Journal of Applied Polymer Science

Industrial melt index prediction with the ensemble anti-outlier just-in-time Gaussian process regression modeling method

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: Several data-driven soft sensors have been applied for online quality prediction in polymerization processes. However, industrial data samples often follow a non-Gaussian distribution and contain some outliers. Additionally, a single model is insufficient to capture all of the characteristics in multiple grades. In this study, the support vector clustering (SVC)-based outlier detection method was first used to better handle the nonlinearity and non-Gaussianity in data samples. Then, SVC was integrated into the just-in-time Gaussian process regression (JGPR) modeling method to enhance the prediction reliability. A similar data set with fewer outliers was constructed to build a more reliable local SVC–JGPR prediction model. Moreover, an ensemble strategy was proposed to combine several local SVC–JGPR models with the prediction uncertainty. Finally, the historical data set was updated repetitively in a reasonable way. The prediction results in the industrial polymerization process show the superiority of the proposed method in terms of prediction accuracy and reliability. © 2015 Wiley Periodicals, Inc. J. Appl. Polym. Sci. **2015**, *132*, 41958.

KEYWORDS: applications; manufacturing; properties and characterization

Received 11 October 2014; accepted 5 January 2015
DOI: 10.1002/app.41958

INTRODUCTION

Nowadays, the increasing reliance on polymers with specific properties for different applications has been documented extensively. Because of the industrial need and increasing social need for the production of polymer grades in quantity and diversity, there has been a growing awareness of the modeling and control of polymer quality for increasing productivity and reducing cost. The melt index (MI) of thermoplastic polymers, including polyethylene and polypropylene, is one of the most important product qualities; it determines the product grade and affects the practical control strategy. MI is defined as the mass rate of extrusion flow through a specified capillary under prescribed conditions of temperature and pressure.[1–3] However, it is usually evaluated offline and infrequently with an analytical procedure; this leaves the process without any real-time quality indicator and usually leads to off-specification products and profit losses. Therefore, the development of an online MI prediction model is important. An alternative method is the development of a detailed mechanistic model between MI and certain easily measurable process variables.[4] However, this approach is time consuming and is often challenged by the engineering activity, mainly because of the relatively high complexity of the kinetic behavior and operation of the polymer plants. It turns out to be infeasible for agile responsive manufacturing because the products are typically short lived and of small volume.

With the rapid development of computer and communication technologies, process data have become widely available in the chemical industry. As a result, in many chemical processes, the increasing data-driven soft-sensor modeling methods have been applied to infer or predict product qualities that are difficult to measure online.[5–9] One main advantage of these soft-sensor models is that they can generally be developed quickly without a substantial understanding of the phenomenology involved. An extensive review of the prediction models for polymer properties and optimal-grade changeover control problems can be found in Ohshima and Tanigaki.[1] For MI prediction, existing common soft-sensor methods include partial least squares,[10,11] various neural networks,[12–16] support vector regression (SVR) and least squares support vector regression (LSSVR),[17–22] Gaussian process regression (GPR),[23,24] relevance vector machine,[25] and other methods.[26–28] SVR-, LSSVR-, and GPR-based soft sensors have attracted more attention recently because of their nonlinear modeling ability.[17–24] However, the selection of suitable parameters for an SVR/LSSVR model is still difficult. Compared with SVR/LSSVR, the GPR model can optimize its parameters automatically.[29–31] Additionally, GPR can simultaneously provide probabilistic information for its prediction; this is an appealing property in the process modeling area.

However, industrial data samples often contain various outliers. The data quality can significantly affect the subsequent modeling methods.[32,33] Therefore, the outliers should be removed before the model is constructed. Compared to the modeling methods development for MI prediction, this issue has been less investigated so far, although better prediction performance can be obtained by outlier detection.[7] There are several methods that have been shown to be able to detect outliers.[32,33] However, most conventional multivariate outlier detection methods assume that the data samples are distributed in an (approximate) Gaussian manner. The assumption may be invalid when the operating conditions change with time, and in fact, many data in complex industrial processes may not be distributed in a Gaussian manner. To address this problem, a support vector clustering (SVC)[34–36] outlier detection method, which can better handle the process nonlinearity and non-Gaussianity, was adopted.

Another disadvantage for most existing data-driven MI soft sensors is that a global model is applied. However, the use of only a single global model does not always give satisfactory predictions in practice, especially for those complicated regions with insufficient data samples. Recently, the just-in-time (JIT) modeling method, which is also named *lazy learning* and originates from the machine learning area, has been applied to chemical process modeling.[37–43] The JIT-based local model is built online with samples that are similar to the historical data set around a query sample when its prediction is required. Consequently, the JIT model can cope with changes in process characteristics and nonlinearity directly.[37–43] In this study, the just-in-time Gaussian process regression (JGPR) modeling method was applied to online MI prediction. Moreover, the SVC outlier detection was integrated into the JGPR modeling method to enhance the reliability of quality prediction in four folds. First, the pattern of the current query sample was recognized with the SVC-based decision boundary. Second, a relevant data set with fewer outliers was adopted to construct a more reliable JGPR prediction model. Third, an ensemble strategy was proposed to further improve the reliability of the JGPR models. Fourth, the historical data set was updated in a repetitive manner to introduce the samples with a new operating mode. These improvements made the JGPR models more reliable for prediction.

Compared with previous data-driven MI soft sensors, two main advantages of the proposed modeling method are summarized as follows. The non-Gaussian outlier detection method and JIT-based local modeling approach are integrated into a relatively unified framework. Both the historical set and the SVC-based decision boundary can be updated in a practical manner. Moreover, the ensemble strategy can improve the reliability of prediction mainly because the local candidate models can be automatically weighted in a reasonable manner.

The remainder of this article is organized as follows. The SVC outlier detection algorithm is described in the next section. The GPR, JGPR, and ensemble SVC-based JGPR soft-sensor modeling methods are proposed in the Ensemble Anti-Outlier Local Prediction Model section. Also, the detailed implementation of the enhanced soft-sensor framework with uncertain data samples is described in this section. The proposed method is evaluated by MI prediction in an industrial process in the Industrial MI Prediction section. Comparison studies with other methods are also discussed. Finally, concluding remarks are made in the Conclusions section.

KERNEL-CLUSTERING-BASED OUTLIER DETECTION

The SVC outlier detection method mainly uses a kernel-based clustering strategy. Its main idea is to map data points $\mathbf{X}=\{\mathbf{x}_k\}_{k=1}^N$ to a high dimensional feature space (H), that is, $\mathbf{x}_k \rightarrow \phi(\mathbf{x}_k)$, and then to find a hypersphere with minimal radius that contains most of the normal data points in H. This sphere, when mapped back to the data space, can be separated into several components, each enclosing a separate cluster of data points.[34] Given a set of input samples$\mathbf{X}=\{\mathbf{x}_k\}_{k=1}^N$, the mathematical formulation of the SVC method is to find the smallest enclosing spherical radius (R),[34] that is

$$\begin{cases} \min & W(R,\mathbf{c},\xi_k)=R^2+\gamma\sum_{k=1}^{N}\xi_k \\ \text{s.t.} & \begin{cases} \|\phi(\mathbf{x}_k)-\mathbf{c}\|^2 \le R^2+\xi_k, \\ \xi_k \ge 0,\ k=1,\cdots,N. \end{cases} \end{cases} \tag{1}$$

where $\mathbf{c}$ is the center of the enclosing hypersphere $\Theta(\mathbf{c}, R)$, respectively, and $\gamma > 0$ is a regularization constant that determines the penalty on the slack variables $\xi_k \ge 0$ and denotes the trade-off between the volume of the sphere and the number of target objects rejected.

The dual optimization problem can be obtained as follows:[34]

$$\begin{cases} \max\ Q(\beta_k)=\sum_{k=1}^{N}\beta_k\langle\phi(\mathbf{x}_k),\phi(\mathbf{x}_k)\rangle-\sum_{j=1}^{N}\sum_{k=1}^{N}\beta_j\beta_k\langle\phi(\mathbf{x}_j),\phi(\mathbf{x}_k)\rangle \\ \qquad\quad =\sum_{k=1}^{N}\beta_k K(\mathbf{x}_k,\mathbf{x}_k)-\sum_{j=1}^{N}\sum_{k=1}^{N}\beta_j\beta_k K(\mathbf{x}_j,\mathbf{x}_k) \\ \text{s.t.}\ \sum_{k=1}^{N}\beta_k=1 \\ 0 \le \beta_k \le C,\ k=1,\cdots,N \end{cases} \tag{2}$$

where β_k is the Lagrange multiplier. The mapping of $\phi(\mathbf{x}_k)$ does not need to be known because it is implicitly defined by the choice of kernel function $K(\mathbf{x}_j,\mathbf{x}_k)=\langle\phi(\mathbf{x}_j),\phi(\mathbf{x}_k)\rangle$. That is, in the SVC method, data points are mapped from the data space to a high dimensional feature space with a kernel function.[34] Here, the common Gaussian kernel function is adopted and defined as follows:[34–36]

$$\begin{aligned} K(\mathbf{x}_i,\mathbf{x}_j) &= \langle\phi(\mathbf{x}_i),\phi(\mathbf{x}_j)\rangle \\ &= e^{-\|\mathbf{x}_i-\mathbf{x}_j\|^2/\sigma^2} \end{aligned} \tag{3}$$

where $\sigma > 0$ is the kernel width. It should be noted that there are only two user-specified parameters to implement SVC, that is, the Gaussian kernel width (σ) and the regularization constant (γ). The shape of the contour is governed by these two parameters.[34–36] The values of σ and γ can be selected simply

with the common cross-validation procedure in the kernel learning area.[34–36]

After the optimization problem is solved, R can be formulated by eq. (4), and the distance between a query sample $D(\mathbf{x}_q)$ and the center of Θ in $\mathbf{x}_q$ can be computed by eq. (5):[34]

$$R=\frac{1}{N}\sum_{k=1}^{N}\sqrt{\|\phi(\mathbf{x}_k)-\mathbf{c}\|^2}$$
$$=\frac{1}{N}\sum_{k=1}^{N}\sqrt{K(\mathbf{x}_k,\mathbf{x}_k)-2\sum_{i=1}^{N}K(\mathbf{x}_i,\mathbf{x}_k)\beta_i+\sum_{i=1}^{N}\sum_{j=1}^{N}\beta_i\beta_jK(\mathbf{x}_i,\mathbf{x}_j)} \quad (4)$$

$$D(\mathbf{x}_q)=\sqrt{\|\phi(\mathbf{x}_q)-\mathbf{c}\|^2}$$
$$=\sqrt{K(\mathbf{x}_q,\mathbf{x}_q)-2\sum_{i=1}^{N}K(\mathbf{x}_q,\mathbf{x}_i)\beta_i+\sum_{i=1}^{N}\sum_{j=1}^{N}\beta_i\beta_jK(\mathbf{x}_i,\mathbf{x}_j)} \quad (5)$$

Therefore, the criterion for judging $\mathbf{x}_q$ that lies in or outside Θ in H is as follows:[34]

$$\begin{cases} D(\mathbf{x}_q) > R, & \textit{outliers} \\ D(\mathbf{x}_q) \le R, & \textit{normal data} \end{cases} \quad (6)$$

That is, a sample is assigned as the normal sample if the distance between it and the sphere center is not larger than R. Otherwise, it is classified as an outlier. On the basis of this simple criterion, the outliers in the modeling set can be recognized and detected. The SVC algorithm has some advantages over other clustering algorithms for its ability to generate cluster boundaries of arbitrary shape and to deal with outliers with a soft margin constant that allows the sphere in the feature space not to enclose all points.[34] Additionally, unlike the Gaussian distribution assumption adopted by most traditional multivariate preprocessing methods for outlier detection, the SVC approach does not assume that data are distributed or approximately distributed in a Gaussian manner. Therefore, SVC is more practical for complicated industrial processes because many process variables are not distributed in a Gaussian manner and often contain various outliers.

ENSEMBLE ANTI-OUTLIER LOCAL PREDICTION MODEL

GPR Soft-Sensor Method

The soft-sensor model development based on the GPR framework can be described as a problem whose aim is to learn a model f that approximates a training set $\{\mathbf{X}\}=\{\mathbf{x}_i\}_{i=1}^{N}$ and $\{\mathbf{Y}\}=\{y_i\}_{i=1}^{N}$ are the input and output datasets with N samples, respectively. A GPR model provides a prediction of the output variable for an input sample through Bayesian inference. For an output variable of $\mathbf{Y}=(y_1,\ldots,y_N)^{\mathrm{T}}$, the GPR model is the regression function with a Gaussian prior distribution and zero mean, or in a discrete form[29]

$$\mathbf{Y}=(y_1,\ldots,y_N)^{\mathrm{T}}\sim G(0,\mathbf{C}) \quad (7)$$

where $\mathbf{C}$ is the $N\times N$ covariance matrix with the ijth element defined by the covariance function, $\mathbf{C}_{ij}=C(\mathbf{x}_i,\mathbf{x}_j)$. A common covariance function can be defined as follows:[29]

$$C(\mathbf{x}_i,\mathbf{x}_j)=a_0+a_1\sum_{d=1}^{D}x_{id}x_{jd}+v_0\exp\left(-\sum_{d=1}^{D}w_d(x_{id}-x_{jd})^2\right)+\delta_{ij}b \quad (8)$$

where x_{id} is the dth component of the vector $\mathbf{x}_i$.$\delta_{ij}=1$ if $i=j$; otherwise, it is equal to zero. $\boldsymbol{\theta}=[a_0,a_1,v_0,w_1,\cdots,w_D,b]^{\mathrm{T}}$ is the hyperparameter vector defining the covariance function. As depicted in eq. (8), the first two terms denote a constant bias and a linear correlation term, respectively. The exponential term takes into account the potentially strong correlation between the outputs for nearby inputs. Additionally, the term b captures the random error effect. By combining both linear and nonlinear terms in the covariance function, the GPR model is capable of handling both linear and nonlinear processes.[29]

With the adoption of a Bayesian approach, the hyperparameter $\boldsymbol{\theta}$ can be estimated by maximization of the following log–likelihood function:[29]

$$L(\boldsymbol{\theta})=-\frac{1}{2}\log(\det(\mathbf{C}))-\frac{1}{2}\mathbf{Y}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{Y}-\frac{N}{2}\log(2\pi) \quad (9)$$

This optimization problem can be solved with the derivative of the log–likelihood with respect to each hyperparameter. Detailed implementations for training a GPR model can be found in the work of Rasmussen and Williams.[29] It should also be noted that the main computational load for training a GPR model is about $O(N^3)$; this is feasible for moderately sized training data sets (less than several thousands) on a conventional computer. For much larger data sets, sparse training strategies may be required to reduce the computational burden.[44,45]

Finally, the GPR model can be obtained once $\boldsymbol{\theta}$ is solved. For a test sample $\mathbf{x}_q$, the predicted output of y_q is also Gaussian with the mean ($\hat{y}_q$) and variance ($\sigma^2_{\hat{y}_q}$) calculated as follows:[29]

$$\hat{y}_q=\mathbf{k}_q^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{Y} \quad (10)$$

$$\sigma^2_{\hat{y}_q}=r_q-\mathbf{k}_q^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{k}_q \quad (11)$$

where $\mathbf{k}_q=[C(\mathbf{x}_q,\mathbf{x}_1),C(\mathbf{x}_q,\mathbf{x}_2),\cdots,C(\mathbf{x}_q,\mathbf{x}_N)]^{\mathrm{T}}$ is the covariance vector between the new input and the training samples, and $r_q=C(\mathbf{x}_q,\mathbf{x}_q)$ is the covariance of the new input. In summary, the vector $\mathbf{k}_q^{\mathrm{T}}\mathbf{C}^{-1}$ denotes a smoothing term that weights the training outputs to make a prediction for the new input sample $\mathbf{x}_q$.[29] In addition, eq. (11) provides a confidence level for the model prediction; this is an appealing property of the GPR method.

The quality of modeling data samples is very important for the construction of GPR-based soft sensors. As shown in eqs. (10) and (11), $\mathbf{C}$ and $\mathbf{k}_q$ are mainly effected by the modeling samples. The prediction would be distorted if outlier samples were preserved in the model. The more outlier samples there are in the model, the worse the prediction is. Therefore, it is essential to construct a GPR model with healthy data samples.

JGPR-Based Local Model

To solve a problem, one should not try to solve a more general problem as an intermediate step.[46] The estimation of the complete density in place of the computation of the boundary around a data set might require too much data and could result

in bad descriptions. For complicated multigrade polymerization processes, the direct application of a global or fixed model may not be an easy task, mainly because the specification of the structure of a global/fixed model is often difficult. Another limitation is that it is difficult to quickly update a global/fixed model when the process dynamics are moved away from the nominal operating area. Additionally, the training samples are not always enough to describe all of the process characteristics in a global/fixed GPR model.

To overcome these problems and construct the local models automatically, the JIT method has been developed as an attractive alternative to nonlinear chemical process modeling and control.[37–43] Several JIT-based nonlinear soft sensors have been proposed previously.[37–43] However, they are all deterministic models [e.g., just-in-time-based support vector regression (JSVR) and just-in-time least squares support vector regression (JLSSVR) models[41] rather than probabilistic ones. In the literature, few probabilistic soft sensors have been applied to multigrade processes with transitional modes. In this section, a JGPR-based online modeling method is proposed for better description of those complex regions. Generally, for $\mathbf{x}_q$, there are three main steps in the construction of a JGPR model online:

1. Determine the relevant samples to form a similar set ($\mathbf{S}_{\text{sim}}$) in the database $\mathbf{S}$ with some defined similarity criteria.
2. Construct a JGPR model $f_{\text{JGPR}}(\mathbf{x}_q)$ online with the relevant set $\mathbf{S}_{\text{sim}}$ and the aforementioned formulations, that is, eqs. (7–9).
3. Obtain the predicted output of y_q, that is, $\hat{y}_q$ and $\sigma^2_{\hat{y}_q}$, for the current $\mathbf{x}_q$, and then discard the JGPR model $f_{\text{JGPR}}(\mathbf{x}_q)$.

This is the basic framework of the JGPR online modeling approach. For a new query sample, a new JGPR model can be built with the same three-step procedure. Generally, the Euclidean distance-based similarity is the most commonly used index.[37,38] The similarity factor (s_{qi}) between the $\mathbf{x}_q$ and the sample $\mathbf{x}_i$ in the dataset is defined as follows:[37,38]

$$s_{qi}=\exp(-d_{qi})=\exp(-\|\mathbf{x}_q-\mathbf{x}_i\|),\ i=1,\cdots,N \tag{12}$$

where d_{qi} is the distance similarity between $\mathbf{x}_q$ and $\mathbf{x}_i$ in the dataset. The value of s_{qi} is bounded between 0 and 1, and when s_{qi} approaches 1, $\mathbf{x}_q$ closely resembles $\mathbf{x}_i$. Although a little better prediction performance can be obtained with the distance-and-angle-based similarity factor than with only the utilization of the Euclidean distance, another parameter for the balance of the distance and angle should be chosen.[39] Additionally, some correlation-based similarity criteria have recently been proposed.[40] All of these similarity criteria can be adopted to construct a JGPR model. However, that was not the main scope of this study, so it was not investigated in detail here.

With the similarity criterion in eq. (12), n similar samples should be selected to construct a JGPR model. However, it is difficult to determine how many similar samples are needed beforehand. In particular, for an industrial process with multiple grades, the relevant sets of query samples for different grades are different. Generally, $n_{\max}$ similar samples can be ranked according to the degree of the similarity. A cumulative similarity factor (CSF or S_{qn}) can be adopted as follows:[41]

$$S_{qn}=\frac{\sum_{i=1}^{n} s_{qi}}{\sum_{i=1}^{n_{\max}} s_{qi}},\ n\le n_{\max} \tag{13}$$

which denotes the cumulative similarity of the n most similar samples compared to the relevant set $\mathbf{S}_{\text{sim}}$. The CSF index can simply compute the cumulative similarity, and then, it can determine the n most similar samples.[41] For example, the choice of S_{qn} = 0.8 means 80% of the most similar samples have been selected. Consequently, the construction of the relevant set $\mathbf{S}_{\text{sim}}$ with the CSF index is efficient.

Compared with JLSSVR/JSVR online modeling methods, two advantages of JGPR-based soft sensors can be obtained. One is that the probabilistic information can be provided for its prediction. The other, the modeling procedures of JGPR through Bayesian inference, is simpler and more straightforward. For JLSSVR/JSVR modeling methods, the kernel function and the related parameters should be carefully chosen. This is not an easy task, especially for complicated industrial processes.

Enhanced SVC–JGPR Soft Sensors

On the basis of the aforementioned outlier detection and modeling methods, the main implementations for the complicated polymerization process can be formulated. The method is simply denoted as SVC–JGPR mainly because it integrates the SVC-based outlier detection and JGPR-based online soft-sensor modeling approach into a unified framework. The step-by-step procedures of the integrated SVC–JGPR modeling method are summarized as follows:

1. Collect the process measured input and output samples of the polymerization process. The initial data set can be noted as $\mathbf{S}=\{\mathbf{X},\mathbf{Y}\}=\{\mathbf{x}_i,y_i\}_{i=1}^{N}$.
2. Apply the SVC method to data points $\mathbf{X}=\{\mathbf{x}_i\}_{i=1}^{N}$, and then, the enclosing hypersphere $\Theta(\mathbf{c},\ R)$ can be obtained. The data samples in $\mathbf{S}$ are relatively enough, and thus, the parameters of SVC can be simply selected with the cross-validation procedure. In this step, eqs. (1–4) are applied to solve the SVC problem.
3. The outliers in $\mathbf{X}=\{\mathbf{x}_i\}_{i=1}^{N}$ can be temporarily moved in the feature space with eqs. (5) and (6) to obtain a tight boundary. Correspondingly, for a refined global modeling data set $\mathbf{S}_0=\{\mathbf{X}_0,\mathbf{Y}_0\}=\{\mathbf{x}_i,y_i\}_{i=1}^{N_0}>$, the SVC and GPR can be combined into an SVC–GPR global soft sensor.
4. For $\mathbf{x}_q$, first judge if it is a normal sample with eqs. (5) and (6). If $D(\mathbf{x}_q) > R$, $\mathbf{x}_q$ is considered a global outlier for the data set $\mathbf{X}_0$. This is unsuitable for prediction in such a situation. Go to step 7. Otherwise, if $D(\mathbf{x}_q) \le R$, $\mathbf{x}_q$ is considered a global normal sample for the data set $\mathbf{X}_0$. Go to step 5.
5. For the online prediction for $\mathbf{x}_q$, first determine the relevant set $\mathbf{S}_q=\{\mathbf{X}_q,\mathbf{Y}_q\}=\{\mathbf{x}_i,y_i\}_{i=1}^{n}$ with eqs. (12) and (13). Apply the SVC method to these data points $\mathbf{X}_q$ to find a local hypersphere $\Theta_q(\mathbf{c}_q,\ R_q)$. The modeling samples are relatively few, and thus, the suitable parameters of SVC are difficult to

choose. With a candidate SVC model, some outliers in $\mathbf{X}_q$ can be temporarily detected in the feature space with this local hypersphere. These removed samples can be treated as the dissimilar pattern for $\mathbf{x}_q$. Finally, with M pairs of candidate parameters $[\gamma_m, \sigma_m]$, $m = 1,\ldots,M$,[34–36] M refined local modeling data sets denoted as $[\gamma_m, \sigma_m]$, $m=1,\cdots,M$, $m = 1,\ldots,M$ for $\mathbf{x}_q$ can be obtained.

6. Altogether, M JGPR-based local models denoted as $f_{\mathrm{JGPR}}(\mathbf{x}_q, \mathbf{S}_{q,m})$, $m=1,\cdots,M$ are built online for prediction with, $\mathbf{S}_{q,m}=\{\mathbf{x}_i, y_i\}_{i=1}^{n_m}$, $m=1,\cdots,M$. Correspondingly, several predicted values ($\hat{y}_{q,m}$, $m=1,\cdots,M$) and their predicted variances ($\sigma^2_{\hat{y}_{q,m}}$, $m=1,\cdots,M$) can be obtained, respectively. Before the actual quality value is obtained, the available information useful for prediction is the predicted variance. Generally, a smaller variance means that the prediction is more reliable. As a simple method, the prediction with the smallest variance can be used as the final prediction:

$$\hat{y}_q = \underset{f_{\mathrm{JGPR}}(\mathbf{x}_q,\mathbf{S}_{q,m})}{\arg} \min_{m=1,\cdots,M} \sigma^2_{\hat{y}_{q,m}} \tag{14}$$

In this study, as an alternative, ensemble learning was further adopted to combine several SVC–JGPR models. For regression, the main idea of ensemble learning typically generates several candidate models, which are combined to make a prediction. Compared to a single regression model, it has been demonstrated that the combined ensemble model can improve the prediction performance and prevent the overfitting problem.[23,47] The proposed ensemble strategy is similar to the weighted rule in the literature.[23] However, the previous modeling approach was global, and the outlier detection was not considered in their study. Here, ensemble learning was investigated in the local learning manner with relatively fewer samples. An ensemble SVC–JGPR [ensemble support vector clustering (ESVC)–JGPR] model was proposed. Generally, if a candidate model has a larger predicted variance, the weight on it should be smaller. Correspondingly, a candidate model should have a larger weight if its predicted variance is smaller. On the basis of the analysis, the final ensemble prediction and the related weights for candidate models could be formulated as follows, respectively:

$$\hat{y}_q = \sum_{m=1}^{M} w_{q,m}\hat{y}_{q,m},\ m = 1,\cdots,M \tag{15}$$

$$w_{q,m} = \frac{\frac{1}{\sigma^2_{\hat{y}_{q,m}}}}{\sum_{m=1}^{M}\frac{1}{\sigma^2_{\hat{y}_{q,m}}}} = \frac{1}{\sigma^2_{\hat{y}_{q,m}}\sum_{m=1}^{M}\frac{1}{E^{\mathrm{FLOO}}_{t,m}}},\ m = 1,\cdots,M \tag{16}$$

7. Once the assay y_q is obtained, $\mathbf{x}_q$ can be reintroduced into $\mathbf{X}$ to judge if it is normal. That is, the initial global data set is updated with $\mathbf{X} = \{\mathbf{X}, \mathbf{x}_q\}$. Before the online prediction of a new query sample, go to steps 2 and 3 to update the refined global modeling data set $\mathbf{S}_0=\{\mathbf{X}_0, \mathbf{Y}_0\}=\{\mathbf{x}_i, y_i\}_{i=1}^{N_0}$. This step can be implemented offline.

It should be noted that the computational load of SVC (including the selection of the model parameters) for larger data sets is not trivial. Fortunately, in this study, the global SVC outlier detection model was solved in an offline manner. Additionally, for $\mathbf{x}_q$, the local SVC model was always solved feasibly with a small sample size much less than several hundreds. When the SVC method is applied to complicated industrial processes, a tight decision boundary for outlier detection may delete a few normal samples, especially when the training samples are not sufficient. For example, for a query sample of the new operating mode, it will likely be treated as an outlier sample. To overcome this problem, as shown in step 7 of SVC–JGPR, the samples in $\mathbf{X}$ should be explored in a repetitive manner to use the useful information as much as possible. Consequently, the samples temporarily treated as outliers in step 3 of SVC–JGPR can go back to the normal pattern when the samples of the new operating mode increase to an extent.

The main advantages of the SVC–JGPR-based soft-sensor method can be summarized as follows. First, the distribution of the process data has no Gaussian assumption; this indicates that the non-Gaussian data variables widely encountered in industrial processes can be handled. Second, mainly because of the nonlinear data description ability of the SVC method, the outlier detection for nonlinear process variables can be addressed with this framework. Through the search of a hypersphere in the feature space, a tight boundary of the data distribution can be formulated for outlier detection. Third, with more reliable data samples for the construction of a local JGPR soft sensor, a better quality prediction can be obtained. Fourth, the SVC is extended to maintain the historical data set. The refined modeling set and its global SVC decision boundary can be updated gradually by the introduction of new data samples.

The main properties of the ESVC–JGPR, SVC–JGPR, SVC–GPR, JGPR, and GPR soft-sensor modeling methods are listed in Table I. The main property of SVC–JGPR is that it can integrate outlier detection and JIT-base local modeling into a relatively unified framework. Moreover, as mentioned previously, both the historical set and the SVC-based decision boundary can be updated in a practical and reasonable manner. Furthermore, compared with SVC–JGPR, the ESVC–JGPR method can achieve more reliable prediction performance mainly because the local models can automatically assign smaller weights than models that are certain about their predictions. These properties make the proposed approach different from the global GPR method.

INDUSTRIAL MI PREDICTION

Data Description

In this section, the ESVC–JGPR and SVC–JGPR soft-sensor modeling methods were explored by online MI prediction in an industrial polyethylene production process in China. All of the data samples were collected from daily process records and corresponding laboratory analysis. The product quality in steady-state grades was commonly analyzed in the laboratory every 4 h for 1 day. Without online analyzers for MI, off-grade products and materials would inevitably be produced in industrial polyethylene production processes. The process input variables correlated with the MI were selected. These input measurable variables included the reactor temperature, pressure, liquid level, and flow rate of the main catalyst.

Table I. Main Property Comparisons for Several Data-Driven Soft-Sensor Models: ESVC–JGPR, SVC–JGPR, SVC–GPR, JGPR, and GPR

Soft-sensor model	Property	Outlier detection (description)	Combination of local models
ESVC–JGPR	Local and unfixed	Yes (for non-Gaussian and nonlinear multivariate data)	Ensemble with the inverse of model uncertainty
SVC–JGPR	Local and unfixed	Yes (for non-Gaussian and nonlinear multivariate data)	Minimization of the local model with predicted variance
SVC–GPR	Global and fixed	Yes (for non-Gaussian and nonlinear multivariate data)	No
JGPR	Local and unfixed	No	No
GPR[29]	Global and fixed	No	No

A set of about 360 samples was investigated in this study. The first 200 samples were treated as the initial historical samples. The remaining set of about 160 samples was used for testing. However, because of commercial secrecy, the details of the industrial data set cannot be provided as supplemental material. From previous research, an alternative data set of 331 samples for an industrial polyethylene production process in a large plant in Taiwan can be obtained via the Internet at http://wave-net.cycu.edu.tw/~cpse/web/Resources.htm.[43] This can be used as a public set for academic research.

To explore the data distribution of the modeling data samples, the normality of four process input variables was evaluated, as shown in Figure 1(a–d), respectively. These four process variables violated the Gaussian distribution, which is represented by the dashed line in Figure 1(a–d). Additionally, as an illustrated in Figure 2, the first to third process input variables were nonlinearly correlated. Moreover, the data samples were distributed irregularly. Some operating areas had more data samples, whereas some areas showed relatively sparse data distribution. Therefore, the use of only a single global model may not be enough to describe all of the process characteristics.

Modeling Software and Performance Assessment

The modeling software was MatLab (Matrix Laboratory). The simulation environment in this case was MatLab V2009b with a CPU main frequency of 2.3 GHz and 4 GB of memory. The basic GPR MatLab codes originally demonstrating the main algorithms from Rasmussen and Williams[29] can be downloaded via the Internet at http://www.gaussianprocess.org/gpml/code/matlab/doc/. Further information for GPR and other related kernel learning algorithms can also be found via the Internet at http://www.gaussianprocess.org/ and http://www.support-vector.net/software.html. All of the algorithms of the ESVC–JGPR, SVC–JGPR, and JGPR soft-sensor modeling methods were written with MatLab codes to generate the results in this study.

Two common performance indices, including root mean square error (RMSE) and relative root mean square error (RE) were adopted to quantitatively evaluate the prediction performance of different soft-sensor models. The RMSE and RE indices are defined, respectively, as follows:

$$\mathrm{RMSE}=\sqrt{\sum_{q=1}^{N_{\mathrm{tst}}}\left(\frac{y_q-\hat{y}_q}{N_{\mathrm{tst}}}\right)^2} \tag{17}$$

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

where $\hat{y}_q$ and y_q are the predicted and actual values, respectively, and N_{tst} is the number of test samples.

RESULTS AND DISCUSSION

Reliable sensor measurements and data collection play crucial roles in process modeling.[7] However, the modeling data set contains some obvious outliers and inconspicuous outliers as they are masked by their adjacent outliers. Several traditional multivariate outlier detection methods are based on Mahalanobis distance (MD). Additionally, there are also some methods that are based on robust distance and other measures.[32,33] Despite differences between these outlier detection methods, most of them are based on the assumption that data samples are distributed or approximately distributed in a Gaussian manner. To show the superiority of SVC, the MD-based outlier detection method was investigated here for comparison. After MD-based preprocessing, the MD and GPR could be simply combined together to form an MD–GPR soft sensor. Finally, six soft-sensor modeling methods altogether, denoted as ESVC–JGPR, SVC–JGPR, JGPR, SVC–GPR, MD–GPR, and GPR for short, were investigated.

Details about the online MI prediction comparisons among the ESVC–JGPR, SVC–JGPR, JGPR, SVC–GPR, MD–GPR, and GPR methods are tabulated in Table II. About 25 samples among all of the 160 test samples were treated as global outlier samples with SVC (step 4 of SVC–JGPR, as discussed earlier) before we made an online prediction. For a justified comparison, they are not included in the results in Table II. The results in Table II clearly show that the ESVC–JGPR method achieved the best prediction performance with the smallest RMSE and RE values.

The corresponding box plots of the relative prediction errors (i.e., $\frac{y_q-\hat{y}_q}{y_q}$) using all five different methods are shown in Figure 3. On each box (e.g., ESVC–JGPR), 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

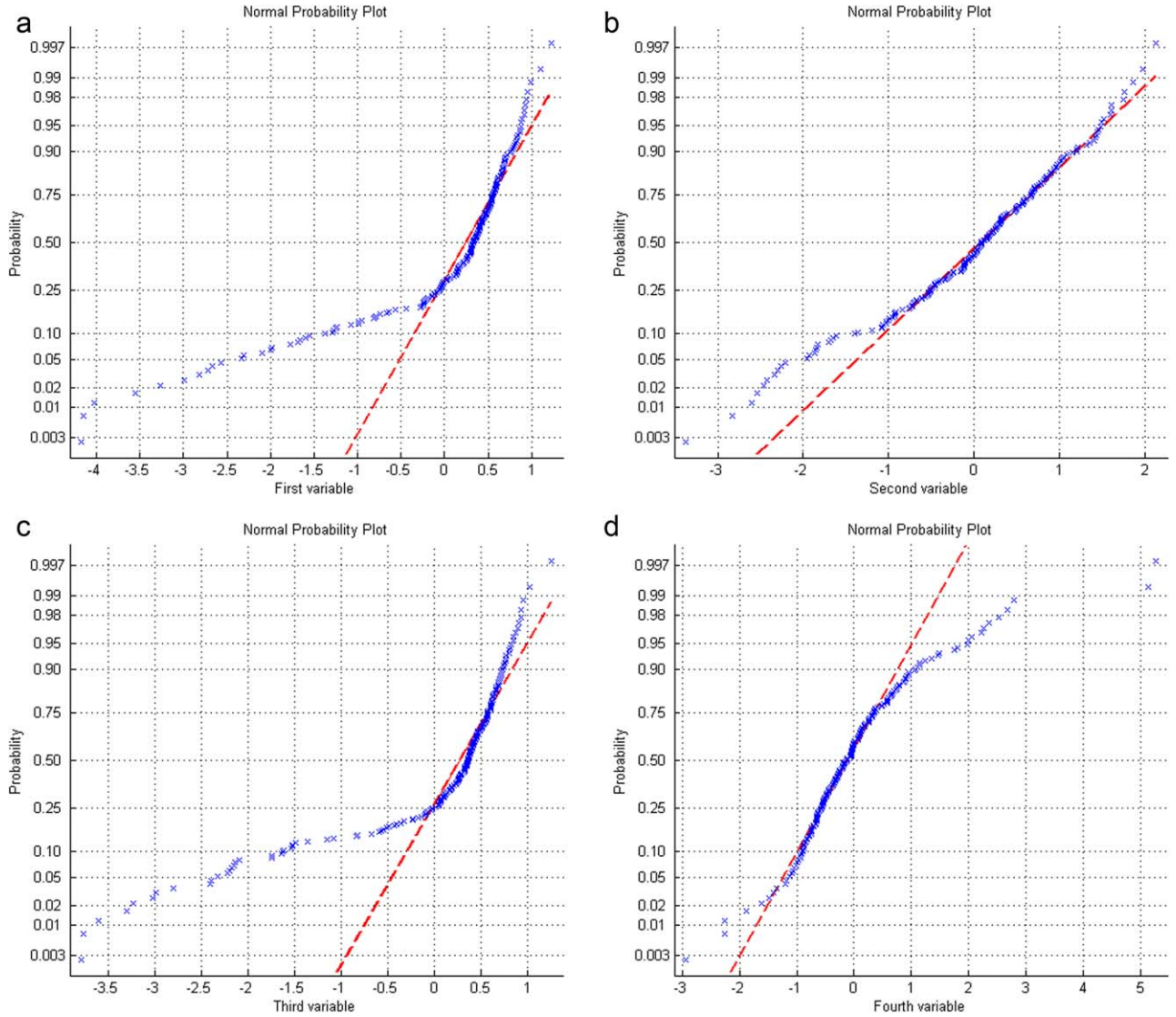

Figure 1. Normality tests of the (a) first, (b) second variable in the industrial polyethylene production process (training set), (c) third, and (d) fourth variables in the industrial polyethylene production process (training set). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

maximum. Several outliers are plotted individually. We found that the proposed ESVC–JGPR method had the narrowest ranges of the relative prediction errors; this implied the best prediction performance. Additionally, the results of the RMSE and RE values in Table II and Figure 3 show that ESVC–JGPR obtained a better distribution of the relative prediction errors than the other approaches because the later ones showed unreliable prediction results for some test samples.

Among all six methods, SVC–GPR, MD–GPR, and GPR were global. It was unsurprising that the single GPR model had the worst prediction performance. The SVC and MD outlier detection methods deleted some outliers and then improved the sequential soft-sensor model. As shown in Table II and Figure 3, SVC–GPR was superior to MD–GPR, with a better distribution of the relative prediction errors. This result indicates that for non-Gaussian-distributed data samples, SVC performed better than traditional MD-based outlier detection methods.

ESVC–JGPR, SVC–JGPR, and JGPR are three methods that were different from the other three aforementioned global methods. They discussed in particular in this section. Exhibiting a better local prediction performance, JGPR had smaller RMSE and RE values than GPR. For online MI prediction, the overall prediction results of the ESVC–JGPR, SVC–JGPR, and JGPR methods are shown in Figure 4. As analyzed previously, the data samples showed uneven, non-Gaussian, and nonlinear relationships (as also shown in Figure 2). For online construction of a local model, similar samples were carefully chosen. With some outliers or some samples with different modes, the prediction performance of a local model might be degraded. Furthermore, the ensemble model showed a more reliable prediction performance than its single local model. Consequently, all of the results shown in Figure 4 and Table II validate that ESVC–JGPR was generally superior to SVC–JGPR and JGPR.

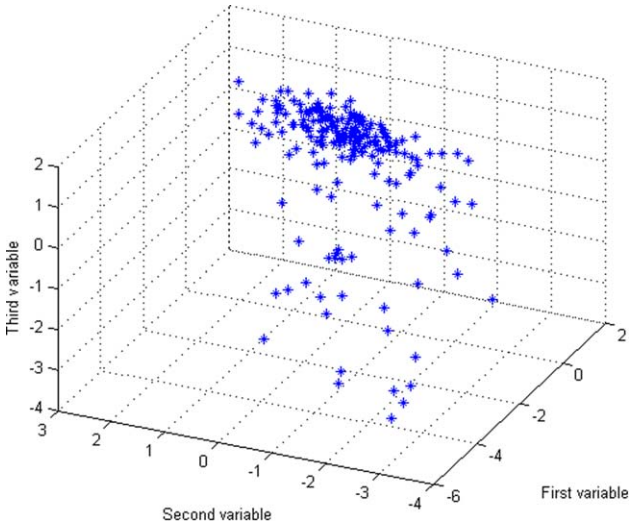

Figure 2. Process input variable relationships in the industrial polyethylene production process (training set). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

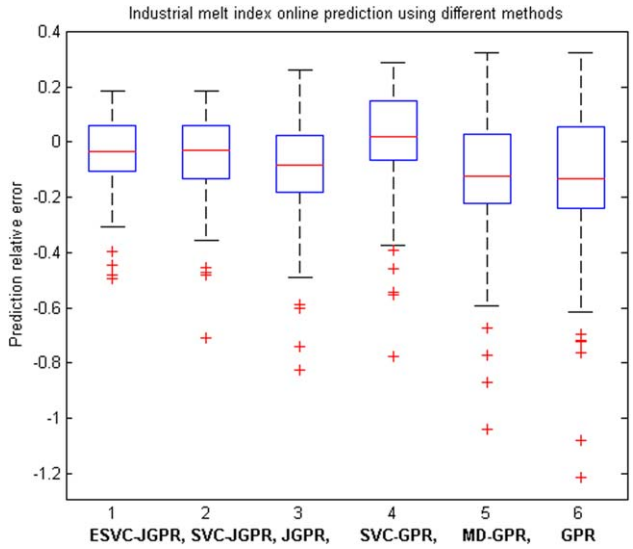

Figure 3. Comparisons between the relative prediction error distributions of the MI in the industrial polyethylene production process via a box plot with the ESVC–JGPR, SVC–JGPR, JGPR, SVC–GPR, MD–GPR, and GPR soft-sensor models. The edges of each box are the first and third quartiles, and the band inside each box shows the median. The whiskers above and below each box show the locations of the minimum and maximum, respectively. Several outliers are plotted individually (test set). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

In contrast to traditional neural networks and SVR-based deterministic modeling methods,[12–22] the GPR-based methods provided probabilistic information for prediction. Comparisons of the predicted variance values for the test samples ($\sigma_{\hat{y}_q}, q=1,\cdots,N_{\text{tst}}$) with two online local modeling methods, ESVC–JGPR and SVC–JGPR, are shown in Figure 5. The ESVC–JGPR method exhibited a smaller prediction uncertainty than the SVC–JGPR method, especially for the test samples after number 70. The $\frac{1}{N_{\text{tst}}}\sum_{q=1}^{N_{\text{tst}}}\sigma_{\hat{y}_q}$ values of ESVC–JGPR and SVC–JGPR were 1.54 and 1.80, respectively. A smaller value of $\frac{1}{N_{\text{tst}}}\sum_{q=1}^{N_{\text{tst}}}\sigma_{\hat{y}_q}$ indicated a smaller uncertainty of the prediction. This means that ESVC–JGPR not only achieved a more accurate prediction performance (smaller RMSE and RE values) but also obtained more reliable prediction. Consequently, the probabilistic information could help operators/engineers use the prediction information in a better way.

It should be noted that with the proposed strategy, the historical data set was updated by the introduction of new samples in a repetitive manner. As a result, some samples previously mistaken as global outliers could be recognized as global normal samples with new operating modes. Nevertheless, the JGPR model could not distinguish samples with new operating modes from outliers. Additionally, the historical data set of JGPR was not refined in a suitable way. For a query sample, the total computational time of ESVC–JGPR for online modeling and prediction was less than 20 s; this was larger than that with the use of only JGPR. However, the online modeling and prediction time was much lower than the 4-h assaying time in the

Table II. Comparisons of the Online Prediction Errors with Six Different Data-Driven Soft-Sensor Models

Soft-sensor model	RMSE	RE (%)
ESVC-JGPR	**4.82**	**14.0**
SVC-JGPR	5.21	15.7
JGPR	6.87	20.6
SVC-GPR	6.90	18.8
MD-GPR	8.48	25.6
GPR[29]	9.66	28.9

The best results are set in boldface type.

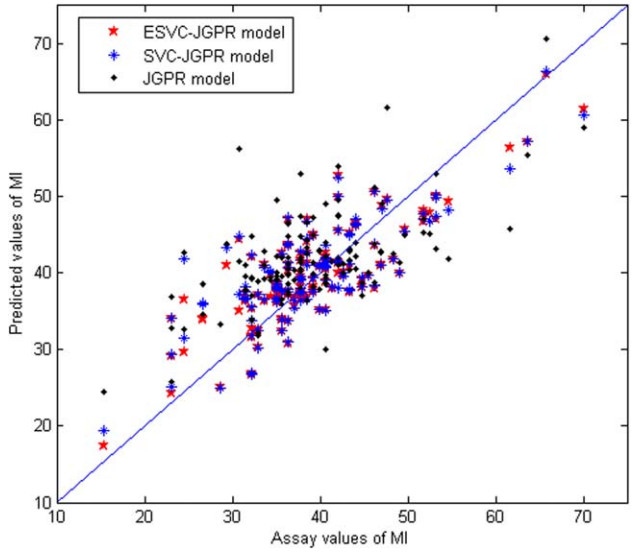

Figure 4. Parity plot based on assay values against the predicted values with three online local modeling methods: ESVC–JGPR, SVC–JGPR, and JGPR (test set). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

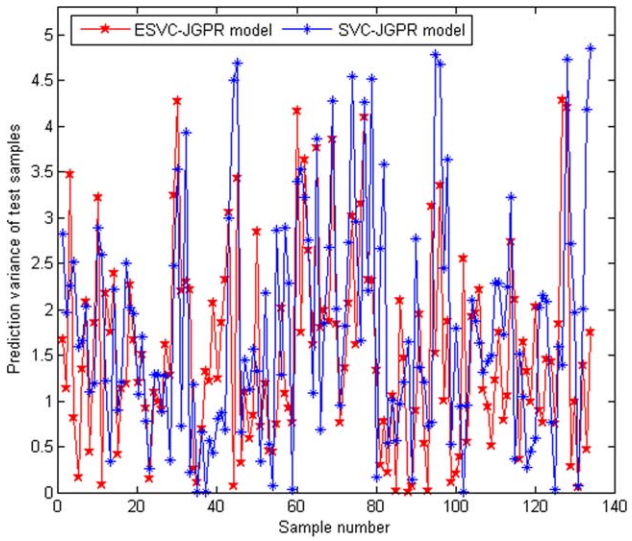

Figure 5. Comparisons of the predicted variances ($\sigma_{\hat{y}_q}, q=1,\cdots,N_{\text{tst}}$) with two online local modeling methods (ESVC–JGPR and SVC–JGPR). The $\frac{1}{N_{\text{tst}}}\sum_{q=1}^{N_{\text{tst}}}\sigma_{\hat{y}_q}$ values of ESVC–JGPR and SVC–JGPR were 1.54 and 1.80, respectively (test set). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com].

laboratory for the offline analysis of the MI. Moreover, from the viewpoint of industrial polymerization processes, an accurate online MI prediction is much more important than the computational loads.

Therefore, from all of the obtained results and comparison analyses, the proposed ESVC–JGPR soft sensor performed better than the other methods in terms of online MI prediction. Note that the incremental methods[48] could be adopted to update the offline SVC global outlier detection model and to then reduce the overall computational burden, especially for a relatively large data set. Additionally, if available, some domain knowledge and expert rules could also be combined into the soft-sensor modeling methods. For example, the number of JGPR-based local models could be reduced when the method is integrated with the operation mode detection in the prediction task.[22] The transitional modes are much more difficult to model, and they were neglected in this study. So, there are still several interesting research directions worth investigating in the future to further enhance the accuracy and transparency of industrial MI soft-sensor models.

CONCLUSIONS

The development of reliable soft sensors for complex industrial polymerization processes is still difficult. In this study, we explored reliable local soft-sensor models for online quality prediction when the modeling data samples were not distributed in a Gaussian manner and contained outliers. Compared to several traditional outlier detection approaches, SVC does not assume that data samples are distributed in a Gaussian manner. The process nonlinearity and non-Gaussianity can be better handled with a tight boundary. Additionally, for local modeling, a relevant data set with fewer outliers can be adopted to build a more reliable local prediction model. ESVC–JGPR and SVC–JGPR were proposed as two local soft sensors. Their advantages were demonstrated and compared with other soft sensors in terms of online industrial MI prediction. ESVC–JGPR is recommended for the ease of determining the model parameters and its better reliability in prediction. It should be noted that the soft-sensor modeling method can also be applied to predict other difficult to measure parameters, such as the polymer quality, polymer melt index, and mixture of initiators. These polymerization processes include the polymerization of methyl methacrylate, the polymerization of nylon 6,6, and rubber-mixing processes.[14,49]

ACKNOWLEDGMENTS

The authors gratefully acknowledge the National Natural Science Foundation of China (contract grant number 61004136), Zhejiang Provincial Natural Science Foundation of China (contract grant number LY12E05023), and Science and Technology Key Program of Quanzhou (contract grant number 2011G8) for their financial support of this project.

REFERENCES

1. Ohshima, M.; Tanigaki, M. J. *Process Control* **2000**, *10*, 135.
2. Richards, J. R.; Congalidis, J. P. *Comput. Chem. Eng.* **2006**, *30*, 1447.
3. Kiparissides, C. J. *Process Control* **2006**, *16*, 205.
4. McAuley, K. B.; MacGregor, J. F. *AIChE J.* **1991**, *37*, 825.
5. Fortuna, L.; Graziani, S.; Rizzo, A.; Xibilia, M. G. Soft Sensors for Monitoring and Control of Industrial Processes; Springer-Verlag: New York, **2007**.
6. Yao, Y.; Gao, F. R. *Annu. Rev. Control* **2009**, *33*, 172.
7. Kadlec, P.; Gabrys, B.; Strandt, S. *Comput. Chem. Eng.* **2009**, *33*, 795.
8. Kano, M.; Ogawa, M. J. *Process Control* **2010**, *20*, 969.
9. Kadlec, P.; Grbic, R.; Gabrys, B. *Comput. Chem. Eng.* **2011**, *35*, 1.
10. Sharmin, R.; Sundararaj, U.; Shah, S.; Griend, L. V.; Sun, Y. J. *Chem. Eng. Sci.* **2006**, *61*, 6372.
11. Ahmed, F.; Nazir, S.; Yeo, Y. K.; Korean, J. *Chem. Eng.* **2009**, *26*, 14.
12. Zhang, J.; Jin, Q.; Xu, Y. M. *Chem. Eng. Technol.* **2006**, *29*, 442.
13. Gonzaga, J. C. B.; Meleiro, L. A. C.; Kiang, C.; Filho, R. M. *Comput. Chem. Eng.* **2009**, *33*, 43.
14. Mat Noor, R. A.; Ahmad, Z.; Mat Don, M.; Uzir, M. H. *Can. J. Chem. Eng.* **2010**, *88*, 1065.
15. Lou, H. C.; Su, H. Y.; Xie, L.; Gu, Y.; Rong, G. *Ind. Eng. Chem. Res.* **2012**, *51*, 8510.
16. Huang, M. Y.; Liu, X. G.; Li, J. B. *J. Appl. Polym. Sci.* **2012**, *126*, 519.
17. Han, I. S.; Han, C. H.; Chung, C. B. *J. Appl. Polym. Sci.* **2005**, *95*, 967.

18. Lee, D. E.; Song, J. H.; Song, H. O.; Yoon, E. S. *Ind. Eng. Chem. Res.* **2005**, *44*, 2101.
19. Shi, J.; Liu, X. G. *J. Appl. Polym. Sci.* **2006**, *101*, 285.
20. Chitralekha, S. B.; Shah, S. L. *Can. J. Chem. Eng.* **2010**, *88*, 696.
21. Kaneko, H.; Funatsu, K. *Ind. Eng. Chem. Res.* **2011**, *50*, 10643.
22. Liu, Y.; Chen, J. J. *Process Control* **2013**, *23*, 793.
23. Chen, T.; Ren, J. H. *Neurocomputing* **2009**, *72*, 1605.
24. Ge, Z. Q.; Chen, T.; Song, Z. H. *Control Eng. Pract.* **2011**, *19*, 423.
25. Jiang, H. Q.; Xiao, Y. D.; Li, J. B.; Liu, X. G. *Chem. Eng. Technol.* **2012**, *35*, 819.
26. Kim, M.; Lee, Y. H.; Han, I. S.; Han, C. H. *Ind. Eng. Chem. Res.* **2005**, *44*, 334.
27. Liu, J. L. *Control Eng. Pract.* **2007**, *15*, 769.
28. Kaneko, H.; Arakawa, M.; Funatsu, K. *Comput. Chem. Eng.* **2011**, *35*, 1135.
29. Rasmussen, C. E.; Williams, C. K. I. Gaussian Processes for Machine Learning; Massachusetts Institute of Technology: Cambridge, MA, **2006**.
30. Gregorcic, G.; Lightbody, G. *Eng. Appl. Artif. Intell.* **2009**, *22*, 522.
31. Ni, W.; Tan, S.; Ng, W.; Brown, S. D. *Ind. Eng. Chem. Res.* **2012**, *51*, 6416.
32. Chiang, L. H.; Pell, R. J.; Seasholtz, M. B. *J. Process Control* **2003**, *13*, 437.
33. Liu, H.; Shah, S. L.; Jiang, W. *Comput. Chem. Eng.* **2004**, *28*, 1635.
34. Ben-Hur, A.; Horn, D.; Siegelmann, H.; Vapnik, V. J. *Machine Learning Res.* **2002**, *2*, 125.
35. Chiang, J. H.; Hao, P. Y. *IEEE Trans. Fuzzy Syst.* **2003**, *11*, 518.
36. Wang, J. S.; Chiang, J. C. *IEEE Trans. Syst. Man Cybernetics Part B* **2008**, *38*, 78.
37. Atkeson, C. G.; Moore, A. W.; Schaal, S. *Artif. Intell. Rev.* **1997**, *11*, 11.
38. Bontempi, G.; Birattari, M.; Bersini, H. *Int. J. Control* **1999**, *72*, 643.
39. Cheng, C.; Chiu, M. S. *Chem. Eng. Sci.* **2004**, *59*, 2801.
40. Fujiwara, K.; Kano, M.; Hasebe, S.; Takinami, A. *AIChE J.* **2009**, *55*, 1754.
41. Liu, Y.; Gao, Z. L.; Li, P.; Wang, H. Q. *Ind. Eng. Chem. Res.* **2012**, *51*, 4313.
42. Chen, K.; Ji, J.; Wang, H. Q.; Liu, Y.; Song, Z. H. *Chem. Eng. Res. Des.* **2011**, *89*, 2117.
43. Liu, Y.; Gao, Z. L.; Chen, J. *Chem. Eng. Sci.* **2013**, *102*, 602.
44. Csato, L.; Opper, M. *Neural Comput.* **2002**, *14*, 641.
45. Ranganathan, A.; Yang, M. H.; Ho, J. *IEEE Trans. Image Processing* **2011**, *20*, 391.
46. Vapnik, V. N. The Nature of Statistical Learning Theory; Springer-Verlag: New York, **1995**.
47. Zhou, Z. H.; Wu, J. X.; Tang, W. *Artif. Intell.* **2002**, *137*, 239.
48. Huang, D.; Wang, C. D.; Lai, J. H. In Proceedings of 2012 21st International Conference on Pattern Recognition (ICPR); IEEE Conference Publications: Tsukuba, Japan, **2012**, p 2339.
49. Liu, Y.; Gao, Z. L. *J. Appl. Polym. Sci.* **2015**, *132*, 1905.