Proceedings of the 9th International Symposium on
Dynamics and Control of Process Systems (DYCOPS 2010),
Leuven, Belgium, July 5-7, 2010
Mayuresh Kothare, Moses Tade, Alain Vande Wouwer, Ilse Smets (Eds.)

Improved Phase-based Calibration Modelling and Quality Prediction by Investigating the Effects of Inter-phase Correlation

Chunhui Zhao*, Furong Gao**

Department of Chemical and Biomolecular Engineering
The Hong Kong University of Science and Technology, HK
**huihuizh@gmail.com*
***kefgao@ust.hk*

Abstract: Phase-based quality analysis and prediction has been widely addressed by employing different calibration modeling techniques in multiphase batch processes. In this paper, a rational analysis scheme is presented to evaluate and understand the effects of the inter-phase correlation on, such as the extraction of the latent information, model structure and quality prediction. This is performed by combining partial least squares and principal component of predictions and implementing it bi-directionally (Bi-PLS-PCP). Within each phase, it separates the process systematic variation into the common and unique parts respectively based on their changes under the influence of the inter-phase correlation. They can then be quantitatively evaluated and made better use of for enhanced process understanding and improved quality prediction. The strength and efficiency of the proposed algorithm are verified on a typical multiphase batch process, injection molding.

Keywords: Multiphase batch process, quality prediction, inter-phase correlation; common and unique information, multiblock PLS.

1. INTRODUCTION

As one important area of statistical analysis, data-based multivariate calibration methods (Martens and Naes, 1994; Burnham and Viveros, et al., 1996; Brereton, 2000; Kleinbaum and Kupper, 2003) have been widely used to establish a quantitative relationship between process measurement ($\mathbf{X}$) and quality property ($\mathbf{Y}$). Among them, latent variable (LV) based methods have a dominating role. They model the correlation pattern between the variables, which allows shrinking of the original data space into a lower-dimensional feature subspace. Fewer uncorrelated LVs can be defined to comprehensively represent the original input variables and used to build a quantitative regression relationship with the concerned quality properties.

The subject of calibration modelling and quality prediction arouses new issues and problems when it refers to multiphase batch processes (Undey and Cinar, 2002; Zhao and Wang, 2008, 2009) where various phases generally operate orderly under the domination of different physical phenomena, revealing different effects on the final qualities. For multiphase batches, it is generally deemed that if the data are handled in a single matrix, the influence of one segment tends to be hidden by the other segments. The resulting tribulation is that the hidden effect could be useful in quality-concerned analysis. It is commonly accepted that 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 than single modelling of all data. The effect of each block can be seen and thus more comprehensive process understanding can be expected. Among the existing phase-based offline quality analysis strategies, two different groups have been developed as the major parallel lines of thought. One is to separately model each individual phase by MPLS algorithm (Undey and Cinar, 2002; Zhao and Wang, 2009). It reveals well the time correlation within the same phase but overlooks the inter-phase correlation. And the other is to model the variable correlation within each phase under the influence of the other phases by multiblock PLS (MBPLS) (Westerhuis and Kourti, 1998, 2001; Qin and Valle, 2001). The objective is to extract the covarying systematic dynamics among phases for quality prediction which might have not been explored when each phase is analyzed individually. It is not difficult to understand that by the two algorithms, the LVs in each data block are extracted differently so that the process variations are also modeled differently to some extent. Actually, both phase-based MPLS and MBPLS involve a risk of losing quality-relevant phase information. Focusing on different types of correlations, comparatively, some key information originally extracted by one algorithm may be hidden when analyzed by the other algorithm whereas some may be newly explored, which depends upon the power comparison between the inner-phase and inter-phase correlations. Sometimes, the missing information may be important for the quality analysis. This is a fundamental problem, which, however, has not been well addressed.

In the present work, it addresses the two following problems: On the one hand, based on the modelling results of MBPLS and phase-based MPLS, a bi-directional PLS-PCP (Bi-PLS-PCP) analysis strategy is developed to reveal the difference between their modelled quality-related process information. It can distinguish different types of process variations in each phase and provide specific model parameters to get a comprehensive understanding of phase behaviours and their roles in quality prediction. On the other hand, based on the evaluation result, the underlying phase-specific underlying information can be extracted more completely and made better use of. Thus possible calibration and prediction improvement can be achieved. The application to injection molding demonstrates the effectiveness of the method.

2. BI-PLS-PCP EVALUATION PROCEDURE AND IMPROVED CALBIRATION MODELING FOR QUALITY PREDICTION

2.1 MBPLS and phase-based MPLS models

MBPLS (Westerhuis and Kourti, 1998, 2001; Qin and Valle, 2001) is an extension of standard PLS which is recommended if additional information is available for blocking the variables into conceptually meaningful blocks/phases. If process variables are obtained from different parts of the process, multiblock regression method gives improved interpretation compared with single-block method because it is possible to zoom into separate blocks using the block scores calculated for each block. The block scores give information on the relation of the specific block with Y in the presence of the other blocks. That is, the covarying information between different blocks is modelled. In contrast, MPLS performed in each isolated block focuses on the separate contribution of each block to qualities and thus extracts the latent variables with no consideration of the inter-block correlation.

Different algorithms can be used to formulate the MBPLS model, and the deflation step plays a crucial role in the difference. Westerhuis and Smilde (2001) have shown that the deflation using block scores led to inferior prediction. However, they also pointed out that since super scores summarized the information contained in all blocks, deflation using super scores mixed variation between the separated blocks and therefore leads to interpretation problem. In order to overcome the above problem, they borrowed the idea of Doyal and MacGregor (1997) that in standard PLS models it was possible to only deflate $\mathbf{Y}$ instead of deflating both $\mathbf{X}$ and $\mathbf{Y}$ and extended it to the MBPLS algorithm, which will be used in the present work.
The MBPLS modeling step is described as below:
(a), Perform regular PLS on $\mathbf{X}=\left[\mathbf{X}_1,\mathbf{X}_2,...,\mathbf{X}_B\right]$ and $\mathbf{Y}$ to obtain the scores $\mathbf{t}_i$ and $\mathbf{u}_i$ for all required number of LVs $i\left(i=1,2,...,A\right)$, as well as the weights $\mathbf{r}_i$ to calculated scores directly from $\mathbf{X}$ and loadings $\mathbf{q}_i$ for $\mathbf{Y}$.
(b), Calculate, for each i,
Direct block weights: $\mathbf{r}_{b,i}=\mathbf{r}_i(b)/\left\|\mathbf{r}_i(b)\right\|$ where $\mathbf{r}_i(b)$ is the part of $\mathbf{r}_i$ that corresponds to block $\mathbf{X}_b$.
Block scores: $\mathbf{t}_{b,i}=\mathbf{X}_b\mathbf{r}_{b,i}$ and $\mathbf{T}_{B,i}=\left[\mathbf{t}_{1,i},\mathbf{t}_{2,i},...,\mathbf{t}_{B,i}\right]$
Super weights: $\mathbf{w}_{T,i}=\mathbf{T}_{B,i}{}^{\mathrm{T}}\mathbf{u}_i/\left\|\mathbf{T}_i{}^{\mathrm{T}}\mathbf{u}_i\right\|$
(c), Deflate residuals:
$\mathbf{Y}_{i+1}=\mathbf{Y}_i-\mathbf{t}_i\mathbf{q}_i{}^{\mathrm{T}}$
Return to step (b).

In the algorithm, obviously $\mathbf{X}$ is repeatedly used and not deflated. Therefore, no information from one block goes into another block. The block scores only give the information of the corresponding data block, which makes interpretation much easier. Their modelled process variation and predicted quality variation can then be calculated from both block and process viewpoints.

For each block:
$$\begin{aligned}\mathbf{p}_{b,i}&=\mathbf{X}_{b,i}{}^{\mathrm{T}}\mathbf{t}_{b,i}/\mathbf{t}_{b,i}{}^{\mathrm{T}}\mathbf{t}_{b,i}\\ \mathbf{X}_{b,i+1}&=\mathbf{X}_{b,i}-\mathbf{t}_{b,i}\mathbf{p}_{b,i}{}^{\mathrm{T}}\\ \mathbf{q}_{b,i}&=\mathbf{Y}_{b,i}{}^{\mathrm{T}}\mathbf{t}_{b,i}/\mathbf{t}_{b,i}{}^{\mathrm{T}}\mathbf{t}_{b,i}\\ \mathbf{Y}_{b,i+1}&=\mathbf{Y}_{b,i}-\mathbf{t}_{b,i}\mathbf{q}_{b,i}{}^{\mathrm{T}}\end{aligned}$$

For the entire process:
$$\begin{aligned}\mathbf{p}_i&=\mathbf{X}_i{}^{\mathrm{T}}\mathbf{t}_i/\mathbf{t}_i{}^{\mathrm{T}}\mathbf{t}_i\\ \mathbf{X}_{i+1}&=\mathbf{X}_i-\mathbf{t}_i\mathbf{p}_i{}^{\mathrm{T}}\\ \mathbf{q}_i&=\mathbf{Y}_i{}^{\mathrm{T}}\mathbf{t}_i/\mathbf{t}_i{}^{\mathrm{T}}\mathbf{t}_i\\ \mathbf{Y}_{i+1}&=\mathbf{Y}_i-\mathbf{t}_i\mathbf{q}_i{}^{\mathrm{T}}\end{aligned}$$

Finally, the following model parameters are output:

Block-related:
$$\begin{aligned}\mathbf{T}_b&=\left[\mathbf{t}_{b,1},\mathbf{t}_{b,2},...,\mathbf{t}_{b,A}\right]\\ \mathbf{R}_b&=\left[\mathbf{r}_{b,1},\mathbf{r}_{b,2},...,\mathbf{r}_{b,A}\right]\\ \mathbf{P}_b&=\left[\mathbf{p}_{b,1},\mathbf{p}_{b,2},...,\mathbf{p}_{b,A}\right]\\ \mathbf{Q}_b&=\left[\mathbf{q}_{b,1},\mathbf{q}_{b,2},...,\mathbf{q}_{b,A}\right]\end{aligned}$$

Super-related:
$$\begin{aligned}\mathbf{T}&=\left[\mathbf{t}_1,\mathbf{t}_2,...,\mathbf{t}_A\right]\\ \mathbf{P}&=\left[\mathbf{p}_1,\mathbf{p}_2,...,\mathbf{p}_A\right]\\ \mathbf{Q}&=\left[\mathbf{q}_1,\mathbf{q}_2,...,\mathbf{q}_A\right]\end{aligned}$$

The phase-based MPLS model parameters can be structured within each local phase:
$$\begin{aligned}\mathbf{T}_{M,b}&=\mathbf{X}_b\mathbf{R}_{M,b}\\ \mathbf{P}_{M,b}{}^{\mathrm{T}}&=\left(\mathbf{T}_{M,b}{}^{\mathrm{T}}\mathbf{T}_{M,b}\right)^{-1}\mathbf{T}_{M,b}{}^{\mathrm{T}}\mathbf{X}_b\\ \mathbf{Q}_{M,b}{}^{\mathrm{T}}&=\left(\mathbf{T}_{M,b}{}^{\mathrm{T}}\mathbf{T}_{M,b}\right)^{-1}\mathbf{T}_{M,b}{}^{\mathrm{T}}\mathbf{Y}\\ \mathbf{X}_b&=\hat{\mathbf{X}}_{M,b}+\mathbf{E}_{M,b}=\mathbf{T}_{M,b}\mathbf{P}_{M,b}{}^{\mathrm{T}}+\mathbf{E}_{M,b}\\ \mathbf{Y}&=\hat{\mathbf{Y}}_{M,b}+\mathbf{F}_{M,b}=\mathbf{T}_{M,b}\mathbf{Q}_{M,b}{}^{\mathrm{T}}+\mathbf{F}_{M,b}\end{aligned}\tag{1}$$
where, the phase scores $\mathbf{T}_{M,b}$ are orthogonal.
The MBPLS model in each phase can be expressed:
$$\begin{aligned}\mathbf{X}_b&=\hat{\mathbf{X}}_{MB,b}+\mathbf{E}_{MB,b}=\mathbf{T}_{MB,b}\mathbf{P}_{MB,b}{}^{\mathrm{T}}+\mathbf{E}_{MB,b}\\ \mathbf{Y}&=\hat{\mathbf{Y}}_{MB,b}+\mathbf{F}_{MB,b}=\mathbf{T}_{MB,b}\mathbf{Q}_{MB,b}{}^{\mathrm{T}}+\mathbf{F}_{MB,b}\end{aligned}\tag{2}$$
where, the block scores can be directly calculated from block data by $\mathbf{T}_{MB,b}=\mathbf{X}_b\mathbf{R}_{MB,b}$; $\hat{\mathbf{X}}_{MB,b}=\mathbf{T}_{MB,b}\mathbf{P}_{MB,b}{}^{\mathrm{T}}$ is the modeled process variation in each block for quality prediction and $\hat{\mathbf{Y}}_{MB,b}=\mathbf{T}_{MB,b}\mathbf{Q}_{MB,b}{}^{\mathrm{T}}$ is the contribution of each block to quality explanation. It should be noted that the block scores ($\mathbf{T}_{MB,b}$) are correlated so that the loadings $\mathbf{P}_{MB,b}$ and $\mathbf{Q}_{MB,b}$ can not be calculated as shown in Eq. (1).

2.2 Bi-PLS-PCP Evaluation procedure

The objective of PLS is to model the variations in both the $\mathbf{X}$ and $\mathbf{Y}$ spaces and relate them, which, however, often leads to difficulty in model interpretation resulting from the existence of large amount of $\mathbf{Y}$-irrelated $\mathbf{X}$ variations. Yu and MacGregor (2004) have suggested that a whole class of methods based on relating the LVs $\mathbf{T}$ and $\mathbf{Y}$, such as canonical correlation analysis (CCA) (Anderson, 2002) and PCP (Langsrud and Naes, 2003), can work as a post-processing to better estimate $\mathbf{Y}$ -correlated $\mathbf{X}$ subspace. In this way, a parsimonious model with the same prediction ability as the standard PLS model can be obtained with improved model interpretability. However, they lacked the proper model structure to describe their non-correlated variation since they only focus on the close covarying patterns between two data tables.

As introduced previously, MBPLS method is a commonly used regression modeling strategy that can well captures the inter-phase correlation, and MPLS performed in each individual phase is regarded to only focus on the inner-phase correlation. They model the phase behavior differently by extracting distinct LVs and may also make different contributions to qualities. Therefore, they provide a rational analysis platform, where their modeled process variations and quality variations can be compared and evaluated in order to understand the effect of inter-phase correlation. The modelling idea of PCP will be combined with PLS and performed in both directions for the evaluation. The basic objective is to analyze the variation information underlying two data tables in both directions and structure their relevant and irrelevant relationship. The evaluation procedure is described as below:

(1), Run PCA on the predicted quality variations by MBPLS and MPLS ($\hat{\mathbf{Y}}_{MB,b}$ and $\hat{\mathbf{Y}}_{M,b}$) respectively:

$$\begin{aligned}\hat{\mathbf{Y}}_{MB,b} &= \mathbf{T}_{MB,b}\mathbf{Q}_{MB,b}{}^{\mathrm{T}} = \mathbf{U}_1\mathbf{Q}_1{}^{\mathrm{T}} \\ \hat{\mathbf{Y}}_{M,b} &= \mathbf{T}_{M,b}\mathbf{Q}_{M,b}{}^{\mathrm{T}} = \mathbf{U}_2\mathbf{Q}_2{}^{\mathrm{T}}\end{aligned} \tag{3}$$

where, it should be noted that by MBPLS, $\hat{\mathbf{Y}}_{MB,b}$ is not orthogonal with $\mathbf{F}_{MB,b}$ as calculated in Eq. (2), which means $\mathbf{U}_1$ extracted from $\hat{\mathbf{Y}}_{MB,b}$ may correlate with $\mathbf{F}_{MB,b}$. Both $\mathbf{U}_1$ and $\mathbf{U}_2$ can be directly extracted from the original $\mathbf{X}$ space as below:

$$\begin{aligned}\mathbf{U}_1 &= \mathbf{T}_{MB,b}\mathbf{Q}_{MB,b}{}^{\mathrm{T}}\mathbf{Q}_1 = \mathbf{X}_b\mathbf{R}_{MB,b}\mathbf{Q}_{MB,b}{}^{\mathrm{T}}\mathbf{Q}_1 = \mathbf{X}_b\boldsymbol{\theta}_1 \\ \mathbf{U}_2 &= \mathbf{T}_{M,b}\mathbf{Q}_{M,b}{}^{\mathrm{T}}\mathbf{Q}_2 = \mathbf{X}_b\mathbf{R}_{M,b}\mathbf{Q}_{B,b}{}^{\mathrm{T}}\mathbf{Q}_2 = \mathbf{X}_b\boldsymbol{\theta}_2\end{aligned} \tag{4}$$

(2), Describe the variations in $\hat{\mathbf{X}}_{MB,b}$ and $\hat{\mathbf{X}}_{M,b}$ driven by $\mathbf{U}_1$ and $\mathbf{U}_2$ respectively:

$$\begin{aligned}\hat{\mathbf{X}}_{MB,by} &= \mathbf{U}_1\widehat{\mathbf{P}}_{1y}{}^{\mathrm{T}},\ \widehat{\mathbf{P}}_{1y}{}^{\mathrm{T}} = \left(\mathbf{U}_1{}^{\mathrm{T}}\mathbf{U}_1\right)^{-1}\mathbf{U}_1{}^{\mathrm{T}}\hat{\mathbf{X}}_{MB,b} \\ \hat{\mathbf{X}}_{M,by} &= \mathbf{U}_2\widehat{\mathbf{P}}_{2y}{}^{\mathrm{T}},\ \widehat{\mathbf{P}}_{2y}{}^{\mathrm{T}} = \left(\mathbf{U}_2{}^{\mathrm{T}}\mathbf{U}_2\right)^{-1}\mathbf{U}_2{}^{\mathrm{T}}\hat{\mathbf{X}}_{M,b} \\ \hat{\mathbf{E}}_{MB,b} &= \mathbf{X}_b - \hat{\mathbf{X}}_{MB,by} \\ \hat{\mathbf{E}}_{M,b} &= \mathbf{X}_b - \hat{\mathbf{X}}_{M,by}\end{aligned} \tag{5}$$

This actually is the PCP post-processing, in which, the original MBPLS and MPLS LVs in each block are further condensed. In this way, the quality-irrelevant process variations are excluded and $\hat{\mathbf{X}}_{MB,by}$ and $\hat{\mathbf{X}}_{M,by}$ only describe the closely quality-related systematic phase behaviours under the influence of inter-phase correlation and those with no such a correlation consideration respectively. $\hat{\mathbf{E}}_{MB,b}$ and $\hat{\mathbf{E}}_{M,b}$ represent the residuals in the original process space $\mathbf{X}_b$ after the PLS-PCP extraction.

3, To reveal the covarying information between $\hat{\mathbf{X}}_{MB,by}$ and $\hat{\mathbf{X}}_{M,by}$ for quality prediction, describe them according to $\mathbf{U}_2$ and $\mathbf{U}_1$ respectively:

$$\begin{aligned}\breve{\mathbf{X}}_{MB,by} &= \mathbf{U}_2\breve{\mathbf{P}}_{2y}{}^{\mathrm{T}} \\ \breve{\mathbf{P}}_{2y}{}^{\mathrm{T}} &= \left(\mathbf{U}_2{}^{\mathrm{T}}\mathbf{U}_2\right)^{-1}\mathbf{U}_2{}^{\mathrm{T}}\hat{\mathbf{X}}_{MB,by} \\ &= \left(\mathbf{U}_2{}^{\mathrm{T}}\mathbf{U}_2\right)^{-1}\mathbf{U}_2{}^{\mathrm{T}}\left(\mathbf{U}_1{}^{\mathrm{T}}\mathbf{U}_1\right)^{-1}\mathbf{U}_1{}^{\mathrm{T}}\hat{\mathbf{X}}_{MB,b} \\ \breve{\mathbf{X}}_{M,by} &= \mathbf{U}_1\breve{\mathbf{P}}_{1y}{}^{\mathrm{T}} \\ \breve{\mathbf{P}}_{1y}{}^{\mathrm{T}} &= \left(\mathbf{U}_1{}^{\mathrm{T}}\mathbf{U}_1\right)^{-1}\mathbf{U}_1{}^{\mathrm{T}}\hat{\mathbf{X}}_{M,by} \\ &= \left(\mathbf{U}_1{}^{\mathrm{T}}\mathbf{U}_1\right)^{-1}\mathbf{U}_1{}^{\mathrm{T}}\left(\mathbf{U}_2{}^{\mathrm{T}}\mathbf{U}_2\right)^{-1}\mathbf{U}_2{}^{\mathrm{T}}\hat{\mathbf{X}}_{M,b}\end{aligned} \tag{6}$$

It is clear that in this way, $\breve{\mathbf{X}}_{MB,by}$ and $\breve{\mathbf{X}}_{M,by}$ denote the parts which are simultaneously driven by both $\mathbf{U}_1$ and $\mathbf{U}_2$ in $\hat{\mathbf{X}}_{MB,by}$ and $\hat{\mathbf{X}}_{M,by}$. They thus represent the common quality-related information which can be modelled no matter whether the inter-phase correlation is considered or not.

4, Run PCA on the residuals $\breve{\mathbf{X}}_{MB,bo}$ and $\breve{\mathbf{X}}_{M,bo}$:

$$\begin{aligned}\breve{\mathbf{X}}_{MB,bo} &= \hat{\mathbf{X}}_{MB,by} - \breve{\mathbf{X}}_{MB,by} = \mathbf{T}_{MB,bo}\mathbf{P}_{MB,bo}{}^{\mathrm{T}} \\ \breve{\mathbf{X}}_{M,bo} &= \hat{\mathbf{X}}_{M,by} - \breve{\mathbf{X}}_{M,by} = \mathbf{T}_{M,bo}\mathbf{P}_{M,bo}{}^{\mathrm{T}}\end{aligned} \tag{7}$$

where, the residual $\breve{\mathbf{X}}_{MB,bo}$ reveal the part in $\hat{\mathbf{X}}_{MB,by}$ which is only driven by $\mathbf{U}_1$, i.e., the unique information in $\hat{\mathbf{X}}_{MB,by}$; the residual $\breve{\mathbf{X}}_{M,bo}$ reveal the part in $\hat{\mathbf{X}}_{M,by}$ which is only driven by $\mathbf{U}_2$, i.e., the unique information in $\hat{\mathbf{X}}_{M,by}$. By PCA, the unique scores $\mathbf{T}_{MB,bo}$ and $\mathbf{T}_{M,bo}$ are then extracted. They are contrary to the common part $\breve{\mathbf{X}}_{MB,by}$ and $\breve{\mathbf{X}}_{M,by}$.

The basic orthogonal properties are as follows:

$\mathbf{T}_{MB,bo}$ is orthogonal with both $\mathbf{U}_2$ and $\hat{\mathbf{X}}_{M,by}$; $\mathbf{T}_{M,bo}$ is orthogonal with both $\mathbf{U}_1$ and $\hat{\mathbf{X}}_{MB,by}$.

Proof:

$$\begin{aligned}&\mathbf{U}_2{}^{\mathrm{T}}\mathbf{T}_{MB,bo} \\ &= \mathbf{U}_2{}^{\mathrm{T}}\left(\hat{\mathbf{X}}_{MB,by} - \breve{\mathbf{X}}_{MB,by}\right)\mathbf{P}_{MB,bo} \\ &= \mathbf{U}_2{}^{\mathrm{T}}\left(\mathbf{I} - \mathbf{U}_2\left(\mathbf{U}_2{}^{\mathrm{T}}\mathbf{U}_2\right)^{-1}\mathbf{U}_2{}^{\mathrm{T}}\right)\hat{\mathbf{X}}_{MB,by}\mathbf{P}_{MB,bo} \\ &= \mathbf{0}\end{aligned} \tag{8}$$

$$\mathbf{T}_{MB,bo}{}^{\mathrm{T}}\hat{\mathbf{X}}_{M,by} = \mathbf{T}_{MB,bo}{}^{\mathrm{T}}\mathbf{U}_2\widehat{\mathbf{P}}_{2y}{}^{\mathrm{T}} = \mathbf{0} \tag{9}$$

In the similar way, it can also prove the orthogonal relationship between $\mathbf{T}_{M,bo}$ and $\mathbf{U}_1$ and that between $\mathbf{T}_{M,bo}$ and $\hat{\mathbf{X}}_{MB,by}$, respectively.

Finally, the variations in $\mathbf{X}$ space are separated into different orthogonal subspaces according to their mutual relationship as well as their respective relationships with qualities:

$$\begin{aligned}\mathbf{X}_b &= \hat{\mathbf{X}}_{MB,by} + \hat{\mathbf{E}}_{MB,b} = \breve{\mathbf{X}}_{MB,by} + \breve{\mathbf{X}}_{MB,bo} + \hat{\mathbf{E}}_{MB,b} \\ \mathbf{X}_b &= \hat{\mathbf{X}}_{M,by} + \hat{\mathbf{E}}_{M,b} = \breve{\mathbf{X}}_{M,by} + \breve{\mathbf{X}}_{M,bo} + \hat{\mathbf{E}}_{M,b}\end{aligned} \tag{10}$$

Therefore, by the analysis procedure in both ways, $\hat{\mathbf{X}}_{MB,by} \leftrightarrow \hat{\mathbf{X}}_{M,by}$, it can handle the different types of structured variations in both data tables, the joint covariation, the $\hat{\mathbf{X}}_{M,by}$ -orthogonal variation ($\breve{\mathbf{X}}_{MB,bo}$) in $\hat{\mathbf{X}}_{MB,by}$, and the

$\hat{\mathbf{X}}_{MB,by}$ -orthogonal variation ($\breve{\mathbf{X}}_{M,bo}$) in $\hat{\mathbf{X}}_{M,by}$. In detail, the $\hat{\mathbf{X}}_{MB,by}$ -orthogonal part will describe variations that is present in $\hat{\mathbf{X}}_{M,by}$ but absent in $\hat{\mathbf{X}}_{MB,by}$, i.e., the variation that has been hidden under the influence of inter-phase correlation. This means that it may have little correlation with the features in other phases. In a similar manner, the $\hat{\mathbf{X}}_{M,by}$ -orthogonal components consist of the structured variation that is only present in $\hat{\mathbf{X}}_{MB,by}$, i.e., the part that has been hidden by MPLS. The hidden information, i.e., the mutual orthogonal variation, reveals the respective missing information in phase-based MPLS and MBPLS modelling. On the contrary, in the case where there is no mutual orthogonal components to be extracted, the modelling performance of phase-based MPLS and MBPLS is equivalent. In this way, the Bi-PLS-PCP methodology facilitates the evaluation of the effect of inter-phase correlation on feature extraction in each phase.

2.3. Improved calibration modelling

The above evaluation procedure creates an overview that can be used to assess the properties of process information modeled by MPLS and MBPLS respectively. Analyzing the correlation and amount of different types of variation allows for verification of the desired modeling method for a full investigation of phase behaviour during quality prediction.

First, the respective contributions of unique scores $\mathbf{T}_{MB,bo}$ and $\mathbf{T}_{M,bo}$ to quality can be checked by directly performing least squares regression:

$$\begin{aligned} \hat{\mathbf{Y}}_{MB,bo} &= \mathbf{T}_{MB,bo}\left(\mathbf{T}_{MB,bo}{}^{\mathrm{T}}\mathbf{T}_{MB,bo}\right)^{-1}\mathbf{T}_{MB,bo}{}^{\mathrm{T}}\mathbf{Y} \\ \hat{\mathbf{Y}}_{M,bo} &= \mathbf{T}_{M,bo}\left(\mathbf{T}_{M,bo}{}^{\mathrm{T}}\mathbf{T}_{M,bo}\right)^{-1}\mathbf{T}_{M,bo}{}^{\mathrm{T}}\mathbf{Y} \end{aligned} \tag{11}$$

Since $\mathbf{T}_{MB,bo}$ are orthogonal to $\hat{\mathbf{X}}_{M,by}$ as well as the orthogonal relationship between $\mathbf{T}_{M,bo}$ and $\hat{\mathbf{X}}_{MB,by}$, both $\mathbf{T}_{MB,bo}$ and $\mathbf{T}_{M,bo}$ can explain additive quality information compared with the original prediction results by $\hat{\mathbf{X}}_{M,by}$ and $\hat{\mathbf{X}}_{MB,by}$. That is, $\hat{\mathbf{Y}}_{MB,bo}$ and $\hat{\mathbf{Y}}_{M,bo}$ are orthogonal with $\hat{\mathbf{Y}}_{M,bo}$ and $\hat{\mathbf{Y}}_{MB,bo}$ respectively, revealing no overlapping information. So as the complementary quality predictions, they should be nicely combined with the original ones. The complete quality prediction information in each phase can thus be prepared as $\left[\hat{\mathbf{Y}}_{MB,b}, \hat{\mathbf{Y}}_{M,bo}\right]$ or $\left[\hat{\mathbf{Y}}_{M,b}, \hat{\mathbf{Y}}_{MB,bo}\right]$, which can then be directly related with the real qualities by simply PLS regression algorithm. So the weights attached to the original and the later quality predictions can be readily calculated:

$$\begin{aligned} \left[\hat{\mathbf{Y}}_{MB,b}, \hat{\mathbf{Y}}_{M,bo}\right]\Theta_{MBb-o} &= \hat{\mathbf{Y}}_{MBb-o} \\ \left[\hat{\mathbf{Y}}_{M,b}, \hat{\mathbf{Y}}_{MB,bo}\right]\Theta_{Mb-o} &= \hat{\mathbf{Y}}_{Mb-o} \end{aligned} \tag{12}$$

Ideally, after the supplementation of later quality prediction information, the final prediction results should be equal for MBPLS and phase-based MPLS. However, because their unique information may not extracted completely more or less, the final quality prediction results $\hat{\mathbf{Y}}_{MBb-o}$ and $\hat{\mathbf{Y}}_{Mb-o}$ may not exactly the same but both show some improvement to a certain extent compared with their original results $\hat{\mathbf{Y}}_{MB,b}$ and $\hat{\mathbf{Y}}_{M,b}$.

3. APPLICATION TO INJECTION MOLDING

In this section, one typical multiphase batch process, injection molding (Yang and Gao, 2000), is used to illustrate the developed method and its performance for the evaluation of inter-phase correlation and improvement of calibration modelling. As a key process in polymer processing, it 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. The material used in this work is high-density polyethylene (HDPE). Twelve process variables are collected online from measurements with a set of sensors. Two dimension indices, product length (mm) and weight (g), are chosen to evaluate the product quality, whose real values can be directly measured by instruments. Totally 25 normal batch runs are conducted under various operation conditions by DOE method.

By the clustering-based phase division method (Zhao and Wang, 2008, 2009), the whole operation cycle can be readily partitioned into five phases. In each phase, for simplicity, only 10 continuous sampling time slices are chosen, which construct multiple data blocks, $\mathbf{X}_b\left(25\times120\right)\left(b=1,2,...5\right)$. They all correspond to the final quality property $\mathbf{Y}(25\times2)$. First, each data set $\left\{\mathbf{X}_b, \mathbf{Y}\right\}$ is regressed by MBPLS and phase-based MPLS to extract the initial LVs and their respective quality contributions within each phase. For the consideration of justice, the number of PLS LVs is set to be the same for both MBPLS and phase-based MPLS, here 4 by cross-validation. The result from MBPLS takes the inter-phase correlation into consideration whereas the other only focuses on the inner-phase correlation. They prepare the analysis platform for the evaluation of effects of inter-phase correlation. By PCP-based post-processing, only the really quality-related variation is kept in each phase, which represents the real quality-related process behaviours with and without consideration of inter-phase correlation respectively. As shown in Figure 1, the values of the first weight vector ($\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$) for the calculation of $\mathbf{U}_1$ and $\mathbf{U}_2$ are plotted through five phases. Generally, for the same phase, the weights for MBPLS and phase-based MPLS are different, which reveals that under the influence of inter-phase correlation, the quality-related phase LVs are extracted differently more or less. Moreover, as previously demonstrated, $\mathbf{T}_{MB,bo}$ and $\mathbf{T}_{M,bo}$ are orthogonal with $\mathbf{U}_2$ and $\mathbf{U}_1$ respectively. They also reveal orthogonal and complementary quality information. In each phase, the original two-dimensional quality prediction results and the later ones construct a four-variable regressor data set $\left[\hat{\mathbf{Y}}_{MB,b}, \hat{\mathbf{Y}}_{M,bo}\right]$ or $\left[\hat{\mathbf{Y}}_{M,b}, \hat{\mathbf{Y}}_{MB,bo}\right]$. By regressing them with the real final quality measurement respectively, in Figure 2, their combination weights Θ_{MB-o} and Θ_{M-o} are plotted. It is clear that the weights distributions are different over phases. For the case of MBPLS, the weights are more stable whereas for the case of phase-based MPLS, in the first three phases, the attached weights are more different. From the result, it can be inferred that for MBPLS, the additive quality information is important to complement the original one whereas for phase-based MPLS, the complementary quality prediction results do not play an important role compared with the original ones.

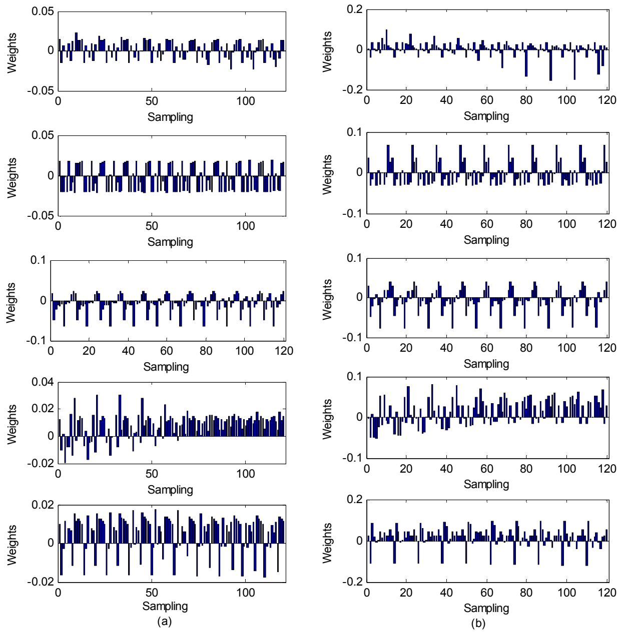

Figure 1 The first weight coefficient for (a) $\mathbf{U}_1$ and (b) $\mathbf{U}_2$ through five phases

After the separation of different types of variation in the original process space, they are quantitatively evaluated. An overview of the analysis results is summarized in Table 1. First, in different phases, they have different contributions to quality prediction. By PCP post-processing after PLS, the modelled process variations are greatly reduced for the case of phase-based MPLS, $R^2\hat{\mathbf{X}}_{M,by} \ll R^2\hat{\mathbf{X}}_{M,b}$, whereas it shows little change for the case of MBPLS, $R^2\hat{\mathbf{X}}_{MB,by} \approx R^2\hat{\mathbf{X}}_{MB,b}$. The unique information ($R^2\breve{\mathbf{X}}_{MB,bo}$ and $R^2\breve{\mathbf{X}}_{M,bo}$) accounts for a certain extent for both the cases of MBPLS and MPLS, which means under the influence of inter-phase correlation, the underlying process information is extracted differently to some extent. However, the unique information does not necessarily mean large contribution to quality property. As shown in Table 1 (b), their modeled quality variations $R^2\hat{\mathbf{Y}}_{MB,bo}$ and $R^2\hat{\mathbf{Y}}_{M,bo}$ are quite different. Although the unique process variation $R^2\breve{\mathbf{X}}_{MB,bo}$ is larger than $R^2\breve{\mathbf{X}}_{M,bo}$, its contribution to quality is little as indicated by $R^2\hat{\mathbf{Y}}_{MB,bo}$. Compare the original quality predictions ($R^2\hat{\mathbf{Y}}_{MB,b}$ and $R^2\hat{\mathbf{Y}}_{M,b}$), localized in each phase, the MBPLS results are much worse than phase-based MPLS, which tells us that the consideration of inter-phase correlation does not necessarily improve the phase-based quality prediction performance for this specific case. After merging the additive quality prediction results with the original ones, the final quality predictions are greatly improved for the case of MBPLS ($R^2\hat{\mathbf{Y}}_{MBb-o}$), which also indicates the importance of $\breve{\mathbf{X}}_{M,bo}$ that is originally hidden in MBPLS. However, for MPLS ($R^2\hat{\mathbf{Y}}_{Mb-o}$), the improvement of quality prediction is not improved much, which also means that $\breve{\mathbf{X}}_{MB,bo}$ is not so important to supplement $\hat{\mathbf{Y}}_{M,b}$. The result agrees well with the case shown in Figure 2.

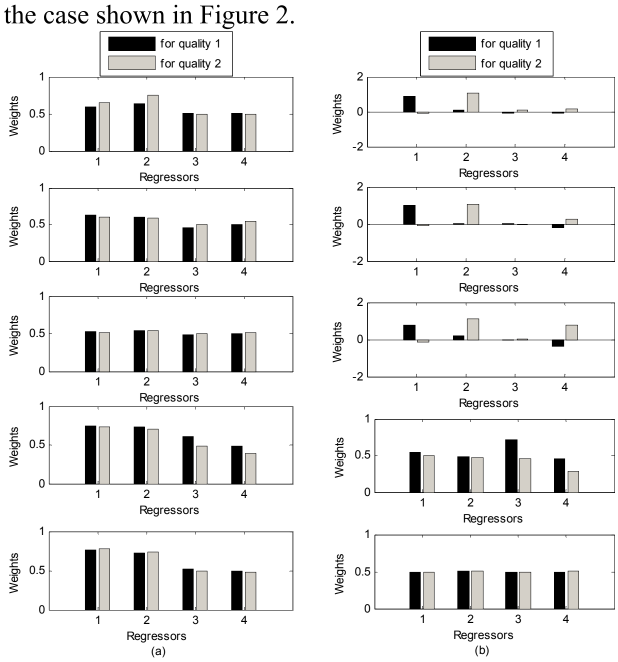

Figure 2 Weights attached to quality predictions (a) Θ_{MBb-o} and (b) Θ_{Mb-o}

For the presented illustrate case, it is shown that MBPLS and phase-based MPLS may reveal the phase behaviours differently more or less. Analysis of the mutual orthogonal variation in the extracted phase information suggests that before and after the consideration of inter-phase correlation, some phase behaviors are hidden and some new underlying information is explored. It also tells one that there is no determinate rule which method is better but more dependent upon whether they can better extract the quality-related LVs to describe the key process behaviours and quality variations.

4. CONCLUSIONS

In the present work, a bi-directional PLS-PCP algorithm is developed. It can provide definite model structures to evaluate the effects of inter-phase correlation on the extraction of process variation used for quality analysis. On the basis of the evaluation results, within each phase, different underlying process variations can be made better use of to improve calibration model and quality prediction. The evaluation method described in this paper can be readily transferred to analyze and evaluate the similarities and differences between other statistical modelling results. Moreover, it aids to decide which method is better and especially provides the possibility to combine the strengths of different methods for better modelling performance.

ACKNOWLEDGE

The work is supported in part by China National 973 program (2009CB320603) and Hong Kong Research Grant Council (613107).

REFERENCES

Anderson T.W. (2002). Canonical correlation analysis and reduced rank regression in autoregressive models. *Annals of Statistics,* **30**(4), 1134-1154.

Brereton RG. (2000). Introduction to multivariate calibration in analytical chemistry. *Analyst*. **125**(11), 2125–2154.

Burnham, A.J., Viveros, R. and MacGregor, J.F., 1996. Frameworks for latent variable multivariate regression. *Journal of Chemometrics,* **10**(1), 31-45.

Dayal, B.S. and MacGregor, J.F. (1997) Improved PLS algorithms. *Journal of Chemometrics,* **11**(1), 73-85.

Kleinbaum, D.G., Kupper, L.L., Muller, K.E., Nizam, A. (2003). *Applied Regression Analysis and Other Multivariable Methods* (3^{rd} *ed.*), California: Wadsworth Publishing Co Inc.

Langsrud O, Naes T. (2003). Optimised score plot by principal components of predictions. *Chemometrics and Intelligent Laboratory Systems,* **68**(1-2), 61-74.

Martens, H. and Naes, T. (1994). *Multivariate Calibration* (2^{nd} *ed.*). Chichester: Wiley.

Qin, S.J., Valle, S. and Piovoso, M.J., 2001. On unifying multiblock analysis with application to decentralized process monitoring. *Journal of Chemometrics,* **15**(9), 715-742.

Undey, C. and Cinar, A. (2002). Statistical monitoring of multistage, multiphase batch processes. *IEEE Control Systems Magazine,* **22**(5), 40-52.

Westerhuis, J.A., Kourti, T. and MacGregor, J.F., 1998. Analysis of multiblock and hierarchical PCA and PLS models. *Journal of Chemometrics,* **12**(5), 301-321.

Westerhuis, J.A. and Smilde, A.K., 2001. Deflation in multiblock PLS. *Journal of Chemometrics,* **15**(5), 485-493.

Yang, Y. and Gao, F. (2000). Adaptive control of the filling velocity of thermoplastics injection molding. *Control Engineering Practice,* **8**(11), 1285-1296.

Yu, H.L. and MacGregor, J.F. (2004) Post processing methods (PLS-CCA): Simple alternatives to preprocessing methods (OSC-PLS). *Chemometrics and Intelligent Laboratory Systems,* **73**(2), 199-205.

Zhao, C.H., Wang, F.L., Mao, Z.Z., Lu, N.Y. and Jia, M.X. (2008). Improved knowledge extraction and phase-based quality prediction for batch processes. *Industrial & Engineering Chemistry Research,* **47**(3), 825-834.

Zhao, C.H., Wang, F.L. and Gao, F.R. (2009). Improved calibration investigation using phase-wise local and cumulative quality interpretation and prediction. *Chemometrics and Intelligent Laboratory Systems,* **95**(2), 107-121.

Table 1 Summary of evaluation and calibration modeling results

(a) modeled process variation (%)

Phase	$R^2\hat{\mathbf{X}}_{MB,b}$	$R^2\hat{\mathbf{X}}_{M,b}$	$R^2\hat{\mathbf{X}}_{MB,by}$	$R^2\hat{\mathbf{X}}_{M,by}$	$R^2\breve{\mathbf{X}}_{MB,by}$	$R^2\breve{\mathbf{X}}_{M,by}$	$R^2\breve{\mathbf{X}}_{MB,bo}$	$R^2\breve{\mathbf{X}}_{M,bo}$
1	79.57	87.86	78.88	52.46	59.51	45.58	24.56	13.12
2	62.68	82.34	47.95	32.30	32.12	25.27	33.02	21.79
3	50.34	65.77	46.30	38.28	38.43	31.26	16.99	18.34
4	77.12	89.38	75.97	23.22	27.13	16.07	64.29	30.79
5	58.84	65.83	47.62	25.15	27.38	15.77	42.51	37.28

(b) modeled quality variation (%)

Phase	$R^2\hat{\mathbf{Y}}_{MB,b}$	$R^2\hat{\mathbf{Y}}_{M,b}$	$R^2\hat{\mathbf{Y}}_{MB,bo}$	$R^2\hat{\mathbf{Y}}_{M,bo}$	$R^2\hat{\mathbf{Y}}_{MBb-o}$	$R^2\hat{\mathbf{Y}}_{Mb-o}$
1	17.80	48.65	0.38	18.38	48.59	48.67
2	51.44	90.35	0.10	13.12	88.12	90.39
3	76.19	90.11	0.05	4.36	89.98	90.14
4	16.48	48.06	1.62	13.54	48.74	49.59
5	15.81	59.28	1.58	24.72	59.77	60.66