Chemical Engineering Science 93 (2013) 96–109

Contents lists available at SciVerse ScienceDirect

Chemical Engineering Science

journal homepage: www.elsevier.com/locate/ces

A Bayesian model averaging based multi-kernel Gaussian process regression framework for nonlinear state estimation and quality prediction of multiphase batch processes with transient dynamics and uncertainty

Jie Yu*, Kuilin Chen, Mudassir M. Rashid

Department of Chemical Engineering, McMaster University, Hamilton, Ontario, Canada L8S 4L7

HIGHLIGHTS

- A novel nonlinear state estimation and quality prediction method is proposed.
- Multi-kernel Gaussian process regression models to deal with multiphase process.
- Bayesian model averaging to characterize between-phase transient dynamics.
- Local models are dynamically and adaptively integrated for transitional stages.
- The proposed method provides accurate and reliable state and quality estimations.

ARTICLE INFO

Article history:
Received 19 October 2012
Received in revised form
2 January 2013
Accepted 31 January 2013
Available online 13 February 2013

Keywords:
Nonlinear state estimation
Quality prediction
Multiphase batch process
Between-phase transient dynamics
Multi-kernel Gaussian process regression
Bayesian model averaging

ABSTRACT

Batch processes are characterized by inherent nonlinearity, multiplicity of operating phases, between-phase transient dynamics and batch-to-batch uncertainty that pose significant challenges for accurate state estimation and quality prediction. Conventional multi-model strategies, however, may be ill-suited for multiphase batch processes because the localized models do not specially take into account the complex transient dynamics between two consecutive operating phases. In this study, a novel Bayesian model averaging based multi-kernel Gaussian process regression (BMA-MKGPR) approach is proposed for state estimation and quality prediction of nonlinear batch processes with multiple operating phases and between-phase transient dynamics. A kernel mixture model strategy is first used to identify the different operating phases of batch processes and then the multi-kernel GPR models are built for all the identified phases. Further, the between-phase transitional stage is determined by the posterior probabilities of measurement samples with respect to the two consecutive phases so that the Bayesian model averaging strategy can be designed to incorporate the two localized GPR models for handling the between-phase transient dynamics. For an arbitrary test sample within the transitional stage, its posterior probabilities with respect to the local models corresponding to the two consecutive phases are set as the adaptive weightings to integrate the corresponding local GPR models for state estimation and quality prediction. The proposed BMA-MKGPR approach is applied to a multiphase batch polymerization process and the result comparison demonstrates that the presented method can effectively handle multiple nonlinear operating phases, between-phase transient dynamics and process uncertainty with fairly high prediction accuracies.

1. Introduction

Batch processes are commonly used in chemical, materials, pharmaceutical, biotechnology and semiconductor industries for the production of low-volume while high-value-added commodities. One of the main challenges in batch process operation is that the product quality measurements are often unavailable during the batch operation and may not be taken through off-line laboratory analysis until the completion of the entire batch. Even though different kinds of analyzers may be implemented to measure quality variables in an online fashion, their reliability is subject to certain degradations over a period of time. Meanwhile,

* Corresponding author. Tel.: +1 905 525 9140x27702; fax: +1 905 521 1350.
E-mail address: jieyu@mcmaster.ca (J. Yu).

0009-2509/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.ces.2013.01.058

online state estimations and quality predictions are crucially important for achieving optimal control and continuous improvement of batch or semi-batch processes (Bogaerts and Wouwer, 2004; Yu and Qin, 2009b; Borchert and Sundmacher, 2012). The accurate estimations of quality and state variables allow for real-time control actions and operational corrections, which can further prevent deteriorated product quality and mitigate the number of rejected batches (Kadlec et al., 2009).

Predictive model based soft sensor techniques have received considerable interest for online state estimation and reliable quality prediction of batch processes. In general, soft sensor methods can be classified into two categories: first-principle model based and data-driven statistical approaches. Though the mechanistic models of processes can lead to Kalman filter or observer based state and quality estimations, the development of first-principle models usually requires in-depth understanding of the physical, chemical and biological relationships in batch processes, which may not be readily obtained in practice (Soroush, 1997; Bosca and Fissore, 2011; Mangold, 2012). Meanwhile, these mechanistic models are typically constructed under the ideal operating conditions from original process design whereas the actual operating scenarios and the corresponding process dynamics can be significant different. Moreover, it is quite time-consuming to develop comprehensive first-principle models that can precisely account for the complex nonlinear dynamics and intricate variable interactions in real processes through a representative set of model equations and the identified parameter values. Although the mechanistic model and Kalman filter based soft sensors have been applied for state estimations, the linear assumption of basic Kalman filter may not be appropriate for many nonlinear batch processes and thus the state estimator designs using nonlinear process models become necessary. The most common solution to the state estimation of nonlinear processes is through the extended Kalman filter (Kozub and MacGregor, 1992; Régnier et al., 1996; Bogaerts and Wouwer, 2004). However, the local linearization of nonlinear process at any given sampling instant may potentially introduce some estimation errors, which can lead to divergence over time. In order to overcome the drawback of the extended Kalman filter based approach, other types of nonlinear state estimators have been employed such as unscented Kalman filter, ensemble Kalman filter and high-order extended Kalman filters (Kandepu et al., 2008). Nevertheless, the development of the required nonlinear first-principle models remains challenging in practice.

Alternately, data-driven soft sensor techniques have attracted increasing attention because the empirical models built from process measurement data can significantly alleviate the requirements of in-depth process knowledge and minimize the model development effort. Modern instrumentation and measurement techniques allow for large amounts of operational data to be collected, stored and analyzed, which enables the data-driven soft sensor as a viable solution for nonlinear state estimation and quality prediction. Traditional data-driven modeling techniques used in state and quality estimation include nonlinear dynamic models such as the nonlinear moving-average models and the nonlinear autoregressive moving-average exogenous models (Fortuna et al., 2009). These time-series models are not well suited for online quality prediction in practice due to their limited capability of handling missing values, inadequate sampling intervals and process outliers. The well-known multivariate statistical methods such as principal component regression (PCR) and partial least squares (PLS) are widely applied to build inferential models for state and quality estimations and are capable of dealing with the high-dimensional and collinear process data (Zhang et al., 2008; Godoy et al., 2011). Furthermore, kernel and recursive PLS algorithms are proposed to tackle the strong nonlinearity and time-varying dynamics in industrial processes (Facco et al., 2009; Galicia et al., 2012). While this type of modified PLS techniques can update the models towards new operating conditions recursively, the recursive PLS methods still cannot cope with the abrupt changes in operating conditions and may adapt to the model shifts too excessively. Though multi-model PLS methods can handle nonlinear batch processes by fitting multiple local linear models, they are unable to account for the between-phase transitional stages that may have different transient dynamics from the individual operating phases (Sharmin et al., 2006; Yu, 2012c).

Meanwhile, machine learning methods such as artificial neural networks (ANN) and support vector machines (SVM) are applied to nonlinear state estimations and quality predictions (Gonzaga et al., 2009; Kadlec et al., 2009). Although ANN is able to model the nonlinear relationships between the process input and output variables, it tends to converge towards the local minimum of multi-peak nonlinear functions so that the obtained model predictions are often suboptimal. Furthermore, the generalization capability of ANN models may not be satisfactory for reliable state and quality predictions. To address these issues, kernel based methods including support vector regression (SVR), least squares support vector regression (LSSVR) and their online updated forms such as adaptive kernel learning networks (AKL) and selected recursive kernel learning (SRKL) are adopted for soft sensor development (Kadlec et al., 2009; Kaneko et al., 2011). With the structural risk minimization criterion, these kernel learning techniques can achieve relatively higher prediction accuracies with even fewer data samples (Yu, 2012a, 2013). These kernel learning based soft sensor techniques typically rely on a global model and aim to achieve a universal generalization performance. However, such strategies may lead to inaccurate state and quality estimations across different operating phases of batch process with shifting dynamics (Yao and Gao, 2009). To overcome this limitation, just-in-time (JIT) learning method is proposed by building local models with selected training samples based on a similarity criterion to predict the output variables (Fujiwara et al., 2009, 2012; Chen et al., 2011). However, the major challenges in JIT learning methods include the construction of an appropriate similarity index and the adaptive selection of local models for state and quality estimations.

As a type of inferential learning technique, Bayesian inference methods are applied to state estimation and quality prediction of batch processes due to its capability of strong statistical learning from process data as well as the efficient utilization of any prior process knowledge (Yu and Qin, 2008, 2009a; Yu, 2012c). The process and measurement uncertainty including missing values, outliers, biases, drifts and sampling delays during batch or continuous operation can be effectively handled by Bayesian inference strategy (Khartibisepehr and Huang, 2008; Yu, 2012a). Meanwhile, Bayesian inference method has been integrated with nonlinear kernel function based support vector regression for inferential modeling and quality prediction of chemical processes, where the parameter values of kernel functions are optimized through Bayesian strategy (Yan et al., 2004). However, such soft sensor model does not take into account the shifting dynamics due to different operating conditions or phases. More recently, a Bayesian inference based Gaussian process regression (GPR) approach is developed for quality prediction of nonlinear and non-Gaussian chemical processes with switching dynamics due to different operating modes, though the transient dynamics that occur during the transitional stage between two consecutive operating modes are neglected (Yu, 2012b) .

In this study, a novel Bayesian model averaging (BMA) based multi-kernel Gaussian process regression (MKGPR) approach is proposed for state estimation and quality prediction of nonlinear

batch processes with multiple operating phases and between-phase transient dynamics. First, the kernel mixture model (KMM) is built from the input measurement data to discriminate the different operating phases in batch processes. Then, MKGPR models combining multiple localized models are constructed to characterize the shifting process dynamics of all the identified operating phases. Further, Bayesian inference strategy is used to estimate the posterior probabilities of new measurement samples with respect to different operating phases. Thus, the between-phase transitional stages with transient dynamics are identified and characterized through Bayesian model averaging strategy, where the posterior probabilities of the adjacent phases are set as adaptive weights to incorporate the local GPR models for state and quality predictions within the transitional stages. It should be noted that the kernel mixture model along with Bayesian inference strategy is needed to accurately isolate different phases as well as between-phase transitional stages in batch processes even though the prior process knowledge of multiple phases may be known. The reasons include that (i) batch processes often have significant batch-to-batch variations and random uncertainty so that each phase may not have the identical duration across different batches, (ii) the KMM method not only identify different phases but also between-phase transitional stages with transient dynamics, and (iii) the quantitative and precise information of phase divisions throughout the entire batch duration is often unavailable in practice.

The remainder of the paper is organized as follows. Section 2 briefly reviews the artificial neural network and support vector regression based soft sensor methods. Then the novel Bayesian model averaging based multi-kernel Gaussian process regression approach for state estimation and quality prediction of multiphase batch processes is developed in Section 3. Section 4 demonstrates the effectiveness of the proposed BMA-MKGPR approach using the simulated batch polymerization process. Finally, the concluding remarks are drawn in Section 5.

2. Preliminaries

2.1. Neural network

Artificial neural network techniques are proven capable of handling process nonlinearity and thus have been widely applied to soft sensor based state and quality estimations (Kadlec et al., 2009). The architecture of the model is in the form of a multi-layer feed-forward network where the nodes are arranged across successive layers and the information flows unidirectionally from the input layer towards the output layer through intermediate hidden layers. The network nodes are essentially the processing units with numerical weights w that represent the relative strengths of the node connections for transferring information from one layer to another. The summation function

$$y_j = \sum_i x_i w_{j,i} \tag{1}$$

computes the weighted sum of all input elements x_i of a node. Further, a sigmoid function

$$y_j^{(T)} = \frac{1}{1+e^{-y_j}} \tag{2}$$

is used to transform the output signal y_j to $y_j^{(T)}$ before it is passed to the next layer. To optimize the network parameter values, the process input and output measurements are used for model learning that involves a forward-propagating step to evaluate the model accuracy followed by backward propagation to update the weights of network nodes (Hecht-Nielsen, 1989; Bishop, 1995).

In this work, a multi-model neural network (MMNN) strategy is used to handle multiphase batch processes by constructing local neural network models for various operating phases. First, a phase identification procedure and Bayesian inference strategy are used to classify the samples into the respective operating phases based on the posterior probabilities. Then, localized neural network models are developed for all the identified phases. Thus, the multi-model neural network models can be built for state estimation and quality prediction of multiphase batch processes.

2.2. Support vector regression

Support vector machine is a type of machine learning technique with attractive merits for pattern classification and model regression (Vapnik, 1998). In the support vector regression framework with N process input variables x_k and the corresponding quality variables y_k, the general regression problem can be formulated as

$$F(X_k) = \langle \omega, \Phi(X_k) \rangle + c \tag{3}$$

where ω is the regression vector in the kernel feature space $\mathcal{F}$, $\Phi(x_k)$ is a nonlinear mapping function from the original measurement space into the high-dimensional feature space $\mathcal{F}$, c is the bias term and $\langle \cdot, \cdot \rangle$ represents the inner product. The SVR model is based on the ϵ-insensitive loss function as follows:

$$L(y_i - F(x_i)) = \begin{cases} 0 & \text{if } |y_i - F(x_i)| \le \epsilon \\ |y_i - F(x_i)| - \epsilon & \text{otherwise} \end{cases} \tag{4}$$

Then the optimization problem can be expressed as

$$\min_{\omega,c,\xi,\xi^*} \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*)$$
$$\text{s.t.} \begin{cases} y_i - \langle \omega, \Phi(x_k) \rangle - c \le \epsilon + \xi_i \\ \langle \omega, \Phi(x_k) \rangle + c - y_i \le \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0, \quad i = 1,2,\ldots,N \end{cases} \tag{5}$$

where ϵ is the maximum value of the tolerable error, ξ_i and ξ_i^* are slack variables, $\|\cdot\|$ denotes the Euclidean norm, and C is a regularization parameter (Gunn, 1998; Smola and Schölkopf, 2004).

The dual form of Eq. (5) becomes a quadratic programming (QP) problem as follows:

$$\min_{\eta,\eta^*} \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\kappa(x_i,x_j)(\eta_i - \eta_i^*)(\eta_j - \eta_j^*) + \epsilon\sum_{i=1}^{N}(\eta_i + \eta_i^*) - \sum_{i=1}^{N} y_i(\eta_i + \eta_i^*)$$
$$\text{s.t.} \begin{cases} 0 \le \eta_i,\ \eta_i^* \le C, \quad i = 1,2,\ldots,N \\ \sum_{i=1}^{N}(\eta_i + \eta_i^*) = 0 \end{cases} \tag{6}$$

where η and η^* are the Lagrange multipliers and $\kappa(x_i,x_j)$ is a kernel function given by $\kappa(x_i,x_j) = \Phi(x_i)^T\Phi(x_j) = \kappa_{ij}$. The solution of the above QP problem results in the optimum values of η_i and η_i^* and the data points corresponding to nonzero values of $(\eta_i - \eta_i^*)$ are termed as support vectors (Drucker et al., 1997). In this study, a multi-kernel support vector regression (MKSVR) approach is implemented where various kernel functions are weighted and summed in order to characterize the multiphase dynamics of batch processes (Qiu and Lane, 2009).

3. Bayesian model averaging based Gaussian process regression approach for state and quality estimation

Batch processes are typically characterized by the strong nonlinearity, time-varying dynamics, batch-to-batch variations, multiple operating phases and random uncertainty, which pose challenges on soft sensor model development, online state

estimations and quality predictions. Though multiple localized models can handle multiphase operation, they may not effectively characterize the transient dynamics during the transitional stages between consecutive phases (Yu, 2011a, 2011b). In order to address the challenges of shifting phases and between-phase transient dynamics, a novel Bayesian model averaging based multi-kernel Gaussian process regression approach is developed for state estimation and quality prediction of multiphase batch processes. First, the different operating phases in batch processes are identified through the estimated kernel mixture models from training data. Then the multi-kernel Gaussian process regression models are built to characterize the identified multiple phases. Further, the between-phase transitional stages are isolated through Bayesian inference strategy so that the Bayesian model averaging method is designed to specifically capture the transient dynamics by using the posterior probability based adaptive weightings on the two consecutive local models. In this way, both the within-phase and between-phase dynamics can be adaptively modeled for more accurate state estimations and quality predictions of batch processes.

For a multiphase batch process, the training data of input process variables are expressed in the form of three-dimensional matrix $\tilde{X} \in \mathbb{R}^{I\times J\times K}$, where I represents the number of batches, J denotes the number of measurement variables and K is the number of sampling instances. Firstly, the dynamic time warping (DTW) technique is used to synchronize batches with unequal lengths (Kassidas et al., 1998). Then, the three-dimensional matrix $\tilde{X}$ is converted into a two-dimensional matrix $X \in \mathbb{R}^{IK\times J}$ through variable-wise unfolding with the row vectors representing various batches at every sampling instant and the columns corresponding to different process variables, as shown in Fig. 1. The unfolded matrix X can be expressed in the following block matrix format:

$$X = [X_{(k=1)}^T \;\; X_{(k=2)}^T \;\; \cdots \;\; X_{(k=K)}^T]^T \tag{7}$$

The corresponding three-dimensional output data matrix of state and quality variables can be unfolded in the same fashion as follows:

$$Y = [Y_{(k=1)}^T \;\; Y_{(k=2)}^T \;\; \cdots \;\; Y_{(k=K)}^T]^T \tag{8}$$

Since the time-series data of batch process need to be divided into multiple operating phases and between-phase transitional stages through kernel mixture model and Bayesian inference strategy, the regular batch-wise unfolding may not be appropriate as different sampling instants will be stacked with all the process or quality variables. Thus, the variable-wise unfolding is used to retain the time-series dynamic trajectories along all different variables for phase division purpose.

With the unfolded data, the kernel mixture model is estimated in order to identify the Q different operating phases with C_q $(q=1,2,\ldots,Q)$ denoting each individual phase and N_q representing the number of samples within each phase. Using Q different kernel functions, the posterior probability of an arbitrary training sample $x(i,k)$ from the i-th batch and k-th sampling instance can be estimated through Bayesian inference strategy as follows:

$$p(\Xi_q|x(i,k)) = \frac{p_q p(x(i,k)|\Xi_q)}{\sum_{q'=1}^{Q} p_{q'} p(x(i,k)|\Xi_{q'})} \tag{9}$$

where Ξ_q represents local kernel density function corresponding to an individual operating phase, p_q denotes the prior probability of this phase and $p(x(i,k)|\Xi_q)$ is the likelihood of sample $x(i,k)$ given the local kernel density Ξ_q. Assume that the Gaussian kernel density functions are used and thus the likelihood function can be expressed as

$$p(x(i,k)|\Xi_q) = \frac{1}{(2\pi)^{J/2}|\Sigma_q|^{1/2}} \cdot \exp\left\{-\frac{1}{2}(x(i,k)-\mu_q)^T \Sigma_q^{-1}(x(i,k)-\mu_q)\right\} \tag{10}$$

where $\mu_q \in \mathbb{R}^J$ is the mean vector representing the center of kernel Ξ_q and $\Sigma_q \in \mathbb{R}^{(J\times J)}$ is the corresponding covariance matrix. The parameters θ of kernel mixture model include the prior probabilities as well as the means and covariances of Gaussian kernel density functions (Titsias and Likas, 2001). In the kernel mixture model learning stage, the following log-likelihood function is set as the objective function as

$$\mathcal{L}(\theta) = \sum_{q=1}^{Q}\sum_{k=1}^{K}\sum_{i=1}^{I} \log p(x(i,k),C_q) \tag{11}$$

and the probability optimization problem can be formulated as follows in order to estimate the model parameter values:

$$\theta^* = \underset{\theta}{\operatorname{argmax}}\, \mathcal{L}(\theta) \tag{12}$$

In this work, the expectation-maximization (EM) algorithm is adopted for model parameters estimation. In the E-step, the model parameters are first initialized and the posterior probabilities are then evaluated for all different kernel density functions using Eq. (9) and the current parameter values $\theta^{(t)}$. In the subsequent M-step, the log-likelihood function is maximized to update the model parameters $\theta^{(t+1)}$ as follows:

$$\mu_q^{(t+1)} = \frac{\sum_{i=1}^{I}\sum_{k=1}^{K} p_q^{(t)}(\Xi_q|x(i,k))x(i,k)}{\sum_{i=1}^{I}\sum_{k=1}^{K} p_q^{(t)}(\Xi_q|x(i,k))} \tag{13}$$

$$\Sigma_q^{(t+1)} = \frac{\sum_{i=1}^{I}\sum_{k=1}^{K} p_q^{(t)}(\Xi_q|x(i,k))(x(i,k)-\mu_q^{(t+1)})(x(i,k)-\mu_q^{(t+1)})^T}{\sum_{i=1}^{I}\sum_{k=1}^{K} p_q^{(t)}(\Xi_q|x(i,k))} \tag{14}$$

$$p_q^{(t+1)} = \frac{\sum_{i=1}^{I}\sum_{k=1}^{K} p_q^{(t)}(\Xi_q|x(i,k))}{IK} \tag{15}$$

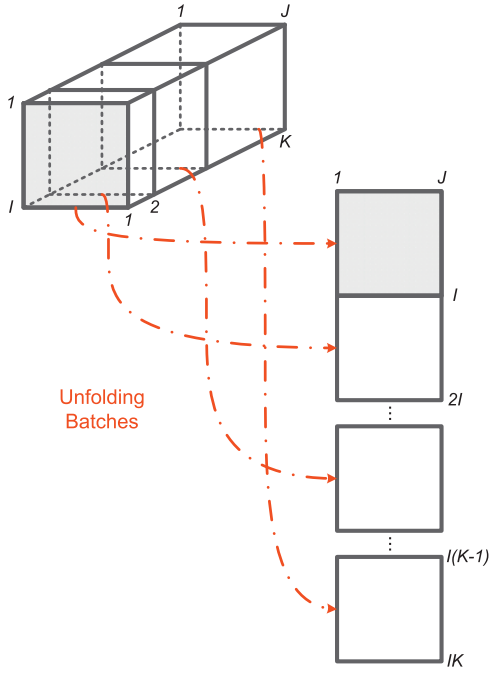

Fig. 1. Illustration of multiway input and output data matrix unfolding.

The EM algorithm is implemented in an iterative way until the parameter values converge to the optimal solution (Titsias and Likas, 2001). The phase identification procedure results in Q subsets of process input data as

$$X = [X^{(1)^T}\ X^{(2)^T}\ \ldots\ X^{(Q)^T}]^T \tag{16}$$

and the corresponding subsets of output data as

$$Y = [Y^{(1)^T}\ Y^{(2)^T}\ \ldots\ Y^{(Q)^T}]^T \tag{17}$$

With different operating phases identified from kernel mixture model, the Bayesian inference based posterior probabilities of test samples can be estimated to classify the measurement samples into different local operating phases. For an arbitrary test sample x_t, it is classified as originating from the q_t-th operating phase with the maximum posterior probability as follows:

$$q_t = \underset{1 \le q \le Q}{\operatorname{argmax}}\ p(\Xi_q|x_t) \tag{18}$$

with the posterior probability satisfying

$$p(\Xi_q|x_t) \ge \alpha \tag{19}$$

where α denotes the specified statistical significance level and is used as the minimum threshold of posterior probability for a nontransient operating phase. The above condition is used to exclude the measurement samples that are located in the transitional stage between two consecutive phases. If all the posterior probabilities are less than the statistical significance level as follows:

$$p(\Xi_q|x_t) < \alpha, \quad q = 1,2,\ldots,Q \tag{20}$$

then the measurement sample is classified as originating from the between-phase transitional stage

$$q_t \in \begin{cases} [q_t\{1\},q_t\{2\}] & \text{if } q_t\{2\} > q_t\{1\} \\ [q_t\{2\},q_t\{1\}] & \text{if } q_t\{2\} < q_t\{1\} \end{cases} \tag{21}$$

where $q_t\{1\}$ and $q_t\{2\}$ correspond to the first and second phases constituting the transitional stage. The first operating phase, $q_t\{1\}$, is determined by searching for the phase with the largest posterior probability as

$$q_t\{1\} = \underset{q}{\operatorname{argmax}}\ p(\Xi_q|x_t) \tag{22}$$

Then the second phase of the transitional stage is identified as one of the two adjacent operating phases of the first phase that has the larger posterior probability

$$q_t\{2\} = \underset{q = q_t\{1\}-1 \text{ or } q_t\{1\}+1}{\operatorname{argmax}} p(\Xi_q|x_t) \tag{23}$$

As the various operating phases and between-phase transitional stages are identified, multi-kernel Gaussian process regression models can be developed for all different operating phases. A Gaussian process is a collection of random variables, any finite number of which have joint Gaussian distributions (Rasmussen and Williams, 2006). Given a set of process input and output measurements $X = [x_1\ x_2\ \ldots\ x_N]$ and $Y = [y_1\ y_2\ \ldots\ y_N]$, the standard linear regression model with Gaussian noise is of the form

$$y_i = f(x_i) + e_i = x_i^T\beta + e_i \tag{24}$$

where $f(\cdot)$ is the regression function, β is the vector of weights of the linear model that has Gaussian prior with zero-mean and covariance matrix Σ_β. Thus the statistical distribution of β is as follows:

$$\beta \sim \mathcal{N}(0,\Sigma_\beta) \tag{25}$$

Meanwhile, e_i is the noise that follows a Gaussian distribution with zero mean and variance σ_e^2

$$e_i \sim \mathcal{N}(0,\sigma_e^2) \tag{26}$$

A Gaussian process is specified by its mean and covariance functions $m(x)$ and $k(x,x^T)$ as follows:

$$m(x) = \mathbb{E}[f(x)] \tag{27}$$

and

$$k(x,x^T) = \mathbb{E}[(f(x)-m(x))(f^T(x)-m^T(x))] \tag{28}$$

Then the conditional probability density function of the output variables given the process input measurements and regression model parameters is denoted as $p(Y|X,\beta)$ and given by

$$p(Y|X,\beta) = \prod_{i=1}^{N} p(y_i|x_i,\beta) \tag{29}$$

$$p(Y|X,\beta) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_e} \exp\left\{-\frac{(y_i - x_i^T\beta)^2}{2\sigma_e^2}\right\} \tag{30}$$

which can be expressed as

$$p(Y|X,\beta) = \frac{1}{(2\pi\sigma_e^2)^{N/2}} \exp\left\{\frac{1}{2\sigma_e^2}|Y - X^T\beta|^2\right\} \tag{31}$$

The above likelihood function follows a Gaussian distribution as

$$p(Y|X,\beta) = \mathcal{N}(X^T\beta,\sigma_e^2 I_{N\times N}) \tag{32}$$

where $I_{N\times N}$ is the identity matrix. Thus, the posterior probability density function can be estimated through Bayesian inference strategy as follows:

$$p(\beta|Y,X) = \frac{p(Y|X,\beta)p(\beta)}{p(Y|X)} \tag{33}$$

where the evidence $p(Y|X)$ is independent of the weights and given by

$$p(Y|X) = \int p(Y|X,\beta)p(\beta)\,\mathrm{d}\beta \tag{34}$$

The posterior probability density functions in Eq. (33) combines the likelihood and prior densities to represent the known information of the model parameters (Rasmussen and Williams, 2006; Chen et al., 2007; Likar and Kocijan, 2007). Further, the mean vector of regression coefficients is given by

$$\overline{\beta} = \sigma_e^{-2}(\sigma_e^{-2}XX^T + \Sigma_\beta^{-1})^{-1}XY = \sigma_e^{-2}A^{-1}XY \tag{35}$$

with $A = \sigma_e^{-2}XX^T + \Sigma_\beta^{-1}$. The posterior density function can be written in the form of a Gaussian with mean $\overline{\beta}$ and covariance A^{-1} as follows:

$$p(\beta|X,Y) \sim \mathcal{N}(\overline{\beta},A^{-1}) \tag{36}$$

A single-kernel Gaussian process regression model may not be able to deal with complex batch processes with multiple operating phases. Therefore, a multi-kernel GPR strategy is utilized to build multiple GPR models in various nonlinear kernel feature spaces. In this way, the local GPR models can be used to characterize the identified operating phases for state estimation and quality prediction. With $\phi(\cdot)$ denoting the nonlinear mapping function, the nonlinear regression model is expressed as

$$y_i = \phi(x_i)^T\beta + e_i \tag{37}$$

and the predictive distribution can be written for the test data X_t and corresponding outputs Y_t as

$$p(Y_t|X_t,X,Y) \sim \mathcal{N}(\sigma_e^{-2}\phi(X)^T A_\phi^{-1}\phi(X)Y, \phi(x_t)^T A_\phi^{-1}\phi(X_t)) \tag{38}$$

with $A_\phi = \sigma_e^{-2}\phi(X)\phi(X)^T + \Sigma_\beta^{-1}$. A kernel function $\kappa(x,x')$ is defined as

$$\kappa(x,x') = \phi(x)^T \Sigma_\beta \phi(x') \tag{39}$$

and can be simplified as

$$\kappa(x,x') = \psi(x)^T \psi(x') \tag{40}$$

where $\psi(x) = \Sigma_\beta^{1/2}\phi(x)$. Substituting the nonlinear mapping function with the kernel function in Eq. (38) yields

$$p(Y_t|X_t,X,Y) \sim \mathcal{N}(\overline{Y}_t, \mathrm{cov}(Y_t)) \tag{41}$$

with the mean $\overline{Y}_t$ and covariance matrix $\mathrm{cov}(Y_t)$ given by

$$\overline{Y}_t = \kappa(X_t,X)[\kappa(X,X) + \sigma_e^2 I_{N\times N}]^{-1} Y \tag{42}$$

and

$$\mathrm{cov}(Y_t) = \kappa(X_t,X_t) - \kappa(X_t,X)[\kappa(X,X) + \sigma_e^2 I_{N\times N}]^{-1}\kappa(X,X_t) \tag{43}$$

Alternatively, Eq. (42) can be written for Q kernel functions corresponding to different operating phase as

$$\widehat{Y}_t^{(q)} = \Lambda \kappa_q(X_t^{(q)}, X^{(q)}) \tag{44}$$

where $\Lambda = [\kappa_q(X^{(q)},X^{(q)}) + \sigma_e^2 I_{N_q\times N_q}]^{-1} Y^{(q)}$ and $\kappa_q(X^{(q)},X^{(q)})$ denotes the local kernel function (Qiu and Lane, 2009; Rasmussen and Williams, 2006; Bishop, 2007). In addition, $X_t^{(q)}$ and $\widehat{Y}_t^{(q)}$ represent the input test data and the corresponding predicted output data for the q-th phase. The Gaussian kernel is the most commonly used kernel function and given by

$$\kappa_q(X_t^{(q)}, X^{(q)}) = \exp\left(-\frac{\|X_t^{(q)} - X^{(q)}\|^2}{2\sigma_q^2}\right) \tag{45}$$

where σ_q^2 is the kernel parameter for the q-th phase. The GPR model parameters Θ can be determined by maximizing the marginal likelihood as

$$\Theta^* = \underset{\Theta}{\mathrm{argmax}} \log p(Y|X,\Theta) \tag{46}$$

with Θ^* as the optimal parameter values (Rasmussen and Williams, 2006). The multi-kernel Gaussian process regression approach with localized kernel functions for different operating phases is used to estimate the predictive distributions of the output data and further predict the state and quality variables in nonlinear multiphase batch processes.

Although the MKGPR approach is suitable for state estimation and quality prediction within multiple operating phases, the between-phase transitional stages with transient dynamics are not well characterized by the localized models. Therefore, Bayesian model averaging strategy is designed to integrate the two local GPR models that represent the adjacent operating phases of each transitional stage. BMA is an effective approach for handling model uncertainty and allows for evaluating the robustness of results to alternative model specifications by estimating the posterior distributions (Raftery et al., 1997; Hoeting et al., 1999; Neuman, 2003). For each identified between-phase transitional stage, the two localized GPR models corresponding to the first and second phases are denoted as M_1 and M_2, respectively. Then the marginal probability distribution of the prediction $\widehat{Y}_t^{(q)}$ under the local model M_l ($l = 1$ or 2) is expressed as

$$p(\widehat{Y}_t^{(l)}|M_l) = \int p(\widehat{Y}_t^{(l)}|\beta_l, M_l) p(\beta_l|M_l)\, \mathrm{d}\beta_l \tag{47}$$

with $p(\beta_l|M_l)$ denoting the prior probability. The posterior probabilities $p(M_l|\widehat{Y}_t^{(l)})$ of the various local models with respect to their predictions $\widehat{Y}_t^{(l)}$ and model priors $p(M_l)$ are given by

$$p(M_l|\widehat{Y}_t^{(l)}) = \frac{p(\widehat{Y}_t^{(l)}|M_l)p(M_l)}{\sum_{l'=1}^{2} p(\widehat{Y}_t^{(l')}|M_{l'})p(M_{l'})} \tag{48}$$

Further, the posterior probabilities can be normalized as follows:

$$\widehat{p}(M_l|\widehat{Y}_t^{(l)}) = \frac{p(M_l|\widehat{Y}_t^{(l)})}{\sum_{l'=1}^{2} p(M_{l'}|\widehat{Y}_t^{(l')})} \tag{49}$$

which leads to dynamic weights for incorporating the two local GPR models into a Bayesian averaged model. Thus, the corresponding state and quality prediction within the between-phase transitional stages is expressed as

$$\widehat{Y}_t = \sum_{l=1}^{2} (\widehat{p}(M_l|\widehat{Y}_t^{(l)}) \cdot \widehat{Y}_t^{(l)}) \tag{50}$$

In this way, the Bayesian inference based posterior probabilities can be used to dynamically adjust the weights of two local models corresponding to the first and second phases of the transitional stages according to the relative dependency of the test sample with respect to the two phases. The above equation is used to predict the state and quality variables within a between-phase transitional stage, which represents a short period of time when the process is shifting from the end of the previous phase to the beginning of the next phase. Because of the remaining effect from the previous phase, the dynamics during this transitional stage may be different from each of the two neighboring phases. Thus, different model prediction strategy is needed to handle the between-phase transitional stages accurately. It should be emphasized that two types of state estimations and quality predictions are designed for test samples within a local operating phase or a between-phase transitional stage. For a batch phase, the state and quality variables are predicted from the localized GPR models corresponding to various kernels. For a between-phase transitional stage with transient dynamics, the Bayesian model averaging strategy is used to integrate two local GPR models corresponding to two consecutive phases through Bayesian posterior probabilities as the adaptive weightings. An illustration of the proposed BMA based multi-kernel GPR approach is shown in Fig. 2.

The step-by-step procedure of the proposed nonlinear state estimation and quality prediction approach is given below and the corresponding flow diagram is shown in Fig. 3.

(1) Collect process input and output data matrices for soft sensor model training.
(2) Synchronize the batch process data using dynamic time warping and further unfold the three-dimension data into two-dimension matrix.
(3) Use the process input data in the training set to estimate the kernel mixture model through EM algorithm and identify all different operating phases in batch process.
(4) Align the identified batch phases to avoid disjoint phases and misclassified samples.
(5) Estimate the posterior probabilities of the input test data with respect to all the identified batch phases.
(6) Specify a statistical significance level as the threshold for the identification of between-phase transient stages.
(7) Classify the input and output samples of the test set with maximum posterior probability greater then the specified significance level into different operating phases.
(8) Classify the test samples with the maximum posterior probabilities below the specified threshold as from the between-phase transitional stage.
(9) Identify the two adjacent phases of the test samples belonging to the transitional stages.
(10) Develop localized Gaussian process regression models for different operating phases using the subsets of input and output data classified into those local phases.

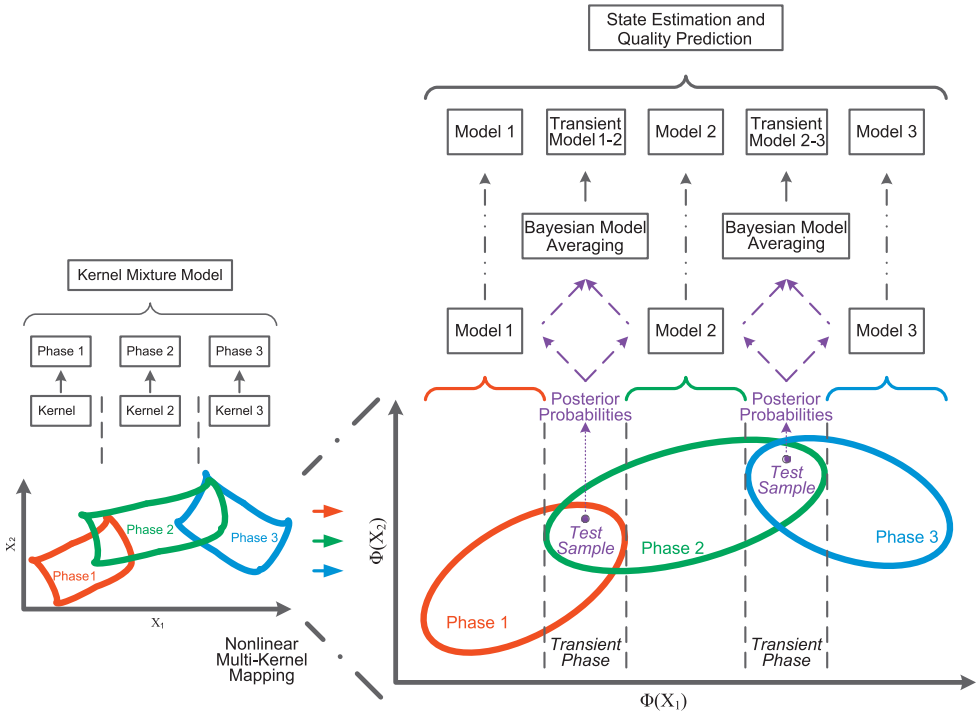

Fig. 2. Illustration of the proposed Bayesian model averaging based multi-kernel Gaussian process regression framework for nonlinear state estimation and quality prediction of multiphase batch processes.

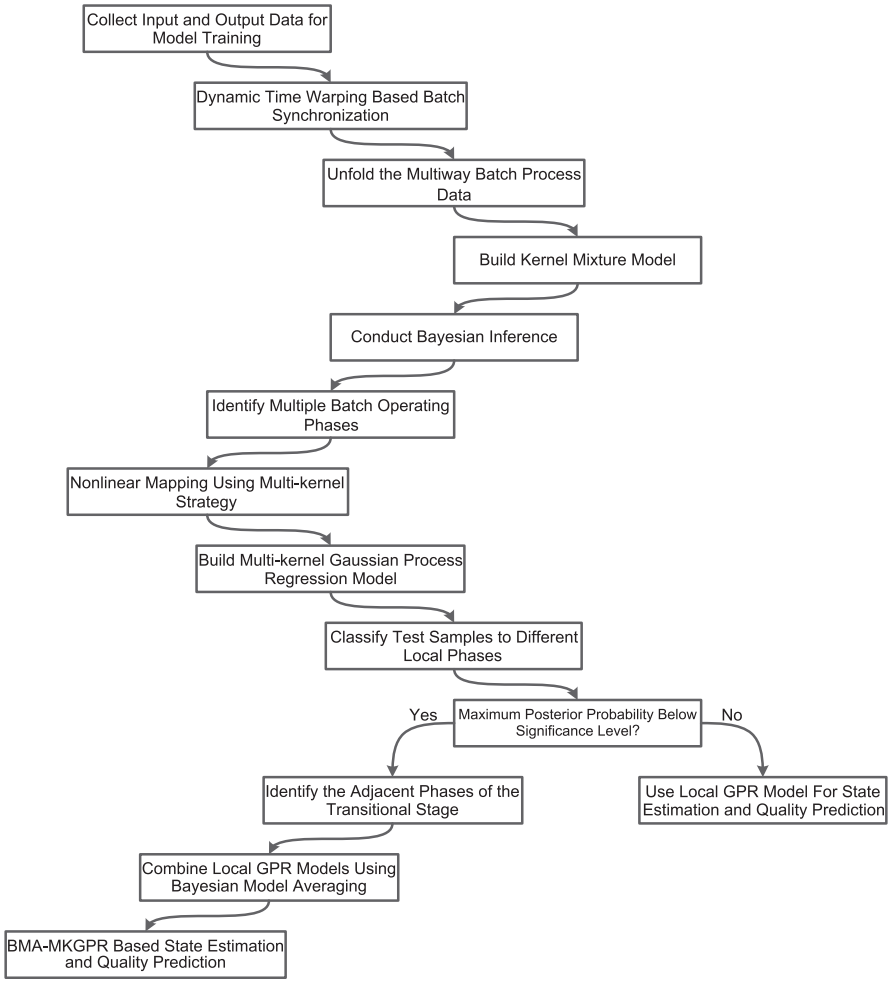

Fig. 3. Flow diagram of the Bayesian model averaging based multi-kernel Gaussian process regression approach for state estimation and quality prediction of nonlinear multiphase batch processes.

(11) Use multi-kernel Gaussian process regression model to predict the state and quality variables for samples belonging to local operating phases.
(12) Integrate the two local GPR models corresponding to the adjacent phases of the transitional stages using Bayesian model averaging.
(13) Conduct the BMA based state estimation and quality prediction for test samples belonging to the between-phase transitional stages.

4. Application example

4.1. Nylon-6,6 batch polymerization process description

In this study, the nylon-6,6 batch polymerization process is used to evaluate the effectiveness of the proposed Bayesian model averaging based multi-kernel Gaussian process regression approach for nonlinear state estimation and quality prediction of multiphase batch process. In the nylon polymerization process, the amine end groups (A) react reversibly with the carboxylic end groups (C) according to the following condensation reaction:

$$A+C \rightleftharpoons L+W$$

The polymerization reaction is treated as a second-order reversible reaction. The product in this process is formed through polymer chain links (L) with the evolution of a water molecule (W). In addition, the carboxylic end groups and polymer chains may decompose to form nonreactive stabilized end groups (SE) according to the degradation reactions as follows:

$$S \rightarrow SE+W$$

and

$$L \rightarrow SE+A$$

The objective of the process is to produce the nylon-6,6 polymers with desired product quality, which is characterized by the number average molecular weight (MW) and the amine end concentration (NH_2). In order to achieve the relatively high molecular weights, the equilibrium of the condensation reaction is shifted towards the formation of polymer chain links and a higher extent of reaction by removing water through vaporization (Russell et al., 1998). Thus, the polymerization takes place in an autoclave reactor with a valve for venting vaporized water and a steam jacket for supplying the heat needed for reaction and vaporization, as shown in Fig. 4. Once the reactant mixture is charged, external heat is supplied through the jacket to drive the polymerization reaction with the vent valve initially being closed to prevent the loss of the volatile reactants. This period of operation is known as the initial heating phase. When a sufficient extent of reaction is achieved, the valve is opened to vent the vaporized water, which marks the boiling phase of the process operation. The primary objective of the boiling stage is to remove the excess water so that high MW levels can be achieved. Once the water has been removed and the boiling stops, the finishing phase begins whereby the characteristics of final product quality are developed and result in viscous polymers with high molecular weights (Russell et al., 1998; Joly and Pinto, 2004).

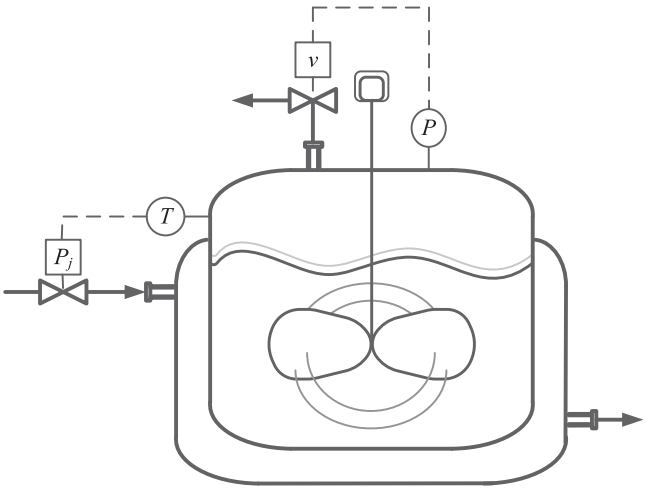

Fig. 4. Process flow diagram of the nylon-6,6 autoclave.

The variables of the nylon-6,6 polymerization process are listed in Table 1 and include four input variables, five state variables and two quality variables. Meanwhile, the initial conditions of this process are shown in Table 2. In this study, all the input variables are used to develop the soft sensor models for estimating and predicting the selected state and quality variables. The two state variables selected to demonstrate the effectiveness of the proposed approach include the concentrations of water and polymer chain links, the former of which is directly related to the molecular weight of the polymer while the latter determines the polymerization reaction equilibrium. Each batch of the process has a duration of 3 h with the sampling period of 1 min. Total 20 training batches are generated to identify the multiple operating phases as well as build the predictive models on state and quality variables. In addition, five test batches are produced and used to assess the performance of the developed models for state estimation and quality prediction. The proposed BMA-MKGPR and BMA-MKSVR approaches are compared to the MMNN, MKSVR and MKGPR methods. It should be noted that the BMA-MKSVR method is the same as the BMA-MKGPR approach except that the phase based local SVR instead of GPR model is built. The statistical significance level used to classify the measurement samples into

Table 1
Input, state and quality variables of the nylon-6,6 polymerization process.

Variable description	Symbol	Type
Steam jacket pressure (psi)	P_j	Input variable
Reactor pressure (psi)	P	Input variable
Reactor temperature (K)	T	Input variable
Vent flow rate (g/h)	v	Input variable
Amine end concentration (mol/g)	NH_2	Quality variable
Number average molecular weight (g/mol)	MW	Quality variable
Concentration of A (mol/L)	C_A	State variable
Concentration of C (mol/L)	C_C	State variable
Concentration of L (mol/L)	C_L	State variable
Concentration of W (mol/L)	C_W	State variable
Concentration of SE (mol/L)	C_{SE}	State variable

Table 2
Initial conditions of the nylon-6,6 batch polymerization process.

Variable	Initial value
Temperature (K)	425
Concentration of A (mol/L)	6
Concentration of C (mol/L)	6
Concentration of L (mol/L)	0
Concentration of W (mol/L)	11.2
Concentration of SE (mol/L)	0

the between-phase transitional stages for the BMA based MKSVR and MKGPR methods is set to $\alpha = 0.05\%$. The model performance indices used in this study include the root-mean-square error (RMSE), mean absolute percentage error (MAPE) and R^2, as defined below

$$\mathrm{RMSE} = \sqrt{\frac{1}{N_t}\sum_{i=1}^{N_t}(y_i - \hat{y}_i)^2} \tag{51}$$

$$\mathrm{MAPE} = \frac{1}{N_t}\sum_{i=1}^{N_t}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\% \tag{52}$$

and

$$R^2 = 1 - \frac{\sum_{i=1}^{N_t}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N_t}(y_i - \bar{y})^2} \tag{53}$$

where N_t denotes the total number of test samples across all the test batches, y_i and $\hat{y}_i$ are the actual and predicted values of each sample, and $\bar{y}$ is the mean value of an output variable.

4.2. Comparison of state estimation and quality prediction results

First, the different operating phases of the batch polymerization process are identified using kernel mixture models and the phase identification results projected onto the three-dimensional latent variable space are shown in Fig. 5. The three identified phases include the initial heating phase, the boiling phase and the finishing phase, which are consistent with the actual process operation. In addition to the three operating phases, the between-phase transitional stages with transient dynamics are also captured by the proposed method.

The averaged batch predictions of different methods on the selected state and quality variables are shown in Fig. 6 and the corresponding average prediction errors are depicted in Fig. 7. Meanwhile, the quantitative model comparison is listed in Table 3. It can be observed from Fig. 6 that the MMNN method leads to the worst prediction accuracies on both state variables. Accordingly, it gives the highest RMSE and MAPE values while the lowest R^2 values for these two state variables. The RMSE values for the concentrations of polymer chain links and water are as high as 766.83 mol/L and 2864.11 mol/L, respectively. In addition, the MAPE values for the two state variables, C_L and C_W, are as high as 27.23% and 21.09%, respectively. The above average percentage errors indicate that the state estimation accuracies are quite low and that the errors are significantly inflated during the between-phase transitional stages. Although the multi-model neural network method is able to handle the process nonlinearity and multiple phases, ANN still suffers from over-fitting issue and poor generalization capability for model predictions. Moreover, the MMNN method does not identify the between-phase transitional stages and thus cannot account for the transient dynamics when the operational phase is shifting. The box plots of model residuals on the two state variables from different methods are shown in Fig. 9(a) and (b). It is readily observed that the red-highlighted outliers from the MMNN method have the widest range implying the poor prediction performance.

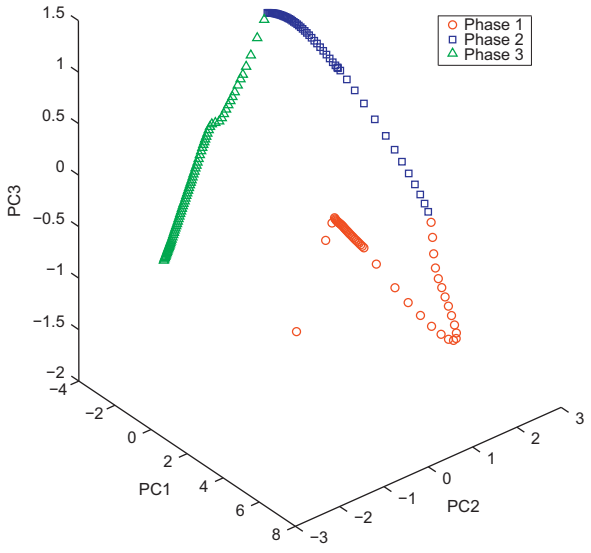

Fig. 5. Phase identification results of the nylon-6,6 polymerization process.

In comparison, both the MKSVR and MKGPR approaches result in relatively lower RMSE and MAPE values but higher R^2 values on the two state variables than the MMNN method. The multi-kernel approach is adopted so that the multiphase operation of batch processes can be well characterized and thus the performance of model predictions may be significantly improved. Usually a single nonlinear kernel function is unable to precisely capture the complex dynamics among multiple operating phases in batch or semi-batch processes. Compared to the MMNN method, the MKSVR approach results in the lower RMSE values of 508.67 mol/L for C_L and 2378.66 mol/L for C_W. Subsequently, the lower MAPE values of 20.65% for C_L and 15.27% for C_W are obtained. The improved performance of the MKSVR method is also demonstrated in terms of the stronger correlations between the actual and predicted values, as shown in Fig. 8. Batch processes may have significant batch-to-batch variations and inherent random uncertainties. Therefore, the multi-kernel Gaussian process regression approach leads to even lower RMSE and MAPE values than the multi-kernel support vector regression method. The RMSE values of MKGPR method are 444.60 mol/L and 1336.08 mol/L for C_L and C_W, respectively. Meanwhile, the corresponding MAPE values are 17.16% and 9.57%. The improved prediction accuracies for MKGPR method are due to the inherently strong capability of Gaussian process regression model in handling process uncertainty.

Although multiple localized models are developed in MMNN, MKSVR and MKGPR methods for different batch phases, the deterministic strategy of classifying process data into a particular phase and then estimating the state variables from the corresponding local models may not be always reliable due to the process uncertainty. Furthermore, the phase based local models are unable to effectively capture the between-phase transient dynamics, which can degrade the accuracy of model predictions. Thus, Bayesian model averaging strategy is designed to adaptively integrate the local model predictions within the transitional stages between two adjacent operating phases. Both BMA-MKSVR and BMA-MKGPR approaches have significantly lower RMSE and MAPE values than the other methods. For instance, the BMA-MKSVR method has the RMSE values of 116.10 mol/L and 1047.04 mol/L for C_L and C_W, respectively. In comparison, the BMA-MKGPR approach performs even better than the BMA-MKSVR method because of its enhanced capability of handling system uncertainty by adopting GPR for localized models.

In addition to the state estimations, the quality variables of the nylon-6,6 batch polymerization process are also predicted by using the five different methods. For the first quality variable of the number average molecular weight, the MMNN method still gives the highest RMSE and MAPE values of 212.46 g/mol and 25.11%, respectively. The poor prediction on the molecular weight by the MMNN method is further observed in Figs. 6 and 7. Similar model performance is obtained on the second quality variable of the amine end group concentration from the MMNN method. As a comparison, the MKSVR and MKGPR methods lead to higher

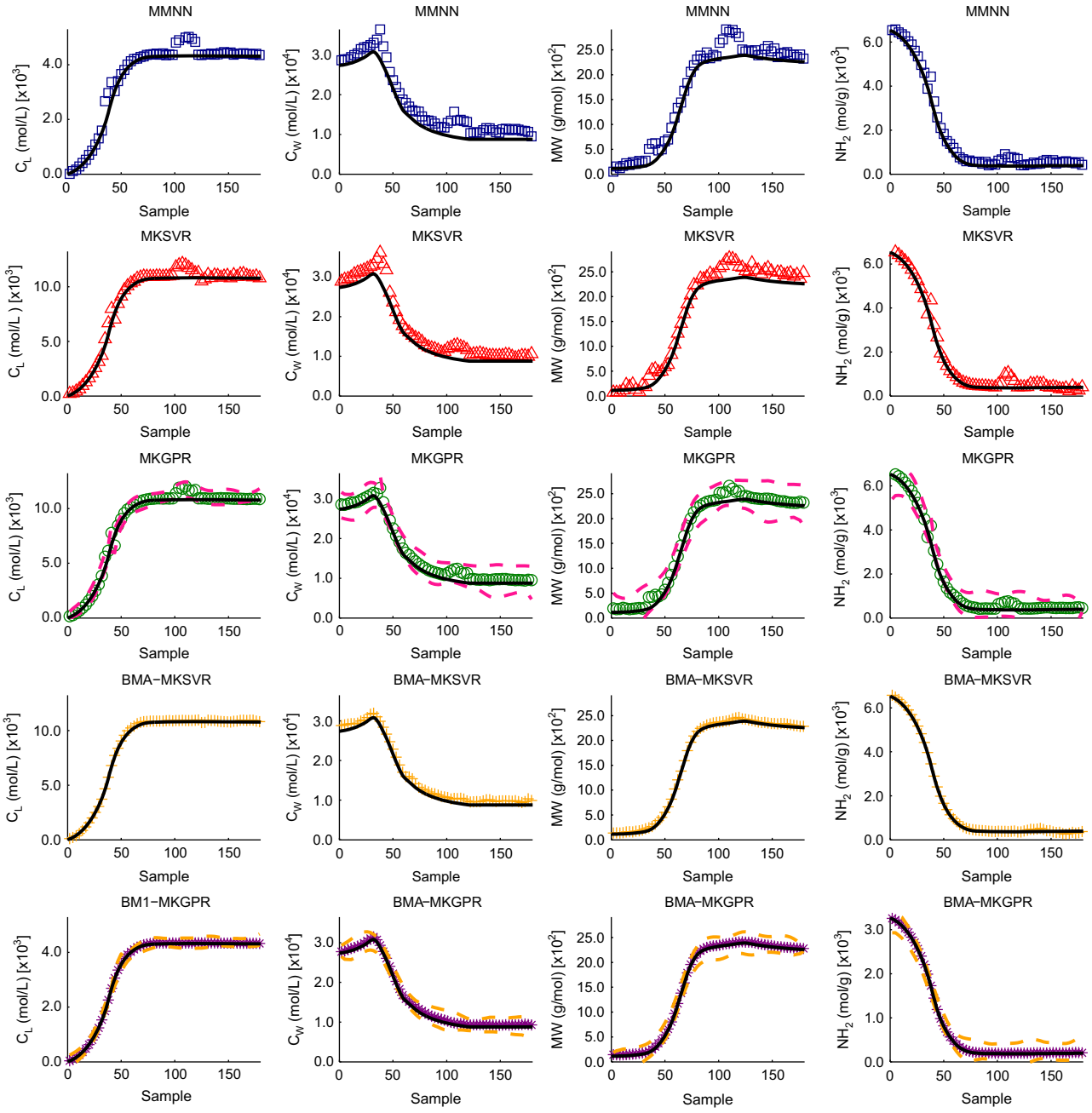

Fig. 6. Trends of actual values (solid line), predicted values (points) and confidence intervals (dashed line) of the state and quality variables from different soft sensor modeling methods.

prediction accuracies on both quality variables in terms of lower RMSE and MAPE values. Nevertheless, it can still be observed from Figs. 6 and 7 that both the MKSVR and MKGPR approaches cause higher prediction errors during the between-phase transitional stages as they are unable to handle the between-phase transient dynamics.

In contrast, the BMA based MKSVR and MKGPR approaches give the lower RMSE and MAPE values than the other methods for both quality variables. For example, the BMA-MKGPR approach yields the lowest RMSE and MAPE values of 24.12 and 5.0% for the quality variable of number average molecular weight. Further, one can see from Fig. 6 that the predicted values of the two quality variables in the trend plots precisely match the actual values for the proposed BMA-MKGPR method without any inflated prediction errors during the between-phase transitional stages. Though the BMA-MKSVR method also provides high model accuracies, the BMA-MKGPR approach performs even better on quality predictions due to the stronger capacity of uncertainty handling in GPR technique compared to SVR method. Therefore, BMA-MKGPR approach is more desirable than BMA-MKSVR method in dealing with the process uncertainty within each local operating phase.

The above state and quality prediction results show that the regular multi-model strategies may not be effective and reliable for multiphase batch processes because they cannot accurately characterize the complex transient dynamics within the transitional stages, where the process operation is shifting from one phase to the next. However, the new BMA-MKSVR and

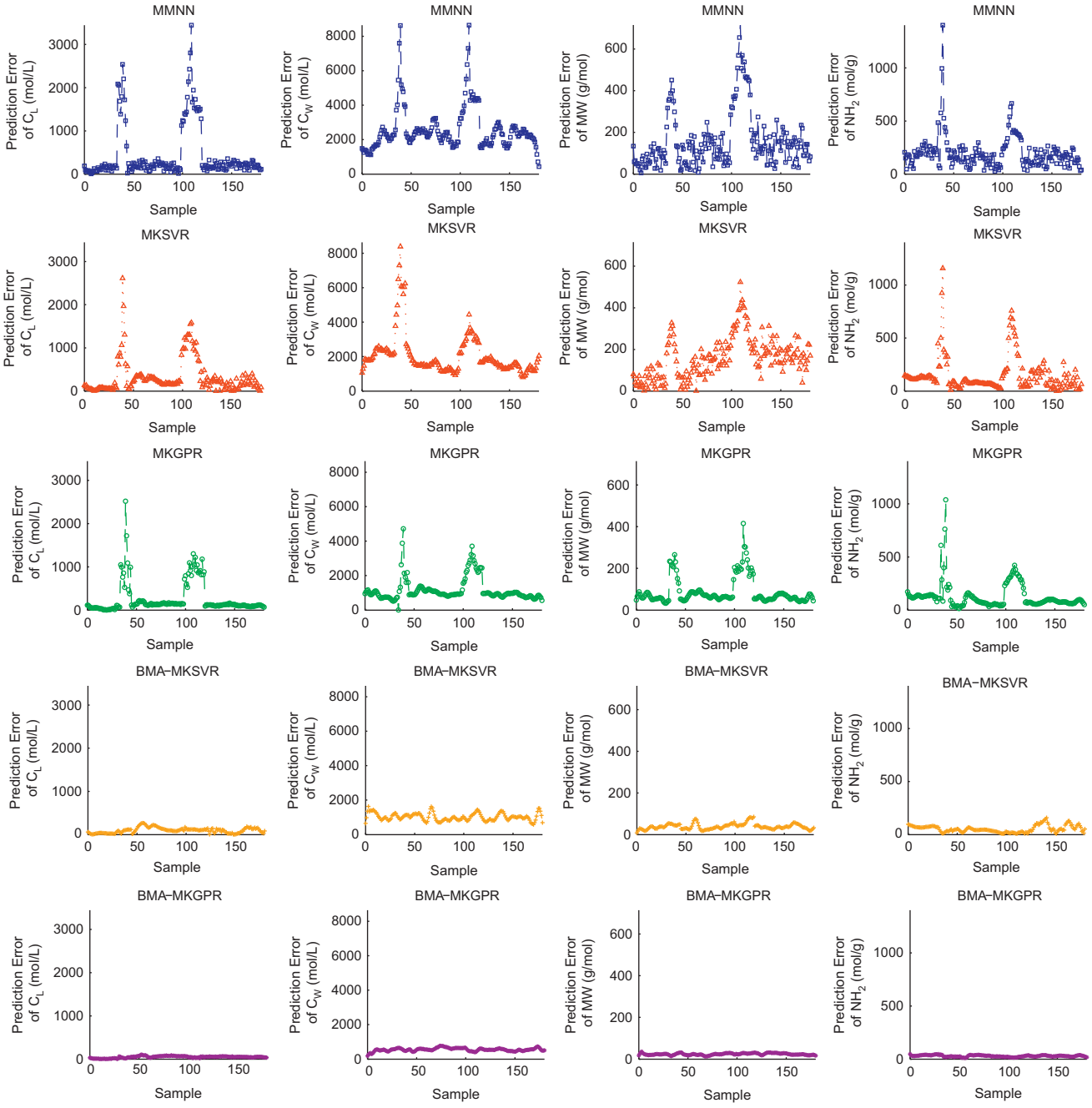

Fig. 7. Prediction errors of the state and quality variables from different soft sensor modeling methods.

BMA-MKGPR approaches are shown to be much more reliable and accurate for state estimation and quality prediction owing to the Bayesian model averaging based local model integration mechanism for capturing between-phase transient dynamics. Through dynamically and adaptively weighting the local models of the adjacent phases using Bayesian inference based posterior probabilities, the proposed approach can precisely characterize the dynamic changes between local phases. The computational load of the proposed BMA-MKGPR method is mainly focused on the off-line model development stage, especially the kernel mixture model based phase identification. However, after the initial model is built from the training batches, the online predictions of new test samples can be conducted very fast. Therefore, the proposed approach is not restricted to small-scale batch processes, but can be potentially applied to large-scale processes.

5. Conclusions

Accurate state estimation and reliable quality prediction are essential for continuous improvement, advanced control and optimization of batch processes. However, the between-phase transitional stages in batch processes are often characterized with complex transient dynamics so that the phase-based multi-model strategies may not be well-suited. In this study, the novel BMA-MKGPR approach is proposed for state estimation and quality prediction of multi-phase nonlinear batch processes. The kernel mixture model and Bayesian inference strategy are first developed and incorporated to identify different operating phases as well as between-phase transitional stages, which can be very challenging for traditional soft sensor methods. Then the multi-kernel local GPR models are built for different operating phases

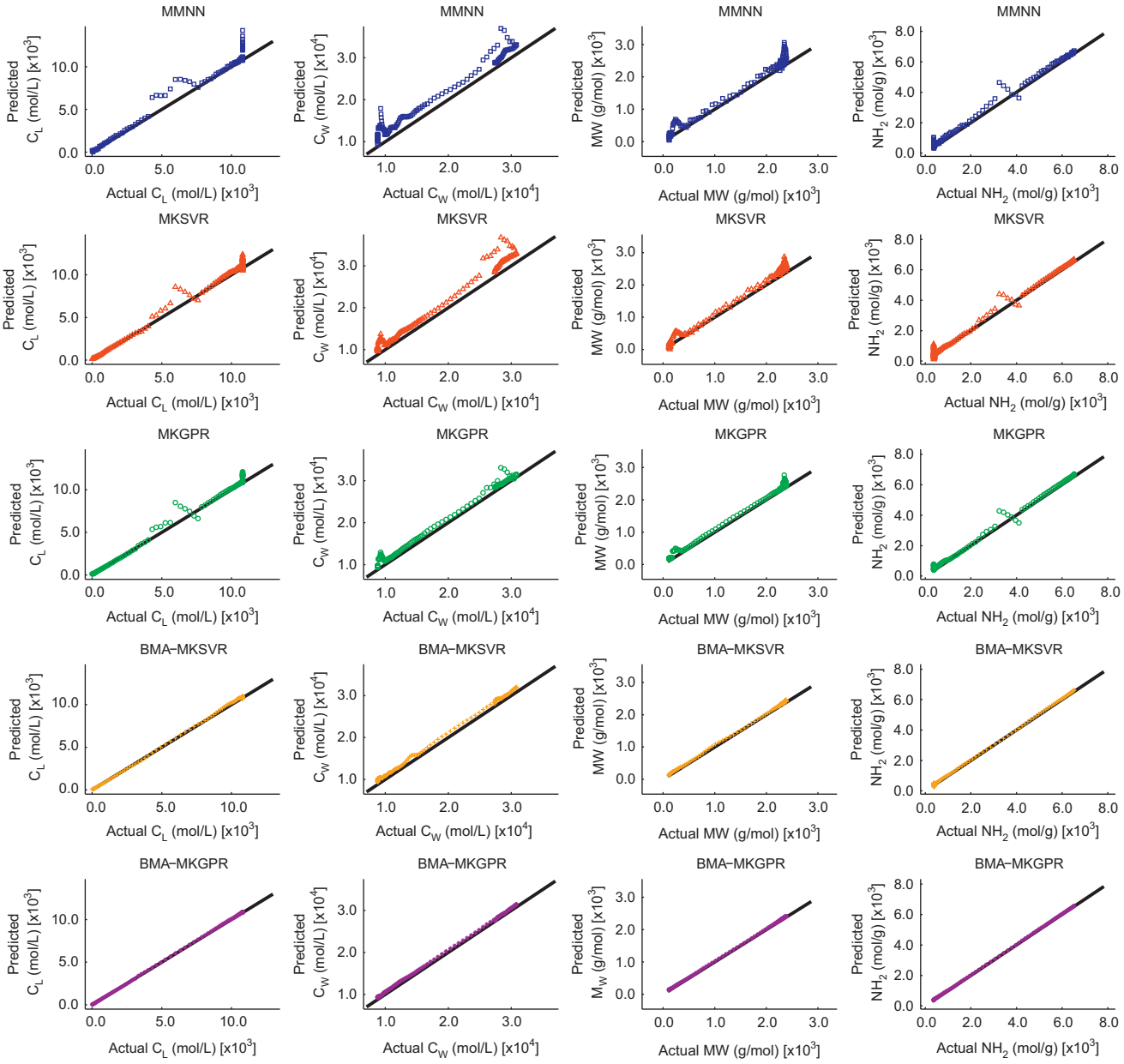

Fig. 8. Correlation trends of predicted values versus actual values of the state and quality variables from different soft sensor modeling methods.

Table 3
Quantitative comparison of state estimation and quality prediction results from different soft sensor modeling methods.

Index	Variable	MMNN	MKSVR	MKGPR	BMA-MKSVR	BMA-MKGPR
RMSE	C_L	766.83	508.67	444.60	116.10	52.50
	C_W	2864.11	2378.66	1336.08	1047.04	557.71
	MW	212.46	187.68	108.19	41.36	24.12
	NH_2	254.35	220.37	177.49	61.68	30.03
MAPE (%)	C_L	27.23	20.65	17.16	8.24	4.15
	C_W	21.09	15.27	9.57	8.24	4.55
	MW	25.11	20.48	17.56	7.48	5.00
	NH_2	33.25	25.09	21.01	9.15	4.99
R^2	C_L	0.9565	0.9809	0.9854	0.9990	0.9998
	C_W	0.8806	0.9176	0.9740	0.9840	0.9955
	MW	0.9496	0.9607	0.9869	0.9981	0.9994
	NH_2	0.9853	0.9890	0.9928	0.9991	0.9998

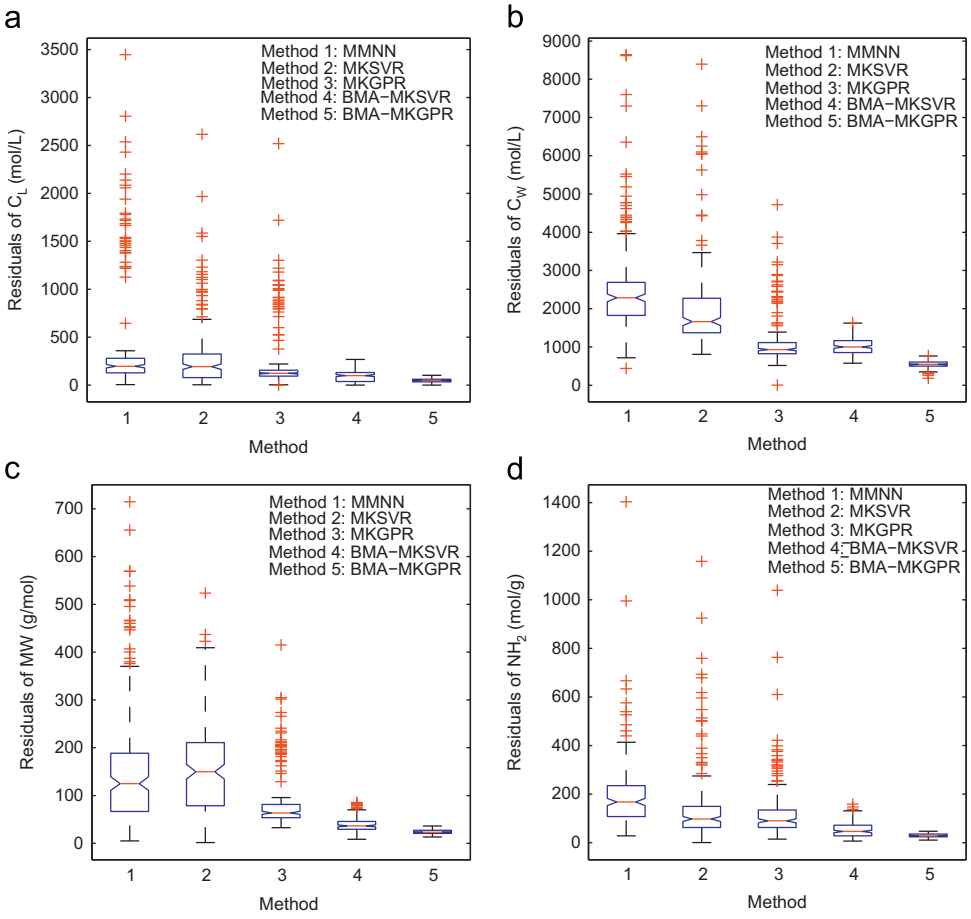

Fig. 9. Prediction residuals for (a) concentration of polymer chain links, (b) concentration of water, (c) molecular weight, and (d) amine end group concentration from different soft sensor modeling methods. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

and further the local models corresponding to the adjacent phases are adaptively integrated through Bayesian model averaging strategy to characterize the between-phase transient dynamics of nonlinear batch processes.

The proposed BMA-MKGPR and BMA-MKSVR approaches are applied to the nylon-6,6 batch polymerization process with three operating phases and its results are compared to those of MMNN, MKSVR and MKGPR methods. The result comparison shows that the developed BMA-MKGPR and BMA-MKSVR approaches have superior state and quality prediction performance as opposed to the other methods. Particularly, the shifting operating phases, between-phase transient dynamics and process uncertainty can be well captured by the proposed approach so that its model accuracies are significantly higher than those of the other methods. Since GPR model has enhanced capability than SVR model in accounting for the stochastic uncertainty in multiphase batch processes, BMA-MKGPR approach has a bit better performance than BMA-MKSVR method and thus is more desirable in practical applications. Future research can be extended to Bayesian model based predictive control and optimization of nonlinear multiphase batch processes with between-phase transient dynamics.

References

Bishop, C.M., 1995. Neural Networks for Pattern Recognition. Oxford University Press, Oxford, UK.

Bishop, C.M., 2007. Pattern Recognition and Machine Learning. Springer, New York, NY, USA.

Bogaerts, P., Wouwer, A.V., 2004. Parameter identification for state estimation-application to bioprocess software sensors. Chem. Eng. Sci. 59, 2465–2476.

Borchert, C., Sundmacher, K., 2012. Morphology evolution of crystal populations: modeling and observation analysis. Chem. Eng. Sci. 70, 87–98.

Bosca, S., Fissore, D., 2011. Design and validation of an innovative soft-sensor for pharmaceuticals freeze-drying monitoring. Chem. Eng. Sci. 66, 5127–5136.

Chen, K., Ji, J., Wang, H., Liu, Y., Song, Z., 2011. Adaptive local kernel-based learning for soft sensor modeling of nonlinear processes. Chem. Eng. Res. Des. 89, 2117–2124.

Chen, T., Morris, J., Martin, E., 2007. Gaussian process regression for multivariate spectroscopic calibration. Chemometrics Intell. Lab. Syst. 87, 59–71.

Drucker, H., Burges, C.J.C., Kaufman, L., Smola, A., Vapnik, V., 1997. Support vector regression machines. Adv. Neural Inf. Process. Syst. 9, 155–161.

Facco, P., Doplicher, F., Bezzo, F., Barolo, M., 2009. Moving average PLS soft sensor for online product quality estimation in an industrial batch polymerization process. J. Proc. Cont. 19, 520–529.

Fortuna, L., Graziani, S., Xibilia, M.G., 2009. Comparison of soft-sensor design methods for industrial plants using small data sets. IEEE Trans. Instrum. Meas. 58, 2444–2451.

Fujiwara, K., Kano, M., Hasebe, S., 2012. Development of correlation-based pattern recognition algorithm and adaptive soft-sensor design. Control Eng. Practice 20, 371–378.

Fujiwara, K., Kano, M., Hasebe, S., Takinami, A., 2009. Soft-sensor development using correlation-based just-in-time modeling. AIChE J. 55, 1754–1765.

Galicia, H.J., He, Q.P., Wang, J., 2012. Comparison of the performance of a reduced-order dynamic PLS soft sensor with different updating schemes for digester control. Control Eng. Practice 20, 747–760.

Godoy, J.L., Minari, R.J., Vega, J.R., Marchetti, J.L., 2011. Multivariate statistical monitoring of an industrial SBR process, soft-sensor for production and rubber quality. Chemometrics Intell. Lab. Syst. 107, 258–268.

Gonzaga, J.C.B., Meleiro, L.A.C., Kiang, C., Maciel Filho, R., 2009. ANN-based soft-sensor for real-time process monitoring and control of an industrial polymerization process. Comput. Chem. Eng. 33, 43–49.

Gunn, S.R., 1998. Support vector machines for classification and regression. ISIS Technical Report.

Hecht-Nielsen, R., 1989. Theory of the backpropagation neural network. In: International Joint Conference on Neural Networks, vol. 1. San Diego, CA, USA, pp. 593–605.

Hoeting, J.A., Madigan, D., Raftery, A.E., Volinsky, C.T., 1999. Bayesian model averaging: a tutorial. Stat. Sci. 14, 382–401.
Joly, M., Pinto, J.M., 2004. Optimal control of product quality for batch nylon-6,6 autoclaves. Chem. Eng. J. 97, 87–101.
Kadlec, P., Gabrys, B., Strandt, S., 2009. Data-driven soft sensors in the process industry. Comput. Chem. Eng. 33, 795–814.
Kandepu, R., Foss, B., Imsland, L., 2008. Applying the unscented Kalman filter for nonlinear state estimation. J. Proc. Cont. 18, 753–768.
Kaneko, H., Arakawa, M., Funatsu, K., 2011. Novel soft sensor method for detecting completion of transition in industrial polymer processes. Comput. Chem. Eng. 35, 1135–1142.
Kassidas, A., MacGregor, J.F., Taylor, P.A., 1998. Synchronization of batch trajectories using dynamic time warping. AIChE J. 44, 864–875.
Khartibisepehr, S., Huang, B., 2008. Dealing with irregular data in soft sensors: Bayesian method and comparative study. Ind. Eng. Chem. Res. 47, 8713–8723.
Kozub, D.J., MacGregor, J.F., 1992. State estimation for semi-batch polymerization reactors. Chem. Eng. Sci. 47, 1047–1062.
Likar, B., Kocijan, J., 2007. Predictive control of a gas–liquid separation plant based on a Gaussian process model. Comput. Chem. Eng. 31, 142–152.
Mangold, M., 2012. Use of a Kalman filter to reconstruct particle size distributions from FBRM measurements. Chem. Eng. Sci. 70, 99–108.
Neuman, S.P., 2003. Maximum likelihood Bayesian averaging of uncertain model predictions. Stoch. Environ. Res. Risk. Assess. 17, 291–305.
Qiu, S., Lane, T., 2009. A framework for multiple kernel support vector regression and its applications to sirna efficacy prediction. IEEE/ACM Trans. Comput. Biol. Bioinf. 6, 190–199.
Raftery, A.E., Madigan, D., Hoeting, J.A., 1997. Bayesian model averaging for linear regression models. J. Am. Stat. Assoc. 92, 179–191.
Rasmussen, C.E., Williams, C.K.I., 2006. Gaussian Processes for Machine Learning. MIT Press, Cambridge, MA, USA.
Régnier, N., Defaye, G., Caralp, L., Vidal, C., 1996. Software sensor based control of exothermic batch reactors. Chem. Eng. Sci. 51, 5125–5136.
Russell, S.A., Robertson, D.G., Lee, J.H., Ogunnaike, B.A., 1998. Control of product quality for batch nylon 6,6 autoclaves. Chem. Eng. Sci. 53, 3685–3702.
Sharmin, R., Sundararaj, U., Shah, S., Griend, L.V., Sun, Y., 2006. Inferential sensors for estimation of polymer quality parameters: industrial application of a PLS-based soft sensor for a LDPE plant. Chem. Eng. Sci. 61, 6372–6384.
Smola, A.J., Schölkopf, B., 2004. A tutorial on support vector regression. Stat. Comput. 14, 199–222.
Soroush, M., 1997. Nonlinear state-observer design with application to reactors. Chem. Eng. Sci. 52, 387–404.
Titsias, M.K., Likas, A.C., 2001. Shared kernel models for class conditional density estimation. IEEE Trans. Neural Net. 12, 987–997.
Vapnik, V., 1998. Statistical Learning Theory. John Wiley and Sons, New York, NY.
Yan, W., Shao, H., Wang, X., 2004. Soft sensing modeling based on support vector machine and Bayesian model selection. Comput. Chem. Eng. 28, 1489–1498.
Yao, Y., Gao, F., 2009. A survey on multistage/multiphase statistical modeling methods for batch processes. Annu. Rev. Control 33, 172–183.
Yu, J., 2011a. Localized Fisher discriminant analysis based complex chemical process monitoring. AIChE J. 57, 1817–1828.
Yu, J., 2011b. Nonlinear bioprocess monitoring using multiway kernel localized Fisher discriminant analysis. Ind. Eng. Chem. Res. 50, 3390–3402.
Yu, J., 2012a. A Bayesian inference based two-stage support vector regression framework for soft sensor development in batch bioprocesses. Comput. Chem. Eng. 41, 134–144.
Yu, J., 2012b. Online quality prediction of nonlinear and non-Gaussian chemical processes with shifting dynamics using finite mixture model based Gaussian process regression approach. Chem. Eng. Sci. 82, 22–30.
Yu, J., 2012c. A particle filter driven dynamic Gaussian mixture model approach for complex process monitoring and fault diagnosis. J. Proc. Cont. 22, 778–788.
Yu, J., 2013. A support vector clustering-based probabilistic method for unsupervised fault detection and classification of complex chemical processes using unlabeled data. AIChE J. 59, 407–419.
Yu, J., Qin, S.J., 2008. Multimode process monitoring with Bayesian inference-based finite Gaussian mixture models. AIChE J. 54, 1811–1829.
Yu, J., Qin, S.J., 2009a. Multiway Gaussian mixture model based multiphase batch process monitoring. Ind. Eng. Chem. Res. 48, 8585–8594.
Yu, J., Qin, S.J., 2009b. Variance component analysis based fault diagnosis of multilayer overlay lithography processes. IIE Trans. 41, 764–775.
Zhang, X., Yan, W., Shao, H., 2008. Nonlinear multivariate quality estimation and prediction based on kernel partial least squares. Ind. Eng. Chem. Res. 47, 1120–1131.