region, respectively

$$\begin{aligned} \tilde{\mathbf{T}}_i^s &= \tilde{\mathbf{X}}_i^s \tilde{\mathbf{R}}_i^s \\ \tilde{\mathbf{Q}}_i^{s\mathrm{T}} &= \left(\tilde{\mathbf{T}}_i^{s\mathrm{T}} \tilde{\mathbf{T}}_i^s\right)^{-1} \tilde{\mathbf{T}}_i^{s\mathrm{T}} \mathbf{Y} \\ \hat{\mathbf{Y}}_i^s &= \tilde{\mathbf{T}}_i^s \tilde{\mathbf{Q}}_i^{s\mathrm{T}} = \tilde{\mathbf{T}}_i^s \left(\tilde{\mathbf{T}}_i^{s\mathrm{T}} \tilde{\mathbf{T}}_i^s\right)^{-1} \tilde{\mathbf{T}}_i^{s\mathrm{T}} \mathbf{Y} \end{aligned} \tag{11}$$

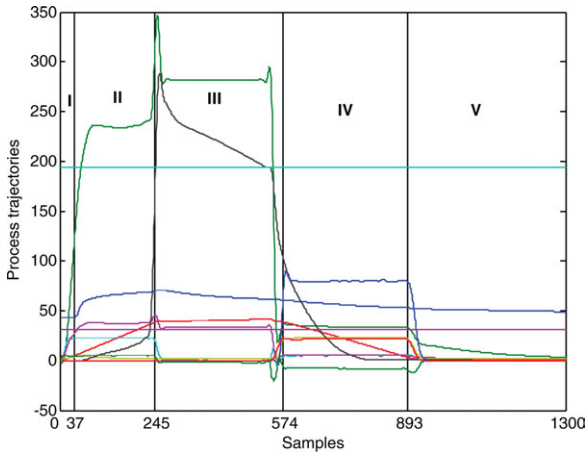

Figure 2. Phase clustering information along with process operation trajectories.

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

$$\begin{aligned}\tilde{\mathbf{T}}_{i-t}^{s} &= \tilde{\mathbf{X}}_{i-t}^{s}\mathbf{R}_{i-t}^{s}\\ \tilde{\mathbf{Q}}_{i-t}^{s\mathrm{T}} &= \left(\tilde{\mathbf{T}}_{i-t}^{s\mathrm{T}}\tilde{\mathbf{T}}_{i-t}^{s}\right)^{-1}\tilde{\mathbf{T}}_{i-t}^{s\mathrm{T}}\mathbf{Y}\\ \hat{\mathbf{Y}}_{i-t}^{s} &= \tilde{\mathbf{T}}_{i-t}^{s}\tilde{\mathbf{Q}}_{i-t}^{s\mathrm{T}} = \tilde{\mathbf{T}}_{i-t}^{s}\left(\tilde{\mathbf{T}}_{i-t}^{s\mathrm{T}}\tilde{\mathbf{T}}_{i-t}^{s}\right)^{-1}\tilde{\mathbf{T}}_{i-t}^{s\mathrm{T}}\mathbf{Y}\end{aligned} \tag{12}$$

where $\tilde{\mathbf{T}}_i^s$ ($N \times R_i^s$) is the specific regression scores in each steady phase calculated by PLS–CCA weights $\tilde{\mathbf{R}}_i^s$ ($JN_i \times R_i^s$) (R_i^s is the number of retained scores). $\tilde{\mathbf{Q}}_i^s$ ($J_y \times R_i^s$) is the loading matrix for quality by the specific scores. $\hat{\mathbf{Y}}_i^s$ ($N \times J_y$) is the predicted quality variation in the specific subspace, which is orthogonal to $\hat{\mathbf{Y}}_i^c$ in the common subspace in each steady phase, revealing that the two different types of phase predictor information contribute to different parts of quality variations. Similarly, the model structures in Eq. 12 are termed with respect to transition region. In general, $\tilde{\mathbf{T}}_i^s$ ($N \times R_i^s$) should be more different between phases, as those common scores are excluded. Moreover, it is clear that the specific scores are orthogonal to those common scores.

In summary, the quality variation can be determined by the joint $\mathbf{T}_t$, which is composed of both between-phase common and specific scores, $\mathbf{T}_t = [\mathbf{T}_g, \tilde{\mathbf{T}}_{i-t}^s]$, with respect to the steady phase and transition region

$$\begin{aligned}\mathbf{Y} &= \hat{\mathbf{Y}}_i^c + \hat{\mathbf{Y}}_i^s + \mathbf{F}_i\\ &= \mathbf{T}_g\tilde{\mathbf{Q}}_i^{c\mathrm{T}} + \tilde{\mathbf{T}}_i^s\tilde{\mathbf{Q}}_i^{s\mathrm{T}} + \mathbf{F}_i\\ &= \mathbf{T}_g\left(\mathbf{T}_g{}^{\mathrm{T}}\mathbf{T}_g\right)^{-1}\mathbf{T}_g^{\mathrm{T}}\mathbf{Y} + \tilde{\mathbf{T}}_i^s\left(\tilde{\mathbf{T}}_i^{s\mathrm{T}}\tilde{\mathbf{T}}_i^s\right)^{-1}\tilde{\mathbf{T}}_i^{s\mathrm{T}}\mathbf{Y} + \mathbf{F}_i\\ &= \left[\mathbf{T}_g, \tilde{\mathbf{T}}_i^s\right]\left(\left[\mathbf{T}_g, \tilde{\mathbf{T}}_i^s\right]^{\mathrm{T}}\left[\mathbf{T}_g, \tilde{\mathbf{T}}_i^s\right]\right)^{-1}\left[\mathbf{T}_g, \tilde{\mathbf{T}}_i^s\right]^{\mathrm{T}}\mathbf{Y} + \mathbf{F}_i\\ &= \mathbf{G}_{\mathbf{T}}\mathbf{Y} + \mathbf{F}_i\end{aligned} \tag{13}$$

$$\begin{aligned}\mathbf{Y} &= \hat{\mathbf{Y}}_{i-t}^c + \hat{\mathbf{Y}}_{i-t}^s + \mathbf{F}_i\\ &= \mathbf{T}_g\tilde{\mathbf{Q}}_{i-t}^{c\ \mathrm{T}} + \tilde{\mathbf{T}}_{i-t}^s\tilde{\mathbf{Q}}_{i-t}^{s\ \mathrm{T}} + \mathbf{F}_i\\ &= \mathbf{T}_g\left(\mathbf{T}_g^{\mathrm{T}}\mathbf{T}_g\right)^{-1}\mathbf{T}_g^{\mathrm{T}}\mathbf{Y} + \tilde{\mathbf{T}}_{i-t}^s\left(\tilde{\mathbf{T}}_{i-t}^{s\ \mathrm{T}}\tilde{\mathbf{T}}_{i-t}^s\right)^{-1}\tilde{\mathbf{T}}_{i-t}^{s\ \mathrm{T}}\mathbf{Y} + \mathbf{F}_i\\ &= \left[\mathbf{T}_g, \tilde{\mathbf{T}}_{i-t}^s\right]\left(\left[\mathbf{T}_g, \tilde{\mathbf{T}}_{i-t}^s\right]^{\mathrm{T}}\left[\mathbf{T}_g, \tilde{\mathbf{T}}_{i-t}^s\right]\right)^{-1}\left[\mathbf{T}_g, \tilde{\mathbf{T}}_{i-t}^s\right]^{\mathrm{T}}\mathbf{Y} + \mathbf{F}_i\\ &= \mathbf{G}_{\mathbf{T}_t}\mathbf{Y} + \mathbf{F}_i\end{aligned} \tag{14}$$

Here, it should be noted that the predicted quality variation by common scores are the same for steady phase and transition region. The difference of quality interpretability lies in the

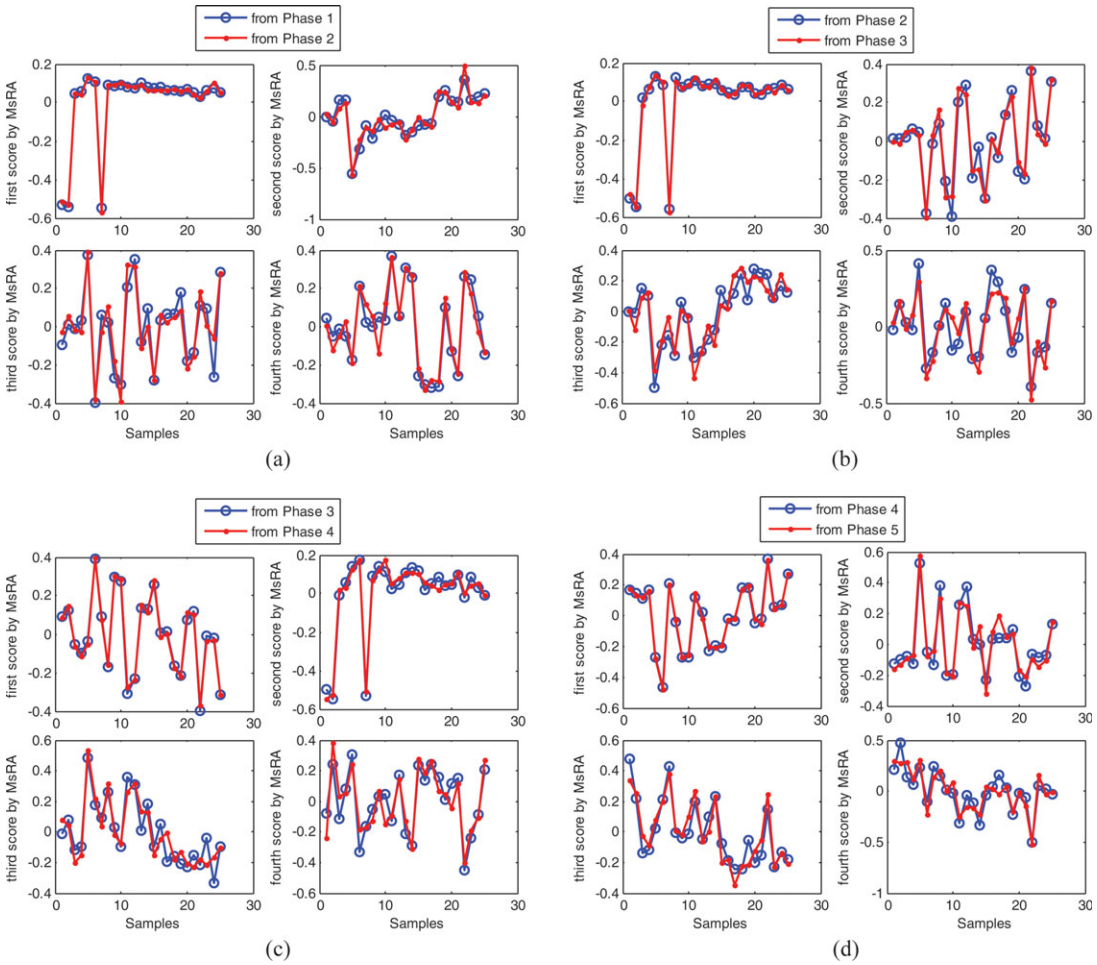

Figure 3. The profiles of the first four between-phase common scores by MsRA algorithm for (a) Phases 1 and 2, (b) Phases 2 and 3, (c) Phases 3 and 4, and (d) Phases 4 and 5.

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

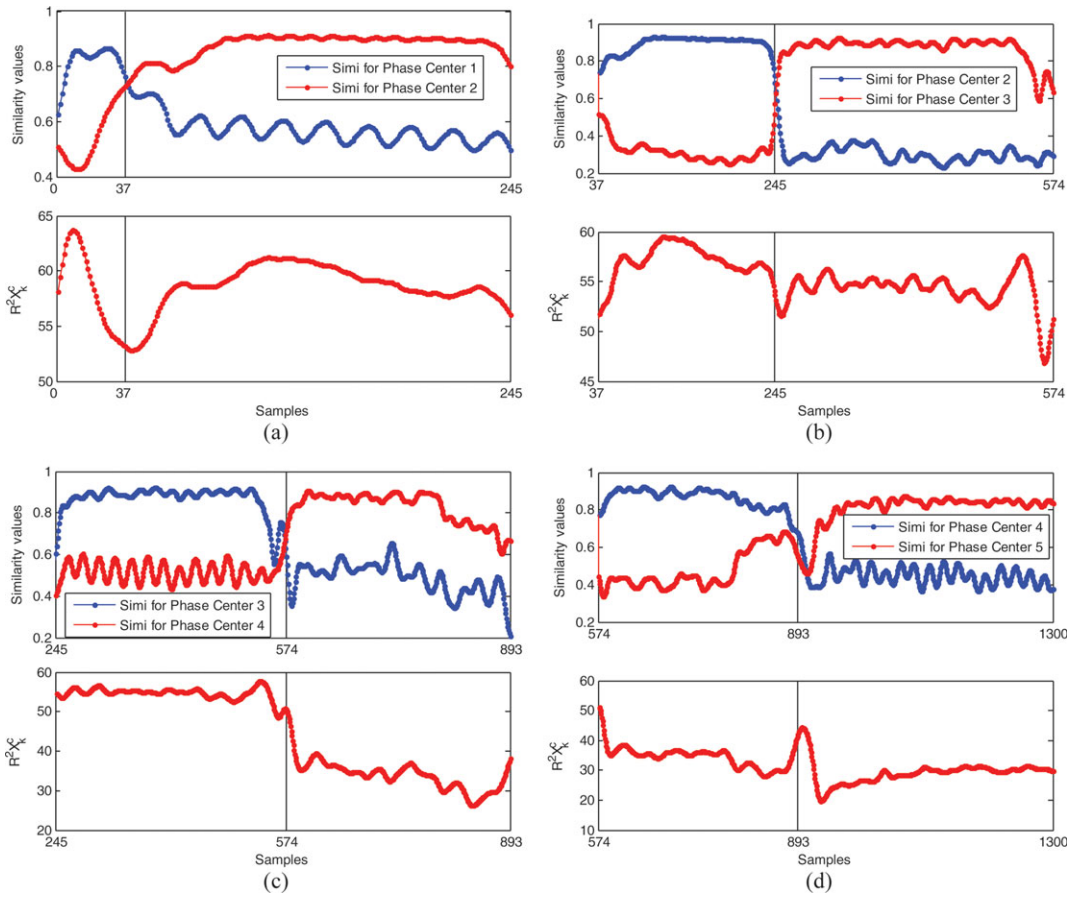

Figure 4. Changing trend of similarity values between time-slice loadings and phase centers as well as the predictor variations (R^2 X_k^c) modeled by common scores for between (a) Phases 1 and 2, (b) Phases 2 and 3, (c) Phases 3 and 4, and (d) Phases 4 and 5 analyses.

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

specific subspace. In this way, the cumulative effects on quality variation in two different subspaces are modeled from the between-phase viewpoint.

To quantitatively evaluate the quality interpretability, the goodness-of-fit of the prediction models is measured by multiple coefficient of determination[37]

$$R^2 Y_{j_y} = \frac{\sum_{m=1}^{I} \hat{y}_{m,j_y}^2}{\sum_{m=1}^{I} y_{m,j_y}^2} \tag{15}$$

where $\hat{y}_{m,j_y}$ is the quality prediction for the j_yth quality index either in steady phase or in the transition region; the subscript m indicates the batch. y_{m,j_y} is the actual measurement.

Illustrations and Discussions

Process description and data preparation

Injection molding,[38,39] a key process in polymer processing, transforms polymer materials into various shapes and types of products. A typical injection molding process consists of three operation phases, injection of molten plastic into the mold, packing-holding of the material under pressure, and cooling of the plastic in the mold until the part becomes sufficiently rigid for ejection. Besides, plastication takes place in the barrel in the early cooling phase, where polymer is melted and conveyed to the barrel front by screw rotation, preparing for next cycle. All key process conditions such as the temperatures, pressures, displacement, and velocity can be measured online by their corresponding transducers, providing abundant process information.

The material used in this work is high-density polyethylene. Twelve process variables and four quality variables are selected for modeling. The process variables can be collected online with a set of sensors. Two dimension indices, product length (mm), and weight (g), whose real values can be

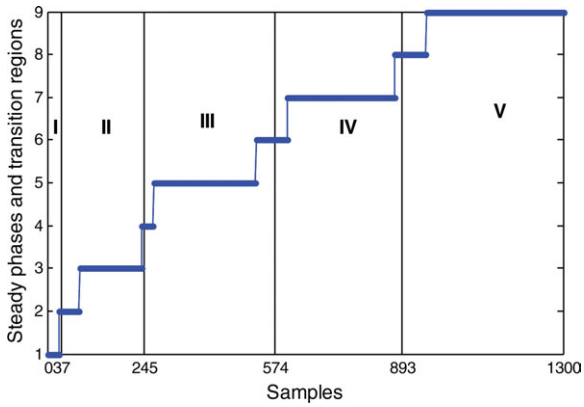

Figure 5. The steady phase and transition region information along time direction.

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

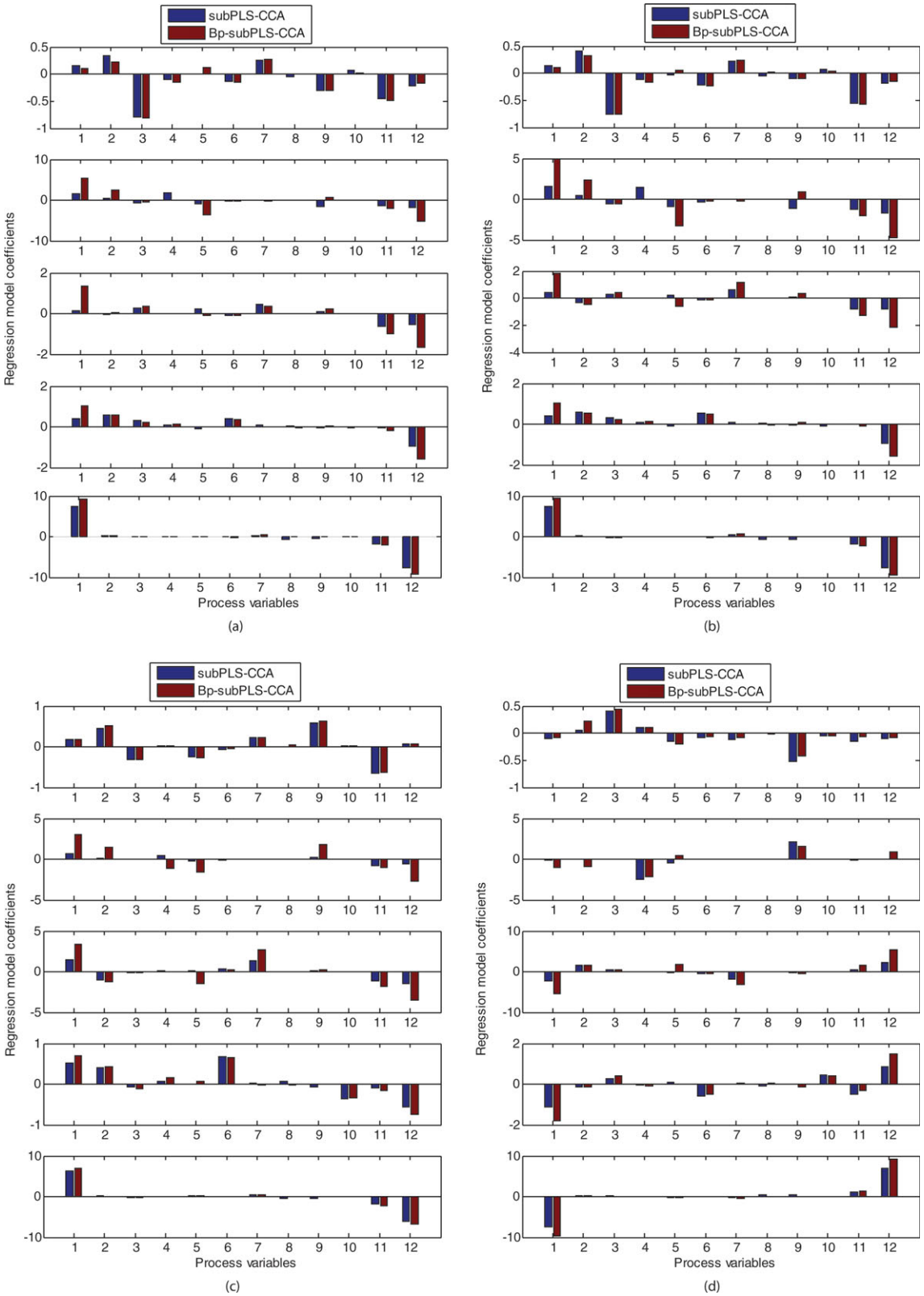

Figure 6. Regression coefficients for process variables for online quality prediction and two different algorithms with respect to (a) the first quality variable, (b) the second quality variable, (c) the third quality variable, and (d) the fourth quality variable (from top to bottom: Phases 1 to 5).

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

directly measured by instruments, and two surface defects, jetting, and record grooves, whose real values can be quantified by a process operator expert before modeling, are chosen to evaluate the product qualities. Totally 32 normal batch runs are conducted under various operation conditions by design of experiment (DOE) method.[21] Using injection

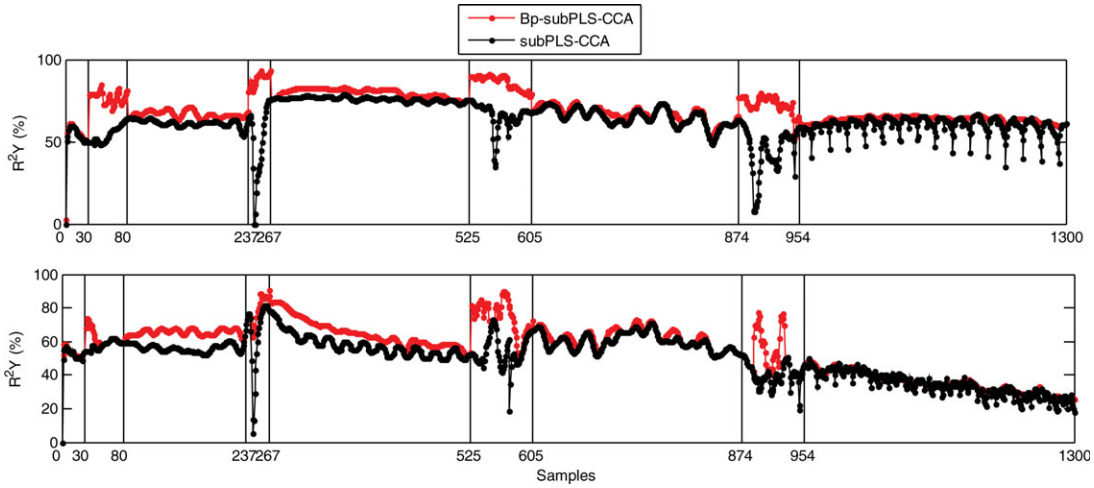

Figure 7. R^2 plot along time direction for online quality prediction with respect to four quality indices (top: training batches; bottom: testing batches).

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

stroke as indicator variable, the reference batches are unified to have even duration (1300 samples in this experiment) by data interpolation, which, thus, results in a regular descriptor array $\underline{\mathbf{X}}$ (32 × 12 × 1300). The measurement data are filtered to remove those obvious noises such as spikes. The qualities are only measured at the end of process, generating the dependent matrix $\mathbf{Y}$(32 × 4). The first 25 batches are used for modeling, whereas the other seven cycles are used for model validation.

Illustration results and analyses

First, 1300 normalized time-slices $\mathbf{X}_k$ (25 × 12) are obtained from $\underline{\mathbf{X}}$ (25 × 12 × 1300) as well as the normalized quality variables $\mathbf{Y}$(25 × 4). Then, 1300 time-slice weight matrices are obtained focusing on the data pairs, $\{\mathbf{X}_k,\mathbf{Y}\}$. They are weighted using the time-varying variances of PLS LVs and then fed to the clustering algorithm.[21] The quality-relevant operation process is partitioned into five main clusters, in which, operation time information is included so that process samples are consecutive within the same clustering. The phase clustering result is shown in Figure 2 along with the time-varying process operation trajectory. The clusters II–V are deemed to be consistent with the real four physical operation phases: injection, packing-holding, plastication, and cooling. Moreover, a short period before cluster II is also separated as one individual phase, which have different characteristics and effects on quality, as illustrated later. Thus, five predictor data blocks are prepared by batch-unfolding, $\mathbf{X}_i$ (25 × JN_i)(i = 1,2,…,5) (where N_i denotes the phase duration), which are associated with the same quality data set $\mathbf{Y}$(25 × 4).

First, the between-phase subspace separation is performed. There are four pairs of between-phase combinations with respect to five phases. The first four between-phase common scores are illustrated in Figure 3 for each between-phase pair over the process. In general, the first two common scores are almost the same between the two neighboring phases. The latter two may be different more or less but still show a similar trend. It is found that the first common score vector for Phases 1 and 2 is quite similar to that for Phases 2 and 3. It is also similar to the second common score for Phases 3 and 4 with respect to their general trend. It is not similar with anyone of those for Phases 4 and 5. This tells that with process evolvement, the phases which are far from each other are more different so the common scores are more different. Based on the extracted common scores, the loadings for each phase are calculated and phase centers are obtained. By Eq. 5, the similarity between time-slice loadings and the phase center is computed as shown in each subplot in Figure 4 for each between-phase pair. Moreover, the predictor variation explained by common scores is evaluated by R^2 X_k^c index and the values are used to help understanding the changing common predictor variation between two neighboring phases as illustrated in each subplot in Figure 4. Note that both similarity and R^2 X_k^c profiles are processed by Kalman filter so that the major changing trend can be captured more clearly by excluding those normal oscillations to a certain extent. The similarity profile with respect to the starting phase is used to confirm the beginning of transition region, which is the time when the steady similarity values begin to show a decreasing trend. The similarity values with respect to the target phase are used to confirm the ending of transition region, which is the time when the increasing similarity begins to present a stable trend. Combining the similarity evaluation results and R^2 X_k^c, steady phases and transition regions are separated as shown in the sketch map in Figure 5. Between different neighboring steady phases, the transition regions may be of different durations and of different characteristics. At the very start of process, the transition region between steady Phases 1 and 2 is longer than steady Phase 1. Actually, the steady Phase 1 may be also called transition region considering that the beginning operation may be oscillatory in each cycle. Here, it should be noted that due to the uncertainty of transition patterns, the transition regions may not be determined definitely. Actually, the analysis of transition regions can be always influenced by subjective factors.

Based on the updated phase division information, phase representative model for each steady phase and transition model for each transition time are developed using PLS-CCA algorithm and used for online quality prediction. Here, it is called between-phase subPLS-CCA (Bp-subPLS-CCA) model. To assess the influence of between-phase transitions on model development, the phase models are also developed

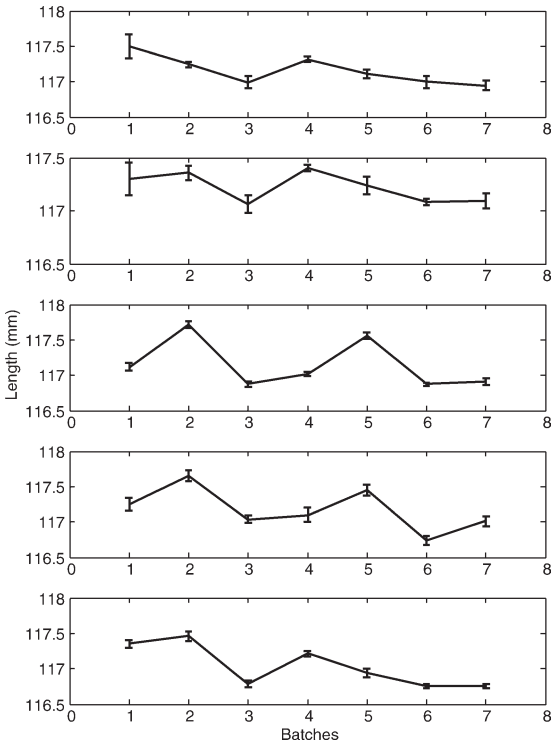

Figure 8. End-of-phase quality prediction results (mean ± STD) for testing batches using the proposed method in different steady phases (from top to bottom: steady Phases 1–5).

based on the original phase division results and PLS-CCA algorithm. That is, transition patterns are not distinguished from steady phase. Instead, they are included in some phases for model development. Here, it is called subPLS-CCA model. The phase representative regression models for the two different methods are compared in Figures 6a–d for four quality indices, respectively. In general, for all quality variables, the model coefficients are more different in Phases 2 and 3. It can be seen clearly that the exclusion of transition patterns from steady phases will result in different phase models more or less. To evaluate whether the exclusion of transition patterns from the steady phase can improve prediction performance, R^2Y values with respect to all quality indices are compared between the two methods and shown in Figure 7. They are plotted along time direction through steady phases and between-phase transition regions over the whole process with respect to both training batches and testing batches. In general, the quality predictions are similar with small normal oscillations in the steady phase and more oscillatory in transition regions. It can be seen that when transition patterns are removed from steady phase, the representability of phase models can be improved, generating more stable and better quality prediction, especially for steady Phases 2 and 3. In transition regions, subPLS-CCA method uses some phase model and bad predictions are observed. Comparatively, the prediction performance by Bp-subPLS-CCA is significantly improved. At some transition time, the prediction performance is better than that based on the neighboring steady phase models.

During online application, the end-of-phase quality values are calculated by averaging all real-time predicted values within the steady phase. Moreover, to reveal the variability or diversity of quality predictions in steady phases or transition regions, the mean and standard deviation (STD) indices are plotted in Figure 8 taking the product length (i.e., the second quality variable) for instance, indicating how much variation or dispersion there is from the average prediction, that is, the end-of-phase value. A low STD indicates that all predictions within the same steady phase tend to be very close to the mean, whereas high STD indicates that the predictions are spread out over a large range of values. For each test batch, in different steady phases, the variability is different more or less. In general, steady Phases 3 and 5 generate the least STD values for all batches, indicating the least variability along time direction.

By performing between-phase modeling, the underlying predictor information is decomposed into two parts. Actually, the underlying predictor information in each phase can be forcefully and completely extracted as ''common'' scores one by one and ordered by descending quality-related between-phase relationship. In this way, no between-phase specific model is developed. However, the latter ''common'' scores may be quite different from each other and are actually pseudo ''common'' scores. To illustrate this, focusing on the original phase division, each phase is modeled only by ''common'' scores. This results in a different regression model (Bpc-MPLS-CCA) in each phase from the conventional calibration methods. It is compared with P-MPLS-CCA model where each single phase is separately modeled by MPLS-CCA. Here, no transition regions are considered. The prediction results are comparatively shown in Table 1 for each quality index. In general, they are similar, revealing that the predictor information extracted only by between-phase ''common'' scores is comparable to that by conventional single phase algorithms so that they make the similar contributions to quality. It should be noted that when one phase is associated with different neighboring phases, different between-phase models are identified, which, however, shows the similar prediction results for the same phase.

Table 1. Offline Prediction Comparison [R^2Y (%)] for Bpc-PLS-CCA Model and P-MPLS-CCA Model

		Phase No.							
Quality index	Method	1 (2)*	2 (1)	2 (3)	3 (2)	3 (4)	4 (3)	4 (5)	5 (4)
1	Bpc-PLS-CCA	35.9	71.2	72.5	89.8	87.8	88.9	89.4	58.9
	P-MPLS-CCA	36.1	70.4		88.3		88.9		59.8
2	Bpc-PLS-CCA	30.6	78.8	72.8	89.9	89.9	93.8	93.7	54.4
	P-MPLS-CCA	30.1	76.8		90.4		93.8		53.0
3	Bpc-PLS-CCA	77.8	56.0	54.8	31.6	28.3	31.9	30.7	32.3
	P-MPLS-CCA	68.6	52.7		22.6		30.7		26.2
4	Bpc-PLS-CCA	99.9	98.8	99.6	98.1	96.9	48.6	48.1	38.5
	P-MPLS-CCA	99.9	99.1		97.1		49.0		31.7

*Each phase is associated with one neighboring phase as indicated by the value in bracket.

Table 2. Online and Offline Quality Prediction Results (R^2Y (%)) in Steady Phases (SP) and Transition Regions

	Region								
Application	SP 1	transition	SP 2	Transition	SP 3	transition	SP 4	transition	SP 5
Offline	74.9	68.6	81.8	82.7	79.9	74.8	79.6	54.9	54.6
Online	51.6	61.7	65.8	79.0	65.1	73.9	62.0	50.6	36.0

Moreover, for different quality indices, different phases are significant for prediction. For example, for the fourth quality index, record grooves, the first three phases (including injection and packing-holding physical operation phases) are more important, which well agrees with the real case.

Then, based on the updated phase information, the offline quality prediction models are developed for steady phases and transition regions with respect to the between-phase subspace separation. To reveal the influence of cumulative effects in each time region (steady phase and transition region) on quality prediction, the online predictions and offline predictions based on between-phase calibration modeling are compared in Table 2. Different from the Bpc-MPLS-CCA in Table 1, here, the predictor and quality are interpreted jointly by common and specific scores for offline quality analysis. The quality prediction for each steady phase is the sum of prediction values in common and specific subspaces, as the common scores and specific scores are orthogonal, revealing complementary quality predictions. Moreover, when each steady phase is associated with different neighboring steady phases, different common scores may be extracted, resulting in different prediction models and performance. The better quality prediction results are chosen and show in Table 2. Compared with end-of-phase quality prediction for real-time application, the offline results are better based on paired t test ($\alpha = 0.5$),[37] revealing that the consideration of cumulative effects can improve the prediction performance. The offline prediction results in transition regions are compared with real-time quality prediction in the same region. When the changing prediction performance in transition region is cumulatively explained by offline analysis, the prediction improvement is also statistically significant in comparison with online quality predictions based on paired t test ($\alpha = 0.5$).

Conclusions

In this article, between-phase transition is analyzed for calibration modeling and quality prediction in batch processes. The separation of between-phase common subspace and specific subspace is the key factor, which provides a good analysis platform for understanding quality-related phase behaviors. Based on this, the transition regions between two neighboring phases can be identified and distinguished from steady phases. Steady-phase models and transition models are thus developed respectively and used for online quality prediction, revealing the changing of prediction power along time direction. Besides the realtime application, the cumulative effects on quality prediction can be jointly captured in two subspaces. The case study concludes with illustrations of how the proposed method performs on one real case, revealing the desirable improvement in quality interpretation.

The proposed method gives an interesting insight into the effects of phase behaviors and between-phase transitions on quality prediction and interpretation. There may be still many specific issues to be investigated, but the promising results reported in the article should encourage further development of this research topic and constitute an important step forward. The other applications, for example, process monitoring of different process variations, could be an interesting but different topic, which should motivate a worthwhile extension of this research line.

Acknowledgments

This work is supported by National Program on Key Basic Research Project (973 Program), under grant 2012CB720505.

Literature Cited

1. Martens H, Naes T. *Multivariate Calibration*, 2nd ed. Chichester: Wiley, 1994.
2. Burnham AJ, Viveros R, MacGregor JF. Frameworks for latent variable multivariate regression. *J Chemometr*. 1996;10:31–45.
3. Doyal BS, MacGregor JF. Improved PLS algorithms. *J Chemometr*. 1997;11:73–85.
4. Cserhati T, Kosa A, Balogh S. Comparison of partial least-square method and canonical correlation analysis in a quantitative structure-retention relationship study. *J Biochem Biophys Methods*. 1998;36:131–141.
5. Brereton RG. Introduction to multivariate calibration in analytical chemistry. *Analyst*. 2000;125:2125–2154.
6. Yu HL, MacGregor JF. Post processing methods (PLS-CCA): Simple alternatives to preprocessing methods (OSC-PLS). *Chemometr Intell Lab Syst*. 2004;73:199–205.
7. Kleinbaum DG, Kupper LL, Muller KE, Nizam A. *Applied Regression Analysis and Other Multivariable Methods*, 3rd ed., California: Wadsworth Publishing, 2003.
8. Hardoon DR, Szedmak S, Taylor JS. Canonical correlation analysis: An overview with application to learning methods. *Neural Comput*. 2004;16:2639–2664.
9. Zhao CH, Gao FR, Wang FL. An improved independent component regression modeling and quantitative calibration procedure. *AICHE J*. 2010;56:1519–1535.
10. Ergon R. Reduced PCR/PLSR models by subspace projections. *Chemomet Intell Lab Syst*. 2006;81:68–73.
11. Yamamoto H, Yamaji H, Fukusaki E, Ohno H, Fukuda H. Canonical correlation analysis for multivariate regression and its application to metabolic fingerprinting. *Biochem Eng J*. 2008;40:199–204.
12. Nomikos P, MacGregor JF. Multi-way partial least squares in monitoring batch processes. *Chemometr Intell Lab Syst*. 1995;30:97–108.
13. Westerhuis JA, Kourti T, MacGregor JF. Analysis of multiblock and hierarchical PCA and PLS models. *J Chemometr*. 1998;12:301–321.
14. Berglund A, Wold S. A serial extension of multiblock PLS. *J Chemometr*. 1999;13:461–471.
15. Qin SJ, Valle S, Piovoso MJ. On unifying multiblock analysis with application to decentralized process monitoring. *J Chemometr*. 2001;15:715–742.
16. Lopes JA, Menezes JC, Westerhuis JA, Smilde AK. Multiblock PLS analysis of an industrial pharmaceutical process. *Biotechnol Bioeng*. 2002;80:419–427.
17. Tenenhaus M, Vinzi VE. PLS regression, PLS path modeling and generalized procrustean analysis: a combined approach for multiblock analysis. *J Chemometr*. 2005;19:145–153.
18. Hoskuldsson A, Svinning K. Modelling of multi-block data. *J Chemometr*. 2006;20:376–385.
19. Reinikainen S, Hoskuldsson A. Multivariate statistical analysis of a multi-step industrial processes. *Anal Chim Acta*. 2007;595:248–256.

20. Zhao CH, Wang FL, Gao FR. Improved calibration investigation using phase-wise local and cumulative quality interpretation and prediction. *Chemometr Intell Lab Syst*. 2009;95:107–121.
21. Lu NY, Gao FR. Stage-based process analysis and quality prediction for batch processes. *Ind Eng Chem Res*. 2005;44:3547–3555.
22. Lu NY, Gao FR. Stage-based online quality control for batch processes. *Ind Eng Chem Res*. 2006;45:2272–2280.
23. Camacho J, Pico J. Multi-phase analysis framework for handling batch processes data. *J Chemometr*. 2008;22:632–643.
24. Zhao CH, Wang FL, Mao ZZ, Lu NY, Jia MX. Improved batch process monitoring and quality prediction based on multiphase statistical analysis. *Ind Eng Chem Res*. 2008;47:835–849.
25. Yao Y, Gao FR. A survey on multistage/multiphase statistical modeling methods for batch processes. *Annu Rev Control*. 2009;33:172–183.
26. Zhao CH, Gao FR. A Two-step multiset regression analysis (MsRA) algorithm. *Ind Eng Chem Res*. 2012;51:1337–1354.
27. Zhao CH, Wang FL, Lu NY, Jia MX. Stage-based soft-transition multiple PCA modeling and on-line monitoring strategy for batch processes. *J Process Control*. 2007;17:728–741.
28. Yao Y, Gao FR. Phase and transition based batch process modeling and online monitoring. *J. Process Control*. 2009;19:816–826.
29. Zhao CH, Gao FR. Between-phase based Statistical analysis and modeling for transition monitoring in multiphase batch processes, *AICHE J*. DOI: 10.1002/aic.12783.
30. Undey C, Cinar A. Statistical monitoring of multistage, multiphase batch processes. *IEEE Control Syst Mag*. 2002;22:40–52.
31. Undey C, Ertunc S, Cinar A. Online batch/fed-batch process performance monitoring, quality prediction, and variable-contribution analysis for diagnosis. *Ind Eng Chem Res*. 2003;42:4645–4658.
32. Camacho J, Pic J. Multi-phase principal component analysis for batch processes modelling. *Chemometr Intell Lab Syst*. 2006;81:127–136.
33. Doan XT, Srinivasan R, Bapat PM, Wangikar PP. Detection of phase shifts in batch fermentation via statistical analysis of the online measurements: a case study with rifamycin B fermentation. *J Biotechnol*. 2007;132:156–166.
34. Maiti SK, Srivastava RK, Bhushan M, Wangikar PP. Real time phase detection based online monitoring of batch fermentation processes. *Process Biochem*. 2009;44:799–811.
35. Grewal MS, Angus PA. *Kalman Filtering Theory and Practice*. Upper Saddle River, NJ USA: Prentice Hall, 1993.
36. Anderson BDO, Moore JB. *Optimal Filtering*. New York: Dover, 2005.
37. Montgomery DC, Runger GC. *Applied Statistics and Probability for Engineers*, 4th ed. New York: Wiley, 2006.
38. Yang Y, Gao FR. Cycle-to-cycle and within-cycle adaptive control of nozzle pressures during packing-holding for thermoplastic injection molding. *Polym Eng Sci*. 1999;39:2042–2064.
39. Yang Y, Gao FR. Adaptive control of the filling velocity of thermoplastics injection molding. *Control Eng. Practice*. 2000;8:1285–1296.

Manuscript received Oct. 10, 2011, and revision received Jan. 18, 2012.