Preprints of the 10th IFAC International Symposium on Dynamics and Control of Process Systems
The International Federation of Automatic Control
December 18-20, 2013. Mumbai, India

Quality-related inner-phase evolution analysis and quality prediction for uneven batch processes

Luping Zhao[1], Chunhui Zhao[*2], Furong Gao[*1,2,3]

[1] Department of Chemical and Biomolecular Engineering,
The Hong Kong University of Science and Technology, Hong Kong, China (email: kelpzhao@ust.hk)
[2] State Key Laboratory of Industrial Control Technology, Department of Control Science and Engineering,
Zhejiang University, Hangzhou, 310027, China (corresponding author, email: chhzhao@zju.edu.cn)
[3] Fok Ying Tung Graduate School, The Hong Kong University of Science and Technology,
Hong Kong, China (corresponding author, email: kefgao@ust.hk)

Abstract: In this paper, a new statistical process analysis and quality prediction method is proposed for multiphase batch processes. A two-level phase division algorithm is designed to capture and trace quality-related inner-phase evolution which in general goes through three statuses sequentially, i.e., transition, steady-phase and transition. Partial least squares (PLS), canonical correlation analysis (CCA) and qualitative trend analysis (QTA) are effectively combined to distinguish different inner-phase process statuses. Their different characteristics are then analyzed respectively for regression modeling and quality analysis. Meanwhile, the uneven-length problem of batch processes is handled in different inner-phase parts so that online quality prediction can be performed at each time. The application to the injection molding process illustrates the feasibility and performance of the proposed algorithm.

Keywords: Multiphase batch processes, inner-phase evolution, quality prediction, uneven batches.

1. INTRODUCTION

As an important type of industrial production, batch processes have been widely applied to fine chemical, biopharmaceutical, food, polymer industries, and metallurgy, to obtain high-value-added products efficiently. The batch process safety and consistent product quality have become a focus of research. Data-based statistical analysis techniques, such as multiway principal component analysis (MPCA) and multiway partial least squares (MPLS) are the popular tools for batch process monitoring and quality prediction (Nomikos et al. (1994, 1995)), after which, different solutions were proposed (Wold et al. (1996), Westerhuis et al. (1998)).

In general, multiple operation steps are included in each batch cycle, resulting in different process segments, called phases. A series of phases comprise a whole batch cycle and each phase has its own characteristic, which requires special attention for multiphase batch processes. Zhao and Lu et al. (2004, 2008, 2013) proposed phase-based PCA/PLS methods, recognizing that phases can reflect the changes of the inherent process correlations. Considering the transition problem between neighboring phases, a soft-transition multiple PCA (STMPCA) method was proposed to detect and model transitions for online process monitoring (Zhao et al. (2007)). To handle the uneven-length problem which widely exists in batch processes, Zhao et al. (2011) proposed an uneven-length batch clustering based modeling algorithm for phase division and process monitoring.

However, these phase-based methods ignore the process trend reflected by inner-phase variations, which may lose important information about process operation. Recently, Zhao et al. (2013) investigated the inner-phase evolution of batch processes by dividing a phase into several parts and different statistical models were developed in different parts for process monitoring. In general, three sequential statuses, i.e., transition, steady part and transition, are the basic structure to describe the process variation within a phase, called inner-phase evolution. In this paper, quality-related inner-phase evolution is investigated by analyzing the changes of process-quality relationships within a phase. Statistical regression modeling and quality prediction on the basis of quality-related inner-phase evolution analysis is addressed for multiphase batch processes.

PLS algorithm has been widely used to approximate the regression relationship between $\mathbf{X}$ and $\mathbf{Y}$ (Wold et al. (1996), Westerhuis et al. (1998)). However, its objective is to maximize their covariance, which may not necessarily mean strong correlation. When the $\mathbf{X}$ space contains large amount of quality-uninformative process variations, PLS often requires many latent variables (LVs) to achieve good fitting. Aalternatively, canonical correlation analysis (CCA) (Cserhati et al. (1998)) is well-suited for relating two data tables. However, since the measurement variables are often high-dimensional and closely correlated, directly applying CCA to the raw input space will lead to an ill-conditioned problem. Yu et al. (2004) developed a PLS-CCA algorithm, in which CCA was implemented on PLS LVs to further condense them. On the other hand, process trend analysis is a useful approach to exploit the temporal information and reason about process state. Since 1980s, qualitative trend analysis (QTA) which is also widely known as dynamic trend analysis has been developed and played an important role in process monitoring and fault analysis (Janusz et al. (1991), Dash et al. (2004)). Therefore, PLS-CCA and QTA methods are effectively combined to trace the quality-related inner-phase evolutions of batch processes. PLS-CCA is used as basic statistical regression analysis tool to obtain statistical information from process data and quality data. Then QTA is used on statistics obtained by PLS-CCA to capture the quality-related inner-phase evolutions. Different modeling strategies are proposed for quality prediction. For online application, quality-related inner-phase evolution is well traced where the affiliation of the current sample point is judged and its corresponding regression model is adopted for quality prediction.

The rest of this paper includes four parts: first, the basic algorithms of PLS-CCA and QTA are briefly revisited in Section 2. Then the proposed method is presented in Section 3, including the description of two-level phase division, modeling in different inner-phase parts and online quality prediction. In Section 4, the application to an injection molding process is presented. At last, the conclusion is drawn.

2. PRELIMINARY

2.1 PLS-CCA algorithm

PLS is a common LV-based regression method. The LVs are linear combinations of the predictor variables that result in maximal covariance with the output variable. The equations for PLS are shown as below,

$$\mathbf{X} = \mathbf{T}_{PLS}\mathbf{P}_{PLS}^{\mathrm{T}} + \mathbf{E}_{PLS} \tag{1}$$

$$\mathbf{Y} = \mathbf{U}_{PLS}\mathbf{Q}_{PLS}^{\mathrm{T}} + \mathbf{F}_{PLS} \tag{2}$$

where $\mathbf{X}$ denotes the predictor data matrix, $\mathbf{Y}$ denotes the output variable data, $\mathbf{T}_{PLS}$ and $\mathbf{U}_{PLS}$ are the score matrices, $\mathbf{P}_{PLS}$ and $\mathbf{Q}_{PLS}$ are the loading matrices, $\mathbf{E}_{PLS}$ and $\mathbf{F}_{PLS}$ are the residual matrices.

Then, CCA is implemented on PLS LVs (Yu et al. (2004)) to further condense them as below,

$$\mathbf{T}_{PLS} = \mathbf{T}_{CCA}\mathbf{P}_{CCA}^{\mathrm{T}} + \mathbf{E}_{CCA} \tag{3}$$

$$\mathbf{Y} = \mathbf{U}_{CCA}\mathbf{Q}_{CCA}^{\mathrm{T}} + \mathbf{F}_{CCA} \tag{4}$$

where $\mathbf{T}_{CCA}$ and $\mathbf{U}_{CCA}$ are the score matrices, $\mathbf{P}_{CCA}$ and $\mathbf{Q}_{CCA}$ are the loading matrices, $\mathbf{E}_{CCA}$ and $\mathbf{F}_{CCA}$ are the residual matrices.

2.2 QTA algorithm

A novel approach proposed to automatically identify the qualitative shapes using a polynomial-fit based interval-halving technique (Dash et al. (2004)) to capture process trends is used in the present work for inner-phase evolution analysis. The fundamental language of QTA is the primitives defined by the first and second derivatives of variables. A trend is represented as a sequence of these seven primitives. The procedure identifies the qualitative trend as a sequence of piecewise unimodals or quadratic segments. The least-order (among constant, first-order and quadratic) polynomial with fit-error statistically insignificant compared to noises (as dictated by F-test) is used to represent the segment. If the fit-error is large even for the quadratic polynomial, then the length is halved and the process is repeated on the first half until fit-error is acceptable. A constrained polynomial fit is used to ensure the continuity of the fitted data and an outlier detection methodology is used to detect any jump changes in the signal. The procedure is recursively applied to the remaining data until the entire data record is covered.

3. METHODOLOGY

3.1 Two-level phase division

To reveal time-varying underlying process characteristics, process division is extended to two levels: conventional phase division and inner-phase division. Instead of tracking the changes of general process characteristics for process monitoring (Zhao et al. (2013)), two-level phase division is performed by focusing on the quality-related process characteristics for the specific purpose of quality prediction.

In phase division, the whole process is divided into multiple phases (C) by indicator variables. And it is assumed that batches belonging to the same operation mode have identical phase length. Those uneven-length batches are first clustered into different groups (G) as indicated by batch lengths.

After the first-level phase division, process data and quality data of the cth phases in the gth group are saved as a three-dimensional matrix $\underline{\mathbf{X}}_{c,g}(I_g \times J_x \times K_{c,g})$ and $\underline{\mathbf{Y}}_g(I_g \times J_y)$, respectively, where I_g, J_x, J_y and $K_{c,g}$ refer to the number of batches, process variables, quality variables and time duration within the cth phases in the gth group. Since only the final quality is available for the batch process investigated, and all process data contribute to the final quality, different phases correspond to the same quality data, $\underline{\mathbf{Y}}_g$. After variable-wise unfolding, the two-dimensional data matrix $\underset{\sim}{\mathbf{X}}_{c,g}(K_{c,g}I_g \times J_x)$ from different groups (where $g = 1,2,...,G$) are put together to obtain a two-dimension phase-representative data matrix $\underset{\sim}{\mathbf{X}}_c(\sum_{g=1}^{G} K_{c,g}I_g \times J_x)$, which are then normalized, denoted as $\mathbf{X}_c$. In this way, the normalized data can keep the process variation information of each group within each phase. Simultaneously, $\underline{\mathbf{Y}}_{c,g}(I_g \times J_y)$ are first repeated $K_{c,g}$ times to get $\underset{\sim}{\mathbf{Y}}_{c,g}(K_{c,g}I_g \times J_y)$, and then $\underset{\sim}{\mathbf{Y}}_{c,g}$ from different groups are put together to obtain $\underset{\sim}{\mathbf{Y}}_c(\sum_{g=1}^{G} K_{c,g}I_g \times J_y)$, which are normalized and denoted as $\mathbf{Y}_c$. They are prepared for quality-related inner-phase evolution extraction in the next step.

In the second-level phase division, i.e., inner-phase division, phases will be divided into different inner-phase parts according to the changes of quality-related process characteristics. In general, a typical phase can be further divided to three parts: initial transition, steady part and terminal transition.

The inner-phase division is conducted by applying the combination of PLS-CCA and QTA on phase-representative data. First, apply PLS-CCA on $\mathbf{X}_c$ and $\mathbf{Y}_c$,

$$\mathbf{X}_c = \mathbf{T}_{c,PLS}\mathbf{P}_{c,PLS}^{\mathrm{T}} + \mathbf{E}_{c,PLS} \tag{5}$$

$$\mathbf{Y}_c = \mathbf{U}_{c,PLS}\mathbf{Q}_{c,PLS}^{\mathrm{T}} + \mathbf{F}_{c,PLS} \tag{6}$$

$$\mathbf{T}_{c,PLS} = \mathbf{T}_{c,CCA}\mathbf{P}_{c,CCA}^{\mathrm{T}} + \mathbf{E}_{c,CCA} \tag{7}$$

$$\mathbf{Y}_c = \mathbf{U}_{c,CCA}\mathbf{Q}_{c,CCA}^{\mathrm{T}} + \mathbf{F}_{c,CCA} \tag{8}$$

$$\mathbf{P}_{c,PLS-CCA} = \mathbf{P}_{c,PLS}\mathbf{P}_{c,CCA} \tag{9}$$

Then by projecting the time intervals $\mathbf{X}_{c,g,k}(I_g \times J_x)$ (where $k = 1,2,...,K_{cg}$) within the cth phases in the gth group onto $\mathbf{P}_{c,PLS-CCA}$, the time-slice PCs are obtained:

$$\mathbf{T}_{c,g,k,PLS-CCA} = \mathbf{X}_{c,g,k}\mathbf{P}_{c,PLS-CCA} \tag{10}$$

where $\mathbf{T}_{c,g,k,PLS-CCA}$ covers the systematic variation information relative to quality $\mathbf{Y}_c$.

The average scores over all batches for the jth quality variable at the same time within the cth phase in the gth group is defined as below,

$$t_{c,g,k,j} = \frac{1}{I_g}\sum_{i=1}^{I_g} t_{c,g,k,i,j} \tag{11}$$

where $t_{c,g,k,i,j}$ is the ith row and jth column of $\mathbf{T}_{c,g,k,PLS-CCA}$, and $t_{c,g,k,j}$ shows the average quality-related variation information relative to the jth quality of the kth time-slice. Then, $t_{c,g,k,j}$ at different time intervals within the same phase ($k = 1,2,...,K_{cg}$) comprise a vector $\mathbf{t}_{c,g,j} = [t_{c,g,1,j},...,t_{c,g,k,j},...,t_{c,g,K_{cg},j}]$, and its gradient, which shows the jth quality-related time-varying evolution of the LV, is denoted as $\Delta\mathbf{t}_{c,g,j} = [\Delta t_{c,g,1,j},...,\Delta t_{c,g,k,j},...,\Delta t_{c,g,K_{cg},j}]$,

$$\Delta t_{c,g,k,j} = \begin{cases} 0, k = 1 \\ t_{c,g,k,j} - t_{c,g,k-1,j}, k = 2,...,K_{c,g} \end{cases} \tag{12}$$

where $j = 1,2,...,J_y$.

The interval-halving algorithm (Dash et al. (2004)) for trend extraction is applied to $\Delta\mathbf{t}_{c,g,j}$. Thus, for each phase, multiple segments are separated using seven primitives of QTA. To judge which part these segments belong to, the deviations of these segments are calculated as:

$$\dot{D}_{c,g,s} = \frac{D_{c,g,s}}{L_{c,g,s}} = \frac{\Delta t_{c,g,k_{end,s},j} - \Delta t_{1,c,g,k_{st,s},j}}{k_{end,c,g,s} - k_{st,c,g,s} + 1} \tag{13}$$

where $\Delta t_{c,g,k_{st,s},j}$ and $\Delta t_{c,g,k_{end,s},j}$ are the score gradient value at the beginning and the end of the sth segment and $D_{c,g,s}$ denotes the difference between them; $\dot{D}_{c,g,s}$ is deviation of the sth segment; $L_{c,g,s}$ is the duration of this segment; $k_{st,c,g,s}$ and $k_{end,c,g,s}$ are the time indices corresponding to the beginning and the end of this segment, respectively. A threshold $\dot{D}_{c,g}^*$ should be defined based on training data so that $\dot{D}_{c,g,s}$ of all segments within the cth phase in the gth group can be divided into two clusters corresponding to steady part and transitions respectively. Here the threshold is defined using the two indices, median (MED) and median absolute deviation (MAD). For each phase, the MED value is dominated by the deviations in steady part and deviations of the segments in the steady part are near MED. The MAD index is also dominated by the segments in steady part since it is the middle value of differences between all deviations and MED. The threshold $\dot{D}_{c,g}^*$ can be defined as MED±αMAD, where α is a constant attached to MAD, termed relaxing factor here. If the deviation $\dot{D}_{c,g,s}$ is smaller than the threshold $\dot{D}_{c,g}^*$, this segment is assigned to steady part, otherwise, it belongs to transitions.

To make it proper for online application, the segments within steady part are further analyzed here. The vector $\Delta\mathbf{t}_{c,g,j}$ within steady part obtained by QTA is denoted as $\Delta\mathbf{t}_{c,g,sp,j}$ revealing the variations between neighbouring time intervals. Then $\Delta\mathbf{t}_{c,sp,j}$ across all G groups are deemed to be normally distributed, from which, 99% confidence region can be readily obtained. For each time interval, compare $\Delta t_{c,k,j}$ with the predefined 99% confidence region. If the time interval shows $\Delta t_{c,k,j}$ beyond the region, it is assigned to steady part; otherwise it is assigned to transitions.

3.2 Regression modeling for different parts

Different statistical regression models should be developed for steady parts and transition parts. A common model for steady part is established based on variable-wise unfolding data in steady part, while time-slice regression models are built for transitions. The details are introduced below.

(1) Model development for a steady part

The steady-part data of each uneven group comprise $\underset{\sim}{\mathbf{X}}_{c,sp}(\sum_{g=1}^{G} K_{c,g,sp}I_g \times J_x)$. They are then normalized, denoted as $\mathbf{X}_{c,sp}$. Corresponding quality data is denoted as $\mathbf{Y}_{c,sp}(\sum_{g=1}^{G} K_{c,g,sp}I_g \times J_y)$. Then, build PLS model on $\mathbf{X}_{c,sp}$ and $\mathbf{Y}_{c,sp}$. When only single quality $\mathbf{y}_{c,sp}(\sum_{g=1}^{G} K_{c,g,sp}I_g \times 1)$ is considered, the regression model is

$$\hat{\mathbf{y}}_{c,sp} = \mathbf{X}_{c,sp}\boldsymbol{\beta}_{c,sp,PLS-CCA} \tag{14}$$

where $\boldsymbol{\beta}_{c,sp,PLS-CCA}$ is the phase-representative regression parameter, $k = 1,2,...,K_{c,g,sp}$.

(2) Model development for a transition part

After the inner-phase division, the transitions from different uneven groups are synchronized by curve fitting method. Time-slice data are obtained by putting time-slice data from different groups together, denoted as $\mathbf{X}_{c,tr,k}(\sum_{g=1}^{G} I_g \times J_x)$ after normalization. Corresponding single quality data is denoted as $\mathbf{y}_{c,tr,k}(\sum_{g=1}^{G} I_g \times 1)$. Then, perform PLS-CCA model on $\mathbf{X}_{c,tr,k}$ and $\mathbf{y}_{c,tr,k}$ to get the time slice regression model,

$$\hat{\mathbf{y}}_{c,tr,k} = \mathbf{X}_{c,tr,k}\boldsymbol{\beta}_{c,tr,k,PLS-CCA} \tag{15}$$

where $\boldsymbol{\beta}_{c,tr,k,PLS-CCA}$ is the time-slice regression parameter, $k = 1,2,...,K_{c,tr}$.

3.3 Online quality prediction

(1) Online identification of inner-phase parts

First, the current phase is judged by indicator variables. For the current phase *c*, new observation is denoted as $\mathbf{x}_{c,new,k}$, which is normalized by the mean and standard deviation calculated from training data in inner-phase division. $\mathbf{x}_{c,new,k}$ is projected onto the subspace spanned by the loading matrix $\mathbf{P}_{c,PLS-CCA}$,

$$\mathbf{t}_{c,new,k} = \mathbf{x}_{c,new,k}\mathbf{P}_{c,PLS-CCA} \tag{16}$$

Then $\Delta t_{c,new,k,j}$ is obtained. To be simple, only one quality is considered, so index *j* is omitted for concision from now on. Then, compare $\Delta t_{c,new,k}$ with the predefined 99% confidence region to assign $\mathbf{x}_{c,new,k}$ to the steady part or transition part.

(2) Quality prediction of inner-phase parts

If the new sample $\mathbf{x}_{c,new,k}$ is judged to belong to a steady part, it is renormalized using the data normalization information from training data used for development of steady-part model, denoted as $\mathbf{x}_{c,new,k,sp}$. The corresponding quality prediction model for steady part is then adopted,

$$\hat{y}_{c,new,k,sp} = \mathbf{x}_{c,new,k,sp}\boldsymbol{\beta}_{c,sp,PLS-CCA} \tag{17}$$

Taking the cumulative effect into account, the prediction up to the new sample $\mathbf{x}_{c,new,i,sp}$ in the steady part is calculated as

$$\hat{y}_{c,new,k,sp,cum} = \frac{\sum_{i=k_{c,sp}^s}^{k} \mathbf{x}_{c,new,i,sp}\boldsymbol{\beta}_{c,sp,PLS-CCA}}{k_{c,sp} - k_{c,sp}^s + 1} \tag{18}$$

where $k_{c,sp}^s$ is the starting time of the steady part.

If the new sample $\mathbf{x}_{c,new,k}$ is judged to belong to transition, it has to wait until all the transition samples within the current transition region are available. At the end of each transition region, all new samples within the transition region are synchronized. Then, the normalized sample at each time $\mathbf{x}_{c,new,k,tr}$ is used for quality prediction as below,

$$\hat{y}_{c,new,tr,k} = \mathbf{x}_{c,new,tr,k}\boldsymbol{\beta}_{c,tr,k,PLS-CCA} \tag{19}$$

In transitions, the cumulative prediction is calculated as

$$\hat{y}_{c,new,k,c,tr,cum} = \frac{\sum_{i=k_{c,tr}^s}^{k} \mathbf{x}_{c,new,i,tr}\boldsymbol{\beta}_{c,tr,k,PLS-CCA}}{k_{c,tr} - k_{c,tr}^s + 1} \tag{20}$$

where $k_{c,tr}^s$ is the starting time of the transition part.

(3) Overall quality prediction

To consider the accumulative effect of different parts to the final quality, the predictions of all considered parts should be combined. Meanwhile, weights are added in regression model based on each sampling interval's contribution to quality. Without losing generality, the main algorithm of the composite regression model for multiphase processes with transitions can be described as,

$$\hat{y}_k = w_1\hat{y}_{k_1^e} + ... + w_p\hat{y}_{k_p^e} + ... + w_{N\text{-}1}\hat{y}_{k_{N-1}^e} + w_N\hat{y}_{k_N} \tag{21}$$

where *N* is the index of current part, *N*-1 is the number of the finished parts, $\hat{y}_{k_p^e}$ represents the cumulative prediction at the last sampling interval in part p $(p = 1,2,...,N-1)$, and $\hat{y}_{k_N}$ is the online prediction in the current part using available process information; $w_p(p = 1,2,...,N-1)$ and w_N are the weights for the finished parts and the current part. Details of weight calculation can be found in Zhao et al. (2012).

4. ILLUSTRAIN AND DISSCUSSION

4.1 Process description

The proposed algorithm is illustrated by an injection molding process. A typical injection molding process consists of three major 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 can be online measured by their corresponding transducers. One dimension index, mass (g) is chosen to evaluate the product quality. The material used in this work is high-density polyethylene (HDPE). Different operation recipes of injection are adopted by setting the injection velocity at 24, 32 and 40 mm/s, respectively, resulting in three different uneven groups regarding the injection phase.

4.2 Two-level phase division

First, conventional phase division is implemented. Using indicator variables, each batch process can be divided into four phases. Screw velocity and SV1 opening are chosen to be indicator variables based on process knowledge.

Second, inner-phase division is performed where injection phase is the focus since the uneven problem exists in this phase. $\Delta\mathbf{t}_{c,g}$ from the three uneven groups are analyzed by QTA, and for concision the results of one group is shown in Fig. 1, which are similar with the results of the other groups. It can be seen that $\Delta\mathbf{t}_{c,g}$ obviously have an evolution trend, represented by two slopes before and after a flat line respectively, revealing the fact that process first evolves to the steady state and then departs from it within each phase. Then, the steady parts of the three groups are identified which are indicated by deviations above the threshold $\dot{D}^*_{c,g}$, illustrated by one group in Fig. 2. Further, $\Delta\mathbf{t}_{c,g,sp}(g=1,2,3)$ are calculated and the 99% confidence region for online inner-phase division is defined, by which, the phase is further divided into initial transition, steady part and terminal transition. Inner-phase division result for injection phase of one group is shown in Fig. 3. Other groups have similar results. According to the analysis of inner-phase evolution, one phase is divided into inner-parts, and then different models will be built for these inner-phase parts separately.

4.3 Online quality prediction and analysis

Since the three parts of injection phase (I, II, III) and the initial transition of packing-holding phase (IV) have important impact on product mass, so the quality prediction will focus on these parts. Different regression models are developed for each part and used for online quality prediction.

Online quality prediction results of one random batch are shown in Fig. 4. Part I, a transition region, shows higher prediction errors and more dynamics than the predictions during part II. During part II, the steady part, quality predictions keep a steady state. And after that, during part III and part IV, which are transitions, quality prediction fluctuates again around the measurement value. Finally, at the end of part IV, the prediction is very near to the measure value. For other test batches, similar results can be obtained.

The median absolute deviations (MAD) of quality predictions for training and testing batches are calculated and listed in Table 1 to evaluate the variability of quality predictions in each part. Three uneven-length groups are denoted as ‘L1’, ‘L2’ and ‘L3’, and ‘All’ indicates the results are evaluated for all these three groups. The mean values of MAD for each group and for all the groups are calculated. It can be seen that the mean values of MAD in I and IV are much higher than the values in II and III. It is reasonable that I and IV, as main transitions, have more dynamic than II, steady part, and III, which has only three points in phase division. Besides, the final prediction results of the proposed method are evaluated by mean squared error (E_{MS}) index as shown in Table 2. The proposed method provides effective quality predictions for uneven-length groups.

5. CONCLUSIONS

In this work, a phase-based quality prediction strategy is proposed by tracing quality-related inner-phase evolution. Two levels of phase divisions are developed to separate transitions from steady part within each phase. Consequently, different statistical regression models are developed in different inner-phase parts for online quality prediction. In the application to an injection molding process, the proposed strategy works well for quality-related evolution analysis of uneven-length phases and meanwhile offers satisfactory online quality prediction performance.

ACKNOWLEDGEMENT

This work is supported by Program for New Century Excellent Talents in University (NCET-12-0492), Zhejiang Provincial Natural Science Foundation of China (LR13F030001), Hong Kong Research Grant Council, General Research Fund (No. 612512), Guangdong Academician Workstation Project (2012B090500010) and Technology Foundation for Selected Overseas Chinese Scholar of Zhejiang Province.

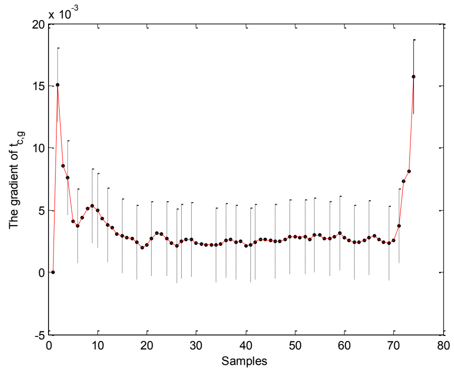

Fig. 1. The QTA segment division results of injection phase (Dot refers to the gradient of $\mathbf{t}_{c,g}$ ($\Delta\mathbf{t}_{c,g}$); red solid line refers to polynomial fitting results of QTA; the vertical dashed line indicates the segments represented by seven primitives.).

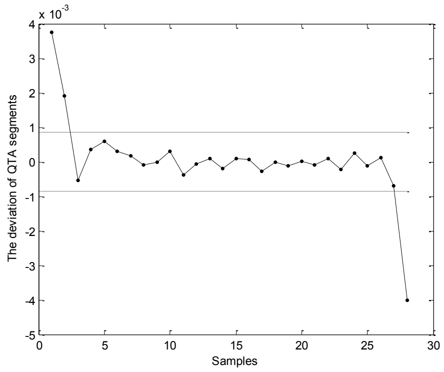

Fig. 2. The deviations of QTA segments within injection phase (Dot refers to the deviation of each QTA segment; horizontal line refers to the threshold $\dot{D}^*_{c,g}$).

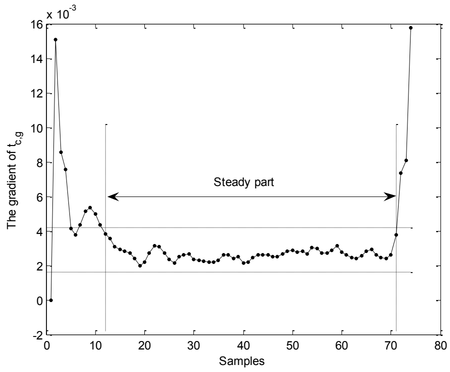

Fig. 3. The inner-phase division results of injection phase (Dot refers to the gradient of $\mathbf{t}_{c,g}$ ($\Delta\mathbf{t}_{c,g}$); horizontal dashed line refers to the 99% confidence region).

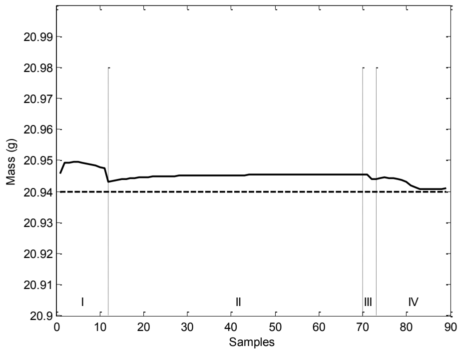

Fig. 4. Online quality predictions of mass for one batch. (Dashed lines refer to measurement values of mass.)

Table 1. Mean value of MAD of quality predictions for injection molding process (10^{-3})

Data / Part	Training data				Testing data			
	All	L1	L2	L3	All	L1	L2	L3
I	2.36	2.43	1.94	2.95	2.31	2.22	2.26	2.62
II	0.49	0.58	0.55	0.30	0.55	0.59	0.56	0.31
III	0.60	0.54	0.58	0.64	0.59	0.68	0.66	0.46
IV	3.65	5.41	2.79	3.88	3.81	5.29	3.12	4.14

Table 2. MSE of quality predictions for injection molding process

MSE	L1	L2	L3
Training data	0.027	0.070	0.106
Testing data	0.029	0.071	0.106

REFERENCES

Cserhati, T., Kosa, A., Balogh, S. (1998). Comparison of partial least-square method and canonical correlation analysis in a quantitative structure-retention relationship study. *J. Biochem. Bioph. Methods*, 36(2–3),131–141.

Dash, S., Maurya, M.R., Rengaswamy, R., Venkatasubramanian, V. (2004). A novel interval-halving framework for automated identification of process trends. *AIChE J.*, 50(1), 149–162.

Janusz, M., Venkatasubramanian, V. (1991). Automatic generation of qualitative description of process trends for fault detection and diagnosis. *Eng. Appl. Artif. Intell.*, 4(5), 329–339.

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

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

Nomikos, P., MacGregor, J.F. (1995). Multi-way partial least squares in monitoring batch processes. *Chemom. Intell. Lab. Syst.*, 30, 97–108.

Wold, S., Kettaneh, N., Tjessem, K. (1996). Hierarchical multiblock PLS and PC models for easier model interpretation and as an alternative to variable selection. *J. Chemom.*, 10, 463–482.

Westerhuis, J.A., Kourti, T., MacGregor, J.F. (1998). Analysis of multiblock and hierarchical PCA and PLS models. *J. Chemom.*, 12, 301–321.

Yu, H.L., MacGregor, J.F. (2004). Post processing methods (PLS-CCA): Simple alternatives to preprocessing methods (OSC-PLS). *Chemom. Intell. Lab. Syst.*, 73(2), 199–205.

Zhao, C.H., Lu, N.Y., Wang, F.L., Jia, M.X. (2007). Stage-based soft-transition multiple PCA modeling and on-line monitoring strategy for batch processes. *J. Process Control*, 17, 728–741.

Zhao, C.H., Wang, F.L., Mao, Z.Z., Lu, N.Y., Jia, M.X. (2008) Quality prediction based on phase-specific average trajectory for batch processes. *AIChE J.*, 54, 693–705.

Zhao, C.H., Mo, S.Y., Gao, F.R., Lu, N.Y., Yao, Y. (2011) Statistical Analysis and Online Monitoring for Handling Multiphase Batch Processes with Varying Durations. *J. Process Control*, 21(6), 817–829.

Zhao, L.P., Zhao, C.H., Gao, F.R. (2012). Phase Transition Analysis Based Quality Prediction for Multi-phase Batch Processes. *Chin. J. Chem. Eng.*, 20(6), 1191–1197.

Zhao, L.P., Zhao, C.H., Gao, F.R. (2013). Inner-phase analysis based statistical modeling and online monitoring for uneven multiphase batch processes. *Ind. Eng. Chem.Res.*, 52(12), 4586–4596.

Zhao, C.H., Sun, Y.X. (2013). Step-wise sequential phase partition (SSPP) algorithm based statistical modeling and online process monitoring. *Chemom. Intell. Lab. Syst.*, 125, 109–120.