Journal of Applied Polymer Science

Real-Time Property Prediction for an Industrial Rubber-Mixing Process with Probabilistic Ensemble Gaussian Process Regression Models

Yi Liu, Zengliang Gao
Engineering Research Center of Process Equipment and Remanufacturing, Ministry of Education, Institute of Process Equipment and Control Engineering, Zhejiang University of Technology, Hangzhou 310014, People's Republic of China
Correspondence to: Y. Liu (E-mail: yliuzju@zjut.edu.cn)

ABSTRACT: In internal rubber-mixing processes, data-driven soft sensors have become increasingly important for providing online measurements for the Mooney viscosity information. Nevertheless, the prediction uncertainty of the model has rarely been explored. Additionally, traditional viscosity prediction models are based on single models and, thus, may not be appropriate for complex processes with multiple recipes and shifting operating conditions. To address both problems simultaneously, we propose a new ensemble Gaussian process regression (EGPR)-based modeling method. First, several local Gaussian process regression (GPR) models were built with the training samples in each subclass. Then, the prediction uncertainty was adopted to evaluate the probabilistic relationship between the new test sample and several local GPR models. Moreover, the prediction value and the prediction variance was generated automatically with Bayesian inference. The prediction results in an industrial rubber-mixing process show the superiority of EGPR in terms of prediction accuracy and reliability. © 2014 Wiley Periodicals, Inc. J. Appl. Polym. Sci. **2015**, *132*, 41432.

KEYWORDS: applications; manufacturing; rubber

Received 11 June 2014; accepted 21 August 2014
DOI: 10.1002/app.41432

INTRODUCTION

The rubber-mixing process is the first and an important production process in the rubber and tire industry. It can be described as a very fast (ca. only 2–5 min), nonlinear, and time-varying batch (or fed-bath) process performed in an internal mixer. The Mooney viscosity, which can be likened to a composite measurement of the viscoelastic behavior of an elastomer and indirectly represents the molecular weight, is one of the key quantities in the end-use properties in the rubber-mixing process.[1–14] To achieve an optimal and uniform product quality, variation in the Mooney viscosity is expected to be small in a fixed narrow distribution. However, it is difficult to accomplish because of several major factors, which are described later:

1. It is still difficult and time-consuming to accurately predict the mechanical properties of the rubber compound in terms of first principles because of the complexity of the internal mixing process. The intrinsically varying nature and different aggregate conditions of raw materials make it difficult to know how these factors will affect the final properties of compounds.[2–5] Moreover, the varying properties of natural or synthetic rubbers introduce an amount of complexity and uncertainty to the mixing process.

2. The rubber compound properties are roughly influenced by manual machine operations and the operators. Any upsets in the feed and additives, for example, carbon black and oil, may change the product properties.[5] Additionally, the extent of batch-to-batch reproducibility depends on the configuration of the mixer and the mixing operating conditions, such as the mixing steps, mixing temperature, mixing pressure, cooling efficiency, and rotor power.[7–14]

3. The Mooney viscosity information can be only obtained from laboratory analysis several hours later after a batch has been discharged. Therefore, a plain feedback control of the viscosity is impossible, and a dominant manner for controlling the viscosity in the rubber industry is mainly based on process operators' long-term experience and expertise.[3,7–14]

To overcome these embarrassments and ensure quality in the final products, various empirical prediction models, including artificial neural networks (ANNs),[3–6] neuro-fuzzy systems,[7] partial least squares (PLS) and linear regression,[8,9] support vector machines (SVMs) and other kernel learning methods,[10–12] and Gaussian process regression (GPR),[13,14] have been applied to the online prediction of the viscosity information. Compared to comprehensive mechanistic models, one main advantage of data-driven soft-sensor models in chemical processes is that they can generally be developed quickly without the need for a substantial understanding of the phenomenology.[15–19] Among

these prediction models, PLS and other multivariate regression methods can only extract the linear information from the modeling data.[8,9] In practice, many product qualities are believed to have nonlinear relationships with the operating conditions. For ANN and neuro-fuzzy systems, the determination of the network complexity is still unsolved.[20] SVMs, least squares support vector machines (LSSVMs), and GPR-based soft-sensors have attracted more attention recently because of their nonlinear modeling ability.[20–32] However, the parameter selection for an SVM/LSSVM model is still difficult in chemical processes. Compared with SVMs/LSSVMs, the GPR model can automatically optimize its parameters with an iterative method.[28] Additionally, GPR can simultaneously provide the probabilistic information for its prediction.[28] However, this interesting property has been investigated less in the application of GPR for Mooney viscosity prediction.

In previous research, most existing data-driven modeling methods for the Mooney viscosity[3–14] have focused on how to construct a global prediction model. Actually, there are many recipes for internal rubber-mixing processes. A single global model is insufficient for capturing all of the process characteristics in different recipes. Although moving-window-based recursive soft sensors can gradually be adapted to new operational conditions, deciding how to choose a suitable moving-window size for complex rubber-mixing processes is difficult.[10,11,17] Additionally, most recursive models may not function well in a new operational region until a sufficient period of time has passed because of the time delay when they adapt themselves to new operational conditions.[18] Recently, three GPR-based mixture models for the multimode chemical processes have been proposed.[30–32] However, prediction uncertainty has rarely been integrated into traditional prediction models, including previous GPR models. Additionally, to the best of our knowledge, the multi-GPR-based modeling method has not been applied to rubber-mixing processes yet.

In this article, considering the aforementioned factors, we propose a novel probabilistic ensemble Gaussian process regression (EGPR) modeling method. First, a fuzzy c-means (FCM) clustering approach[33,34] was used to cluster all of the training samples into several subclasses. Then, several single GPR models were trained with each subclass of samples. Moreover, they were further assembled to improve the prediction reliability. In contrast to conventional ensemble strategies,[35–37] the uncertainty information of each single GPR model was analyzed, and a Bayesian strategy was adopted. Finally, the prediction value, the prediction variance and the status information are all obtained. They can be provided for operators.

Briefly, this article is organized as follows. The FCM clustering algorithm and several local GPR models for subclasses are formulated in the EGPR Prediction Model section. Sequentially, the novel EGPR-based soft sensor for the online evaluation and prediction of a test sample is proposed in this section. The EGPR method is evaluated by the Mooney viscosity prediction in an industrial process in the Industrial Application: Mooney Viscosity Online Prediction section. Comparison studies with other methods are also investigated. Finally, concluding remarks are made in the final section.

EGPR PREDICTION MODEL

FCM Clustering Approach

There are several recipes in an internal rubber-mixing process. Each recipe has different sized samples. The samples in different recipes may exhibit different characteristics. Additionally, for a special recipe, samples with different operating conditions often show distinguished properties. It is unsuitable to construct a global model with all of the samples. Also, the samples should be pretreated to divide them into some subgroups that can be modeled in a relatively simple way.

The FCM clustering algorithm divides a set of samples into several clusters such that samples within a given cluster have a higher degree of similarity, whereas samples belonging to different clusters have a higher degree of dissimilarity.[33,34] It has been widely applied to solve many clustering problems and has shown superiority to k means.[33,34] Therefore, FCM was adopted to preprocess the training samples before the construction of a prediction model. A set of training samples, $X=(x_1, x_2, \ldots x_N)$, was organized into L clusters by the minimization of the objective function J as follows:[33,34]

$$J(\mathbf{U},\mathbf{c})=\sum_{i=1}^{L}\sum_{j=1}^{N}u_{ij}^{m}d_{ij}^{2} \qquad st\sum_{i=1}^{L}u_{ij}=1,\forall j=1,2,\ldots N \tag{1}$$

where st is subject to; u_{ij} is the relationship value of the jth sample in the ith cluster; $\mathbf{U}$ is the related fuzzy partition matrix consisting of u_{ij} within the interval of [0, 1]; $\mathbf{c}=(c_1,c_2,\ldots c_L)$ is the cluster center matrix; $d_{ij}=\|c_i-x_j\|$ is the Euclidean distance-based similarity of x_j and c_i; $m \in [1,\infty]$ is the weighting parameter, and the fuzziness of the clustering increases with increasing m. In addition, the necessary conditions for minimizing eq. (1) are the update equations as follows:[33,34]

$$c_i=\frac{\sum_{j=1}^{N}u_{ij}^{m}x_j}{\sum_{j=1}^{N}u_{ij}^{m}} \tag{2}$$

$$u_{ij}=\frac{1}{\sum_{l=1}^{L}\left(\frac{d_{ij}}{d_{lj}}\right)^{2/(m-1)}} \tag{3}$$

Detailed implementations of the iterative clustering scheme for FCM are found in the literature.[33,34] Finally, the initial training set X can be clustered into L subclasses, which are described as $X=\{X_1, X_2, \ldots X_L\}$.

Single GPR-Based Prediction Model for Subclasses

Generally, soft-sensor development in a nonlinear process based on the GPR framework can be described as a problem whose aim is to learn a model f that approximates a training set $S=\{X, Y\}$, where $X=\{x_i\}_{i=1}^{N}$ and $Y=\{y_i\}_{i=1}^{N}$ are the input and output datasets with N samples, respectively. The initial set S can be clustered into L subclasses denoted as $S=\{S_1, S_2, \ldots S_L\}$. Each subclass with N_l samples can be described as $S_l=\{X_l, Y_l\}=\{x_{l,i}, y_{l,i}\}_{i=1}^{N_l}$ and $N=\sum_{i=1}^{L}N_l$. For the lth subclass, a general nonlinear GPR model provides a prediction of the output variable for an input sample through Bayesian

inference. For an output variable $Y_l = (y_{l,1}, \ldots, y_{l,Nl})^T$, the GPR model is the regression function with a Gaussian prior distribution and zero mean or, in a discrete form:[28,29]

$$Y_l = \left(y_{l,1}, \ldots, y_{l,N_l}\right)^T \approx G(0, \mathbf{C}_l) \tag{4}$$

where $\mathbf{C}_l$ is the $N_l \times N_l$ covariance matrix with the ijth element defined by the covariance function $\mathbf{C}_{l,ij} = C_l^*(x_{l,i}, x_{l,j})$. A common covariance function can be defined as follows:[28]

$$C_l\left(x_{l,i}, x_{l,j}\right) = a_{l,0} + a_{l,1}\sum_{d=1}^{D} x_{l,i.d}x_{l,j.d} + v_{l,0} \exp\left[-\sum_{d=1}^{D} w_{ld}\left(x_{l,i.d} - x_{l,j.d}\right)^2\right] + \delta_{ij}b_l \tag{5}$$

where $x_{l,i,d}$ is the dth component of the vector $\mathbf{x}_{l,i}$. The value of δ_{ij} is 1 if $i = j$; otherwise, it is equal to zero. $\boldsymbol{\theta}_l = [a_{l,0}, a_{l,1}, v_{l,0}, w_{l,1}, \ldots, w_{l,D}, b_l]^T$ describes the hyperparameters vector defining the covariance function. When both the linear and nonlinear terms in the covariance function are combined, the GPR model is capable of handling both linear and nonlinear processes.[28,29]

Adopting a Bayesian approach to train a GPR model, the values of the hyperparameters ($\boldsymbol{\theta}_l$) can be estimated by the maximization of the following log likelihood function:[28]

$$J(\boldsymbol{\theta}_l) = -\frac{1}{2}\log[\det(\mathbf{C}_l)] - \frac{1}{2}Y_l^T\mathbf{C}_l^{-1}Y_l - \frac{N_l}{2}\log(2\pi) \tag{6}$$

This optimization problem can be solved with the derivative of the log likelihood with respect to each hyperparameter given as follows:[28]

$$\frac{\partial J}{\partial \boldsymbol{\theta}_l} = -\frac{1}{2}\mathrm{tr}\left(\mathbf{C}_l^{-1}\frac{\partial \mathbf{C}_l}{\partial \boldsymbol{\theta}_l}\right) + \frac{1}{2}Y_l^T\mathbf{C}_l^{-1}\frac{\partial \mathbf{C}_l}{\partial \boldsymbol{\theta}_l}\mathbf{C}_l^{-1}Y_l \tag{7}$$

where $\frac{\partial \mathbf{C}_l}{\partial \boldsymbol{\theta}_l}$ can be obtained from the covariance function. Detailed implementations for training a GPR model can be found in Rasmussen and Williams.[28] Also, the main computational load for training a GPR model is about $O\left(N_l^3\right)$, which is always feasible for a moderate size of training data set (less than several thousands) on a conventional computer. For larger data sets, sparse training strategies may be required to reduce the overall computational burden.[38]

Finally, this single GPR model (GPR_l) for subclass S_l can be obtained once $\boldsymbol{\theta}_l$ is determined. As for a new test sample x_t, the predicted output of y_t is also Gaussian with the mean ($\hat{y}_{l,t}$) and variance ($\sigma^2_{\hat{y}_{l,t}}$), calculated as follows:[28]

$$\hat{y}_{l,t} = k_{l,t}^T\mathbf{C}_l^{-1}Y_l \tag{8}$$

$$\sigma^2_{\hat{y}_{l,t}} = k_{l,t} - k_{l,t}^T\mathbf{C}_l^{-1}k_{l,t} \tag{9}$$

where $\mathbf{k}_{l,t} = [C_l(x_t, x_{b1}), C_l(x_t, x_{b2}), \ldots C_l(x_t, x_{l,N_l})]^T$ is the covariance vector between the new input and the training samples and $k_{l,t} = C(x_t, x_t)$ is the covariance of the new input. In summary, eq. (9) provides a confidence level on the prediction, which is an appealing property of the GPR method.

Consequently, several single GPR models, denoted as GPR_l, $l = 1, \ldots, L$, can be constructed offline for different subclasses with the aforementioned formulations, that is, eqs. (4–7). For the online prediction of a new sample x_t, both $\hat{y}_{l,t}$ and $\sigma^2_{\hat{y}_{l,t}}$ can be obtained with eqs. (8) and (9).

Prediction Variance-Based Bayesian Ensemble of the GPR Models

For a rubber-mixing process with several recipes, after the FCM-based clustering, several GPR-based local prediction models can be built. For a new test sample x_t, however, its prediction model is not supposed to be known beforehand. It is important to evaluate to which model it is most suitable (or how to combine these existing local models in a reasonable way). Without any process or expert knowledge, this should be done purely by the data-driven method. Compared with SVM-based deterministic models, prediction variances provide useful information for the description of a model's reliability. However, the prediction uncertainty has been neglected in previous mixture GPR models for chemical processes. Additionally, the applications of mixture GPR models to rubber-mixing processes have never been investigated. In this section, a prediction variance-based Bayesian method is proposed for exploring the reliability of existing GPR-based prediction models. Moreover, a probabilistic information-based ensemble strategy is formulated to obtain a more reliably predictive result.

To evaluate the relationship between a local GPR model and the new test sample x_t, a prediction variance-based Bayesian inference is proposed to determine the probability of x_t with each GPR, $l = 1, \ldots, L$ model, that is, $P(\mathrm{GPR}|x_t)$, $l = 1, \ldots, L$, which is given as follows:[30,32]

$$P(\mathrm{GPR}_l|x_t) = \frac{P(x_t|\mathrm{GPR}_l)P(\mathrm{GPR}_l)}{P(x_t)} = \frac{P(x_t|\mathrm{GPR}_l)P(\mathrm{GPR}_l)}{\sum_{l=1}^{L}[P(x_t|\mathrm{GPR}_l)P(\mathrm{GPR}_l)]}, l = 1, \ldots, L \tag{10}$$

where $P(\mathrm{GPR}_l)$, $l = 1, L$, and $P(x_t|\mathrm{GPR}_l)$, $l = 1, \ldots, L$ are the prior probability and the conditional probability, respectively. To calculate the posterior probability value, these two terms at the right side of eq. (10) should be defined. Without any process or expert knowledge, the prior probability for each local GPR model can be simply defined as follows:[30,32]

$$P(\mathrm{GPR}_l) = \frac{N_l}{N}, l = 1, \ldots, L \tag{11}$$

To determine the other terms in eq. (10), first, a relative root prediction variance (RPV) item of this test sample for each single GPR model can be further modified as follows:

$$v_{l,x_t} = \frac{\sqrt{\sigma^2_{\hat{y}_{t,l}}}}{\hat{y}_{t,l}} \times 100\% = \frac{\sigma_{\hat{y}_{t,l}}}{\hat{y}_{t,l}} \times 100\%, l = 1, \ldots, L \tag{12}$$

where the actual value of y_t is unknown, so it is substituted by its prediction with the related GPR_l, $l = 1, \ldots, L$ models, noted as $\hat{y}_{t,l}$. The item $\sigma_{\hat{y}t,l}$ can be used to describe the relative prediction uncertainty for a sample with this special local model. The value of $\sigma_{\hat{y}t,l}$ is relatively large if the test sample x_t is predicted with an unsuitable model. Consequently, a larger value of $v_{l,xt}$ generally means a larger uncertainty when this special GPR model for online prediction is adopted. Because of this reason, without a loss of generality, the conditional probability

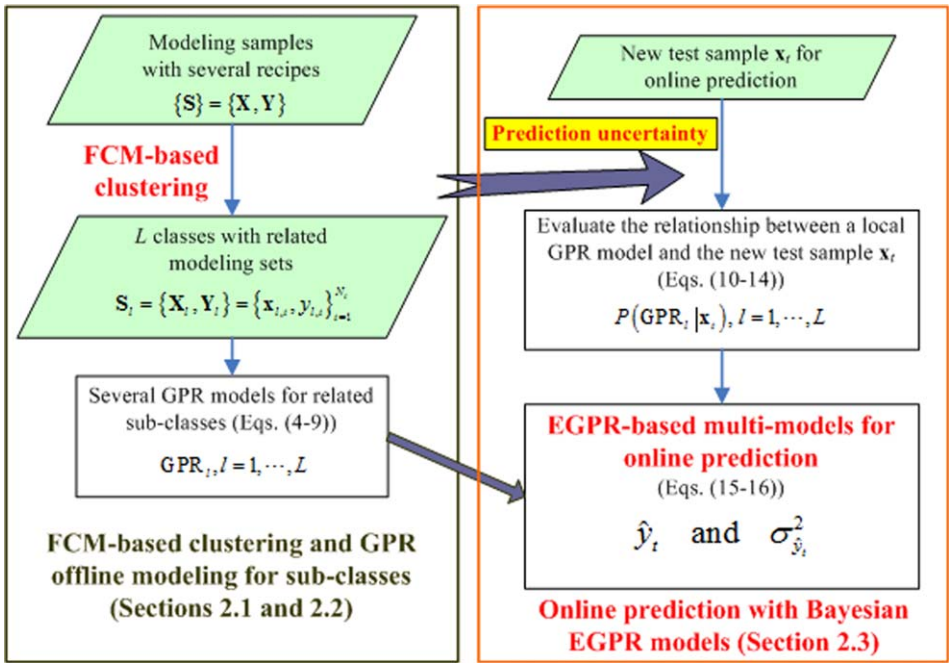

Figure 1. Online prediction of the rubber properties with the Bayesian EGPR-based soft-sensor modeling method for a rubber-mixing process with several recipes. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

$[P(x_t|\mathrm{GPR}_l)]$ can be defined on the basis of an inverse relationship of $v_{l,xt}$:

$$P(x_t|\mathrm{GPR}_l)=\frac{1}{v_{l,x_t}}, l=1, \ \ldots, L \tag{13}$$

Consequently, eq. (10) becomes

$$P(\mathrm{GPR}_l|x_t)=\frac{N_l}{v_{l,x_t}\sum_{l=1}^{L}\left(N_l/v_{l,x_t}\right)}, l=1,\ldots,L \tag{14}$$

On the basis of the probabilistic analysis approach, the use of the prediction variance of local GPR models can determine the probability between x_t and each local GPR model. This probabilistic information of local predictions can be combined to form the final prediction result. Therefore, an EGPR modeling method for multirecipe rubber-mixing processes is formulated with the aforementioned probability analysis. The mean and variance values of the final predictive distribution can be calculated as follows:

$$\hat{y}_t=\sum_{l=1}^{L}P(\mathrm{GPR}_l|x_t)\hat{y}_{l,t} \tag{15}$$

$$\sigma_{\hat{y}_t}^2=\sum_{l=1}^{L}P(\mathrm{GPR}_l|x_t)\sigma_{\hat{y}_{l,t}}^2 \tag{16}$$

In summary, the algorithm implement of the EGPR modeling method is illustrated in Figure 1. The left part mainly includes an offline clustering and modeling stage. Other clustering algorithms, for example, the finite Gaussian mixture model,[32] can be used for clustering. The right part is the online prediction method. It is relatively a simple task to construct several single local GPR models, whereas the determination of how to predict a new test sample with suitable models is difficult, especially for those samples in recipe transitions or new operating modes with complex characteristics. From an engineering standpoint, this method can be simply implemented, and the interpretation of the prediction result is also straightforward. Although the single GPR model is only efficient for quality prediction in its specific region, the ensemble GPR model can handle multiple recipes and different operational modes. Through the evaluation of the new test sample, a probabilistic prediction result can be generated. Therefore, the EGPR method is more appropriate for online prediction of the Mooney viscosity with multiple recipes in rubber-mixing processes.

INDUSTRIAL APPLICATION: MOONEY VISCOSITY ONLINE PREDICTION

The flowchart of the modeling and online prediction of the Mooney viscosity in an industrial rubber-mixing process is shown in Figure 2. The manufacturer was located in southeast China. Because commercial secrets, further details are not provided. The rubber compound samples for the laboratory assay were taken from a continuous rubber sheet roughly according to the end discharge time. The Mooney viscosity assay results were obtained via a shearing disk viscometer at a standard-prescribed elevated temperature. The Mooney viscosity basically reflects the degree of polymerization and the molecular weight of the mixed rubber. For example, the measurement result is reported in the form 50ML, 1 + 4 (100°C), where 50 M is the Mooney viscosity value (where M is its unit), L indicates the use of the large 1.5-in. rotor, 1 is the time in minutes for which the specimen is permitted to warm in the machine before the motor is started and the reading is taken, 4 is the time in minutes that the rotor runs, and 100°C is the temperature of the test.[12,13] According to the physical and chemical

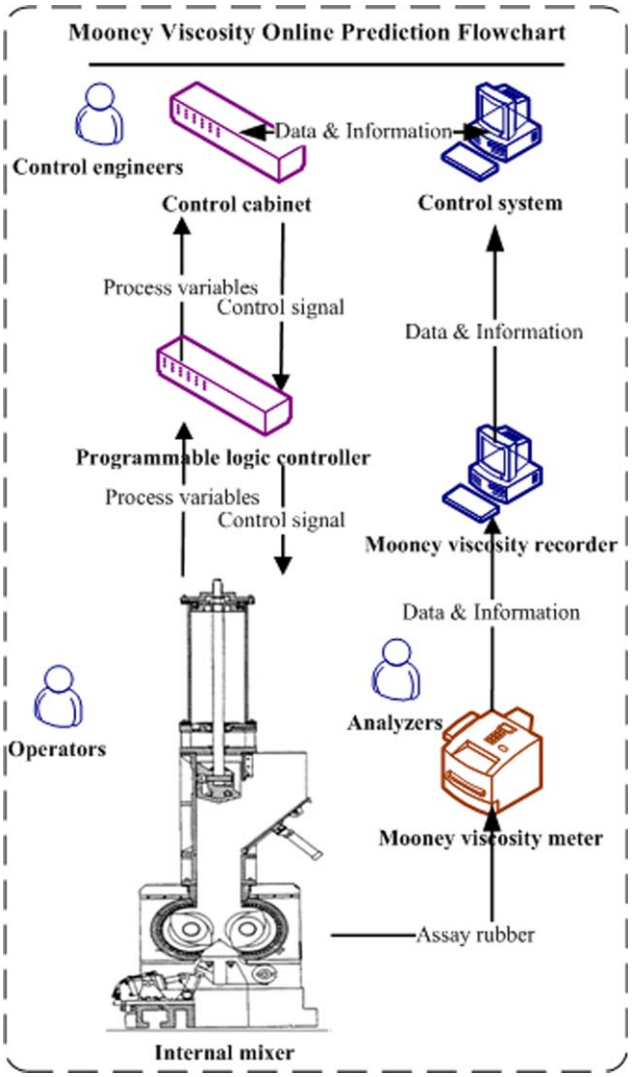

Figure 2. Mooney viscosity online prediction flowchart in an industrial rubber-mixing process. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

mechanisms, the Mooney viscosity can be measured after the mixed rubber has been conditioned for a given time (ca. 4 h). Therefore, the measurement time delay is much larger than the period of mixing process.

As mentioned previously, the properties of rubber materials and additives play an important role in the Mooney viscosity. Additionally, the measured variables of the process, including the mixing temperature, mixing energy, mixing power, mixing pressure, and mixing duration in the chamber of the internal mixer, can be selected as secondary variables according to technology analysis.[8–13] Therefore, all of the information obtained from the database during the process can be used to form the input variables of the prediction model. Additionally, as prior requirements, the reliable sensor measurements and data collections play crucial roles in the process modeling.[16] After simple preprocessing of the modeling set with 3-σ criterion, most of the outlier samples and missing values are removed. Finally, about 200 samples of six recipes collected in a product line were investigated in this study. Four recipes of the samples were used for training and the remaining two recipes were for testing.

In this study, the GPR and SVM methods were investigated for comparison, mainly because they were very popular in the chemical process modeling area.[9–13,20–32] The Gaussian kernel function $\{K(x_1, x_2) = \exp[-\|x_1 - x_2\|/\sigma]$ (with σ as a positive parameter)$\}$ may be the most common kernel in the SVM method[9–13,20–27] and thus is used for SVM. The parameters of SVM were selected with the fivefold cross-validation approach. The simulation environment in this case was MatLab V2009b with a CPU main frequency of 2.3 GHz and 4 GB of memory.

To quantitatively evaluate the prediction performance of these prediction models, three performance indices, including the root mean square error (RMSE), relative root mean square error (RE), and maximal absolute error (MAE), are defined as follows:

$$\mathrm{RMSE}=\sqrt{\sum_{i=1}^{N_{\mathrm{tst}}}\left(\hat{y}_i-y_i\right)^2/N_{\mathrm{tst}}} \tag{17}$$

$$\mathrm{RE}=\sqrt{\sum_{i=1}^{N_{\mathrm{tst}}}\left(\frac{\hat{y}_i-y_i}{y_i}\right)^2/N_{\mathrm{tst}}} \tag{18}$$

$$\mathrm{MAE}=\max|\hat{y}_i-y_i|,\ \text{for}\ i=1,\cdots,N_{\mathrm{tst}} \tag{19}$$

where $\hat{y}_i$ denotes the prediction of y_i, that is, the Mooney viscosity of the compound, and N_{tst} is the number of testing samples.

To investigate the data distribution of the training and testing samples, as an illustrated case shown in Figure 3, the relationships between the first to third process input variables were nonlinearly correlated with each other. Additionally, the training samples were distributed irregularly with multiple groups. Some areas had more data samples, whereas other areas showed sparse data distribution. Therefore, only a single global model was not enough to describe all of the process characteristics distributed

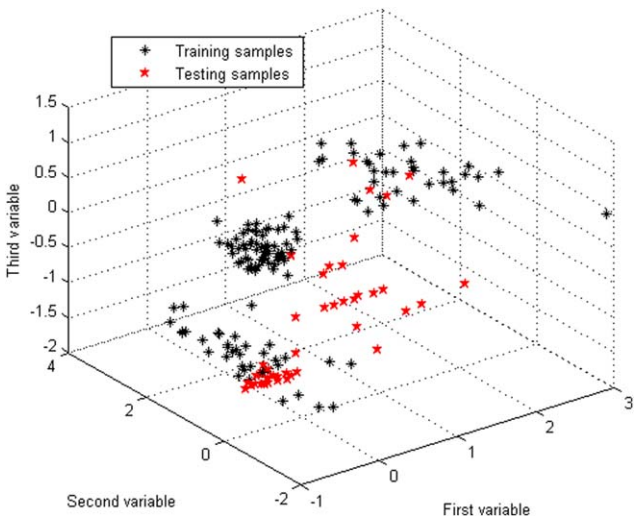

Figure 3. Process input variable relationships in the industrial rubber-mixing process (the training and testing data sets). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

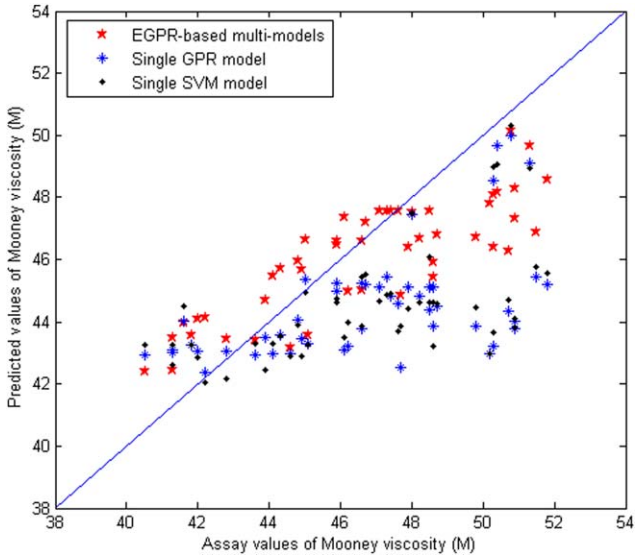

Figure 4. Parity plot based on assay values against the prediction values in the test set with EGPR-based multimodels and single GPR/SVM models. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

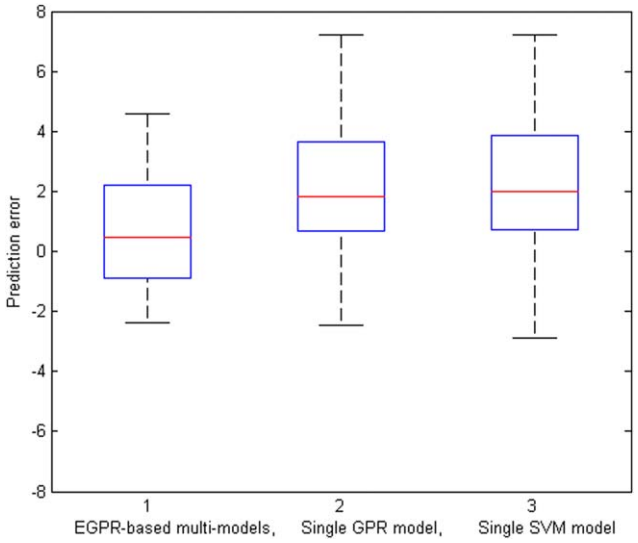

Figure 5. Comparison between the prediction error distributions of the Mooney viscosity with a box plot with EGPR-based multimodels, the single GPR model, and the single SVM model. The edges of each box are the first and third quartiles, and the band inside the box shows the median. The whiskers above and below the box show the locations of the minimum and maximum. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

in multiple recipes. To investigate the soft-sensor modeling methods, as also shown in Figure 3, the test samples exhibited different process operating conditions compared with the training samples.

In this case, after the FCM clustering, four clusters were obtained. The distribution of each subclass was separated from other subclasses. Generally, each class had its own characteristics; this was conducive to the modeling of the sub-GPR models. Then, four sub-GPR models were simply trained with the samples in each subclass (formulated in the Single GPR-Based Prediction Model for Subclasses section). For online prediction of a test sample, the results produced by these sub-GPR models were combined by the Bayesian method described in the Prediction Variance-Based Bayesian Ensemble of the GPR Models section.

The overall prediction results of the EGPR-based multimodels and the single GPR/SVM models are shown in Figure 4. The online prediction times for the test samples were less than 1 s for all methods. Obviously, the ensemble model showed an improved prediction performance over its single model. The corresponding box plots of the prediction errors with all three different prediction models are shown in Figure 5. On each box (e.g., EGPR-based multimodels), the edges of the box are the first and third quartiles, and the band inside the box shows the second quartile (i.e., the median). The whiskers above and below the box show the locations of the minimum and maximum. We found that the proposed EGPR method had the narrowest ranges of prediction errors; this implied the best prediction performance among the three methods. Additionally, the details of the online prediction comparisons of the Mooney viscosity among the EGPR, GPR, and SVM methods are tabulated in Table I. The results of the RMSE, RE, and MAE indices in Table I show that EGPR obtained a better distribution of the prediction errors compared to the other approaches because the GPR and SVM models showed unreliable prediction results in some test samples, as also shown in Figure 4.

For the online prediction of the Mooney viscosity, the prediction values of EGPR, GPR, and SVM methods for test samples are shown in Figure 6. As analyzed previously, the viscosity information often exhibited different characteristics because it was very sensitive to process conditions and operations from raw materials to the end discharge.[1–14] In this case, samples 1–29 belonged to one recipe. Samples 30–49 were in the other recipe. As shown in Figure 6, the tendencies of the EGPR prediction values and the assay values were more similar. However, neither the GPR nor SVM method could capture the tendencies in these two recipes.

To further show the prediction performance of EGPR, two recent soft sensors were adopted for comparison. One was a mixture GPR-based multimodel, called *combined local Gaussian process regression* (CLGPR).[30] The principle component analysis was used to combine the local GPR models for online prediction. The CLGPR model exhibited better prediction performance than single local GPR models through applications to a

Table I. Mooney Viscosity Online Prediction: Performance Comparison with Popular Global Soft-Sensor Models

Method	RMSE	RE (%)	MAE
EGPR	**2.01**	**4.17**	**4.02**
GPR[12,28]	3.44	6.99	11.81
SVM[9]	3.37	6.88	11.38

The best prediction performance is bolded and underlined.

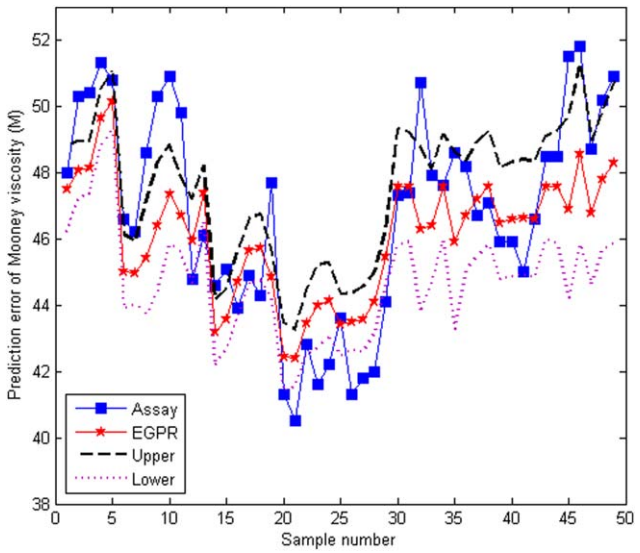

Figure 6. Comparison results of EGPR, GPR, and SVM soft-sensor models for the online prediction of the Mooney viscosity in the industrial rubber-mixing process (the test data set). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

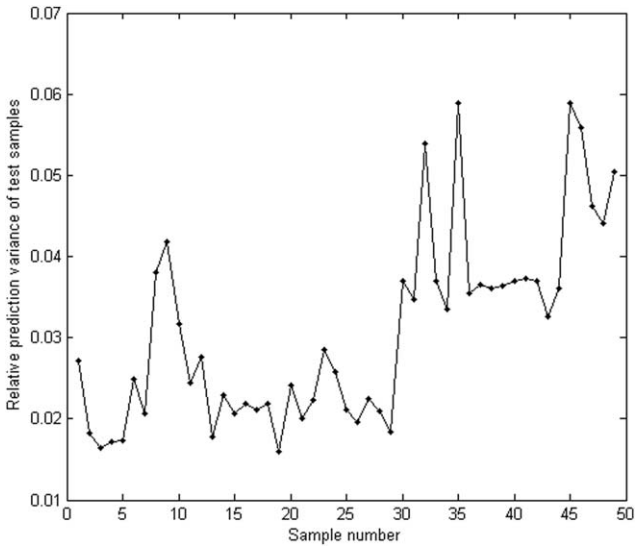

Figure 7. Online prediction of the Mooney viscosity in the industrial rubber-mixing process with the EGPR prediction model (the test data set).

polypropylene production process.[30] However, the prediction uncertainty was omitted. If some test samples show different characteristics from the training samples, the prediction accuracy will be degraded. The other method for comparison is integration of the independent component regression (ICR) and GPR into an ICR–GPR prediction model. The ICR–GPR model shows better prediction performance than the traditional ICR and GPR models.[13] Nevertheless, ICR–GPR is a fixed global model. The non-Gaussian information that exists in different recipes may be different. It is relatively difficult to capture all the non-Gaussian and nonlinear information with only a model. Additionally, the probabilistic information of GPR was not used in previous research. The details about online prediction comparisons of the Mooney viscosity among EGPR, CLGPR, and ICR–GPR methods are listed in Table II. Both of CLGPR and ICR–GPR methods exhibit a better prediction performance than only a GPR model. Moreover, the results of the RMSE, RE, and MAE indices show that the EGPR model can achieve a more accurate prediction performance than the other methods.

Finally, the detailed results of the online Mooney viscosity prediction for the test data set with the EGPR prediction model are shown in Figure 7. The upper line and lower line show $\hat{y}_i + \sigma_{\hat{y}i}$ and $\hat{y}_i - \sigma_{\hat{y}i}$, respectively. Correspondingly, the comparisons of the RPV values of the EGPR-based soft sensor for online prediction of all the test samples are shown in Figure 8 [defined in eq. (12)]. The proposed RPV item can provide additional information for the evaluation of prediction models. As shown in Figures 7 and 8, the test samples 30–49 generally exhibited a larger prediction uncertainty (cf. test samples 1–29). Actually, this recipe, which included several shifts, was much different from the previous recipes in this process. As also shown in Figure 3, these test samples were much different from those training samples. With the proposed simple RPV index, operators and engineers can determine whether the prediction is good or bad before the laboratory analysis results are available. In our opinion, it is unnecessary to further analyze the samples in the

Table II. Mooney Viscosity Online Prediction: Performance Comparison with Two Recent Prediction Models

Method	RMSE	RE (%)	MAE
EGPR	**2.01**	**4.17**	**4.02**
CLGPR[30]	2.47	5.26	6.23
ICR-GPR[13]	3.05	6.12	9.41

The best prediction performance is bolded and underlined.

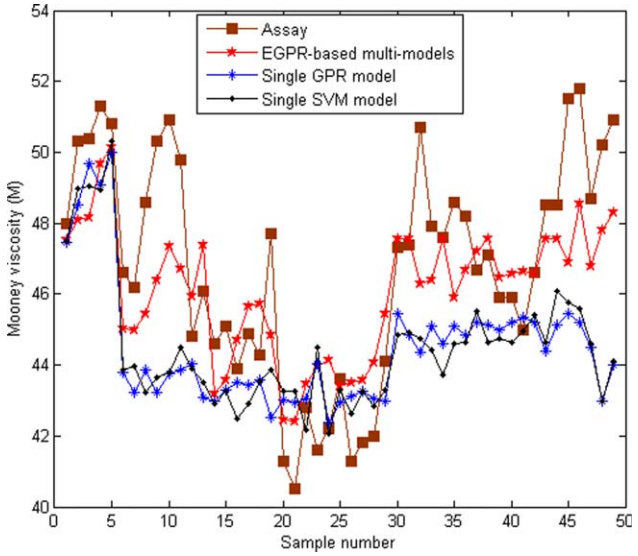

Figure 8. Comparisons of the RPV values of the EGPR-based soft sensor for the online prediction of all of the test samples in the industrial rubber-mixing process. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

laboratory if the RPV values are always very small because this indicates that the prediction model performs well for these samples.

Unlike the traditional soft sensors, for example, PLS, ANN, and SVM/LSSVM,[3–14] the GPR-based soft sensor can provide probabilistic information for its prediction.[28] The proposed RPV index is important for soft sensors because the probabilistic information can help operators/engineers use the prediction in a better way. Therefore, from all of the obtained results, the proposed EGPR method showed a better and more reliable prediction performance than the other soft sensors in terms of the online Mooney viscosity prediction for an industrial rubber-mixing process.

One assumption is that the outlier samples have been removed from the modeling data set. The determination of how to automatically detect and reconcile both input and output measurement biases and misalignments with some novel strategies, for example, the Correntropy concept,[39] is one of our future directions. Additionally, the feature extraction can be further investigated to improve prediction performance. If available, some domain knowledge and expert rules can also be combined into the soft-sensor modeling methods. So, there are still several interesting research directions worth investigating in the future to further enhance the accuracy and transparency of a reliable prediction model for rubber-mixing processes.

CONCLUSIONS

In this study, we developed a novel ensemble probabilistic prediction model for the Mooney viscosity in rubber-mixing processes. Two main distinguishing characteristics can be summarized. First, the EGPR-based multimodeling method could better handle multiple recipes with different process characteristics. Second, the prediction uncertainty was analyzed and integrated into the EGPR method to enhance the prediction reliability. Consequently, the EGPR model could effectively reduce the variance error of prediction compared to the use of only a global GPR model. The superiority of EGPR was demonstrated through an online Mooney viscosity prediction of an industrial rubber-mixing process with multiple recipes. Compared with several existing approaches, better and more reliable prediction performance of EGPR was obtained.

ACKNOWLEDGMENTS

The authors thank the National Natural Science Foundation of China (contract grant number 61004136) for the financial support.

REFERENCES

1. Mark, J. E.; Erman, B.; Eirich, F. R. The Science and Technology of Rubber; Elsevier Academic: San Diego, CA, **2005**.
2. Ryzko, P.; Haberstroh, E. *Macromol. Mater. Eng.* **2000**, *284–285*, 64.
3. Marcos, A. G.; Espinoza, A. V. P.; Elías, F. A.; Forcada, A. G. *Int. J. Comput. Integr. Manuf.* **2007**, *20*, 828.
4. Vijayabaskar, V.; Gupta, R.; Chakrabarti, P. P.; Bhowmick, A. K. *J. Appl. Polym. Sci.* **2006**, *100*, 2227.
5. Padmavathi, G.; Mandan, M. G.; Mitra, S. P.; Chaudhuri, K. K. *Comput. Chem. Eng.* **2005**, *29*, 1677.
6. Deshpande, P. B.; Yerrapragada, S. S. *Control Eng.* **1997**, *44*, 55.
7. Merikoski, S.; Laurikkala, M.; Koivisto, H. In Proceedings of Advances in Neural Networks and Applications; Mastorakis, N., Ed.; World Scientific and Engineering Society Press: Greece, **2001**; p 287.
8. Song, K.; Gao, Y. C.; Wang, H. Q.; Li, P. In Proceedings of the Fifth World Congress on Intelligent Control and Automation, **2004**; p 3295.
9. Xie, Y. C.; Wang, H. Q.; Gao, Y. C.; Li, P. In Proceedings of the 2003 American Control Conference; IEEE Conference Publications: Denver, USA, **2003**; p 3673.
10. Yang, D. C.; Liu, Y.; Fan, Y. G.; Wang, H. Q. In Proceedings of the 48th IEEE Conference on Decision and Control; IEEE Conference Publications: Shanghai, China, **2009**; p 404.
11. Gao, Y. C.; Ji, J.; Wang, H. Q.; Li, P. In Proceedings of the IEEE International Conference on Intelligent Computing and Intelligent Systems, IEEE Conference Publications: Shanghai, China, **2010**; p 470.
12. Zhang, Z.; Song, K.; Tong, T. P.; Wu, F. *Chemom. Intell. Lab. Syst.* **2012**, *112*, 17.
13. Song, K.; Wu, F.; Tong, T.; Wang, X. *J. Chemom.* **2012**, *11*, 557.
14. Sodupe Ortega, E.; Urraca Valle, R.; Antoñanzas Torres, J.; Alía Martínez, M. J.; Sanz García, A.; Martínez de Pisón Ascacíbar, F. J. In Proceedings of the 17th International Congress on Project Management and Engineering; Logroño, Spain, **2013**.
15. Fortuna, L.; Graziani, S.; Rizzo, A.; Xibilia, M. G. Soft Sensors for Monitoring and Control of Industrial Processes; Springer-Verlag: New York, **2007**.
16. Kadlec, P.; Gabrys, B.; Strandt, S. *Comput. Chem. Eng.* **2009**, *33*, 795.
17. Kadlec, P.; Grbic, R.; Gabrys, B. *Comput. Chem. Eng.* **2011**, *35*, 1.
18. Kano, M.; Ogawa, M. *J. Process. Control* **2010**, *20*, 969.
19. Yao, Y.; Gao, F. R. *Annu. Rev. Control* **2009**, *33*, 172.
20. Schölkopf, B.; Smola, A. J. Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond; MIT Press: Cambridge, MA, **2002**.
21. Liu, Y.; Hu, N. P.; Wang, H. Q.; Li, P. *Ind. Eng. Chem. Res.* **2009**, *48*, 5731.
22. Han, I. S.; Han, C. H.; Chung, C. B. *J. Appl. Polym. Sci.* **2005**, *95*, 967.
23. Shi, J.; Liu, X. G. *J. Appl. Polym. Sci.* **2006**, *101*, 285.
24. Huang, M. Y.; Liu, X. G.; Li, J. B. *J. Appl. Polym. Sci.* **2012**, *126*, 519.
25. Martínez-de-Pisón, F. J.; Barreto, C.; Pernía, A.; Alba, F. *J Mater. Process. Tech.* **2008**, *197*, 161.

26. Chitralekha, S. B.; Shah, S. L. *Can. J. Chem. Eng.* **2010**, *88*, 696.

27. Kaneko, H.; Funatsu, K. *Ind. Eng. Chem. Res.* **2011**, *50*, 10643.

28. Rasmussen, C. E.; Williams, C. K. I. Gaussian Processes for Machine Learning. MIT Press: Cambridge, MA, **2006**.

29. Chen, T.; Ren, J. H. *Neurocomputing* **2009**, *72*, 1605.

30. Ge, Z. Q.; Chen, T.; Song, Z. H. *Control Eng. Pract.* **2011**, *19*, 423.

31. Ou, X. L.; Martin, E. *Neural Comput. Appl.* **2008**, *17*, 471.

32. Yu, J. *Chem. Eng. Sci.* **2012**, *82*, 22.

33. Miyamoto, S.; Ichihashi, H.; Honda, K. Algorithms for Fuzzy Clustering, Methods in c-Means Clustering with Applications. Springer: New York, **2008**.

34. Tsai, D. M.; Lin, C. C. *Pattern Recogn.* **2011**, *44*, 1750.

35. Zhou, Z. H.; Wu, J. X.; Tang, W. *Artif. Intell.* **2002**, *137*, 239.

36. Kim, H. C.; Pang, S. N.; Je, H. M.; Kim, D. J.; Bang, S. Y. *Pattern Recogn.* **2003**, *36*, 2757.

37. Sattlecker, M.; Baker, R.; Stone, N.; Bessant, C. *Chemom. Intell. Lab. Syst.* **2011**, *107*, 363.

38. Ranganathan, A.; Yang, M. H.; Ho, J. *IEEE Trans. Image Process.* **2011**, *20*, 391.

39. Liu, Y.; Chen, J. *Ind. Eng. Chem. Res.* **2014**, *53*, 5248.