AIChE

Nonlinear Quality Prediction for Multiphase Batch Processes

Zhiqiang Ge
Dept. of Chemical and Biomolecular Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong

State Key Laboratory of Industrial Control Technology, Institute of Industrial Process Control, Zhejiang University, Hangzhou 310027, Zhejiang, P.R. China

Zhihuan Song
State Key Laboratory of Industrial Control Technology, Institute of Industrial Process Control, Zhejiang University, Hangzhou 310027, Zhejiang, P.R. China

Furong Gao
Dept. of Chemical and Biomolecular Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong

State Key Laboratory of Industrial Control Technology, Institute of Industrial Process Control, Zhejiang University, Hangzhou 310027, Zhejiang, P.R. China

DOI 10.1002/aic.12717
Published online July 25, 2011 in Wiley Online Library (wileyonlinelibrary.com).

Typically, a multiphase batch process comprises several steady phases and transition periods. In steady phases, the data characteristics remain similar during the phase and have a significant repeatability from batch to batch; thus most data nonlinearities can be removed through the batch normalization step. In contrast, in each transition period, process observations vary with time and from batch to batch, so nonlinearities in the data may not be eliminated through batch normalization. To improve quality prediction performance, an efficient nonlinear modeling method—relevance vector machine (RVM) was introduced. RVMs were formulated for each transition period of the batch process, and for combining the results of different process phases. For process analysis, a phase contribution index and a variable contribution index are defined. Furthermore, detailed performance analyses on the prediction uncertainty and variation were also provided. The effectiveness of the proposed method is confirmed by an industrial example. © 2011 American Institute of Chemical Engineers *AIChE J*, 58: 1778–1787, 2012

Keywords: multiphase batch processes, process transitions, nonlinearity, quality prediction, process analysis

Introduction

Because batch processes lack online measurements of the quality variable, significant efforts have been made in recent years developing quality prediction methods. Among those methods, multivariate statistical analysis models have attracted much attention. They can be directly derived from process data and can handle high dimensionality and correlated variables. Multiway principal component analysis (MPCA) and multiway partial least squares (MPLS) are two of the most widely used batch modeling techniques.[1,2] Different extensions of MPCA/MPLS have been proposed to improve the modeling of batch processes and the interpretation of the batch results,[3–7] but these multiway methods cannot reveal the multiphase characteristics which may be present in the data from many batch processes.[8–11] Batch processes which have multiple operation phases may exhibit different data behavior in different phases.

Phase-separated modeling approaches build a separate statistical model for each phase of a batch process.[12–17] However, most of them have assumed that the process can be broken into steady phases, and interphase transition behavior has rarely been considered. In practice, many batch processes may have significant transition behavior from phase to phase, especially in the start-up and ending periods of the process. The information in these transition periods may influence the quality prediction performance of any model, so incorporating the transition data is important. Here, the steady phase is defined as a part of the batch process among which the data characteristic and variable correlation keep similar. In contrast, the transition period is defined as the part in which the data characteristic and variable correlation change more significantly than the steady phase. Typically, the transition period corresponds to the start-up, ending of the process, or lies between two steady phases.

Recently, a soft-transition MPCA (STMPCA) method has been proposed for batch process monitoring in which transition information is represented in terms of a weighted combination of the two adjacent steady phase models.[18] A similarity factor-based method was later developed for identifying the transition

Correspondence concerning this article should be addressed to F. Gao at kef gao@ust.hk.

phases.[19] However, these methods focus on process monitoring; the quality prediction problem has not been considered, and a common shortcoming of these two methods is that the number of phases should be specified a priori, which may not be in accordance with the real process behavior. Recent work in our laboratory has suggested a new phase division method based on the repeatability information of a batch process.[20] After the process has been divided into steady and transition phases, corresponding phase models have been developed for quality prediction.

So far, however, while MPCA and MPLS are inherently linear modeling approaches, most phase-based quality prediction methods have also been restricted to linear cases. For those batch processes which have significant nonlinear data correlations, a proper nonlinear statistical model will be helpful to improve modeling and prediction. In each steady phase, where the data characteristics remain similar during the phase and from batch to batch, most data nonlinearities can be removed through batch scaling and normalization, i.e., the mean trajectory of different batches can be removed. In these cases, a linear regression model (using PLS, for example) is sufficient for quality modeling in each of the steady phases. The transition periods, in contrast, often exhibit more uncertainties, which makes modeling more complicated, and any nonlinearity may not easily be eliminated through batch normalization. As a result, the linear model may not function well in the transition period. To improve modeling performance, it is necessary to construct a nonlinear regression model in each transition period.

In the past years, there has been some work on nonlinear regression modeling using approaches such as artificial neural networks (ANNs) and support vector machines (SVMs).[21–23] Indeed, the traditional PLS method has also been extended to nonlinear counterparts as neural network PLS (NNPLS), kernel PLS (KPLS), etc.[24–27] Compared to PLS- and ANN-based regression methods, SVM is a more promising nonlinear modeling method and can work well with limited training data samples. Vapnik has provided a detailed derivation and theoretical justification of the use of SVMs.[22] However, SVM has some significant drawbacks. First, it does not allow for the use of an arbitrary kernel function, because the kernel function should satisfy Mercer's conditions. Second, the SVM method needs to determine the error/margin trade-off parameter and the insensitivity parameter, which will generally introduce a cross-validation procedure.

Recently, another kernel learning method termed the relevance vector machine (RVM) has been proposed.[28] This shares a similar functional form with SVMs, but has no limitations on the type of kernel function. By introducing a prior distribution over the weight parameter, a Bayesian approach has been adopted for RVM learning. While SVM can only provide a point estimate of the predicted variable, RVM gives a probabilistic prediction result which provides detailed uncertainty information about the prediction. The uncertainty of the prediction result is important for understanding the process and interpreting the results because it can provide information about the reliability of the predicted values in specific phases and at specific time points. Additionally, compared to SVM, RVM modeling leads to sparser models, thus faster computation on test data while maintaining comparable generalization errors.

With both linear PLS and nonlinear RVM models, prediction results can be generated for both steady phases and transition periods. But another important issue which has not been properly considered is the relationships between different phases and how to efficiently combine phase prediction results. Several combination approaches involving weighted parameters and linear regression methods have been proposed,[10] but more detailed relationships can be constructed based on nonlinear models. Particularly when incorporating transition information gives a nonlinear prediction result, the relationships between the steady phases and transition periods are probably also nonlinear. Therefore, a nonlinear regression model is then more appropriate for combining different phases of the process. Furthermore, for the purpose of quality control, the phases which are most important for the final quality need to be determined, and the key variables in those phases need to be identified. This study set out to develop a strategy for identifying critical phases and key variables using a nonlinear quality prediction algorithm.

The remainder of this article is organized as follows. In section 2, preliminaries of PLS and RVM are given. Detailed description of the nonlinear quality prediction algorithm is presented in section 3, followed by an industrial application study of the injection molding process in the next section. Finally, conclusions are made.

Terminology

Partial least squares

Denote the process and quality variable datasets as $\{\mathbf{X},\mathbf{Y}\}$. PLS intends to decompose $\mathbf{X}$ and $\mathbf{Y}$ into scores matrix $\mathbf{T}$, $\mathbf{U}$, and loading matrices $\mathbf{P}$ and $\mathbf{Q}$, given as[2]

$$\begin{aligned}\mathbf{X} &= \mathbf{T}\mathbf{P}^T + \mathbf{E}\\ \mathbf{T} &= \mathbf{X}\mathbf{W}(\mathbf{P}^T\mathbf{W})^{-1}\end{aligned} \tag{1}$$

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

where $\mathbf{W}$ is the weight matrix of the PLS model, $\mathbf{E}$ and $\mathbf{F}$ are residuals matrices of the process and quality variables. For a new data sample $\mathbf{x}_{\text{new}}$, the predicted output variables can be calculated as

$$\hat{y}_{\text{new}} = [\mathbf{W}(\mathbf{P}^T\mathbf{W})^{-1}\mathbf{Q}^T]^T\mathbf{x}_{\text{new}} \tag{3}$$

Support vector machine for regression

Given the training dataset $\{\mathbf{x}_i,\ y_i\}_{i\ =\ 1,2,\cdots n}$, the regression SVM metho aims to solve the following optimization problem

$$\begin{aligned}&\min\left\{\frac{1}{2}\|\mathbf{w}\| + \frac{C}{n}\sum_{i=1}^{n}(\xi_i - \xi_i^*)\right\}\\ &\text{subject to}: f(\mathbf{x}_i) - y_i \le \varepsilon + \xi_i\\ &\qquad\qquad y_i - f(\mathbf{x}_i) \le \varepsilon + \xi_i^*\\ &\qquad\qquad \xi_i, \xi_i^* \ge 0\end{aligned} \tag{4}$$

where $C > 0$ is the predefined regularizing parameter, ε is the insensitive parameter, ξ_i and ξ_i^* are slack variables for each data sample. Based on the Lagrange multiplier method and QP algorithm, the SVM regression function can e derived as follows

$$f(\mathbf{x}) = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)K(\mathbf{x}, \mathbf{x}_i) + b \tag{5}$$

$$b = y_j - \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)K(\mathbf{x}_j, \mathbf{x}_i) + \varepsilon \tag{6}$$

where α_i and α_i^* are the optimized Lagrange multipliers for each support vector, $K(\cdot)$ is the kernel function, l is the number of support vectors selected in the model. By introducing the kernel function, SVM model becomes flexible to deal with complicated nonlinear regression problem.

Relevance vector machine

The main idea of the RVM algorithm was originally proposed by Tipping.[28] It has been used for data classification and regression in areas such as machine learning, pattern recognition and many others. Based on the introduction of a prior distribution for each parameter weight, the RVM algorithm can be developed on an SVM framework through Bayesian methods. Denote the training dataset as $\{\mathbf{x}_i, y_i\}_{i=1,2,\cdots n}$. The nonlinear relationship between the process and quality variables can then be expressed as

$$y = f(\mathbf{x}, \mathbf{w}) + e \tag{7}$$

where e is the random error, which is assumed to be independent and Gaussian distributed with a zero mean and variance σ^2, $e \sim N(0,\sigma^2)$, $\mathbf{w}$ is the weighted parameter. Similar to the SVM algorithm, the nonlinear function $f(\mathbf{x})$ can be expressed as a linearly weighted sum of some basis functions[28]

$$f(\mathbf{x}, \mathbf{w}) = \sum_{i=1}^{h} w_i K(\mathbf{x}, \mathbf{x}_i) + w_0 = \mathbf{w}^T \psi(\mathbf{x}) \tag{8}$$

where $\mathbf{w} = [w_0, w_1, w_2, \cdots, w_h]^T$ is the weighted parameter vector of the basis functions w and $\psi(\mathbf{x}) = [1, K(\mathbf{x},\mathbf{x}_1), K(\mathbf{x},\mathbf{x}_2), \cdots, K(\mathbf{x},\mathbf{x}_h)]^T$.

Unlike the SVM, the RVM can give a probabilistic prediction for the quality variable based on the weighted parameter vector $\mathbf{w}$ and the noise variance σ^2. The probabilistic prediction can be expressed as $p(y|\mathbf{x}) = N(f(\mathbf{x},\mathbf{w})\sigma^2)$. By carrying out the following optimization, the optimal values of α (which is a parameter in the priori distribution of $\mathbf{w}$ in Eq. 11) and σ^2 can be obtained

$$\begin{aligned} \max L(\alpha, \sigma^2) &= \max \log\{p(\mathbf{y}|\mathbf{X}, \mathbf{w}, \sigma^2)\} \\ &= \max \log\left\{\int p(\mathbf{y}|\mathbf{X}, \mathbf{w}, \sigma^2) p(\mathbf{w}|\alpha) d\mathbf{w}\right\} \end{aligned} \tag{9}$$

$$p(\mathbf{y}|\mathbf{x}, \mathbf{w}, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\|\mathbf{y} - \psi(\mathbf{x})\mathbf{w}\|\right\} \tag{10}$$

$$p(\mathbf{w}|\alpha) = \prod_{i=0}^{n} N(w_i|0, \alpha_i^{-1}) = \frac{1}{(2\pi)^{(n+1)/2}} \prod_{i=0}^{n} \alpha_i^{1/2} \exp\left(-\frac{\alpha_i w_i^2}{2}\right) \tag{11}$$

where $\mathbf{y} = (y_1, y_2, \cdots, y_n)$, $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n)$. The posterior distribution of $\mathbf{w}$ can be determined from

$$\begin{aligned} p(\mathbf{w}|\mathbf{y}, \alpha, \sigma^2) &= \frac{p(\mathbf{y}|\mathbf{w}, \sigma^2) p(\mathbf{w}|\alpha)}{p(\mathbf{y}|\alpha, \sigma^2)} \\ &= \frac{1}{(2\pi)^{(n+1)/2}} |\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{w}-\mu)^T \Sigma^{-1} (\mathbf{w}-\mu)\right\} \end{aligned} \tag{12}$$

which is also a Gaussian distribution with mean and variance satisfying

$$\mu = \sigma^{-2} \Sigma \psi^T(\mathbf{x}) \mathbf{y} \tag{13}$$

$$\Sigma = (\sigma^{-2} \Psi^T(\mathbf{x}) \Psi(\mathbf{x}) + \mathbf{A})^{-1} \tag{14}$$

where $\Psi(\mathbf{x}) = [\psi_1(\mathbf{x}), \psi_2(\mathbf{x}), \cdots, \psi_n(\mathbf{x})]$, and $\mathbf{A}$ is a diagonal matrix with elements $\mathbf{A} = \mathrm{diag}(\alpha_0, \alpha_1, \cdots, \alpha_n)$. As a result, many elements of $\mathbf{w}$ become zero and the optimal mean $\mathbf{w}$ value will have very few nonzero elements. By analogy with the term "support vector" used in SVM methods, the non-zero $\mathbf{w}$ element training data are called a "relevance vector" in the RVM method.

Methodology

A detailed quality prediction algorithm will now be demonstrated which includes phase-based prediction models for steady phases and transition periods, nonlinear regression modeling for combining the results, critical phase and key variable identification, and uncertainty and performance analyses.

Assume that the batch process has already been divided into steady phases and transition periods. The historical batch process dataset can be represented as a three-way data matrix $\mathbf{X}(I \times J \times K)$, where I is the number of batches, J is the number of variables, and the duration of each batch is K. Through the batch direction, the three-way data matrix can be unfolded into a two-dimensional matrix $\mathbf{X}(I \times JK)$, which can be further decomposed into different time slices

$$\mathbf{X}(I \times JK) = [\mathbf{X}_1(I \times J) \;\; \mathbf{X}_2(I \times J) \;\; \cdots \;\; \mathbf{X}_K(I \times J)] \tag{15}$$

Suppose the batch process has been divided into S steady phases and T transition periods. The dataset can then be partitioned as follows

$$\begin{aligned} \mathbf{X}(I \times JK) = [&\mathbf{X}_1^{\mathrm{st}}(I \times JK_1^{\mathrm{st}}) \;\; \mathbf{X}_1^{\mathrm{tr}}(I \times JK_1^{\mathrm{tr}}) \;\; \mathbf{X}_2^{\mathrm{st}}(I \times JK_2^{\mathrm{st}}) \;\; \cdots \\ &\mathbf{X}_T^{\mathrm{tr}}(I \times JK_T^{\mathrm{tr}}) \;\; \mathbf{X}_S^{\mathrm{st}}(I \times JK_S^{\mathrm{st}})] \end{aligned} \tag{16}$$

where the st and tr superscripts represent steady phases and transition periods, respectively, $K_1^{\mathrm{st}}, K_2^{\mathrm{st}}, \cdots, K_S^{\mathrm{st}}$ are the steady phase time slices and $K_1^{\mathrm{tr}}, K_2^{\mathrm{tr}}, \cdots, K_T^{\mathrm{tr}}$ are the transition time intervals.

Quality prediction modeling for the steady phases and transition periods

To build a prediction model for each steady phase, the PLS prediction model for each time slice should be developed first.[2]

$$\begin{aligned} \mathbf{X}_{s,k}^{\mathrm{st}} &= \mathbf{T}_{s,k}^{\mathrm{st}} \mathbf{P}_{s,k}^{\mathrm{st}\,T} + \mathbf{E}_{s,k}^{\mathrm{st}} \\ \mathbf{Y} &= \mathbf{U}_{s,k}^{\mathrm{st}} \mathbf{Q}_{s,k}^{\mathrm{st}\,T} + \mathbf{F}_{s,k}^{\mathrm{st}} \\ \mathbf{R}_{s,k}^{\mathrm{st}} &= \mathbf{W}_{s,k}^{\mathrm{st}} (\mathbf{P}_{s,k}^{\mathrm{st}\,T} \mathbf{W}_{s,k}^{\mathrm{st}})^{-1} \mathbf{Q}_{s,k}^{\mathrm{st}\,T} \end{aligned} \tag{17}$$

where $s = 1,2,\cdots,S$ and $k = 1,2,\cdots,K_s^{\mathrm{st}}$. $\mathbf{P}_{s,k}^{\mathrm{st}}$ and $\mathbf{Q}_{s,k}^{\mathrm{st}}$ are the loading matrices, $\mathbf{T}_{s,k}^{\mathrm{st}}$ and $\mathbf{U}_{s,k}^{\mathrm{st}}$ are the score matrices, $\mathbf{E}_{s,k}^{\mathrm{st}}$ and $\mathbf{F}_{s,k}^{\mathrm{st}}$ are the residual matrices, and $A_{s,k}^{\mathrm{st}}$ is the retained number of latent variables in each PLS model. Conventionally, the number of latent variables can be determined by cumulative percentage of explained variance of the process data (CPV), Predicted residual sum of squares (PRESS), Cross-validation (CV), etc. In this article, the CPV method is used, thus the number of latent variables is determined if the CPV value of the PLS model exceeds a cutoff value, e.g., 85%. $\mathbf{W}_{s,k}^{\mathrm{st}}$ is the weighting matrix of each PLS model, and $\mathbf{R}_{s,k}^{\mathrm{st}}$ is its regression matrix. In the next step, the representative PLS regression model can be developed, the regression matrix of which is the mean value of all steady phase slices

$$\mathbf{R}_s^{\mathrm{st}^*} = \frac{1}{K_s^{\mathrm{st}}} \sum_{k=1}^{K_s^{\mathrm{st}}} \mathbf{R}_{s,k}^{\mathrm{st}} \tag{18}$$

Similarly, in each transition period the RVM model is constructed for each time slice in the first step[28]

$$\mathbf{y} = f(\mathbf{x}, \mathbf{w}_{t,k}^{\mathrm{tr}}) + \mathbf{e}_{t,k}^{\mathrm{tr}} = \sum_{i=1}^{I} w_{t,k,i}^{\mathrm{tr}} K(\mathbf{x}, \mathbf{x}_{t,k,i}^{\mathrm{tr}}) + w_{t,k,0}^{\mathrm{tr}} + \mathbf{e}_{t,k}^{\mathrm{tr}} = \mathbf{w}_{t,k}^{\mathrm{tr}\,T} \psi_{t,k}^{\mathrm{tr}}(\mathbf{x}) + \mathbf{e}_{t,k}^{\mathrm{tr}} \tag{19}$$

where $t = 1,2,\cdots,T$ and $k = 1,2,\cdots,K_t^{\mathrm{tr}} \cdot \mathbf{w}_{t,k\mathrm{tr}} = [w_{t,k,0}^{\mathrm{tr}}, w_{t,k,1}^{\mathrm{tr}}, w_{t,k,2}^{\mathrm{tr}},\cdots,w_{t,k,I}^{\mathrm{tr}}]^T$ is the weighted parameter vector of the basis functions $\psi_{t,k}^{\mathrm{tr}}(\mathbf{x}) = [1,K(\mathbf{x},\mathbf{x}_{t,k,I}^{\mathrm{tr}}), K(\mathbf{x},\mathbf{x}_{t,k,2}^{\mathrm{tr}}),\cdots,K(\mathbf{x},\mathbf{x}_{t,k,I}{}^{\mathrm{tr}})]^T$. The prior distribution of the weighted parameter vector for each RVM model is

$$p(\mathbf{w}_{t,k}^{\mathrm{tr}}|\alpha_{t,k}^{\mathrm{tr}}) = \prod_{i=0}^{I} N(w_{t,k,i}^{\mathrm{tr}}|0, \alpha_{t,k,i}^{\mathrm{tr}\,-1}) = \frac{1}{(2\pi)^{(I+1)/2}} \prod_{i=0}^{I} \alpha_{t,k,i}^{\mathrm{tr}\,1/2} \exp\left(-\frac{\alpha_{t,k,i}^{\mathrm{tr}} w_{t,k,i}^{\mathrm{tr}\,2}}{2}\right) \tag{20}$$

The posterior distribution of $\mathbf{W}_{t,k}^{\mathrm{tr}}$ can now be determined based on the maximum likelihood algorithm, the mean and variance of which are given by

$$\mathbf{w}_{t,k}^{\mathrm{tr}} = \sigma_{t,k}^{\mathrm{tr}\,-2} \Sigma_{\mathbf{w},t,k}^{\mathrm{tr}} \Psi_{t,k}^{\mathrm{tr}\,T}(\mathbf{x})\mathbf{y} \tag{21}$$

$$\Sigma_{\mathbf{w},t,k}^{\mathrm{tr}} = (\sigma_{t,k}^{\mathrm{tr}\,-2} \Psi_{t,k}^{\mathrm{tr}\,T}(\mathbf{x}) \Psi_{t,k}^{\mathrm{tr}}(\mathbf{x}) + \mathbf{A}_{t,k}^{\mathrm{tr}})^{-1} \tag{22}$$

where $\Psi_{t,k}^{\mathrm{tr}}(\mathbf{x}_{t,k}^{\mathrm{tr}}) = [\psi_{t,k}^{\mathrm{tr}}(\mathbf{x}_{t,k,1}^{\mathrm{tr}}),\psi_{t,k}^{\mathrm{tr}}(\mathbf{x}_{t,k,2}^{\mathrm{tr}}), \cdots, \psi_{t,k}^{\mathrm{tr}}(\mathbf{x}_{t,k,I}^{\mathrm{tr}})]$, and $\mathbf{A}_{t,k}^{\mathrm{tr}}$ is a diagonal matrix with elements $\mathbf{A}_{t,k}^{\mathrm{tr}} = \mathrm{diag}(\alpha_{t,k,0}^{\mathrm{tr}}, \alpha_{t,k,1}^{\mathrm{tr}},\cdots,\alpha_{t,k,I}^{\mathrm{tr}})$.

Based on the RVM model for each time slice, the RVM model for each transition period can be developed. The mean weighted parameter is

$$\mathbf{w}_t^{\mathrm{tr}^*} = \frac{1}{K_t^{\mathrm{tr}}} \sum_{k=1}^{K_t^{\mathrm{tr}}} \mathbf{w}_{t,k}^{\mathrm{tr}} = \frac{1}{K_t^{\mathrm{tr}}} \sum_{k=1}^{K_t^{\mathrm{tr}}} \sigma_{t,k}^{\mathrm{tr}\,-2} \Sigma_{\mathbf{w},t,k}^{\mathrm{tr}} \Psi_{t,k}^{\mathrm{tr}\,T}(\mathbf{x})\mathbf{y} \tag{23}$$

Nonlinear results combination

To predict the quality of a new batch, the prediction result in a specific steady phase or transition period can be calculated by averaging the results of all available information about that phase or period

$$\hat{\mathbf{y}}_{\mathrm{new},s}^{kc} = \frac{1}{kc} \sum_{k=1}^{kc} \hat{\mathbf{y}}_{\mathrm{new},s}^{k} = \frac{1}{kc} \sum_{k=1}^{kc} \mathbf{R}_s^{st^*\,T} \mathbf{x}_{\mathrm{new}}^{k} \tag{24}$$

$$\hat{\mathbf{y}}_{\mathrm{new},t}^{kc} = \frac{1}{kc} \sum_{k=1}^{kc} \hat{\mathbf{y}}_{\mathrm{new},t}^{k} = \frac{1}{kc} \sum_{k=1}^{kc} \mathbf{w}_t^{tr^*\,T} \psi_{t,k}^{\mathrm{tr}}(\mathbf{x}_{\mathrm{new}}^{k}) \tag{25}$$

where $s = 1,2,\cdots,S$ and $t = 1,2,\cdots,T$. $kc = 1,2,\cdots,K_s^{\mathrm{st}}$ (K_t^{tr}) is the current time point in the corresponding phase $\psi_{t,k}^{\mathrm{tr}}(\mathbf{x}_{\mathrm{new}}^{k}) = [1,K(\mathbf{x}_{\mathrm{new}}^{k}, \mathbf{x}_{t,k,1}^{k}),K(\mathbf{x}_{\mathrm{new}}^{k},\mathbf{x}_{t,k,2}^{\mathrm{tr}}),\cdots,K(\mathbf{x}_{\mathrm{new}}^{k},\mathbf{x}_{t,k,I}^{\mathrm{tr}})]^T$.

A series of nonlinear cumulative regression models must now be developed to combine the different phase and period prediction results. Generally, when s steady phases and t transition periods are being incorporated for quality prediction, the cumulative RVM model is

$$\mathbf{y} = f(\hat{\mathbf{y}}_{s+t-1}, \mathbf{w}_{s+t-1}) + \mathbf{e}_{s+t-1} = \sum_{i=1}^{I} w_{s+t-1,i} K_{s+t-1}(\hat{\mathbf{y}}_{s+t-1}, \hat{\mathbf{y}}_{s+t-1,i}^{\mathrm{tr}}) + w_{s+t-1,0} + \mathbf{e}_{s+t-1} = \mathbf{w}_{s+t-1}^{T} \psi_{s+t-1}(\hat{\mathbf{y}}_{s+t-1}^{\mathrm{tr}}) + \mathbf{e}_{s+t-1} \tag{26}$$

where $\hat{\mathbf{y}}_{s+t-1,i}^{\mathrm{tr}}$, $i = 1,2,\cdots,I$ and $\hat{\mathbf{y}}_{s+t-1}$, are the phase prediction values of the training batches and the test batch, respectively. $s = 1,2,\cdots,S$, $t = 1,2,\cdots,T$, and the basis function is $\psi_{s+t-1}(\hat{\mathbf{y}}_{s+t-1}) = [1,K(\hat{\mathbf{y}}_{s+t-1},\hat{\mathbf{y}}_{s+t-1,1}^{\mathrm{tr}}), K(\hat{\mathbf{y}}_{s+t-1}, \hat{\mathbf{y}}_{s+t-1,2}^{\mathrm{tr}}),\cdots,K(\hat{\mathbf{y}}_{s+t-1}, \hat{\mathbf{y}}_{s+t-1,I}^{\mathrm{tr}})]^T$. As a result, a total of $S + T - 1$ cumulative RVM regression models must be developed to combine the results. Given the prior distribution of the weighted parameter vector for each RVM model, the posterior distribution of $\mathbf{w}_{s+t-1}$ can be determined as

$$p(\mathbf{w}_{s+t-1}|\mathbf{y}, \hat{\mathbf{y}}_{s+t-1}^{\mathrm{tr}}, \alpha_{s+t-1}, \sigma_{s+t-1}^2) = \frac{p(\mathbf{y}|\hat{\mathbf{y}}_{s+t-1}^{\mathrm{tr}}, \mathbf{w}_{s+t-1}, \sigma_{s+t-1}{}^2) p(\mathbf{w}_{s+t-1}|\alpha_{s+t-1})}{p(\mathbf{y}|\alpha_{s+t-1}, \sigma_{s+t-1}{}^2)} \tag{27}$$

which is also Gaussian distributed with mean

$$\mathbf{w}_{s+t-1} = \sigma_{s+t-1}^{-2} \Sigma_{\mathbf{w},s+t-1} \Psi_{s+t-1}^{T}(\hat{\mathbf{y}}_{s+t-1}^{\mathrm{tr}})\mathbf{y} \tag{28}$$

and variance

$$\Sigma_{\mathbf{w},s+t-1} = (\sigma_{s+t-1}^{-2} \Psi_{s+t-1}^{T}(\hat{\mathbf{y}}_{s+t-1}^{\mathrm{tr}}) \Psi_{s+t-1}(\hat{\mathbf{y}}_{s+t-1}^{\mathrm{tr}}) + \mathbf{A}_{s+t-1})^{-1} \tag{29}$$

where $\Psi_{s+t-1}(\hat{\mathbf{y}}_{s+t-1}^{\mathrm{tr}}) = [\psi_{s+t-1}(\hat{\mathbf{y}}_{s+t-1,1}^{\mathrm{tr}}), \psi_{s+t-1}(\hat{\mathbf{y}}_{s+t-1,2}^{\mathrm{tr}}),\cdots,\psi_{s+t-1}(\hat{\mathbf{y}}_{s+t-1,I}^{\mathrm{tr}})]$, and $\mathbf{A}_{s+t-1}$ is a diagonal matrix with elements $\mathbf{A}_{s+t-1} = \mathrm{diag}(\alpha_{s+t-1,0}, \alpha_{s+t-1,1},\cdots,\alpha_{s+t-1,I})$.

Using these nonlinear regression models, the phase and period prediction results can be combined on the second level. Suppose an online quality prediction is to be made in transition period t. The prediction results in the previous phases/periods and the current transition period for the new data sample $\mathbf{x}_{\mathrm{new}}$ should have been calculated in the first step. They are

$$\hat{\mathbf{y}}_{\mathrm{new},1}^{\mathrm{st}} = \frac{1}{K_1^{\mathrm{st}}} \sum_{k=1}^{K_1^{\mathrm{st}}} \hat{\mathbf{y}}_{\mathrm{new},1}^{\mathrm{st},k} = \frac{1}{K_1^{\mathrm{st}}} \sum_{k=1}^{K_1^{\mathrm{st}}} \mathbf{R}_1^{\mathrm{st}^*\,T} \mathbf{x}_{\mathrm{new}}^{k}$$

$$\hat{\mathbf{y}}_{\mathrm{new},1}^{\mathrm{tr}} = \frac{1}{K_1^{\mathrm{tr}} - K_1^{\mathrm{st}}} \times \sum_{k=K_1^{\mathrm{tr}}-K_1^{\mathrm{st}}+1}^{K_1^{\mathrm{tr}}} \hat{\mathbf{y}}_{\mathrm{new},1}^{\mathrm{tr},k} = \frac{1}{K_1^{\mathrm{tr}} - K_1^{\mathrm{st}}} \sum_{k=K_1^{\mathrm{tr}}-K_1^{\mathrm{st}}+1}^{K_1^{\mathrm{tr}}} \mathbf{w}_1^{\mathrm{tr}^*\,T} \psi_{1,k}^{\mathrm{tr}}(\mathbf{x}_{\mathrm{new}})$$

$$\vdots$$

$$\hat{\mathbf{y}}_{\mathrm{new},s}^{\mathrm{st}} = \frac{1}{K_s^{\mathrm{st}} - \sum_{j=1}^{s-1} K_j^{\mathrm{st}} - \sum_{j=1}^{t-1} K_j^{\mathrm{tr}}} \sum_{k=K_s^{\mathrm{st}}-\sum_{j=1}^{s-1} K_j^{\mathrm{st}}-\sum_{j=1}^{t-1} K_j^{\mathrm{tr}}+1}^{K_s^{\mathrm{st}}} \times \hat{\mathbf{y}}_{\mathrm{new},s}^{\mathrm{st},k} = \frac{1}{K_s^{\mathrm{st}} - \sum_{j=1}^{s-1} K_j^{\mathrm{st}} - \sum_{j=1}^{t-1} K_j^{\mathrm{tr}}} \times \sum_{k=K_s^{\mathrm{st}}-\sum_{j=1}^{s-1} K_j^{\mathrm{st}}-\sum_{j=1}^{t-1} K_j^{\mathrm{tr}}+1}^{K_s^{\mathrm{st}}} \mathbf{R}_s^{\mathrm{st}^*\,T} \mathbf{x}_{\mathrm{new}}^{k}$$

$$\hat{\mathbf{y}}_{\text{new},t}^{\text{tr},kc} = \frac{1}{kc - \sum_{j=1}^{s} K_j^{\text{st}} - \sum_{j=1}^{t-1} K_j^{\text{tr}}} \sum_{k=kc-\sum_{j=1}^{s} K_j^{\text{st}} - \sum_{j=1}^{t-1} K_j^{\text{tr}}+1}^{kc}$$

$$\times \hat{\mathbf{y}}_{\text{new},t}^{\text{tr},k} = \frac{1}{kc - \sum_{j=1}^{s} K_j^{\text{st}} - \sum_{j=1}^{t-1} K_j^{\text{tr}}} \times \sum_{k=kc-\sum_{j=1}^{s} K_j^{\text{st}} - \sum_{j=1}^{t-1} K_j^{\text{tr}}+1}^{kc} \mathbf{w}_t^{\text{tr}^*\,T} \psi_{t,k}^{\text{tr}}(\mathbf{x}_{\text{new}}) \quad (30)$$

The prediction can now be made by combining these phase and period prediction results

$$\hat{\mathbf{y}}_{\text{new}}^{ph} = [\hat{\mathbf{y}}_{\text{new},1}^{st}\ \hat{\mathbf{y}}_{\text{new},1}^{tr} \cdots\ \hat{\mathbf{y}}_{\text{new},1}^{st}\ \hat{\mathbf{y}}_{\text{new},t}^{tr,kc}]$$

$$\psi_{s+t-1}^{kc}(\hat{\mathbf{y}}_{\text{new}}^{ph}) = [1, K(\hat{\mathbf{y}}_{\text{new}}^{ph}, \hat{\mathbf{y}}_1^{ph}), K(\hat{\mathbf{y}}_{\text{new}}^{ph}, \hat{\mathbf{y}}_2^{ph}), \cdots, K(\hat{\mathbf{y}}_{\text{new}}^{ph}, \hat{\mathbf{y}}_I^{ph})]^T$$

$$\hat{\mathbf{y}}_{\text{new}}^{kc} = \mathbf{w}_{s+t-1}^T \psi_{s+t-1}^{kc}(\hat{\mathbf{y}}_{\text{new}}^{ph}) \quad (31)$$

where $\hat{\mathbf{y}}_{\text{new}}^{ph}$ is a combined vector of different phase or period prediction results $\hat{\mathbf{y}}_i^{ph}$, $i = 1,2,\cdots,I$ from the training batches.

Critical phase/period determination and key variable identification

To quantify the importance of each phase and period in determining the final quality, the phase/period contribution index (PCI) is defined as

$$\psi_j^{ph}(y_{\text{new}}^{ph,j}) = [1, K(\hat{y}_{\text{new}}^{ph,j}, \hat{y}_1^{ph,j}), K(\hat{y}_{\text{new}}^{ph,j}, \hat{y}_2^{ph,j}), \cdots, K(\hat{y}_{\text{new}}^{ph,j}, \hat{y}_I^{ph,j})]^T$$

$$\text{PCI}(j) = \frac{\left\| \mathbf{w}_{S+T-1}^T \psi_j^{ph}(\hat{y}_{\text{new}}^{ph,j}) \right\|}{\sum_{j=1}^{S+T} \left\| \mathbf{w}_{S+T-1}^T \psi_j^{ph}(\hat{y}_{\text{new}}^{ph,j}) \right\|} \quad (32)$$

where $j = 1,2,\cdots,S+T$ is a phase/period index and $\hat{y}^{ph,i}$ and $\hat{y}_i^{ph,i}$, $i = 1,2,\cdots,I$ are the quality predictions for the j-th phase or period of the current batch and the training batches respectively. The contribution index ranges from zero to one, and $\sum_{j=1}^{S+T} PCI(j) = 1$. Using this index, the contribution of each phase and period to the final quality prediction can be quantified. Those phases/periods with the large *PCI* values should be determined as the critical-to-quality phases/periods in the batch process. In contrast, others which have relatively smaller *PCI* values should be categorized as minor-contributed phases/periods. Once the most important phases and periods have been determined, it is important for quality control that the key variables in those phases/periods should also be identified. By analogy, variable contribution indices for the key steady phases and transition periods can be defined as

$$\Xi_{s,k,j}^{vb}(x_{\text{new},k}^{vb,j}) = [0, \cdots, x_{\text{new},k}^{vb,j}, \cdots, 0]^T$$

$$\text{VCI}_s(k,j) = \frac{\left\| \mathbf{R}_s^{\text{st}^*\,T} \Xi_{s,k,j}^{vb}(x_{\text{new},k}^{vb,j}) \right\|}{\sum_{j=1}^{J} \left\| \mathbf{R}_s^{\text{st}^*\,T} \Xi_{s,k,j}^{vb}(x_{\text{new,k}}^{vb,j}) \right\|} \quad (33)$$

$$\psi_{t,k,j}^{vb}(x_{\text{new},k}^{vb,j}) = [1, K(x_{\text{new},k}^{vb,j}, x_{t,k,1}^{tr,j}), K(x_{\text{new},k}^{vb,j}, x_{t,k,2}^{tr,j}), \cdots, K(x_{\text{new},k}^{vb,j}, x_{t,k,I}^{\text{tr},j})]^T$$

$$\text{VCI}_t(k,j) = \frac{\left\| \mathbf{w}_t^T \psi_{t,k,j}^{vb}(x_{\text{new},k}^{vb,j}) \right\|}{\sum_{j=1}^{J} \left\| \mathbf{w}_t^T \psi_{t,k,j}^{vb}(x_{\text{new},k}^{vb,j}) \right\|} \quad (34)$$

where $j = 1,2,\cdots,J$ is the variable index, $k = 1,2,\cdots,K_s^{\text{st}}(K_t^{\text{tr}})$ is the sample time point in each phase/period, and $x_{\text{new},k}^{vb,j}$ and $x_{t,k,i}^{tr,j}$, $i = 1,2,\cdots,I$ are the process variables of interest in the current batch and the training batches respectively. For simplicity, the key variables for quality prediction in each phase/period can be determined by the following two mean variable contribution indices

$$\text{MVCI}_s(j) = \frac{1}{K_s^{\text{st}}} \sum_{k=1}^{K_s^{\text{st}}} \text{VCI}_s(k,j) \quad (35)$$

$$\text{MVCI}_t(j) = \frac{1}{K_t^{\text{tr}}} \sum_{k=1}^{K_t^{\text{tr}}} \text{VCI}_t(k,j) \quad (36)$$

where $s = 1,2,\cdots,S$, and $t = 1,2,\cdots,T$. As before, $\sum_{j=1}^{J} \text{MVCI}_s(j) = 1$ and $\sum_{j=1}^{J} \text{MVCI}_t(j) = 1$. Using these two indices, the key variables for quality prediction in each critical-to-quality phase and period can be successfully identified.

Prediction uncertainty

Up to now, the mean values of the weighted parameters in the RVM model have been used for quality prediction in each transition period and in the phase/period combination step. Actually, the weighted parameter is an uncertainty vector which follows a Gaussian distribution, with its variance given in Eqs. 22 and 29. The RVM model can therefore provide a probabilistic prediction result for each quality variable. More specifically, the final prediction result in transition phase t can be calculated as

$$p(\hat{\mathbf{y}}_{\text{new}}^{kc} | \hat{\mathbf{y}}_{\text{new}}^{ph}, \mathbf{y}, \alpha_{s+t-1}, \sigma_{s+t-1}^2) = \int p(\hat{\mathbf{y}}_{\text{new}}^{kc} | \hat{\mathbf{y}}_{\text{new}}^{ph}, \mathbf{w}_{s+t-1}, \sigma_{s+t-1}^2) \times p(\mathbf{w}_{s+t-1} | \hat{\mathbf{y}}_{\text{new}}^{ph}, \mathbf{y}, \alpha_{s+t-1}, \sigma_{s+t-1}^2) d\mathbf{w}_{s+t-1} \quad (37)$$

where the distribution of the weighted parameter $\mathbf{w}_{s+t-1}$ is given in Eq. 27. Hence the distribution of the predicted value $\hat{\mathbf{y}}_{\text{new}}^{kc}$ is also Gaussian, with mean

$$\mu_{y,\text{new}} = \bar{\mathbf{w}}_{s+t-1}^T \psi_{s+t-1}^{kc}(\hat{\mathbf{y}}_{\text{new}}^{ph}) \quad (38)$$

and variance

$$\Sigma_{y,\text{new}} = \sigma_{s+t-1}^2 + \psi^T(\hat{\mathbf{y}}_{\text{new}}^{ph}) \Sigma_{\mathbf{w},s+t-1} \psi(\hat{\mathbf{y}}_{\text{new}}^{ph}) \quad (39)$$

where $\psi_{s+t-1}^{kc}(\hat{\mathbf{y}}_{\text{new}}^{ph}) = [1, K(\hat{\mathbf{y}}_{\text{new}}^{ph}, \hat{\mathbf{y}}_1^{ph}), K(\hat{\mathbf{y}}_{\text{new}}^{ph}, \hat{\mathbf{y}}_2^{ph}), \cdots, K(\hat{\mathbf{y}}_{\text{new}}^{ph}, \hat{\mathbf{y}}_1^{ph})]^T$. As can be seen in Eq. 39, the uncertainty in the quality prediction comes from two sources. The first is the noise variance in the process data, while the second arises from the uncertainty of the prediction model. Compared to the noise, which is inevitable, the uncertainty in the prediction model is related to the process data itself and the model's structure.

In each transition period the prediction uncertainty can be obtained similarly. Through combining the prediction results for different phases and periods, the uncertainty of the combined result may be increased as more uncertain information is incorporated, especially the prediction results in the transition phases. Therefore, in each of the steady phase the prediction uncertainty is given by the cumulative RVM model as $Pu_{\text{new}} = \sum_{y,\text{new}}$. But in each of transition period the uncertainty is

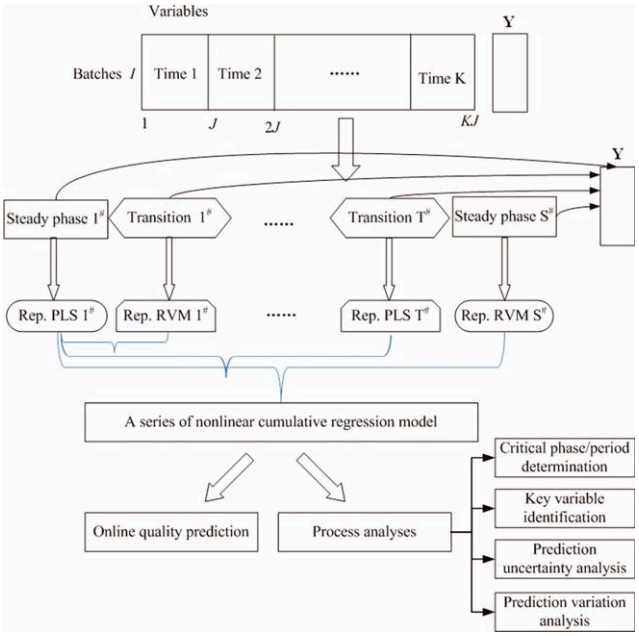

Figure 1. Flowchart of the nonlinear quality prediction and process analysis method.

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

$$Pu_{\text{new}} = \Sigma_{y,\text{new}} + \Sigma^{\text{tr}}_{y,\text{new}} = \sigma^2_{s+t-1} + \psi^T(\hat{\mathbf{y}}^{ph}_{\text{new}})\Sigma_{\mathbf{w},s+t-1}\psi(\hat{\mathbf{y}}^{ph}_{\text{new}}) + \sigma^{\text{tr}\,2}_{t,k} + \psi^T(\mathbf{x}^{\text{tr}}_{\text{new}})\Sigma^{\text{tr}}_{\mathbf{w},t,k}\psi(\mathbf{x}^{\text{tr}}_{\text{new}}) \quad (40)$$

where $\sum^{\text{tr}}_{\text{w},t,k}$ is given in Eq. 22, and $\sigma^{\text{tr2}}_{t,k}$ is the noise variance in the transition period.

Performance evaluation

Given a test dataset covering I_{te} batches, a root mean square error (RMSE) criterion can be used to assess the performance of the method at each time point of the batch process.

$$\text{RMSE}(kc) = \sqrt{\frac{\sum_{i=1}^{I_{te}} \left\|\mathbf{y}_i - \hat{\mathbf{y}}^{kc}_i\right\|^2}{I_{te}}} \quad (41)$$

where $kc = 1,2,\cdots,K$. $\mathbf{y}_i$, $i = 1,2,\cdots,I_{te}$ is the real measured value for each test batch, and $\hat{\mathbf{y}}^{kc}_i$ is the predicted value for test batch i at time point kc.

Another performance index is the batch prediction variance (BPV), which can be defined to evaluate the variation of the predicted results for each batch.

Table 1. Variables Selected for Process Monitoring

No.	Variables	Unit
1	Valve 1 opening	%
2	Valve 2 opening	%
3	Screw stroke	Mm
4	Screw velocity	Mm/s
5	Ejector stroke	Mm
6	Mold stroke	Mm
7	Mold velocity	Mm/s
8	Injection press	Bar
9	Temperature 3	–
10	Temperature 2	–
11	Temperature 1	–

Table 2. Phase Division Results of the Injection Molding Process

Transition Periods	Time Points	Steady Phases	Time Points
Transition period1	1–18		
		Steady phase 1	19–117
Transition period2	118–119		
		Steady phase 2	120–367
Transition period3	368–369		
		Steady phase 3	370–609
Transition period4	610–635		

$$\text{BPV}(i) = \frac{1}{K}\sum_{k=1}^{K}\hat{\mathbf{y}}^2_i(k) - \left[\frac{1}{K}\sum_{k=1}^{K}\hat{\mathbf{y}}_i(k)\right]^2 \quad (42)$$

where $i = 1,2,\cdots,I_{te}$, and $k = 1,2,\cdots,K$. Similarly, a phase/period prediction variance (PPV) can be defined as

$$\begin{aligned}\text{PPV}_s(i) &= \frac{1}{K}\sum_{k=1}^{K^{\text{st}}_s}\hat{\mathbf{y}}^2_i(k) - \left[\frac{1}{K}\sum_{k=1}^{K^{\text{st}}_s}\hat{\mathbf{y}}_i(k)\right]^2 \\ \text{PPV}_t(i) &= \frac{1}{K}\sum_{k=1}^{K^{\text{tr}}_t}\hat{\mathbf{y}}^2_i(k) - \left[\frac{1}{K}\sum_{k=1}^{K^{\text{tr}}_t}\hat{\mathbf{y}}_i(k)\right]^2\end{aligned} \quad (43)$$

where $s = 1,2,\cdots,S$, and $t = 1,2,\cdots,T$. These indices can be used to evaluate the prediction variations in each of the steady phase and transition period, and also among batches. Compared to the steady phases, the prediction variation may be more significant in the transition periods. This is because the data in the transition period varies more making the model uncertainty also more significant.

Figure 1 presents a summary description of the proposed nonlinear quality prediction and process analysis method.

Further discussions

In the present work, we have previously assumed that the batch process has already been divided into different steady phases and transition periods. Although several efficient phase division methods have been developed in the past years, this is still an open question, thus needs further investigations. Particularly, when the termination times of different batches are varying, it is a challenge to divide the

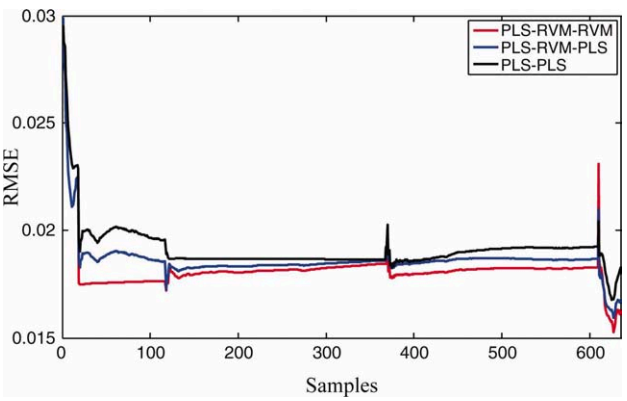

Figure 2. Quality prediction of different methods.

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

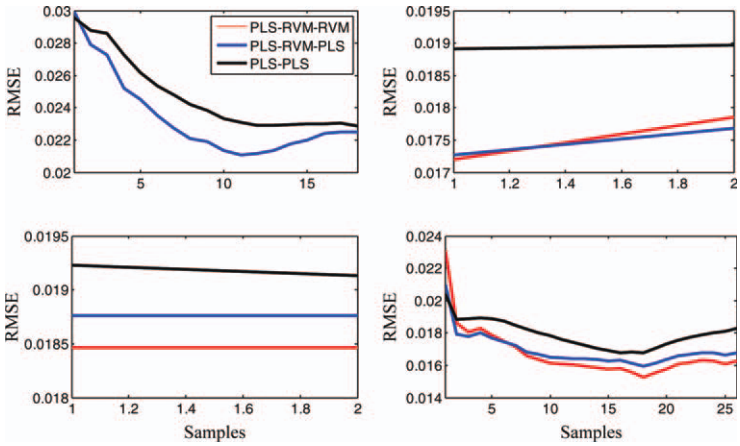

Figure 3. Detailed prediction results of the three methods in transition phases.

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

process into different phases, and the duration time of each phase is also very difficult to determine. Also, how to incorporate the nonlinear data behavior for phase division in the batch process is another research issue, which may improve the phase division result, especially when the nonlinearities of phases/transition periods are different from each other.

An injection molding example

Injection molding is a common multiphase batch process. Its operation phases might be defined as injection, packing-holding, plastication and cooling. The weight of the final product might be selected as the key quality variable. Online process measurements commonly include temperatures, pressures and the screw velocity. Some typical values from bench scale molding experiments are tabulated in Table 1. A detailed description of the equipment and the experimental design can be found in reference.[10] A dataset covering 150 batches has been collected, among which 100 batches were used for model training and the other 50 for testing. Each batch had 635 sampling time points. Based on the repeatability factors, the process was divided into 3 steady and 4 transition periods, as shown in Table 2.

A linear or nonlinear prediction model was developed for each phase and period. For each steady phase a PLS model with 4 latent variables was constructed. In each transition period the representative RVM model had a kernel parameter of 5.5. Three methods for evaluating the quality predictions

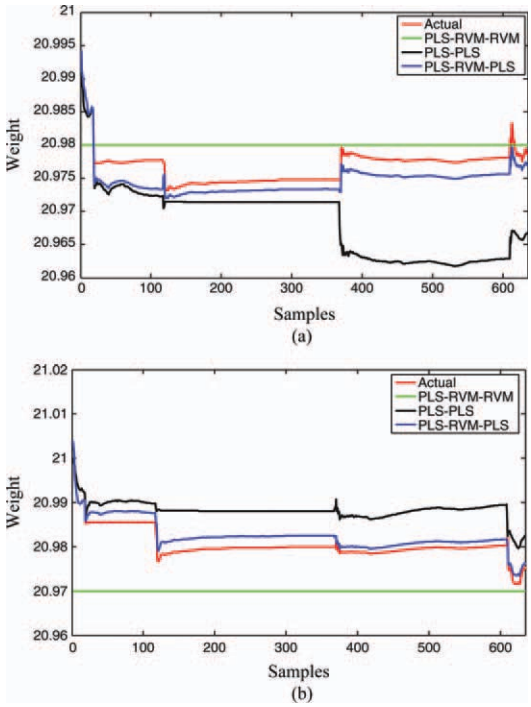

Figure 4. Quality predictions for two test batches, (a) First batch; (b) Second batch.

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

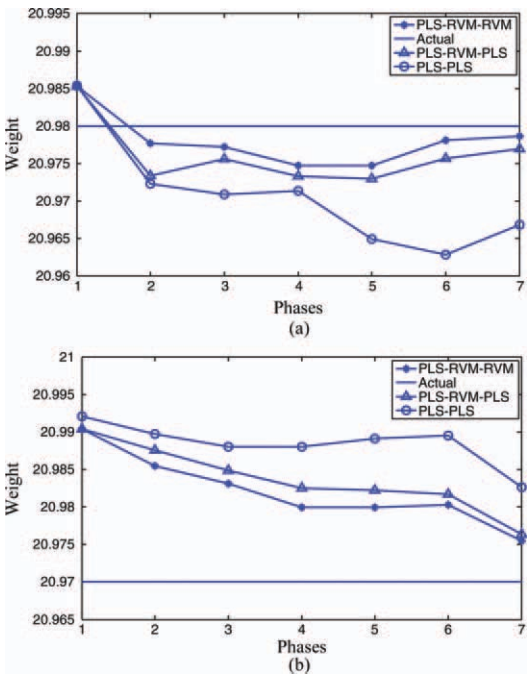

Figure 5. End of phase quality predictions for two test batches, (a) First batch; (b) Second batch.

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

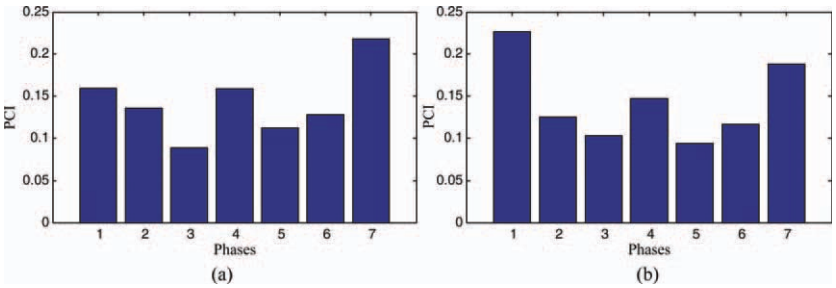

Figure 6. Critical-to-quality phase evaluation results, (a) Training batches; (b) Test batches.

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

were compared. In the first method, each of the steady phases and transition periods was modeled with a linear PLS model, and the phase prediction results were combined through the linear PLS model to the phase level. In the second method, an RVM model was used for modeling each transition period, and the phase prediction results were combined through a linear PLS model. The nonlinear information in the transition periods was better represented by this second method, and that improved quality prediction performance. In the third method, the modeling approaches for the steady phases and transition periods were the same as in the second method, but the phase combination model was a

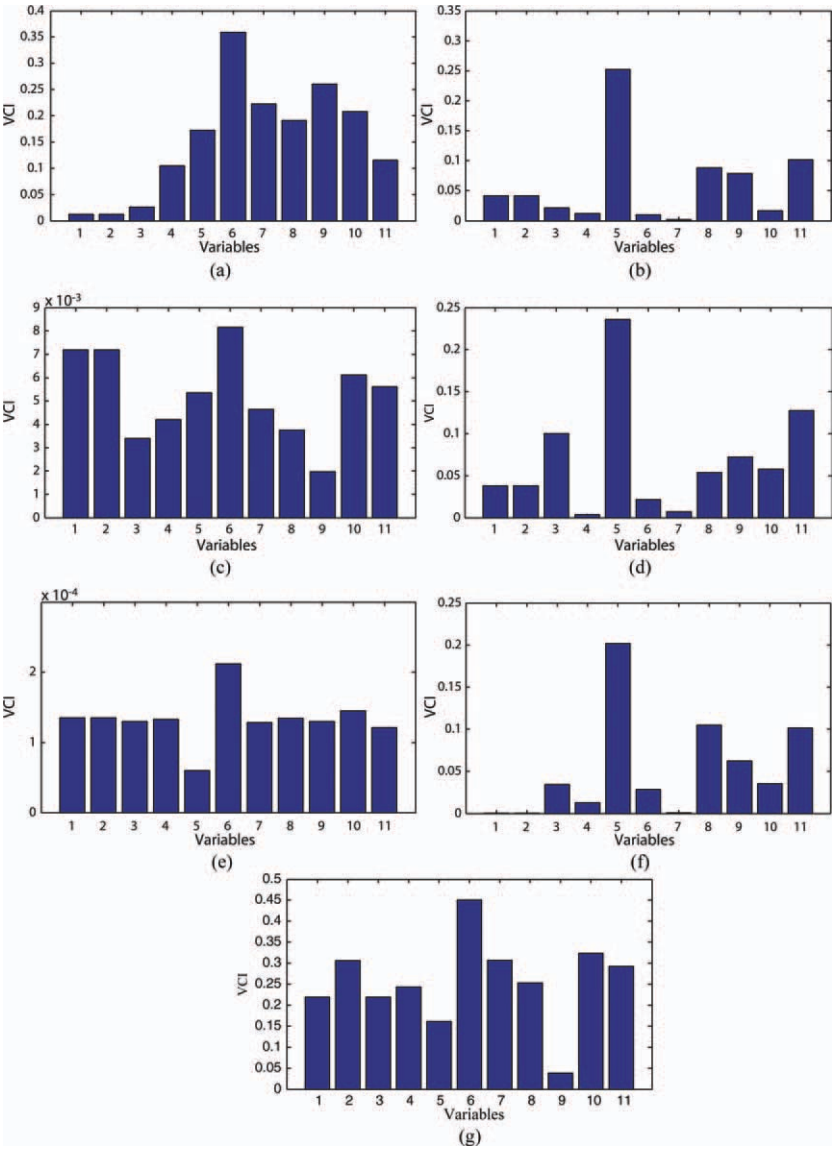

Figure 7. Key variable identification results, (a) Transition phase 1; (b) Steady phase 1; (c) Transition Phase 2; (d) Steady phase 2; (e) Transition phase 3; (f) Steady phase 3; (g) Transition phase 4.

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

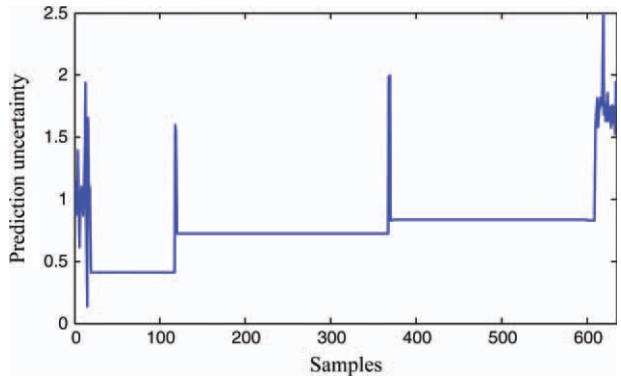

Figure 8. Mean values of the quality prediction uncertainty for the test batches.

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

nonlinear model. This further improved the quality prediction performance. These three methods are termed the PLS-PLS, PLS-RVM-PLS, and PLS-RVM-RVM approaches.

The RMSE values for the test batches using these three methods are presented in Figure 2. The results with the PLS-RVM-RVM method were consistently the best. Quality prediction in each transition period was particularly improved compared to the linear PLS method, detailed results of which are provided in Figure 3. The performance has been enhanced through the nonlinear combination of the prediction results for different operational phases. In the first transition period, the prediction results of the PLS-RVM-RVM and PLS-RVM-PLS methods are the same because there was no results combination in this period. Figure 4 compares the prediction results of the three methods for two test batches together with the measured values for comparison. The nonlinear method clearly gave more accurate predictions than the linear method. Detailed prediction results at the ends of different steady and transition phases/periods are shown in Figure 5, and again the nonlinear treatment produced the best results for both batches.

To determine the importance of each phase, the phase contribution index for each phase and period was calculated. The mean values for the training and test batches are given in Figures 6a, b, respectively. Both figures suggest that the first, fourth and final phases/periods were the most important for quality. The first and last of these are transition periods, while the fourth one corresponds to a steady phase. Transition periods apparently have an important impact on the variation of product quality. Next, the importance of each process variable in each phase and period was evaluated using the variable contribution index. The key results are presented in Figure 7. Clearly the sixth process variable (Table 1), the mold stroke, is a key variable common to all four transition periods. Similarly, in the three steady phases, a common key variable is the fifth variable in Table 1, the ejector stroke.

Finally, the quality prediction uncertainty and variation were evaluated using the uncertainty and the variance indices previously defined. Figure 8 shows the mean values of the prediction uncertainty index for all of the test batches plotted against time. The uncertainty of the transition periods is clearly much more significant than that of the steady phases. However, the prediction uncertainties in the previous phases/periods are accumulated in the phase combination step, so the uncertainty increases slightly in the late phases/periods. The batch prediction variances of the three methods are given in Table 3, where the nonlinear method shows the smallest prediction variance. The detailed phase prediction variances of the three methods are tabulated in Table 4. Compared to the results in steady phases, the prediction variances in transition periods are much bigger, especially in the first and last transition periods. Variation in the prediction results is more significant in the transition periods than in the steady phases. Still, the prediction variation of the PLS-RVM-RVM method is much smaller than those of the other methods.

Table 3. Batch Prediction Variance of Different Methods for the Testing Dataset

Methods	PLS-PLS	PLS-RVM-PLS	PLS-RVM-RVM
BPV	7.9086e-006	9.8092e-006	6.0172e-006

Table 4. Phase Prediction Variance of Different Methods for the Testing Dataset

Methods/ Phases	1	2	3	4	5	6	7
PLS-PLS	0.0914	0.0097	0.0007	0.0001	0.0055	0.0089	0.0251
PLS-RVM-PLS	0.1023	0.0051	0.0021	0.0011	0.0001	0.0037	0.0350
PLS-RVM-RVM	0.1023	0.0003	0.0080	0.0017	0.0001	0.0027	0.0804

Conclusions

In this study an RVM model has been adapted for nonlinear quality prediction with multiphase batch processes. Compared to phase-based PLS prediction models, nonlinear regression modeling improves quality prediction performance. Introduction of a series of nonlinear accumulative regression models for combining the results of different phases of the process further improves the prediction accuracy. To evaluate the importance of each phase for quality, a phase contribution index has been defined. Similarly, a variable contribution index has been proposed to identify the key variables in each steady and transition phase of the process. When applied to injection molding data, the nonlinear methods improved the quality prediction performance, but also our understanding of the injection molding process.

Acknowledgments

This work was supported in part by project National 973 (2009CB320603), the National Natural Science Foundation of China (NSFC) (61004134), NSFC and Hong Kong Research Grant Council Joint Project (N-HKUST 639109) and also by the China Postdoctoral Science Foundation (20090461370).

Literature Cited

1. Nomikos P, MacGregor JF. Monitoring batch processes using multiway principal component analysis. *AIChE J*. 1994;44:1361–1375.
2. Nomikos P, MacGregor JF. Multi-way partial least square in monitoring batch processes. *Chem Intel Lab Syst*. 1995;30:97–108.
3. Chen J, Liu KC. On-line batch process monitoring using dynamic PCA and dynamic PLS models. *Chem Eng Sci*. 2003;57:63–75.
4. Ramaker HJ, van Sprang ENM, Westerhuis JA, Smilde AK. Fault detection properties of global, local and time evolving models for batch process monitoring. *J Process Control*. 2005;15:799–805.
5. Chen JH, Chen HH. On-line batch process monitoring using MHMT-based MPCA. *Chem Eng Sci* 2006;61:3223–3239.
6. van Sprang ENM, Ramaker HJ, Westerhuis JA, Guiden SP, Smilde AK. Critical evaluation of approaches for on-line batch process monitoring. *Chem Eng Sci*. 2008;57:3979–3991.

7. Chen T, Zhang J. On-line multivariate statistic monitoring of batch processes using Gaussian mixture model. *Comput Chem Eng*. 2010; 34:500–507.
8. Undey C, Cinar A. Statistical monitoring of multistage, multiphase batch processes. *IEEE Contr Syst Mag*. 2002;22:53–63.
9. Lu NY, Gao FR, Wang FL. A sub-PCA modeling and online monitoring strategy for batch processes. *AIChE J*. 2004;50:255–259.
10. Lu NY, Gao FR. Stage-based process analysis and quality prediction for batch processes. *Ind Eng Chem Res*. 2005;44:3547–3555.
11. Yao Y, Gao FR. A survey on multistage/multiphase statistical modeling methods for batch processes. *Annu Rev Control*. 2009;33:172–183.
12. Muthuswamy K, Srinivasan R. Phase-based supervisory control for fermentation process development. *J Process Control*. 2003;13:367–382.
13. Liu J, Wang DSH. Fault detection and classification for a two-stage batch process. *J Chemometrics*. 2008;22:385–398.
14. Camacho J, Pico J, Ferrer A. Multi-phase analysis framework for handling batch process data. *J Chemometrics* 2008;22:632–643.
15. Zhao CH, Wang FL, Mao ZZ, Lu NY, Jia MX. Quality prediction based on phase-specific average trajectory for batch processes. *AIChE J*. 2008;54:693–705.
16. Yu J, Qin SJ. Multiway Gaussian mixture model based multiphase batch process monitoring. *Ind Eng Chem Res*. 2009;48:8585–8594.
17. Zhao CH, Gao FR, Wang FL. Phase-phased joint modeling and spectroscopy analysis for batch process monitoring. *Ind Eng Chem Res*. 2010;49:669–681.
18. Zhao CH, Wang FL, Lu NY, Jia MX. Stage-based soft-trainsition multiple PCA modeling. *J Process Control*. 2007;17:728–741.
19. Yao Y, Gao FR. Phase and transition based batch process modeling and online monitoring. *J Process Control*. 2009;19:816–826.
20. Ge ZQ, Zhao LP, Yao Y, Song ZH, Gao FR. Utilizing transition information in online quality prediction of multiphase batch processes. *J Process Control*. In press.
21. Willis MJ, Montegue GA, Massimo CD, Tham MT, Morris AJ. Artificial neural networks in process estimation and control. *Automatica* 1992;28:1181–1187.
22. Vapnik VN. *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1995.
23. Suykens JAK, Gestel TV, Brabanter JD, Moor BD, Vandewalle J. *Least Squares Support Vector Machines*. Singapore: World Scientific, 2002.
24. Adebiyi OA, Corripio AB. Dynamic neural networks partial least squares (DNNPLS) identification of multivariable processes. *Comput Chem Eng*. 2003;27:143–155.
25. Li CF, Ye H, Wang GZ, Zhang J. A recursive nonlinear PLS algorithm for adaptive nonlinear process modeling. *Chem Eng Tech*. 2005;28:141–152.
26. Bylesjo M, Rantalainen M, Nicholson JK, Holmes E, Trygg J. K-OPLS package: Kernel-based orthogonal projections to latent structures for prediction and interpretation in feature space. *BMC Bioinformatics* 2008;9:106.
27. Zhang YW, Teng YD, Zhang Y. Complex process quality prediction using modified kernel partial least squares. *Chem Eng Sci*. 2010;65:2153–2158.
28. Tipping ME. Sparse Bayesian learning and the relevance vector machine. *J Machine Learning Res*. 2001;1:211–244.

Manuscript received Feb. 20, 2011, and revision received May 12, 2011.