2013 American Control Conference (ACC)
Washington, DC, USA, June 17-19, 2013

Multi-Kernel Gaussian Process Regression and Bayesian Model Averaging Based Nonlinear State Estimation and Quality Prediction of Multiphase Batch Processes

Jie Yu*, Kuilin Chen, Junichi Mori and Mudassir M. Rashid

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 specifically characterize 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. The new approach is applied to a simulated batch polymerization process and the result comparison shows that it can effective handle multiple nonlinear operating phases, between-phase transient dynamics and process uncertainty with high prediction accuracies.**

I. INTRODUCTION

Batch processes are widely used in chemical, materials, pharmaceutical, biotechnology and semiconductor industries for producing low-volume while high-value-added commodities. In batch operation, product quality attributes are often not measured until the end of batch duration. However, precise quality and state estimations are critically important for achieving optimal control, which can continuously improve product quality, increase production yield and optimize process economics [1]. Predictive model based soft sensor techniques have received considerable interest for online state and quality prediction of batch processes [2], [**?**].

Soft sensor methods can be classified into two categories: mechanistic model based and data-driven statistical approaches. The former class of approaches require in-depth understanding of the physical and chemical mechanisms within batch processes so that the first-principle models can be established for Kalman filter or observer based state and quality estimations [4]. Although the mechanistic model and Kalman filter based soft sensors have been applied to state and quality estimations, the linear assumption of regular Kalman filter may not be appropriate for many nonlinear batch processes and thus the state estimator designs using nonlinear process models such as the extended Kalman filter are needed [1]. Other nonlinear state estimation techniques that have been attempted in batch or continuous processes include unscented Kalman filter, ensemble Kalman filter and high-order extended Kalman filters [5]. The process knowledge requirement and time-consuming model development are not desirable for this type of techniques.

Alternately, multivariate statistical methods such as principal component regression (PCR) and partial least squares (PLS) have been widely applied to build inferential sensors for state and quality estimations and are capable of dealing with collinear process data [6]. Furthermore, kernel and recursive PLS algorithms are proposed to deal with the nonlinearity and time-varying dynamics in industrial processes [7], [8]. While these modified PLS techniques can update the models towards new operating conditions recursively, they cannot cope with the abrupt operational changes. Meanwhile, machine learning methods such as artificial neural networks (ANN) and support vector regression (SVR) are applied to nonlinear state estimations and quality predictions [2], [9]. Though they are able to model the nonlinear input-output relationships, ANNs are prone to the restrictions of local minima that can result in degraded predictions. Meanwhile, SVR methods are typically global modeling approaches and may lead to inaccurate estimations for shifting operation conditions or phases.

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, kernel mixture model is built to discriminate different batch phases. Then, MKGPR models are constructed to characterize the shifting dynamics of multiple phases. Further, the between-phase transitional stages with transient dynamics are identified and characterized through Bayesian model averaging strategy.

The remainder of the article is organized as follows. Section II briefly reviews the ANN and SVR based soft sensor methods. Then the novel BMA-MKGPR approach is developed in Section III. Section IV demonstrates the effectiveness of the proposed method using the simulated batch polymerization process. The conclusions of this work are summarized in Section V.

II. PRELIMINARIES

A. Artificial Neural Networks

Artificial neural networks are able to handle process nonlinearity and thus have been applied to soft sensor based state and quality estimations. The model architecture is in

*Jie Yu is Assistant Professor with Department of Chemical Engineering, McMaster University, 1280 Main Street West, Ontario, Canada L8S 4L7 jieyu@mcmaster.ca

978-1-4799-0178-4/$31.00 ©2013 AACC

the form of a multi-layer feed-forward network where the nodes are arranged in successive layers and the information flows unidirectionally from the input layer towards the output layer [10]. The weights w of different nodes correspond to the relative strengths of the connections from one layer to another and can be updated through backward-propagation algorithm. The summation function $y_j = \sum_i x_i w_{j,i}$ computes the weighted sum of all input elements x_i of a node. Further, a sigmoid function

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

is used to transform the output before it is passed to the next layer. 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 phases.

B. Support Vector Regression

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

$$F(x_k) = \langle \omega, \Phi(x_k) \rangle + c \tag{2}$$

where ω is a vector in the kernel feature space $\mathscr{F}$, $\Phi(x_k)$ is a nonlinear mapping from the original measurement space to the feature space $\mathscr{F}$, c is the bias term and $\langle \cdot, \cdot \rangle$ represents for the inner product. The SVR model is based on the ε-insensitive loss function that is aimed at minimizing the empirical risk. Then the optimization problem can be expressed as

$$\min_{\omega,c,\xi,\xi^*} \quad \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*) \tag{3}$$
$$\text{s.t.} \begin{cases} y_i - \langle \omega, \Phi(x_k) \rangle - c \le \varepsilon + \xi_i \\ \langle \omega, \Phi(x_k) \rangle + c - y_i \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0, \qquad i = 1,2,\ldots,N \end{cases}$$

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 [11]. The dual form of Eq. 3 becomes a quadratic programming (QP) problem and its solution results in support vectors. In this study, a multi-kernel support vector regression (MKSVR) approach [12] is implemented where various kernel functions are weighted and summed to characterize the multiphase dynamics of batch processes.

III. BMA-MKGPR APPROACH FOR STATE ESTIMATION AND QUALITY PREDICTION

Although localized models can handle multiphase operation, they may not effectively characterize the transient dynamics in the transitional stages between consecutive phases of batch processes. To address this challenge, 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 are identified through kernel mixture models. Then the multi-kernel Gaussian process regression models are built for multiple phases. Further, the between-phase transitional stages are determined through Bayesian inference based posterior probabilities so that the Bayesian model averaging strategy is designed to characterize the transient dynamics through adaptive weightings on the consecutive local models.

For a multiphase batch process, the training data of input variables are expressed as $\widetilde{X} \in \mathbb{R}^{I \times J \times K}$, where I represents the number of batches, J denotes the number of measurement variables and K is the number of sampling instances. The original batches can be synchronized through dynamic time warping and then the three-dimensional matrix is unfolded into a two-dimensional matrix $X \in \mathbb{R}^{IK \times J}$. Further, the kernel mixture model is applied to the unfolded data 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 density functions, the conditional densities of an arbitrary training sample $x(i,k)$ from the i-th batch and k-th sampling instance can be expressed as

$$p(x(i,k) \mid C_q) = \sum_{q=1}^{Q} p_q p(x(i,k) \mid \Xi_q) \tag{4}$$

where Ξ_q represents local kernel density function corresponding to an individual phase and p_q denotes the prior probability of the local kernel density function. Further, the posterior probabilities $P(\Xi_q \mid x(i,k))$ can be estimated from Bayesian inference to represent the conditional probability that the sample comes from the q-th local kernel density function

$$p(\Xi_q \mid x(i,k)) = \frac{\rho_q p(x(i,k) \mid \Xi_q)}{\sum_{q'=1}^{Q} \rho_{q'} p(x(i,k) \mid \Xi_{q'})} \tag{5}$$

Assume that the Gaussian density functions of the general form below are used

$$\Xi_q = \frac{1}{(2\pi)^{J/2}|\Sigma_q|^{1/2}} \cdot \exp\left\{ -\frac{1}{2}(x(i,k) - \mu_q)^T \Sigma_q^{-1} (x(i,k) - \mu_q) \right\} \tag{6}$$

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 of kernel mixture model denoted by θ include the prior probabilities as well as the means and covariances of kernel density functions. The following log-likelihood function can be maximized to learn the model parameter values

$$\theta^* = \underset{\theta}{\operatorname{argmax}} \ \mathscr{L}(\theta) = \underset{\theta}{\operatorname{argmax}} \left\{ \sum_{q=1}^{Q} \sum_{k=1}^{K} \sum_{i=1}^{I} \log p\left(x(i,k), C_q\right) \right\} \tag{7}$$

In this work, the expectation-maximization (EM) algorithm is adopted for model parameters estimation. In the E-step, the model parameters are initialized and the posterior

probabilities are then evaluated for all different kernel density functions using Eq. 5. Subsequently, the model parameters are updated in the M-step to maximize the log-likelihood function. The EM algorithm is implemented in an iterative way until the parameter values converge to the optimal solution [13]. The phase identification procedure results in the subsets of input data $X = [X^{(1)^T} X^{(2)^T} \ldots X^{(Q)^T}]^T$ and the corresponding output data $Y = [Y^{(1)^T} Y^{(2)^T} \ldots Y^{(Q)^T}]^T$ that represent the Q different phases. Then the Bayesian inference based posterior probabilities can be computed to classify the measurement samples into various phases. For an arbitrary test sample x_t, it is classified into the q_t-th phase with the maximum posterior probability as follows

$$q_t = \underset{q}{\operatorname{argmax}} \; p(\Xi_q \mid x_t) \tag{8}$$

if the maximum posterior probability satisfies

$$p_{\max}(\Xi_q \mid x_t) \geq \alpha \tag{9}$$

where α is the specified statistical significance level and is set to 0.05 in this work. The above criterion is used to exclude the samples from the between-phase transitional stages. If all the posterior probabilities with respect to different phases satisfy

$$p(\Xi_q \mid x_t) < \alpha, \quad q = 1, 2, \ldots, Q \tag{10}$$

then the sample is classified into a between-phase transitional stage as

$$q_t \in \begin{cases} [q_t\{1\}, q_t\{2\}] & \text{if } q_t\{2\} > q_t\{1\} \\ [q_t\{2\}, q_t\{1\}] & \text{if } q_t\{2\} < q_t\{1\} \end{cases} \tag{11}$$

where $q_t\{1\}$ and $q_t\{2\}$ correspond to the first and second phases constituting the transitional stage. The first phase $q_t\{1\}$ is set to the operating phase with the largest posterior probability

$$q_t\{1\} = \underset{q}{\operatorname{argmax}} \; p(\Xi_q \mid x_t) \tag{12}$$

Further, the second phase is identified as one of the two adjacent operating phases of the first phase that has the larger posterior probability

$$q_t\{2\} = \underset{q \; = \; q_t\{1\}-1 \text{ or } q_t\{1\}+1}{\operatorname{argmax}} p(\Xi_q \mid x_t) \tag{13}$$

As the various phases are determined, multiple localized Gaussian process regression models can be developed to characterize the shifting dynamics. A Gaussian process is a collection of random variables, any finite number of which have joint Gaussian probability distributions [14]. 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

$$y_i = f(x_i) + e_i = x_i^T \beta + e_i \tag{14}$$

where $f(\cdot)$ is the regression function, β is the vector of weights in the linear model that have zero-mean Gaussian prior probability density with covariance matrix Σ_β. Meanwhile, e_i is the Gaussian noise with zero mean and variance σ_e^2. A Gaussian process is specified by its mean and covariance functions $m(x) = \mathbb{E}[f(x)]$ and $k(x, x^T) = \mathbb{E}[(f(x) - m(x))(f^T(x) - m^T(x))]$. Then the conditional probability density function of the output variables is

$$p(Y \mid X, \beta) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_e} \exp\left\{ -\frac{(y_i - x_i^T \beta)^2}{2\sigma_e^2} \right\} \tag{15}$$

The above likelihood function follows a Gaussian distribution of $\mathcal{N}(X^T\beta, \sigma_e^2 I_{N\times N})$ with $I_{N\times N}$ being an identity matrix. Thus, the posterior probability density function can be computed through Bayesian inference strategy as follows

$$p(\beta \mid Y, X) = \frac{p(Y \mid X, \beta) p(\beta)}{p(Y \mid X)} \tag{16}$$

where the evidence is given by

$$p(Y \mid X) = \int p(Y \mid X, \beta) p(\beta) \mathrm{d}\beta \tag{17}$$

Further, the mean of the weights is computed as

$$\bar{\beta} = \sigma_e^{-2} (\sigma_e^{-2} X X^T + \Sigma_\beta^{-1})^{-1} X Y = \sigma_e^{-2} A^{-1} X Y \tag{18}$$

with $A = \sigma_e^{-2} X X^T + \Sigma_\beta^{-1}$. The posterior probability function is Gaussian as $p(\beta \mid X, Y) \sim \mathcal{N}(\bar{\beta}, A^{-1})$.

Further, a multi-kernel GPR strategy is developed to build multiple localized GPR models for different batch phases. With $\phi(\cdot)$ denoting the nonlinear mapping function, the nonlinear regression model is expressed as

$$y_i = \phi(x_i)^T \beta + e_i \tag{19}$$

A kernel function $\kappa(x, x')$ can be defined as follows to replace the nonlinear mapping function

$$\kappa(x, x') = \phi(x)^T \Sigma_\beta \phi(x') \tag{20}$$

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

$$p(Y_t \mid X_t, X, Y) \sim \mathcal{N}(\bar{Y}_t, \mathrm{cov}(Y_t)) \tag{21}$$

where

$$\bar{Y}_t = \kappa(X_t, X) \left[\kappa(X, X) + \sigma_e^2 I_{N\times N} \right]^{-1} Y \tag{22}$$

and

$$\mathrm{cov}(Y_t) = \kappa(X_t, X_t) - \kappa(X_t, X) \left[\kappa(X, X) + \sigma_e^2 I_{N\times N} \right]^{-1} \kappa(X, X_t) \tag{23}$$

Alternatively, Eq. 22 can be rewritten for Q different kernel functions corresponding to various phases as

$$\bar{Y}_t^{(q)} = \Lambda \kappa_q(X_t^{(q)}, X^{(q)}) \tag{24}$$

where $\Lambda = \left[\kappa_q(X^{(q)}, X^{(q)}) + \sigma_e^2 I_{N_q \times N_q} \right]^{-1} Y^{(q)}$, $\kappa_q(X^{(q)}, X^{(q)})$ denotes the local kernel function, and $X_t^{(q)}$ and $\bar{Y}_t^{(q)}$ are the input and output test data for q-th operating phase [12], [14]. The following Gaussian kernel function is adopted in the model developments

$$\kappa_q(X_t^{(q)}, X^{(q)}) = \exp\left(-\frac{\| X_t^{(q)} - X^{(q)} \|^2}{2\sigma_q^2} \right) \tag{25}$$

where σ_q^2 is the Gaussian kernel width for the q-th operating phase. The GPR model parameters Θ including all the kernel widths can be obtained by maximizing the following likelihood function

$$\Theta^* = \underset{\Theta}{\operatorname{argmax}} \ \log p(Y \mid X, \Theta) \tag{26}$$

Although the developed MKGPR approach can handle different operating phases, the between-phase transient dynamics is not well characterized by the localized models. Therefore, Bayesian model averaging strategy [15] is further designed to integrate the local GPR models that represent the adjacent operating phases comprising each transitional stage. For each between-phase transitional stage, the two local GPR models corresponding to the first and second phases are denoted as M_1 and M_2, respectively. Then the conditional probability distribution of the prediction $\widehat{Y}_t^{(l)}$ given a local model M_l $(l = 1 \text{ or } 2)$ is expressed as

$$p(\widehat{Y}_t^{(l)} \mid M_l) = \int p(\widehat{Y}_t^{(l)} \mid \beta_l, M_l) p(\beta_l \mid M_l) \mathrm{d}\beta_l \tag{27}$$

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

$$p(M_l \mid \widehat{Y}_t^{(l)}) = \frac{p(\widehat{Y}_t^{(l)} \mid M_l) p(M_l)}{\sum_{l'=1}^{2} p(\widehat{Y}_t^{(l')} \mid M_{l'}) p(M_{l'})} \tag{28}$$

Further, the posterior probabilities can be normalized to $\widehat{p}(M_l \mid \widehat{Y}_t^{(l)})$ and used as dynamic weights for incorporating the two local GPR models into an averaged model for state and quality predictions as follows

$$\widehat{Y}_t = \sum_{l=1}^{2} \widehat{p}(M_l \mid \widehat{Y}_t^{(l)}) \cdot \widehat{Y}_t^{(l)} \tag{29}$$

where $\widehat{Y}_t$ is the predicted output value in a transient stage.

The step-by-step procedure of the proposed state estimation and quality prediction approach is given below and an illustrative diagram is shown in Fig. 1.

i. Synchronize and unfold the collected batch process data;
ii. Construct kernel mixture model using the training data and identify the multiple operating phases through EM algorithm;
iii. Classify the test samples into the local operating phases or the between-phase transitional stages through posterior probabilities;
iv. Develop multi-kernel Gaussian process regression models for different operating phases;
v. Conduct Bayesian model averaging to integrate the local GPR models corresponding to the adjacent operating phases of the transient stages;
vi. Perform the BMA-MKGPR based state estimation and quality prediction for test samples.

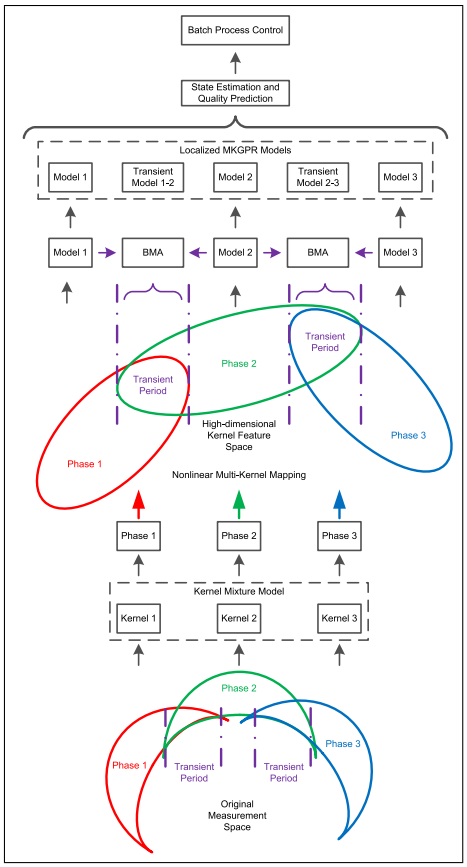

Fig. 1. Illustration of Proposed State Estimation and Quality Prediction Approach

IV. APPLICATION EXAMPLE

A. Nylon-6,6 Batch Polymerization Process

In this study, the nylon-6,6 batch polymerization process is used to evaluate the effectiveness of the proposed BMA-MKGPR approach for nonlinear state estimation and quality prediction. In the nylon polymerization process, the amine end groups react with the carboxylic end groups according to a second-order reversible condensation reaction. The product is formed through polymer chain links with the evolution of a water molecule. This process is to produce the nylon-6,6 polymers with desired product quality, which is evaluated by the number average molecular weight and amine end concentration. 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 [16]. The polymerization occurs 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. 2. 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 is the boiling phase of the batch process. The finishing phase begins when the water has been removed and the boiling step stops, which results in viscous polymers with high molecular weights.

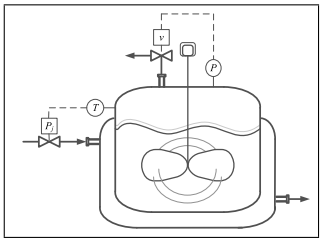

Fig. 2. Process Flow Diagram of the Nylon-6,6 Autoclave

The four input, five state and two quality variables of the nylon polymerization process are listed in Table I. In this study, all the input variables are used to develop the soft sensor models. The three selected state variables are the concentrations of amine end groups, carboxylic end groups and stabilized end groups while the chosen quality variable is the number average molecular weight. Each batch has the duration of 3 hours and the sampling period is 1 minute. Total 20 training batches and five test batches are generated for model development and performance assessment. The proposed BMA-MKSVR and BMA-MKGPR approaches are compared against the MMNN, MKSVR and MKGPR methods. It should be noted that the BMA-MKSVR method is the same as the BMA-MKGRP approach except that the phase based local SVR instead of GPR model is adopted. The model performance indices include the root-mean-square error (RMSE), mean absolute percentage error (MAPE) and R^2, as defined below

$$\text{RMSE} = \sqrt{\frac{1}{K_t}\sum_{k=1}^{K_t}(y_k - \hat{y}_k)^2} \tag{30}$$

$$\text{MAPE} = \frac{1}{K_t}\sum_{k=1}^{K}\left|\frac{y_k - \hat{y}_k}{y_k}\right| \times 100\% \tag{31}$$

$$R^2 = 1 - \frac{\sum_{k=1}^{K_t}(y_k - \hat{y}_k)^2}{\sum_{k=1}^{K_t}(y_k - \bar{y})^2} \tag{32}$$

where K_t is the total number of test samples, y_k and $\hat{y}_k$ are the actual and predicted output values at the k-th sampling instance, and $\bar{y}$ is the mean output value.

B. State Estimation and Quality Prediction Results

In this study, the selected state and quality variables are predicted to demonstrate the effectiveness of the presented approach. The average prediction errors of different methods are depicted in Fig. 3. Meanwhile, the quantitative model comparison is listed in Table II. It is readily observed from Fig. 3 that the MMNN method leads to the worst prediction errors and thus cannot accurately estimate the state and quality variables. Meanwhile, it results in the highest RMSE and MAPE values while the lowest R^2 values across different output variables. For instance, its MAPE value

TABLE I
INPUT, STATE AND QUALITY VARIABLES OF THE BATCH 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 Amine End Groups (mol/L)	C_A	State Variable
Concentration of Carboxylic End Groups (mol/L)	C_C	State Variable
Concentration of Polymer Chain Links (mol/L)	C_L	State Variable
Concentration of Water (mol/L)	C_W	State Variable
Concentration of Stabilized End Groups (mol/L)	C_{SE}	State Variable

for the concentration of amine end groups is as high as 31.88%, indicating that the state estimation accuracy is very low. Similarly, MMNN based predictions on the molecular weight do not show satisfactory performance because of its highest RMSE and MAPE values, which are 212.46 g/mol and 25.11% respectively. The poor prediction errors on molecular weight by MMNN method is also observed in Fig. 3 and 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 itself suffers from overfitting issue and poor generalization capability. Moreover, the MMNN method does not identify the between-phase transitional stages and thus cannot account for the transient dynamics when the operating phase is shifting.

In comparison, both the MKSVR and MKGPR approaches result in lower RMSE and MAPE values but higher R^2 values on different state and quality variables than the MMNN method. For example, the MAPE values on the concentration of stabilized end groups are 20.74% and 17.85% for MKSVR and MKGPR methods, respectively. They are much lower than that of the MMNN method and it reveals that the SVR and GPR models have stronger generalization capability than ANN model. Meanwhile, the multi-kernel strategy can deal with the shifting phases so that the prediction errors in local phases relatively low. However, neither MKSVR nor MKGPR method specifically identifies the between-phase transitional stages. Therefore, the prediction errors become much worse in the transitional periods as the transient dynamics are not precisely characterized. Though multiple models are developed for different phases, the strategy of classifying process samples into particular phases is insufficient as the between-phase transitional stages may have different transient dynamics from the regular phases. The further comparison between the MKSVR and MKGPR methods indicates that the MKGPR model provides slightly better prediction results on different performance indices. This is due to the enhanced capability of GPR model in accounting for stochastic uncertainty in multiphase batch processes.

The prediction error trends on different output variables in Fig. 3 show that the BMA-MKSVR and BMA-MKGPR

approaches lead to much smaller prediction errors throughout the entire batch than the other methods. Particularly, BMA based model predictions during the between-phase transitional stages exhibit dramatically reduced errors and improved accuracies. Further, as shown in Table II, BMA based approaches yield the significantly lower RMSE and MAPE values while the higher R^2 values across different state and quality variables. The improved performance can be attributed to the identification of between-phase transitional stages as well as the Bayesian inference based dynamic model averaging. Because GPR model can handle process uncertainty more efficiently, BMA-GPR approach shows even higher prediction accuracies than BMA-SVR method.

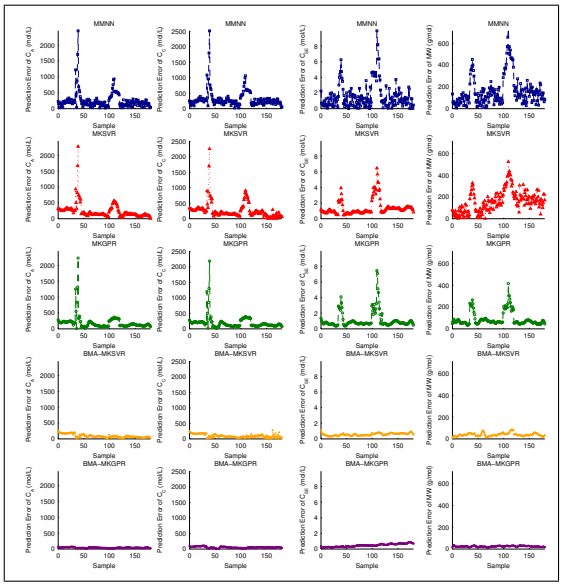

Fig. 3. Prediction Errors of State and Quality Variables by Different Soft Sensor Methods

TABLE II
QUANTITATIVE COMPARISON ON STATE AND QUALITY VARIABLES BY DIFFERENT SOFT SENSOR METHODS

Index	Variable	MMNN	MKSVR	MKGPR	BMA-MKSVR	BMA-MKGPR
RMSE	C_A	362.64	333.65	287.44	104.24	41.27
	C_C	404.87	365.61	287.84	107.85	56.45
	C_{SE}	2.30	1.66	1.50	0.59	0.50
	MW	212.46	187.68	108.19	41.36	24.12
MAPE	C_A	31.88	25.26	21.78	9.24	5.28
	C_C	31.31	23.08	17.81	8.35	5.17
	C_{SE}	24.27	20.74	17.85	9.47	5.22
	MW	25.11	20.48	17.56	7.48	5.00
R^2	C_A	0.9906	0.9920	0.9941	0.9992	0.9999
	C_C	0.9880	0.9902	0.9939	0.9991	0.9998
	C_{SE}	0.9897	0.9946	0.9956	0.9993	0.9995
	MW	0.9496	0.9607	0.9869	0.9981	0.9994

V. CONCLUSIONS

Accurate state estimation and quality prediction are essential for advanced control and optimization on batch processes, which can be very complicated due to the strong nonlinearity, inherent dynamics, and multiplicity of operating phases. Moreover, the between-phase transitional stages are often present with complex transient dynamics so that the phase-based multi-model approaches may not be well-suited. In this study, a novel BMA-GPR approach is developed for state estimation and quality prediction of multiphase batch processes. The kernel mixture model is used to identify different operating phases and then the Bayesian inference strategy is designed to isolate local phases as well as between-phase transitional stages. Further, multi-kernel local GPR models are integrated with Bayesian model averaging to specifically characterize between-phase transient dynamics.

The proposed method is applied to a simulated batch polymerization process and the results show that the BMA-GPR approach has superior prediction accuracies as opposed to the other methods. Particularly, the shifting operating phases and between-phase transient dynamics are well characterized by the present approach. Future research can be focused on Bayesian model averaging based predictive control design for nonlinear multiphase batch processes.

REFERENCES

[1] P. Bogaerts and A. Wouwer, "Parameter identification for state estimation-application to bioprocess software sensors," *Chem. Eng. Sci.*, vol. 59, pp. 2465–2476, 2004.

[2] P. Kadlec, B. Gabrys, and S. Strandt, "Data-driven soft sensors in the process industry," *Comput. Chem. Eng.*, vol. 33, pp. 795–814, 2009.

[3] J. Yu, "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.*, vol. 82, pp. 22–30, 2012.

[4] M. Soroush, "Nonlinear state-observer design with application to reactors," *Chem. Eng. Sci.*, vol. 52, pp. 387–404, 1997.

[5] R. Kandepu, B. Foss, and L. Imsland, "Applying the unscented Kalman filter for nonlinear state estimation," *J. Proc. Cont.*, vol. 18, pp. 753–768, 2008.

[6] X. Zhang, W. Yan, and H. Shao, "Nonlinear multivariate quality estimation and prediction based on kernel partial least squares," *Ind. Eng. Chem. Res.*, vol. 47, pp. 1120–1131, 2008.

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

[8] J. Yu, "A multiway Gaussian mixture model based adaptive kernel partial least squares regression method for soft sensor estimation and reliable quality prediction of nonlinear multiphase batch processes," *Ind. Eng. Chem. Res.*, vol. 51, pp. 13227–13237, 2012.

[9] J. Yu, "A Bayesian inference based two-stage support vector regression framework for soft sensor development in batch bioprocesses," *Comput. Chem. Eng.*, vol. 41, pp. 134–144, 2012.

[10] C. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2007.

[11] A. Smola and B. Schölkopf, "A tutorial on support vector regression," *Stat. Comput.*, vol. 14, pp. 199–222, 2004.

[12] S. Qiu and T. Lane, "A framework for multiple kernel support vector regression and its applications to siRNA efficacy prediction," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 6, pp. 190–199, 2009.

[13] M. Titsias and A. Likas, "Shared kernel models for class conditional density estimation," *IEEE Trans. Neural Net.*, vol. 12, pp. 987–997, 2001.

[14] C. Rasmussen and C. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA, USA: MIT Press, 2006.

[15] S. Neuman, "Maximum likelihood Bayesian averaging of uncertain model predictions," *Stoch. Environ. Res. Risk. Assess.*, vol. 17, pp. 291–305, 2003.

[16] S. Russell, D. Robertson, J. Lee, and B. Ogunnaike, "Control of product quality for batch nylon 6,6 autoclaves," *Chem. Eng. Sci.*, vol. 53, pp. 3685–3702, 1998.