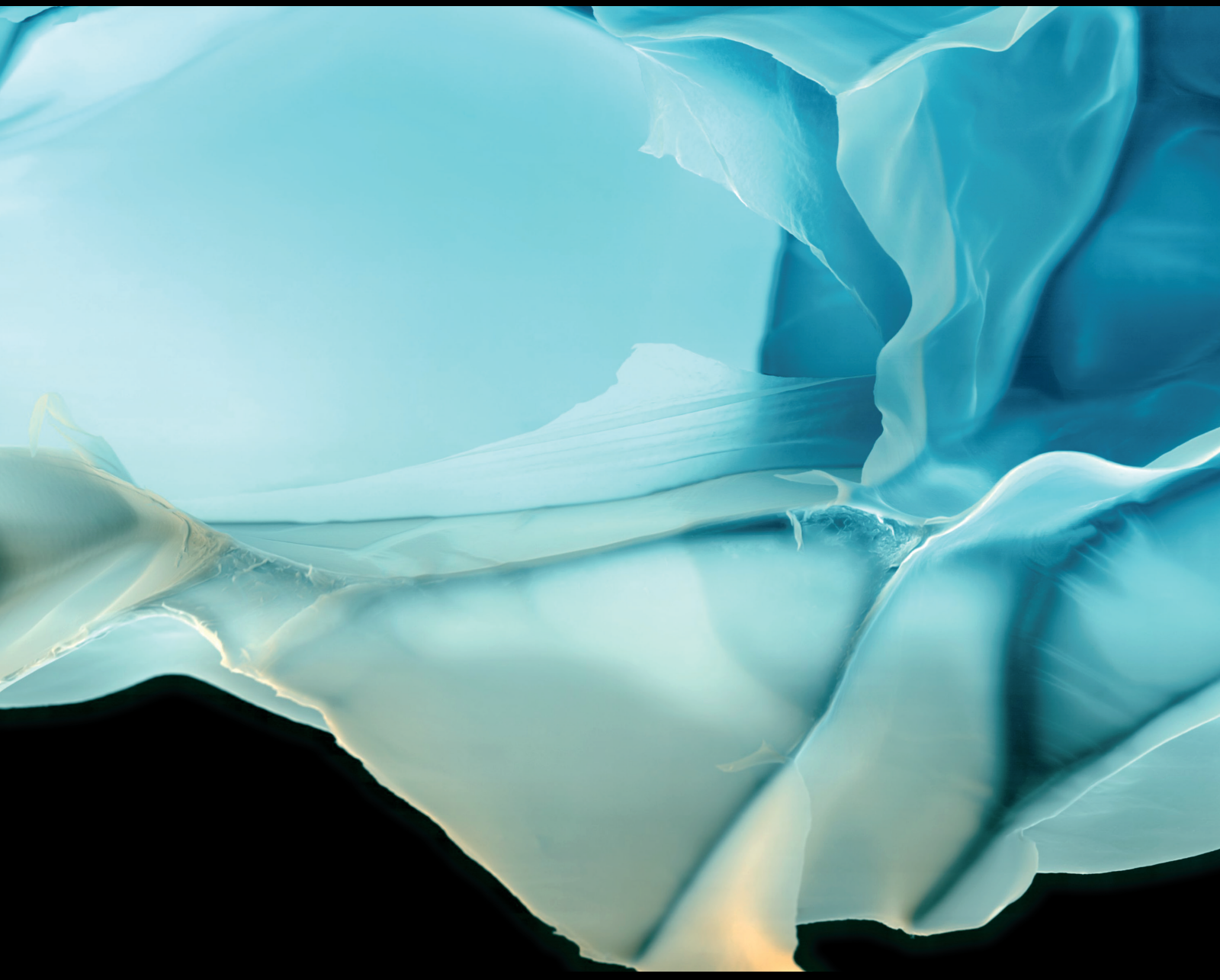
Advances in Polymer Technology

Machine Learning for Advanced Polymer Manufacturing, Processing, and Testing

Lead Guest Editor: Yi Liu
Guest Editors: Yuan Yao and Tao Chen



Machine Learning for Advanced Polymer Manufacturing, Processing, and Testing

Advances in Polymer Technology

Machine Learning for Advanced Polymer Manufacturing, Processing, and Testing

Lead Guest Editor: Yi Liu
Guest Editors: Yuan Yao and Tao Chen

Contents

Hindawi
Advances in Polymer Technology
Volume 2020, Article ID 6575326, 14 pages
https://doi.org/10.1155/2020/6575326

Research Article

Ensemble Just-In-Time Learning-Based Soft Sensor for Mooney Viscosity Prediction in an Industrial Rubber Mixing Process

Huaiping Jin, Jiangang Li, Meng Wang, Bin Qian, Biao Yang, Zheng Li, and Lixian Shi

Department of Automation, Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China

Correspondence should be addressed to Huaiping Jin; jinhuaiping@gmail.com

Received 30 May 2019; Accepted 3 March 2020; Published 27 March 2020

Guest Editor: Yi Liu

The lack of online sensors for Mooney viscosity measurement has posed significant challenges for enabling efficient monitoring, control, and optimization of industrial rubber mixing process. To obtain real-time and accurate estimations of Mooney viscosity, a novel soft sensor method, referred to as multimodal perturbation- (MP-) based ensemble just-in-time learning Gaussian process regression (MP-EJITGPR), is proposed by exploiting ensemble JIT learning. This method employs perturbations on similarity measure and input variables for generating the diversity of JIT learners. Furthermore, a set of accurate and diverse JIT learners are built through an evolutionary multiobjective optimization by balancing the accuracy and diversity objectives explicitly. Moreover, all base JIT learners are combined adaptively using a finite mixture mechanism. The proposed method is applied to an industrial rubber mixing process for Mooney viscosity prediction, and the experimental results demonstrate its effectiveness and superiority over traditional soft sensor methods.

1. Introduction

Rubber mixing is a crucial step in rubber and tire industry. The quality of rubber products highly depends on the exact mixing of raw materials and additives. The Mooney viscosity, indicating the molecular weight and viscoelastic behavior of an elastomer, has been recognized as an important quality index for producing nonvulcanized rubbery materials [1, 2]. In most rubber and tire factories, however, the Mooney viscosity can only be determined through manual analysis, which often takes 4~6 h after a batch has been discharged, while the duration of a batch run of mixing process is only about 2~5 min. Therefore, in recent years, soft sensor methods have been widely applied to provide real-time estimations of the Mooney viscosity to obtain the optimal and uniform rubber product quality [3–12].

In general, it is time-consuming and even impossible to build accurate first-principles soft sensors for Mooney viscosity due to the lack of in-depth chemical and physical knowledge of rubber mixing process. Alternatively, data-driven soft sensors have attracted much attention because of the availability of large amounts of data and advanced data mining and analytics tools [13–18]. The early attempts to data-driven soft sensors for quality estimation mainly focus on global modeling techniques, such as multivariate statistical techniques [19, 20], artificial neural networks [3, 21], support vector regression [22], and Gaussian process regression [4, 5]. Recently, deep learning methods have also been introduced to soft sensor applications [23].

However, global soft sensor methods cannot always function well because they may fail to handle local process characteristics effectively and perform model adaptation efficiently. Thus, in practical applications, local learning-based soft sensor methods are more appealing for providing accurate predictions. Compared to global modeling, local learning exhibits two outstanding advantages. First, global soft sensors are usually based on the underlying assumption of a constant operating phase and conditions throughout the entire duration of a production process, whereas industrial processes are often characterized by strong nonlinearity, and multiple operation phases or modes. Thus, from the process industry viewpoint, local learning is more suitable for

handling complex process characteristics. Second, since local learning is completely localized, the local models can be built and updated independently of each other, which greatly simplifies the incremental adaptation, inclusion, or removal of local models and receptive fields.

Generally, there are two categories of local learning methods: ensemble methods [24–27] and just-in-time learning (JIT) methods [6, 7, 9, 12]. These methods employ the divide-and-conquer philosophy to model the relationships between the inputs and output by building a set of locally valid models. In particular, JIT learning, known as a representative local learning paradigm, has gained growing attention in soft sensor applications for Mooney viscosity estimation due to its strong capability of handling nonlinearity, time-varying behavior, multiphase and multimode process characteristics, etc. [9, 11, 12]. However, traditional JIT soft sensors attempt to build a globally optimal encapsulation of local modeling techniques, similarity measures, input variables, model hyper-parameters, etc., while the diversity of JIT learning is ignored. To tackle this problem, various ensemble JIT learning (EJIT) soft sensors have been developed [6, 8, 25, 28–31].

The basic idea of EJIT modeling is to build multiple component JIT learners and then combine their predictions. For instance, Liu et al. (2012) [28] employed heterogeneous predictive models to build base JIT models and then combined them through a simple averaging rule. Liu and Gao (2015) [6] developed an EJIT soft sensor by using diverse similar data sets, which are obtained by assigning diverse hyperparameters to the support vector clustering for outlier detection. Kaneko and Funatsu (2016) [25] developed an ensemble locally weighted partial least squares (LWPLS) soft sensor, where diverse subsets are first built using moving window method and then multiple of the most relevant ones to the query state are selected to build diverse LWPLS models, which are finally integrated via Bayes' theorem. Liu et al. (2015) [29] built an EJIT kernel learning framework through perturbing the hyperparameters of local learning methods. Yuan et al. (2018) [30] developed an EJIT soft sensor by using different similarity measures. Besides, we proposed an EJIT soft sensor by perturbing the input features for building diverse input subspaces [8]. Recently, we developed an EJIT soft sensor by employing multiple weighted Euclidean distance- (WED-) based similarity measures, which are optimized through an EMO approach [31]. These studies show that it is feasible and effective to enhance the prediction accuracy of JIT soft sensors by introducing ensemble learning.

However, it remains challenging to build high-performance EJIT soft sensors due to the following issues. First, many current EJIT soft sensor methods only consider single perturbation, such as perturbing training data [25], similarity measure [30, 31], perturbing input variables [8], perturbing local modeling technique [28], or perturbing model parameters [29]. In practice, the diversity of JIT learning is often originated from multiple factors. Second, most of the current EJIT methods construct base JIT learners in a heuristic way. In such situations, the accuracy and diversity objectives of JIT learners are difficult to achieve a good tradeoff. Finally, most methods employ nonadaptive weightings for the combination of base JIT learners, which will limit the prediction performance of EJIT soft sensors.

To address the aforementioned issues, a novel EJIT soft sensor method, referred to as multimodal perturbation-based EJITGPR (MP-EJITGPR), is proposed for enabling accurate predictions of Mooney viscosity. This method works through integrating perturbation on similarity measures and perturbation on input variables together. With the multimodal perturbation mechanism, a set of accurate and diverse JIT learners are built by balancing the accuracy and diversity objectives explicitly through an evolutionary multiobjective optimization (EMO) approach. Then, a finite mixture mechanism- (FMM-) based weighting stagey is used to achieve an adaptive combination of base learners. In summary, the main contributions of this study are as follows:

(1) A multimodal perturbation mechanism is proposed by utilizing heterogeneous similarity measures and building diverse input subspaces, which allows enhancing the diversity of base JIT learners efficiently

(2) The generation of accurate and diverse JIT learners is formulated as a multiobjective optimization problem and then solved by an EMO approach

(3) The combination of base JIT learners is achieved through the finite mixture mechanism, which enables adaptive assignments of weights

(4) A novel EJIT soft sensor modeling framework is built by integrating the multimodal perturbation mechanism-based diversity creation, the EMO-based generation of base JIT learners, and the FMM-based adaptive combination of base JIT learners

The rest of the paper proceeds as follows. Section 2 briefly introduces JIT learning and Gaussian process regression. Section 3 details the proposed MP-EJITGPR soft sensor method and its implementation procedure. The application of MP-EJITGPR for Mooney viscosity prediction in an industrial rubber mixing process is reported in Section 4. Finally, conclusions are drawn in Section 5.

2. Preliminaries

2.1. Just-In-Time Learning. JIT learning [32], also known as lazy learning [33] and locally weighted learning [34], refers to a family of algorithms in which all historical data are stored in a database and local models are built dynamically by retrieving the most similar data to the query state. Compared to conventional global modeling methods, JIT learning has the following features:

(1) All available modeling data are stored at a database. And only those samples most similar to the query point are used for modeling for each run of prediction.

(2) Only when an estimation is required, a local model is built dynamically based on samples with high similarities to the query point.

(3) The constructed local model is discarded after the estimation is given.

2.2. Gaussian Process Regression. A Gaussian process is a collection of random variables, any finite number of which follows joint Gaussian distributions [35]. Considering a data set $D = \{\mathbf{X}, \mathbf{y}\} = \{\mathbf{x}_i, y_i\}_{i=1}^{n}$, the regression model can be formulated as

$$y = f(\mathbf{x}) + \varepsilon, \tag{1}$$

where $f(\cdot)$ represents an unknown regression function and ε denotes the Gaussian noise with zero mean and variance σ_n^2.

A Gaussian process is completely specified by its mean function $m(\mathbf{x})$ and covariance function $C(\mathbf{x}, \mathbf{x}')$:

$$f(\mathbf{x}) \sim \mathrm{GP}\left(m(\mathbf{x}), C(\mathbf{x}, \mathbf{x}')\right). \tag{2}$$

Since the modeling data is usually normalized to be zero mean, the output observations follow a Gaussian distribution as

$$\mathbf{y} \sim \mathrm{GP}(0, \mathbf{C}), \tag{3}$$

where $\mathbf{C}$ is an $n \times n$ covariance matrix with $\mathbf{C}_{ij} = C(\mathbf{x}_i, \mathbf{x}_j)$ representing the ijth element. In this study, a Matérn covariance function with noise term is adopted:

$$C\left(\mathbf{x}_i, \mathbf{x}_j\right) = \sigma_f^2 \left(1 + \frac{\sqrt{3}\left\|\mathbf{x}_i - \mathbf{x}_j\right\|}{l}\right) \exp\left(-\frac{\sqrt{3}\left\|\mathbf{x}_i - \mathbf{x}_j\right\|}{l}\right) + \sigma_n^2 \delta_{ij}, \tag{4}$$

where $\boldsymbol{\Theta} = \left\{\sigma_f^2, l, \sigma_n^2\right\}$ are the hyperparameters, l is the input scale, σ_n^2 is the noise variance, σ_f^2 is the output scale, and $\delta_{ij} = \begin{cases} 1; i = j \\ 0; i \neq j \end{cases}$. The hyperparameters $\boldsymbol{\Theta}$ are determined by Bayesian inference.

Given a query data $\mathbf{x}_{\mathrm{new}}$, the training outputs $\mathbf{y}$ and the test output y_{new} follow a joint Gaussian distribution as follows:

$$\begin{bmatrix} \mathbf{y} \\ y_{\mathrm{new}} \end{bmatrix} \sim N\left(0, \begin{bmatrix} \mathbf{C} & \mathbf{k}_{\mathrm{new}} \\ \mathbf{k}_{\mathrm{new}}^{\mathrm{T}} & C\left(\mathbf{x}_{\mathrm{new}}, \mathbf{x}_{\mathrm{new}}\right) \end{bmatrix}\right), \tag{5}$$

where $\mathbf{k}_{\mathrm{new}} = [C(\mathbf{x}_{\mathrm{new}}, \mathbf{x}_1), \ldots, C(\mathbf{x}_{\mathrm{new}}, \mathbf{x}_n)]^{\mathrm{T}}$. Then, the prediction output $\hat{y}_{\mathrm{new}}$ and variance σ_{new}^2 can be calculated as

$$\begin{cases} \hat{y}_{\mathrm{new}} = \mathbf{k}_{\mathrm{new}}^{\mathrm{T}} \mathbf{C}^{-1} y, \\ \sigma_{\mathrm{new}}^2 = C\left(\mathbf{x}_{\mathrm{new}}, \mathbf{x}_{\mathrm{new}}\right) - \mathbf{k}_{\mathrm{new}}^{\mathbf{T}} \mathbf{C}^{-1} \mathbf{k}_{\mathrm{new}}, \end{cases} \tag{6}$$

3. Proposed MP-EJITGPR Soft Sensor

In this section, a multimodal perturbation-based ensemble just-in-time learning Gaussian process regression (MP-EJITGPR) is presented. First, data preprocessing is conducted on the three-way data matrices of industrial rubber mixing process. Then, heterogeneous similarity measures are defined. Furthermore, by introducing the multimodal perturbation mechanism, a set of accurate and diverse JIT learners, which are equipped with heterogeneous similarity measures and diverse subspaces, are generated through an EMO approach. Next, a finite mixture mechanism is employed to achieve an adaptive combination of base JIT learners. Finally, the implementation procedure of MP-EJITGPR is provided.

3.1. Data Preprocessing. Typically, the online data of those easy-to-measure variables in industrial rubber mixing process can be arranged in a three-way matrix $\underline{X}(I \times J \times K)$ consisting of J process variables measured at K points for I batches. Meanwhile, the quality variable (i.e., Mooney viscosity), which is only available at the end of the batch, i.e., time point K, can be expressed as $\mathbf{y}_K(I \times 1)$. Before soft sensor modeling, it is desirable to perform essential processing. First, the process data is first preprocessed using a simple 3σ rule for outlier detection. Then, the three-way data matrix $\underline{X}(I \times J \times K)$ is unfolded into a two-way matrix, which allows utilizing the standard regression techniques for building the predictive model between online measured variables and the end-use quality variable. Generally, $\underline{X}$ can be unfolded in six different ways [36], among which batchwise unfolding is employed in this study, as illustrated in Figure 1. In practice, this way of unfolding has been actually recognized to be the most meaningful one for analysis and monitoring of batch processes. With this way of unfolding, all potential input variables at different time instants can be obtained for predicting the final quality variable. Moreover, since the dimensions and magnitudes of various process variables are significantly different from each other, another crucial step to guarantee the reliability and accuracy of soft sensors is data normalization, which is achieved by scaling the unfolded data matrices to zero mean and unit variance in this study.

3.2. Definition of Heterogeneous Similarity Measures. Similarity measure plays a central role in JIT modeling. In contrast to traditional modeling methods utilizing all available data, the JIT method is to construct a local model based on a small data set with high similarities to the query data. Thus, the key to building highly accurate JIT soft sensors is to define appropriate similarity metrics. Despite the availability of various similarity measures, it is hopeless to pursue a similarity evaluation criterion, which is consistently better than other metrics across different application scenarios. Consequently, it is a common practice to select one of the best similarity measures for a given task, which is usually time-consuming and even impossible. In practice, different similarity measures can provide different insights into similarity evaluation between data points. Thus, in this work, heterogeneous similarity measures will be combined for JIT learning, including Euclidean distance- (ED-) based similarity, cosine similarity (cosine), covariance weighted distance- (CWD-) based similarity, and correlation coefficient- (CC-) based similarity.

The ED similarity measure is the most commonly used metric for JIT learning due to its simplicity and efficiency. It is defined based on the Euclidean distance between points in space; that is,

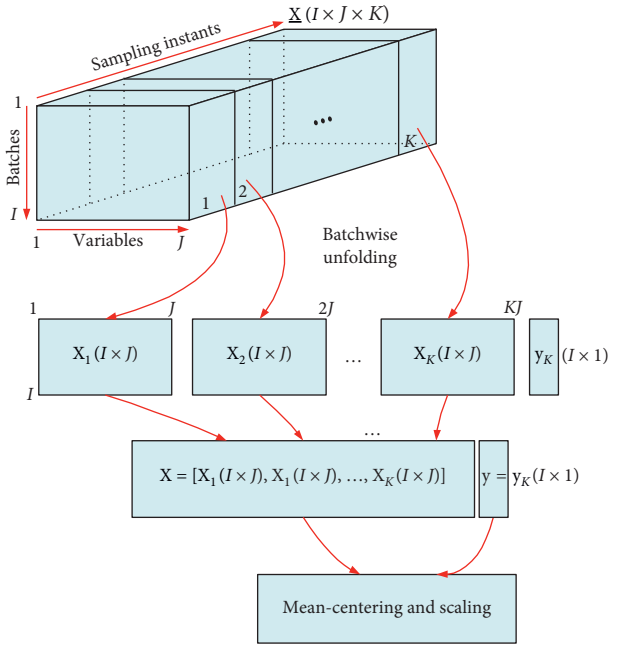


Figure 1: Batchwise unfolding of three-way data matrices.

$$\omega_i = \exp\left(-\frac{d_i}{\sigma_d \varphi}\right), \qquad d_i = \left\|\mathbf{x}_i - \mathbf{x}_{\text{new}}\right\|, \tag{7}$$

where σ_d is the standard deviation of d_i $(i = 1, 2, \ldots, n)$ and φ is a localization parameter.

One disadvantage of ED similarity is that the differences among input variables are ignored. To address this issue, various weighted distance-based similarity measures have been proposed, among which the CWD similarity measure is defined by considering the relationships among input variables and among input and output variables. [37] That is, CWD similarity is defined by using the weighted distance metric:

$$d_i = \sqrt{\left(\mathbf{x}_i - \mathbf{x}_{\text{new}}\right)^{\mathrm{T}} \mathrm{H} \left(\mathbf{x}_i - \mathbf{x}_{\text{new}}\right)}, \qquad \mathrm{H} = \frac{\left(\mathbf{X}^{\mathrm{T}}\mathbf{y}\right)^{\mathrm{T}} \left(\mathbf{X}^{\mathrm{T}}\mathbf{y}\right)}{\left\|\mathbf{X}^{\mathrm{T}}\mathbf{y}^2\right\|}, \tag{8}$$

where H denotes a weighting matrix, and $\mathbf{X}$ and $\mathbf{y}$ are the input and output matrices, respectively.

Alternatively, by exploiting the angle between two vectors in space, the cosine similarity measure can be defined as

$$\omega_i = \exp\left(\frac{\cos_i}{\sigma_{cos} \varphi}\right), \qquad \cos_i = \frac{\mathbf{x}_i^T \mathbf{x}_{\text{new}}}{\left\|\mathbf{x}_i\right\| \left\|\mathbf{x}_{\text{new}}\right\|}, \tag{9}$$

where $\cos_i$ denotes the cosine value of the angle between two vectors.

In addition to the distance and angle criteria, the relevance between two vectors can also be used to evaluate the similarity between samples. For the sake of simplicity, and without loss of generality, the frequently used correlation coefficient (CC) criterion is used to define similarity measure as follows:

$$\omega_i = \exp\left(\frac{r_i}{\sigma_r \varphi}\right), \qquad r_i = \frac{\mathrm{Cov}\left(\mathbf{x}_i, \mathbf{x}_{\text{new}}\right)}{\sqrt{\mathrm{Var}\left(\mathbf{x}_i\right)\mathrm{Var}\left(\mathbf{x}_{\text{new}}\right)}}, \tag{10}$$

where $\mathrm{Cov}(\cdot,\cdot)$ and $\mathrm{Var}(\cdot)$ are used to compute the covariance and variance, respectively.

It is noteworthy that the above similarity measures are defined from different points of view, thus behaving differently in different applications. So, one promising solution towards further improving the prediction performance of JIT learning soft sensors is to use heterogeneous similarity measures together, which has not been well investigated and will be discussed in the following section.

3.3. Generation of Base JIT Learners through Evolutionary Multiobjective Optimization. When considering JIT learning as a base learner for ensemble learning, EJIT modeling is essentially one ensemble method. It has been proven both theoretically and empirically that both accuracy and diversity of base learners are crucial to guarantee high ensemble performance [38]. According to the famous bias-variance decomposition [39] and error-ambiguity decomposition [40] theories, the more accurate and the more diverse base learners are, the better the ensemble is. Hence, the success of developing high-performance EJIT soft sensors lies in generating accurate and diverse base JIT learners.

Among various ensemble learning soft sensors, perturbing training data remains dominant for creating diversity, such as clustering [41], moving window [24, 27], bootstrapping sampling [42, 43], and sequential sampling [44]. However, such data manipulation strategies do not always function well for EJIT modeling because JIT learning only relies on a small subset of relevant samples for each run of prediction and is less sensitive to the randomness injected to the database. Moreover, the perturbation on input variables is often ignored in developing ensemble soft sensors. In addition, many of the current methods for base learner generation are based on heuristic mechanisms without measuring or ensuring diversity among base learners explicitly.

Thus, in this study, the diversity generation is achieved through the multimodal perturbation mechanism, i.e., perturbing similarity measure and input variables together. Then, the generation of accurate and diverse base JIT learners is formulated as a multiobjective optimization problem (MOP). Finally, the MOP problem is solved by using an EMO approach, which leads to a tradeoff between accuracy and diversity objectives. In the following, the

decision variables, optimization objectives, and the adopted EMO approach are detailed.

Suppose M JIT learners, each of which is characterized by one similarity measure and one input subspace, are required for constructing an EJIT model. Since M heterogeneous similarity measures have been defined in advance, the decision variables $\mathbf{z}$ only include the selection variables, which indicate whether an input variable is selected or not in order to form a subspace. That is, the decision vector $\mathbf{z}$ can be expressed as follows:

$$\mathbf{z} = \left[\overbrace{\theta_1^1, \theta_2^1, \cdots, \theta_D^1}^{\text{Simi 1, JIT learner 1}}, \ldots, \overbrace{\theta_1^m, \theta_2^m, \cdots, \theta_D^m}^{\text{Simi } m\text{, JIT learner } m}, \ldots, \overbrace{\theta_1^M, \theta_2^M, \cdots, \theta_D^M}^{\text{Simi } M\text{, JIT learner } M}\right], \tag{11}$$

where M and D denote the number of base JIT learners and that of potential input variables for building subspaces, respectively. θ is a binary variable, which indicates inclusion and exclusion of an input variable by using "1" and "0", respectively.

Furthermore, the accuracy and diversity objectives are defined. In this study, the ensemble accuracy is given as the average of individual accuracies. Given the training data set $D_{\text{trn}} = \{\mathbf{X}_{\text{trn}}, \mathbf{y}_{\text{trn}}\}$ and an independent validation set $D_{\text{val}} = \{\mathbf{X}_{\text{val}}, \mathbf{y}_{\text{val}}\}$, the accuracy objective is estimated as

$$\text{RMSE}_{\text{avg,val}} = \frac{\sum_{m=1}^{M} \text{RMSE}_{\text{val}}^m}{M}, \tag{12}$$

where $\text{RMSE}_{\text{val}}^m$ denotes the root-mean-squared error (RMSE) on the validation set for the m th JIT learner.

In comparison with the accuracy objective, measuring ensemble diversity is not straightforward. Up to now, there is no generally accepted formal formulation and measures for ensemble diversity. Thus, in this study, the standard deviation of individual prediction outputs is used to evaluate the ensemble diversity. By applying the base JIT learners, $\{f_{\text{JIT}}^m\}_{m=1}^M$, the prediction outputs of individual JIT learners can be given as

$$\widehat{\mathbf{Y}}_{\text{val}} = \left[\widehat{\mathbf{y}}_{\text{val}}^1, \cdots, \widehat{\mathbf{y}}_{\text{val}}^m, \cdots, \widehat{\mathbf{y}}_{\text{val}}^M\right], \tag{13}$$

where $\widehat{\mathbf{y}}_{\text{val}}^m$ is a column vector denoting the prediction output vector on the validation data using the mth JIT learner.

Let italic R be a row vector denoting the prediction outputs from various base JIT learners for the i th query data in validation set. Then, the ensemble diversity can be defined as follows:

$$\sigma_{\text{avg,val}} = \frac{\sum_{i=1}^{N_{\text{val}}} \sigma_{\text{val},i}}{N_{\text{val}}}, \tag{14}$$

where N_{val} is the number of validation samples, and $\sigma_{\text{val},i}$ is the standard deviation of $\widehat{\mathbf{y}}_{\text{val},i}$.

To build accurate and diverse base JIT learners, small $\text{RMSE}_{\text{avg,val}}$ and large $\sigma_{\text{avg,val}}$ are desirable. Thus, generating accurate and diverse JIT learners can be formulated as a biobjective optimization problem:

$$\min \left\{\text{RMSE}_{\text{avg,val}}, 1/\sigma_{\text{avg,val}}\right\}. \tag{15}$$

To solve the MOP problem in equation (15), one of the most famous EMO algorithms, i.e., NSGA-II (nondominated sorting genetic algorithm II), is employed. Details about the NSGA-II algorithm can be found in [45]. First, the decision vector $\mathbf{z}$ is coded as a binary string and an initial population is created. Then, the Pareto-optimal solutions can be obtained by performing the following procedure:

(i) Step 1: generate an initial population with N_{pop} individuals.

(ii) Step 2: evaluate each individual in the population by calculating the accuracy and diversity objectives, i.e., $\text{RMSE}_{\text{avg,val}}$ and $1/\sigma_{\text{avg,val}}$.

(iii) Step 3: repeat the following steps until the stopping condition is reached.

  (a) Assign a front number to all solutions by using nondominated sorting and calculated the crowding distance.
  (b) Generate an offspring population by using the binary tournament selection, recombination, and mutation operations.
  (c) Evaluate each solution as that in Step (2).
  (d) Merge the current and offspring populations such that elitism is ensured.
  (e) Sort the solutions from the merged population according to the nondominated sorting method.
  (f) Choose the first N_{pop} solutions from the merged population and increase the generation counter.

(iv) Step 4: find the Pareto-optimal solutions from the combined population in the last generation by applying the nondominated sorting method.

The outcome of this step is a set of Pareto-optimal solutions, one of which is selected for the ensemble construction of MP-EJITGPR modeling.

3.4. Adaptive Combination of Base JIT Learners by Finite Mixture Mechanism. With the utilization of multimodal perturbation and EMO optimization, a group of accurate and diverse JIT learners $\{f_{\text{JIT}}^m\}_{m=1}^M$ can be obtained, where f_{JIT} is built by the GPR method in this paper. When a query data is requested to be predicted, each base JIT learner makes a prediction for the output variable. To get the final predictions, these individual predictions have to be combined.

Generally, there are two classes of combination methods: nonadaptive and adaptive weightings. In the former, weights

assigned to base learners remain unchanged once deployed into the real-life operation, whereas in the latter weights are assigned adaptively to accommodate the query process state. The simplest nonadaptive weighting method is the simple averaging rule, which provides the average of individual predictions as the final prediction. Another popular nonadaptive weighting is to determine weights according to their prediction capability on training set or validation set. For instance, weights can be determined by using linear regression methods, such as PCR and PLS. Besides, the combination of base learners can be achieved by learning, i.e., stacking, which usually leads to a nonadaptive combination model. However, a deficiency of nonadaptive combination methods is that they tend to assign larger weights to the models that exhibit excellent prediction on the training set or validation set, which may lead to overestimation or underestimation of weights and thus deteriorate the generalization capability of ensemble models. Therefore, adaptive combination strategies are highly appealing.

In this study, a finite mixture mechanism-based adaptive weighting method is proposed to achieve the combination of base learners. For a new query data $\mathbf{x}_{\text{new}}$, the predictive distribution of the mth output $y_{m,\text{new}}$ of the target variable is estimated from the mth JITGPR model and $y_{m,\text{new}}$ follows a Gaussian distribution as follows:

$$y_{m,\text{new}} \sim \mathcal{N}\left(\mathbb{E}\left(y_{m,\text{new}}\right), \text{Var}\left(y_{m,\text{new}}\right)\right),\ m = 1, 2, \cdots, M, \quad (16)$$

where $\mathbb{E}(y_{m,\text{new}})$ and $\text{Var}(y_{m,\text{new}})$ are the prediction mean and variance, respectively.

Assume the local predictions $y_{1,\text{new}}, \ldots, y_{m,\text{new}}, \ldots, y_{M,\text{new}}$ are independent realizations of the overall output variable y_{new}. That is to say, y_{new} arises from a finite mixture distribution of $y_{1,\text{new}}, \ldots, y_{m,\text{new}}, \ldots, y_{M,\text{new}}$. Thus, by applying the finite mixture mechanism [26, 46], the mean and variance of the predictive distribution of the target variable can be estimated by combining all local predictions:

$$\begin{cases} \widehat{y}_{\text{new}} = \sum_{m=1}^{M} \omega_{m,new} \widehat{y}_{m,\text{new}}, \\ \sigma^2_{\text{new}} = \sum_{m=1}^{M} \omega_{m,new} \left\{ \sigma^2_{m,\text{new}} + \left(\widehat{y}_{m,\text{new}} - \widehat{y}_{\text{new}} \right)^2 \right\}, \end{cases} \quad (17)$$

where $\widehat{y}_{m,\text{new}}$ and $\sigma^2_{m,\text{new}}$ represent the prediction output and variance using the m th JITGPR model, respectively; and ω_m denotes the mixture weights satisfying the following constraints:

$$0 \le \omega_{m,\text{new}} \le 1, \sum_{m=1}^{M} \omega_{m,\text{new}} = 1. \quad (18)$$

Since the prediction uncertainty can effectively indicate the confidence level of the output predictions, we assume the mixture weights are inversely proportional to the prediction variances from individual JITGPR models. Thus, $\omega_{m,\text{new}}$ can be estimated as follows:

$$\omega_{m,\text{new}} = \frac{\left(\widehat{y}_{m,\text{new}} / \sigma_{m,\text{new}} \right)^p}{\sum_{m=1}^{M} \left(\widehat{y}_{m,\text{new}} / \sigma_{m,\text{new}} \right)^p}, \quad (19)$$

where p is an adjustable parameter.

The proposed FMM-based combination strategy allows predictions from individual JITGPR models to be combined adaptively at each run of prediction.

3.5. Implementation Procedure. The step-by-step procedure of the proposed MP-EJITGPR soft sensor method for Mooney viscosity prediction is summarized below and the schematic diagram of this approach is illustrated in Figure 2.

3.5.1. Offline Optimization Phase

(a) Collect the process data of the batch process for model training and validation

(b) Data processing is performed, including outlier detection, data unfolding, mean-centering, and scaling

(c) Formulate the generation of accurate and diverse JITGPR models as a multiobjective optimization problem (MOP)

(d) Solve the MOP problem using the EMO approach, i.e., NSGA-II

(e) By using one of the best-performing Pareto-optimal solutions, a set of JITGPR models characterized by heterogeneous similarity measures and diverse input subspaces are constructed

3.5.2. Online Prediction Phase

(a) For any query data, M relevant subsets are obtained by using the heterogeneous similarity measures and diverse input subspaces.

(b) M JITGPR models are constructed and further provide the prediction outputs and variances of the target variable.

(c) The final prediction output and variance are produced using the proposed FMM-based combination method and then those built JITGPR models are discarded. When new query data comes, go to Step (a).

It is worth noting that the computational load of the proposed MP-EJITGPR method is mainly focused on the offline optimization phase, especially the NSGA-II-based EMO optimization. However, once the learning configurations for generating diverse JIT models have been obtained, the online prediction for new test samples can be conducted fast. This is because, on one hand, if the similarity measures are defined appropriately, only a small number of samples are selected for online local modeling, which enables fast training of diverse JIT models for each query data. On the other hand, the finite mixture mechanism-based adaptive combination is very efficient because only simple calculations are involved. Therefore, the proposed approach can be applied for providing real-time estimations of Mooney viscosity in an industrial rubber mixing process.

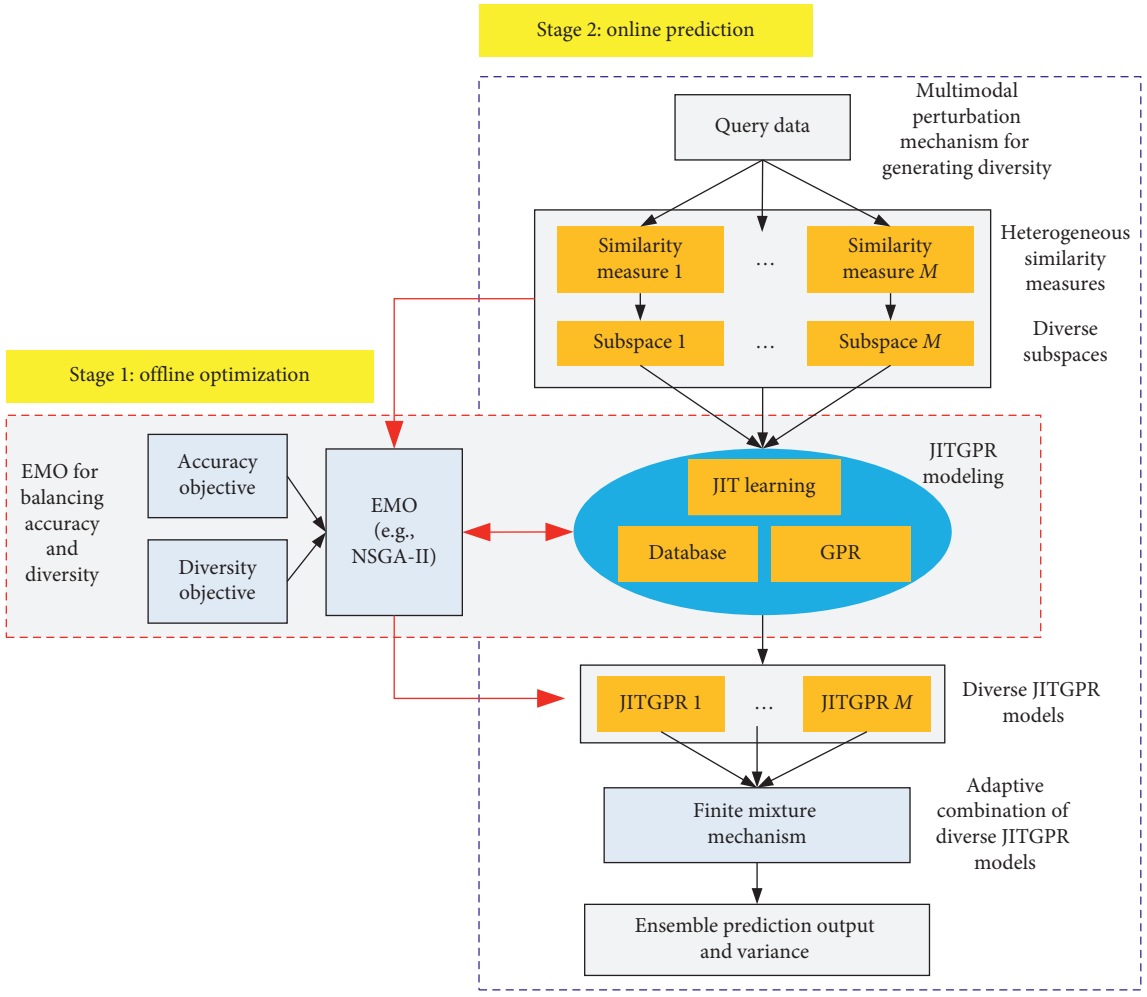


Figure 2: Schematic diagram of the proposed MP-EJITGPR soft sensing approach.

4. Application to an Industrial Rubber Mixing Process

The effectiveness and superiority of the proposed MP-EJITGPR soft sensor for Mooney viscosity prediction are demonstrated through an industrial rubber mixing process in China. The methods for comparison are as follows:

(1) PLS (partial least squares regression): global PLS model

(2) GPR (Gaussian process regression): global GPR model

(3) GMMGPR : Gaussian mixture model- (GMM-) based ensemble GPR models

(4) JITGPR (ED similarity): JIT learning GPR using ED similarity measure

(5) JITGPR (cosine similarity): JIT learning GPR using cosine similarity measure

(6) JITGPR (CWD similarity): JIT learning GPR using CWD similarity measure

(7) JITGPR (CC similarity): JIT learning GPR using CC similarity measure

(8) SP-EJITGPR (SAR): similarity perturbation-based EJITGPR with simple averaging rule (SAR) for combination

(9) SP-EJITGPR (PLS stacking): similarity perturbation-based EJITGPR with PLS stacking for combination

(10) SP-EJITGPR (GPR stacking): similarity perturbation-based EJITGPR with GPR stacking for combination

(11) SP-EJITGPR (FMM): similarity perturbation-based EJITGPR with the finite mixture mechanism for combination

(12) MP-EJITGPR (SAR): multimodal perturbation-based EJITGPR with a simple averaging rule for combination

(13) MP-EJITGPR (PLS stacking): multimodal perturbation-based EJITGPR with PLS stacking for combination

(14) MP-EJITGPR (GPR stacking): multimodal perturbation-based EJITGPR with GPR stacking for combination

(15) MP-EJITGPR (FMM) (the proposed method): multimodal perturbation-based EJITGPR with the finite mixture mechanism for combination

A total of 15 soft sensor methods are involved, and their characteristics are compared in Table 1. These methods can be roughly categorized into five classes: (1) global methods, i.e., PLS and GPR; (2) ensemble methods through training data perturbation, i.e., GMMGPR; (3) JITGPR methods using single similarity measure; (4) EJITGPR methods using similarity perturbation and ensemble learning, i.e., SP-EJITGPR; (5) EJITGPR methods using multimodal perturbation and ensemble learning, i.e., MP-EJITGPR. For those EJIT methods, four types of combination methods are investigated, i.e., simple averaging rule (SAR), PLS stacking, GPR stacking, and finite mixture mechanism (FMM).

The modeling data are split into three parts: training set for modeling learning, validation set for parameter tuning and EMO optimization, and testing set for model evaluation. Moreover, some critical parameters should be predetermined. In detail, the number of principal components for PLS is selected based on the prediction accuracy on the validation set. The local modeling size l and the adjustable parameter p in equation (19) are determined by trial and error. In addition, the optimization settings for NSGA-II are given as follows: population size $N_{\text{pop}} = 100$, and maximum generation size $N_{\text{gen}}^{\max} = 100$.

To assess the prediction performance of soft sensors, three indices, namely, root-mean-square error (RMSE), relative RMSE (RRMSE), and coefficient of determination (R^2), are used:

$$\begin{aligned} \text{RMSE} &= \sqrt{\frac{1}{n_{\text{test}}}\sum_{i=1}^{n_{\text{test}}}\left(\widehat{y}_i - y_i\right)^2}, \\ \text{RRMSE} &= \sqrt{\frac{1}{n_{\text{test}}}\sum_{i=1}^{n_{\text{test}}}\left(\frac{\widehat{y}_i - y_i}{y_i}\right)^2} \times 100\%, \\ R^2 &= 1 - \frac{\sum_{i=1}^{n_{\text{test}}}\left(\widehat{y}_i - y_i\right)^2}{\sum_{i=1}^{n_{\text{test}}}\left(y_i - \overline{y}\right)^2}, \end{aligned} \tag{20}$$

where y_i and $\widehat{y}_i$ denote the actual and predicted outputs, respectively; $\overline{y}$ represents the mean value; and n_{test} is the number of testing samples.

The computer configurations for experiments are as follows. OS : Windows 10 (64 bit); CPU : Intel (R) Core(TM) i7-6700 (3.4 GHz × 2); RAM : 8G byte; and the simulation software is MATLAB R2016a. The MATLAB codes for running GPR regression can be downloaded from the website: http://www.gaussianprocess.org/gpml/code/matlab/doc/.

4.1. Process Description. Rubber mixing is a crucial step in the rubber and tire industry. [7, 8] The industrial rubber mixing process in this study is practiced in a Chinese tire company. The industrial production site is illustrated in Figure 3. The rubber mixing process lasts for 2 min, during which various raw materials are fed into the raw rubber to produce synthetic rubber according to the technical formula and then various complex chemical reactions take place in an internal mixer. Generally, Mooney viscosity is a crucial index for monitoring the product quality of the rubber mixing process. However, in a practical production process, Mooney viscosity is only measured through a viscometer with a large delay of 4~6 h after one batch has discharged. Consequently, it becomes challenging to assure the optimal and uniform quality of mixed rubber. Fortunately, soft sensor technology provides the possibility of estimating Mooney viscosity in real-time. Thus, we attempt to build a high-performing soft sensor for Mooney viscosity prediction in this study. The process variables used for soft sensor modeling include temperature in the mixer chamber, motor power, ram pressure, motor speed, and energy.

4.2. Prediction Results of Mooney Viscosity. The modeling data have been collected from DCS system and laboratory analysis and are then preprocessed by using a simple 3σ rule for outlier detection, batchwise unfolding, zero mean centering, and one variance scaling. With a sampling interval of 2 s, a total of 1172 batches are selected from three internal mixers and are further divided into three sets: 822 batches as the training set, 175 batches as the validation set, and 175 batches as the testing set. By considering the time instants 0 s, 14 s, 18 s, 22 s, ..., 118 s, a total of 140 delayed and nondelayed variables are obtained as potential input variables and the Mooney viscosity is chosen as the output variable.

The prediction results of Mooney viscosity from different soft sensor methods are presented in Table 2. It is readily observed that PLS leads to the poorest prediction performance among those methods in terms of RMSE, RRMSE, and R^2. This is mainly because PLS cannot effectively handle the nonlinearity of rubber mixing process. In comparison, other nonlinear soft sensor methods achieve much better prediction accuracy than PLS. Though GPR obtains significant accuracy improvement, it still produces high prediction errors due to its failure in dealing with local process characteristics. Instead of relying on a global model, GMMGPR and various JITGPR methods employ local learning philosophy, thus obtaining much better performance than global GPR. Although GMMGPR performs well in this case study, JITGPR methods are more appealing to provide better prediction performance.

However, the prediction accuracy of JITGPR methods is highly related to the similarity measure definition. As can be seen in Table 2, different prediction performance is obtained by using different similarity measures. In real applications, it is difficult to determine which similarity measure performs best in advance. Thus, a promising idea is to fully exploit the advantages of multiple similarity measures for JIT learning by using ensemble methods. As expected, by introducing ensemble learning, SP-EJITGPR can deliver better prediction results than single similarity measure-based JITGPR methods. It is noteworthy that, however, inappropriate

Table 1: Characteristics of soft sensor methods for comparison.

No.	Method	Model structure	Learning type	Diversity generation mechanism
1	PLS	Single	Global	—
2	GPR	Single	Global	—
3	GMMGPR	Ensemble	Local	Training data perturbation
4	JITGPR (ED similarity)	Single	Local	—
5	JITGPR (cosine similarity)	Single	Local	—
6	JITGPR (CWD similarity)	Single	Local	—
7	JITGPR (CC similarity)	Single	Local	—
8	SP-EJITGPR (SAR)	Ensemble	Local	Similarity perturbation
9	SP-EJITGPR (PLS stacking)	Ensemble	Local	Similarity perturbation
10	SP-EJITGPR (GPR stacking)	Ensemble	Local	Similarity perturbation
11	SP-EJITGPR (FMM)	Ensemble	Local	Similarity perturbation
12	MP-EJITGPR (SAR)	Ensemble	Local	Multimodal perturbation
13	MP-EJITGPR (PLS stacking)	Ensemble	Local	Multimodal perturbation
14	MP-EJITGPR (GPR stacking)	Ensemble	Local	Multimodal perturbation
15	MP-EJITGPR (FMM)	Ensemble	Local	Multimodal perturbation

Figure 3: Industrial rubber mixing process.

Table 2: Comparison of Mooney viscosity prediction results using different soft sensors ($l = 10$ for JIT learning).

No.	Method	RMSE	RRMSE (%)	R^2
1	PLS	7.3298	11.7026	0.8002
2	GPR	4.2628	5.9354	0.9324
3	GMMGPR	3.4606	4.8077	0.9555
4	JITGPR (ED similarity)	3.1561	4.2700	0.9630
5	JITGPR (cosine similarity)	3.2053	4.3494	0.9618
6	JITGPR (CWD similarity)	3.6552	5.2370	0.9503
7	JITGPR (CC similarity)	3.2029	4.3313	0.9618
8	SP-EJITGPR (SAR)	3.2127	4.4105	0.9616
9	SP-EJITGPR (PLS stacking)	3.9916	5.4067	0.9407
10	SP-EJITGPR (GPR stacking)	3.7792	5.0508	0.9469
11	SP-EJITGPR (FMM)	3.0670	4.2073	0.9650
12	MP-EJITGPR (SAR)	4.3769	6.0332	0.9288
13	MP-EJITGPR (PLS stacking)	4.1966	5.3943	0.9345
14	MP-EJITGPR (GPR stacking)	3.7819	4.9735	0.9468
15	MP-EJITGPR (FMM)	**2.9202**	**3.9085**	**0.9683**

ED: Euclidean distance; CWD: covariance weighted distance; CC: correlation coefficient; SAR: simple averaging rule; FMM: finite mixture mechanism; SP: similarity perturbation; and MP: multimodal perturbation.

combination methods can lead to performance degradation instead of improvements. Among the compared combination methods, the proposed FMM-based combination successfully obtains significant performance enhancement, while simple averaging rule, PLS stacking, and GPR stacking lead to performance degradation. These results reveal that the integration of heterogeneous similarity measures and the FMM-based adaptive combination method significantly allows improving the prediction accuracy of JIT learning soft sensors.

Apart from the similarity measure, input variable selection is also critical to guarantee the performance of JIT learning. Thus, it is interesting to explore whether EJITGPR model using only similarity perturbation can be further improved or not by performing perturbations on similarity measure and input variables simultaneously. As we have expected, when the FMM-based adaptive combination is employed, EJITGPR model using multimodal perturbation, i.e., MP-EJITGPR (FMM), performs better than SP-EJITGPR methods. Once again, the simple averaging rule, PLS stacking, and GPR stacking does not function well in this study because they are nonadaptive. The above observations show that the proposed MP-EJITGPR (FMM) soft sensor method is the best among the compared methods. In addition, as illustrated in Figure 4, the superior prediction performance of MP-EJITGPR (FMM) is further verified by a good agreement between the predicted and actual trend plots of Mooney viscosity.

To further investigate the estimation performance of MP-EJITGPR (FMM) soft sensor, the prediction RMSE values of JITGPR, SP-EJITGPR, and MP-EJITGPR under different local modeling sizes are compared in Figure 5. It can be found that the increase of local modeling samples can lead to prediction accuracy reduction in most cases for all compared methods. In particular, for this case study, the prediction accuracy of JITGPR methods and SP-EJITGPR using small local modeling sizes is significantly better than that using large local modeling sizes. In comparison, the proposed MP-EJITGPR (FMM) is much less sensitive to local modeling size than other methods. Therefore, the proposed approach is more desirable than other traditional JIT methods in providing accurate and reliable predictions.

Compared to the traditional JIT learning soft sensors, the outstanding prediction performance of MP-EJITGPR

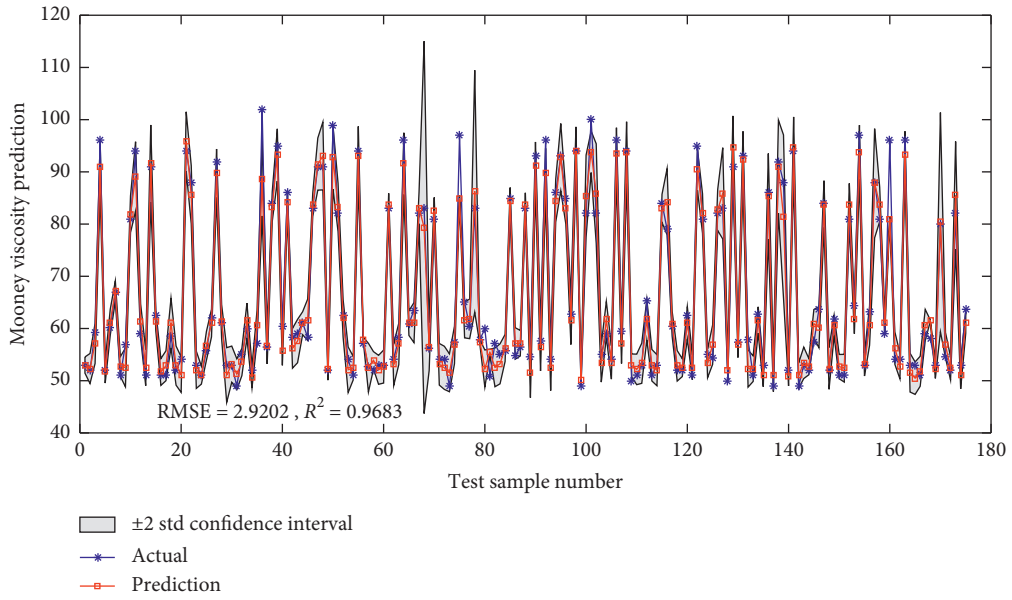


Figure 4: Trend plots of Mooney viscosity predictions using the proposed MP-EJITGPR (FMM) approach (l=10).

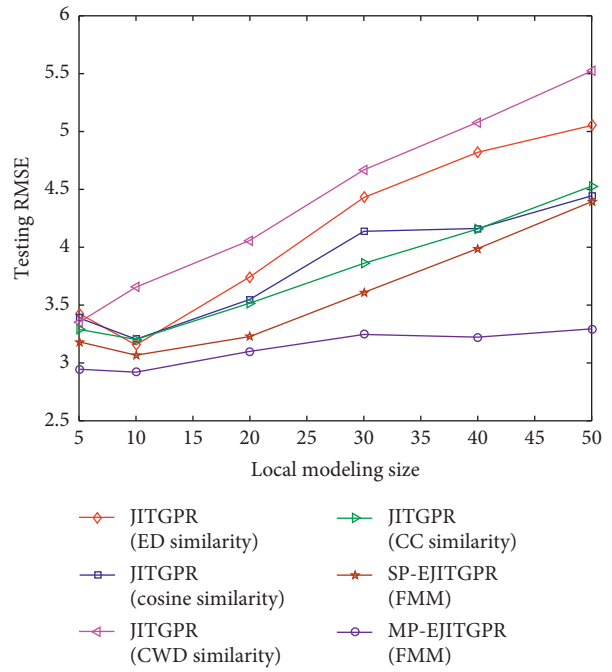


Figure 5: Comparison of prediction performance using single and ensemble JITGPR models.

(FMM) is mainly due to the effective cooperation of multimodal perturbation, EMO optimization, and adaptive combination for ensemble construction. On the one hand, the utilization of heterogeneous similarity measures, i.e., ED, cosine, CWD, and CC similarity metrics, and input variable selection for constructing subspaces shown in Figure 6, can be helpful to generate accurate and diverse base JITGPR models. On the other hand, the accuracy and diversity objectives of base JITGPR models can be well balanced by using an EMO approach. Additionally, the FMM-based adaptive combination scheme allows the proposed MP-EJITGPR method to accommodate the query process state by dynamically assigned weights to base JITGPR models, as illustrated in Figure 7.

Moreover, the real-time performance of MP-EJITGPR (FMM) for online prediction is analyzed. The average CPU time for each run of prediction under different local modeling sizes is shown in Figure 8. Clearly, the online computational load becomes large with the increase of local modeling samples. However, only a small relevant subset is

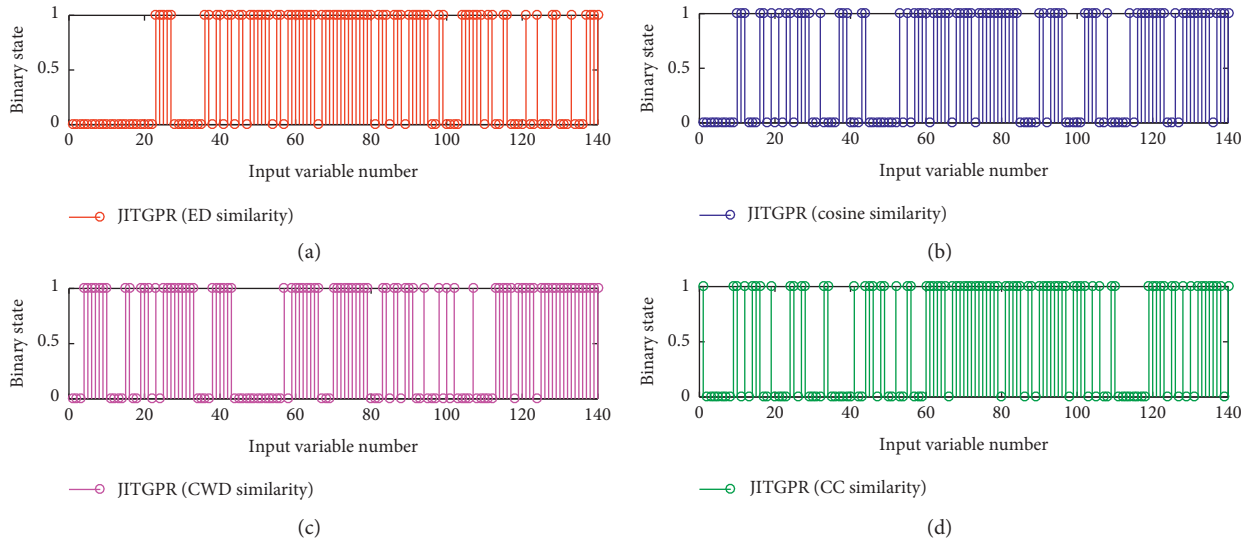


Figure 6: Input variable selection results for various similarity measures using EMO optimization. (a) JITGPR (ED similarity). (b) JITGPR (cosine similarity). (c) JITGPR (CWD similarity). (d) JITGPR (CC similarity).

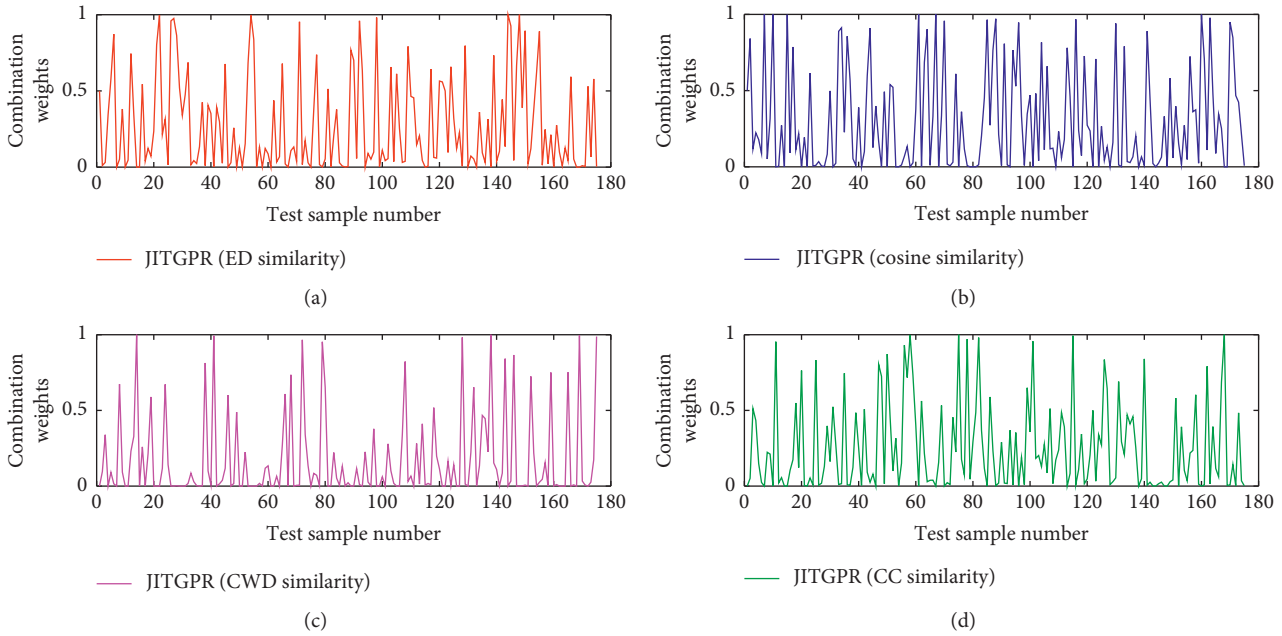


Figure 7: Combination weights assigned to different JITGPR models in MP-EJITGPR (FMM) method. (a) JITGPR (ED similarity). (b) JITGPR (cosine similarity). (c) JITGPR (CWD similarity). (d) JITGPR (CC similarity).

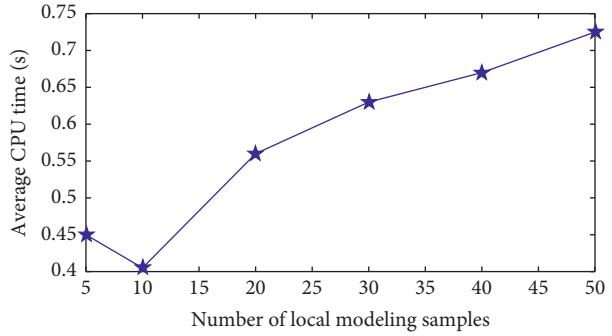


Figure 8: Average CPU time for online prediction using the proposed MP-EJITGPR (FMM) method.

required for local modeling, and the prediction taking less than 1 s is completely acceptable in practical application.

The obtained application results confirm that the proposed MP-EJITGPR (FMM) soft sensor method outperforms the other traditional JIT soft sensors, implying that it is more suitable for providing accurate predictions of Mooney viscosity in an industrial rubber mixing process.

5. Conclusions

In this paper, a new soft sensor method MP-EJITGPR is proposed for facilitating accurate estimations of Mooney viscosity in an industrial rubber mixing process. This method enables to enhance the diversity of base JIT learners through the multimodal perturbation mechanism, i.e., perturbing similarity measure and input variables. Moreover, a group of accurate and diverse base JIT learners is generated by employing an EMO approach to achieve a tradeoff between the accuracy and diversity objectives explicitly. In addition, a finite mixture mechanism is exploited to achieve an adaptive combination of base JIT learners. By integrating the multimodal perturbation-based diversity generation, the EMO optimization-based generation of base JIT learners, and the FMM-based adaptive combination of base learners for EJIT modeling, the proposed MP-EJITGPR method allows providing marked improvement of prediction performance over its conventional counterparts in nonlinear process modeling. The superiority and effectiveness of the proposed approach are demonstrated through the Mooney viscosity prediction of an industrial rubber mixing process.

Besides the presented case study, the proposed method has the potential of addressing other nonlinear modeling issues in process industry. In future research, more efforts are encouraged to extend the library of heterogeneous similarity measures and improve the diversity generation mechanism for building high-performance JIT soft sensors. Moreover, although this paper mainly focuses on manipulating input variables for building diverse input spaces based on evolutional multiobjective optimization approach, exploiting feature extraction by deep learning and making use of unlabeled data by semisupervised learning for improving the prediction performance of soft sensors are also interesting [47]. These will be investigated in the future.

Nomenclature

CC:	Correlation coefficient
CWD:	Covariance weighted distance
ED:	Euclidean distance
EJIT:	Ensemble just-in-time learning
EJITGPR:	Ensemble just-in-time learning Gaussian process regression
EMO:	Evolutionary multiobjective optimization
FMM:	Finite mixture mechanism
GPR:	Gaussian process regression
JIT:	Just-in-time learning
JITGPR:	Just-in-time learning Gaussian process regression
MOP:	Multiobjective optimization problem
MP-EJITGPR:	Multimodal perturbation-based ensemble just-in-time learning Gaussian process regression
NSGA-II:	Nondominated sorting genetic algorithm II
SAR:	Simple averaging rule
SP-EJITGPR:	Similarity perturbation-based ensemble just-in-time learning Gaussian process regression.

Data Availability

The industrial data set of rubber mixing process involved in the present study is not allowed to be disclosed for reasons of commercial confidentiality.

Conflicts of Interest

The authors declare no conflicts of financial interest.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (NSFC) (61763020, 61863020, and 61563025), the Applied Basic Research Project of Yunnan Province (2018FD040), and the Scientific Research Foundation Project of Yunnan Education Department (2017ZZX149).

References

[1] E. Ehabé, F. Bonfils, C. Aymard, A. K. Akinlabi, and J. Sainte Beuve, "Modelling of Mooney viscosity relaxation in natural rubber," *Polymer Testing*, vol. 24, no. 5, pp. 620–627, 2005.

[2] J. E. Mark, B. Erman, and M. Roland, *The Science and Technology of Rubber. The Science and Technology of Rubber*, Academic Press, Roland, IA, USA, 2013.

[3] V. Vijayabaskar, R. Gupta, P. P. Chakrabarti, and A. K. Bhowmick, "Prediction of properties of rubber by using artificial neural networks," *Journal of Applied Polymer Science*, vol. 100, no. 3, pp. 2227–2237, 2006.

[4] Z. Zhang, K. Song, T.-P. Tong, and F. Wu, "A novel nonlinear adaptive Mooney-viscosity model based on DRPLS-GP algorithm for rubber mixing process," *Chemometrics and Intelligent Laboratory Systems*, vol. 112, pp. 17–23, 2012.

[5] K. Song, F. Wu, T.-P. Tong, and X.-J. Wang, "A real-time Mooney-viscosity prediction model of the mixed rubber based on the independent component regression-Gaussian process algorithm," *Journal of Chemometrics*, vol. 26, no. 11-12, pp. 557–564, 2012.

[6] Y. Liu and Z. Gao, "Real-time property prediction for an industrial rubber-mixing process with probabilistic ensemble Gaussian process regression models," *Journal of Applied Polymer Science*, vol. 132, 2015.

[7] Y. Liu, Y. Fan, L. Zhou, F. Jin, and Z. Gao, "Ensemble correntropy-based Mooney viscosity prediction model for an industrial rubber mixing process," *Chemical Engineering & Technology*, vol. 39, no. 10, pp. 1804–1812, 2016.

[8] K. Yang, H. Jin, X. Chen, J. Dai, L. Wang, and D. Zhang, "Soft sensor development for online quality prediction of industrial batch rubber mixing process using ensemble just-in-time Gaussian process regression models," *Chemometrics and Intelligent Laboratory Systems*, vol. 155, pp. 170–182, 2016.

[9] W. Jin, Y. Liu, and Z. Gao, "Fast property prediction in an industrial rubber mixing process with local ELM model," *Journal of Applied Polymer Science*, vol. 134, no. 41, p. 45391, 2017.

[10] W. Zheng, X. Gao, Y. Liu, L. Wang, J. Yang, and Z. Gao, "Industrial Mooney viscosity prediction using fast semi-supervised empirical model," *Chemometrics and Intelligent Laboratory Systems*, vol. 171, pp. 86–92, 2017.

[11] W. Zheng, Y. Liu, Z. Gao, and J. Yang, "Just-in-time semi-supervised soft sensor for quality prediction in industrial rubber mixers," *Chemometrics and Intelligent Laboratory Systems*, vol. 180, pp. 36–41, 2018.

[12] B. Pan, H. Jin, L. Wang et al., "Just-in-time learning based soft sensor with variable selection and weighting optimized by evolutionary optimization for quality prediction of nonlinear processes," *Chemical Engineering Research and Design*, vol. 144, pp. 285–299, 2019.

[13] L. Fortuna, S. Graziani, A. Rizzo, and M. G. Xibilia, *Soft Sensors for Monitoring and Control of Industrial Processes*, Springer Science & Business Media, Heidelberg, Germany, 2007.

[14] M. Kano and Y. Nakagawa, "Data-based process monitoring, process control, and quality improvement: recent developments and applications in steel industry," *Computers & Chemical Engineering*, vol. 32, no. 1-2, pp. 12–24, 2008.

[15] P. Kadlec, B. Gabrys, and S. Strandt, "Data-driven soft sensors in the process industry," *Computers & Chemical Engineering*, vol. 33, no. 4, pp. 795–814, 2009.

[16] Y. Shen, X. Li, H. Gao, and O. Kaynak, "Data-based techniques focused on modern industry: an overview," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 1, pp. 657–667, 2015.

[17] Z. Ge, Z. Song, S. X. Ding, and B. Huang, "Data mining and analytics in the process industry: the role of machine learning," *IEEE Access*, vol. 5, pp. 20590–20616, 2017.

[18] Y. Jiang and S. Yin, "Recent advances in key-performance-indicator oriented prognosis and diagnosis with a matlab toolbox: Db-kit," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 5, pp. 2849–2858, 2018.

[19] H. Kaneko, M. Arakawa, and K. Funatsu, "Development of a new soft sensor method using independent component analysis and partial least squares," *AIChE Journal*, vol. 55, no. 1, pp. 87–98, 2010.

[20] L. Zhou, Y. Wang, Z. Ge, and Z. Song, "Multirate factor analysis models for fault detection in multirate processes," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 7, pp. 4076–4085, 2019.

[21] G. Padmavathi, M. G. Mandan, S. P. Mitra, and K. K. Chaudhuri, "Neural modelling of Mooney viscosity of polybutadiene rubber," *Computers & Chemical Engineering*, vol. 29, no. 7, pp. 1677–1685, 2005.

[22] K. Desai, Y. Badhe, S. S. Tambe, and B. D. Kulkarni, "Soft-sensor development for fed-batch bioreactors using support vector regression," *Biochemical Engineering Journal*, vol. 27, no. 3, pp. 225–239, 2006.

[23] X. Yuan, L. Li, and Y. Wang, "Nonlinear dynamic soft sensor modeling with supervised long short-term memory network," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 5, pp. 3168–3176, 2020.

[24] P. Kadlec and B. Gabrys, "Local learning-based adaptive soft sensor for catalyst activation prediction," *AIChE Journal*, vol. 57, no. 5, pp. 1288–1301, 2011.

[25] H. Kaneko and K. Funatsu, "Ensemble locally weighted partial least squares as a just-in-time modeling method," *AIChE Journal*, vol. 62, no. 3, pp. 717–725, 2016.

[26] H. Jin, X. Chen, W. Li, Y. Kai, and W. Lei, "Adaptive soft sensor development based on online ensemble Gaussian process regression for nonlinear time-varying batch processes," *Industrial & Engineering Chemistry Research*, vol. 54, no. 30, pp. 7320–7345, 2015.

[27] H. Jin, X. Chen, J. Yang, H. Zhang, L. Wang, and L. Wu, "Multi-model adaptive soft sensor modeling method using local learning and online support vector regression for nonlinear time-variant batch processes," *Chemical Engineering Science*, vol. 131, pp. 282–303, 2015.

[28] Y. Liu, D. Huang, and Y. Li, "Development of interval soft sensors using enhanced just-in-time learning and inductive confidence predictor," *Industrial & Engineering Chemistry Research*, vol. 51, no. 8, pp. 3356–3367, 2012.

[29] Y. Liu, Z. Zhang, and J. Chen, "Ensemble local kernel learning for online prediction of distributed product outputs in chemical processes," *Chemical Engineering Science*, vol. 137, pp. 140–151, 2015.

[30] X. Yuan, J. Zhou, Y. Wang, and C. Yang, "Multi-similarity measurement driven ensemble just-in-time learning for soft sensing of industrial processes," *Journal of Chemometrics*, vol. 32, no. 9, p. e3040, 2018.

[31] H. Jin, B. Pan, X. Chen, and B. Qian, "Ensemble just-in-time learning framework through evolutionary multi-objective optimization for soft sensor development of nonlinear industrial processes," *Chemometrics and Intelligent Laboratory Systems*, vol. 184, pp. 153–166, 2019.

[32] G. Cybenko, "Just-in-time learning and estimation," in *Identification, Adaptation, Learning: The Science of Learning Models from Data*, S. Bittanti and G. Picci, Eds., Springer-Verlag, Berlin, Germany, vol. 153 of Nato ASI Series F Computer and Systems Sciences, pp. 423–434, 1996.

[33] D. W. Aha, *Lazy Learning*, Springer Science & Business Media, Berlin, Germany, 2013.

[34] C. G. Atkeson, A. W. Moore, and S. Schaal, "Locally weighted learning," in *Artificial Intelligence Review*, pp. 11–73, Springer, Berlin, Germany, 1997.

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

[36] J. A. Westerhuis, T. Kourti, and J. F. MacGregor, "Comparing alternative approaches for multivariate statistical analysis of batch process data," *Journal of Chemometrics*, vol. 13, no. 3-4, pp. 397–413, 1999.

[37] K. Hazama and M. Kano, "Covariance-based locally weighted partial least squares for high-performance adaptive modeling," *Chemometrics and Intelligent Laboratory Systems*, vol. 146, pp. 55–62, 2015.

[38] Z.-H. Zhou, *Ensemble Methods: Foundations and Algorithms*, Chapman and Hall/CRC, Boca Raton, FL, USA, 2012.

[39] S. Geman, E. Bienenstock, and R. Doursat, "Neural networks and the bias/variance dilemma," *Neural Computation*, vol. 4, no. 1, pp. 1–58, 1992.

[40] A. Krogh and J. Vedelsby, "Neural network ensembles, cross validation and active learning," *Advances in Neural Information Processing Systems*, vol. 5, pp. 231–238, 1995.

[41] J. Yu, "Online quality prediction of nonlinear and non-Gaussian chemical processes with shifting dynamics using finite mixture model based Gaussian process regression approach," *Chemical Engineering Science*, vol. 82, pp. 22–30, 2012.

[42] T. Chen and J. Ren, "Bagging for Gaussian process regression," *Neurocomputing*, vol. 72, no. 7–9, pp. 1605–1610, 2009.
[43] Z. Ge and Z. Song, "Ensemble independent component regression models and soft sensing application," *Chemometrics and Intelligent Laboratory Systems*, vol. 130, pp. 115–122, 2014.
[44] H.-X. Tian, Y.-D. Liu, K. Li, R.-R. Yang, and B. Meng, "A new AdaBoost.IR soft sensor method for robust operation optimization of ladle furnace refining," *ISIJ International*, vol. 57, no. 5, pp. 841–850, 2017.
[45] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: nsga-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, 2002.
[46] S. Frühwirth-Schnatter, *Finite Mixture and Markov Switching Models*, Springer Science & Business Media, Berlin, Germany, 2006.
[47] S. Zheng, K. Liu, Y. Xu, H. Chen, X. Zhang, and Y. Liu, "Robust soft sensor with deep kernel learning for quality prediction in rubber mixing processes," *Sensors*, vol. 20, no. 3, p. 695, 2020.

Hindawi
Advances in Polymer Technology
Volume 2020, Article ID 4682689, 12 pages
https://doi.org/10.1155/2020/4682689

Research Article

Defect Detection in Composite Products Based on Sparse Moving Window Principal Component Thermography

Jing Jie, Shiqing Dai, Beiping Hou, Miao Zhang, and Le Zhou

School of Automation and Electrical Engineering, Zhejiang University of Science & Technology, Hangzhou 310023, Zhejiang, China

Correspondence should be addressed to Le Zhou; zhoule@zust.edu.cn

Received 31 August 2019; Revised 18 December 2019; Accepted 9 January 2020; Published 7 February 2020

Guest Editor: Yuan Yao

As a nondestructive testing (NDT) technology, pulsed thermography (PT) has been widely used in the defect detection of the composite products due to its efficiency and large detection range. To enhance the distinction between defective and defect-free region and eliminate the influence of the measurement noise and nonuniform background of the thermal image generated by PT, a number of thermographic data analysis approaches have been proposed. However, these traditional methods only consider the correlations among the pixel while leave the time series correlations unmodeled. In this paper, a sparse moving window principal component thermography (SMWPCT) method is proposed to incorporate several thermal images using the moving window strategy. Also, the sparse trick is used to provide clearer and more interpretable results because of the structure sparsity. The effectiveness of the method is verified by the defect detection experiment of carbon fiber-reinforced plastic specimens.

1. Introduction

The nondestructive testing (NDT) [1, 2] is a method to detect the presence of defects or unevenness in the tested objects by utilizing the characteristics of light, heat, magnetism, or electricity, which do not affect the performance of the testing objects. The nondestructive nature of the inspection makes NDT more and more popular, such as the detection of defects on the surface and subsurface of composite materials. Among different types of NDT methods, pulsed thermography has been widely used and studied owing to its fast detection speed, noncontact, and nonpollution [3]. Through scanning, recording, or observing the change of the surface temperature field, which is caused by the difference of heat transfer to the deep layer, the NDT will realize the detection of the surface and internal workpiece defects or analyze the internal structure [4–6].

For enhancing the detection efficiency and visibility, several signal processing and data analysis methods have been proposed. Among them, the most straightforward approach is to reduce the measurement noises and eliminate the nonuniform background. Thermographic signal reconstruction (TSR) method [7] converts time domain signals into frequency domain signals, which separated the nonuniform noise and significantly reduce noise interference by removing noise part in the data reconstruction. Besides, the differential absolute contrast (DAC) [8] and mathematical morphology (MM) [9] are also proposed for the same purpose, which are both based on the reconstruction of defect-free images so that the nonuniform background has been eliminated by extracting the features of the original thermal images. However, both TSR and DAC have processed the time domain data while they ignored the spatial information. Hence, penalized least square (PELS) [10] has been developed to utilize the time series thermal images and their spatial information simultaneously. Furthermore, Chang et al. have decomposed the thermal images into high-frequency noises, low-frequency background, and signal information using the signal decomposition, which is called the multimaintenance ensemble empirical mode decomposition (MEEMD) method [11]. Even though PELS and MEEMD have been proven the effective signal processing methods for NDT, they are still quite time-consuming, and the model selection is hard to achieve due to a

large number of model parameters. In addition, the noise reduction and background removal methods are only suitable for defect detection of single temperature attenuation curve or single thermal images.

Another kind of approach is the thermographic data analysis method, which extracts the principal features from multiple thermal images and automatically recognize the defects using these features or loading matrixes. Recently, the data analysis and feature extraction technologies have been widely used in process modeling, monitoring, and optimization areas [12–17]. Utilizing their advantages, the main information can be maintained with few features and the minimum reconstruction errors [18–20]. The higher-order statistics (HOS) [21] extracts the features of red-hot image sequences and compresses the feature information into a unique image for defect detection. The pulsed phase thermography (PPT) method separates one-dimensional Fourier for each pixel of the thermal imaging sequence and judge defects according to amplitude and phase [22]. Besides, principal component thermography (PCT) [23] applied principal component analysis (PCA) [24] to the thermal image data processing, which has the advantages of feature extraction, data compression, and noise reduction. Compared with the original data, the obtained PCT feature map has a significant improvement in defect significance. Similar to PCT, several extension works have been made for improvement of the defect detection performance, such as stable principal component pursuit (SPCP) [25], sparse principal component thermography (SPCT) [26, 27], and independent component thermography (ICT) [28].

Among them, SPCT is an improved algorithm for PCT, which can obtain the sparse principal components by applying L_1 constraint [29]. Indeed, the components containing few features will enhance the interpretability and visibility of the detection results. Moreover, an improved method called CCIPCT is proposed to use a shorter computational cost to estimate the covariance matrix and singular value decomposition (SVD) when calculating the principal component. The main advantage of CCIPCT is its faster performance in higher sequential acquisition [30]. However, PCT, SPCT, and CCIPCT only considered the correlations within the signal image and ignored the dynamics of heat transfer. The pulse thermal imaging defect detection method mainly determines the defect location based on the inconsistent heat transfer rates of different areas inside the object. Hence, there are strong cross-correlations among the adjacent temporal thermal images. In this paper, sparse moving window principal component thermal imaging (SMWPCT) is proposed to extract both cross-correlations from the temporal and spatial scales. For this purpose, the moving window strategy is utilized to cover several adjacent temporal thermal images within a period of time. The principal features of these moving windows are further extracted by sparse PCT, which will provide clearer and more interpretable detection results.

The rest of this paper is structured as follows. In the second section, the structure of the data collected from the pulse thermal imaging is briefly introduced. Then, the moving window strategy-based SPCT is proposed with the detailed algorithm in the third section. Next, the feasibility and effectiveness of the proposed method are demonstrated by the carbon fiber-reinforced plastic (CFRP) specimen. Finally, some conclusions are made.

2. Thermographic Data Preprocessing

The pulse thermal imaging data are collected as follows. Firstly, the tested object is firstly heated by the flash lamps using the pulse signals. After that, it is cooled naturally. During the whole heating and cooling period, the thermal images are acquired by the infrared camera, which is shown in Figure 1. The collected thermal data are the grayscale image and the pixel value of the image which represents the heating degree of the corresponding position. If there is a defect area within the object and the material is nonuniform, it will result in the discontinuity of heat conduction inside the object. Hence, the temperature of the defect area will be abnormally higher or lower than that of the surrounding area, which results that the pixel value of the defect area is different from that of the surrounding area on the thermal image.

For some obvious defects inside the testing object, it is also apparent to find out the defect areas directly from the original thermal images. However, it is born to be an unattractive work to check every image on visual observations. Also, some tiny defects are more difficult to distinguish by sight. On the contrary, the uneven heating will lead to uneven background, and the presence of measurement noise is inevitable, which further increases the detection difficulty. Finally, suppose there are n frames of thermal images have been collected, in which each image consists of $h \times w$ pixels. For most data analysis models, the three-dimensional data cannot be directly applied. Therefore, the thermal image data processing methods are needed to reduce the number of thermal images detected by vision and improve the significance of defects.

For the three-dimensional matrix $h \times w \times n$, the common data processing method is to expand the original data to a two-dimensional matrix. According to the chronological order, each thermal image can be converted to a row vector of length $h \times w$, which is shown in Figure 2. Then, a two-dimensional matrix with $n \times hw$ size can be obtained, in which each row represents the original thermal image data, and each element represents an image pixel value. Finally, the normalization of the processing data is made, which firstly subtracts the average value of each row and then divides it by the standard deviation.

3. Sparse Moving Window Principal Component Thermography

In this section, the traditional PCT method is firstly extended using the moving window strategy. After that, the sparse moving window principal component thermography (SMWPCT) method is proposed, which follows the detailed model parameter derivation. In PCT, the original thermal image data are projected along the maximum orthogonal direction, in which the principal components are extracted. Since PCT is built based on the preprocessing data two-

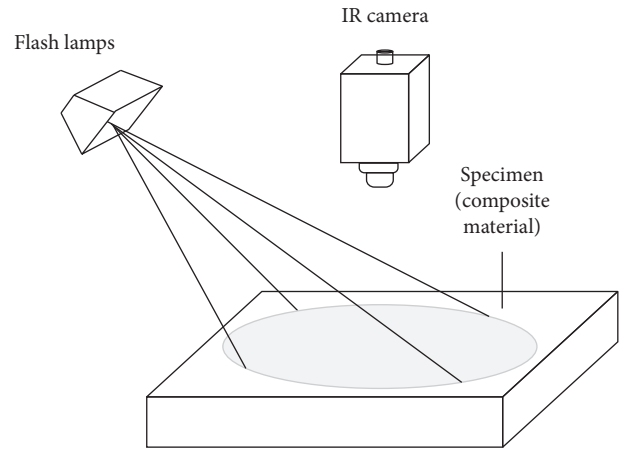


Figure 1: Thermal imaging data acquisition.

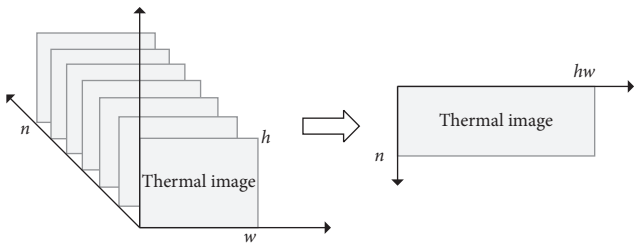


Figure 2: Three-dimensional expansion.

dimensional matrix, it indicates that the thermal image at a sampling interval is independent with its historical thermal values. When the sampling interval is large, such assumption is valid. However, the pulse thermography sampling interval is extremely short. Hence, it becomes necessary to consider both cross-correlations between the pixels in different regions of the single thermal image and the autocorrelations of pixels in the same region at different sampling intervals.

For this purpose, the moving window strategy is introduced before PCT is used, which is named as moving window PCT (MWPCT). In MWPCT, the augmented matrix containing several past values is constructed. Using the moving window, the time series correlations among the thermal images can be effectively extracted, and it is helpful to distinguish the defect region and the normal region. Assume that the preprocessed thermal image data matrix $\mathbf{X}$ contains n thermal images, and each image is composed of $m = h \times w$ pixel values. Hence, $\mathbf{X} \in R^{n \times m}$ can be expressed as follows:

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_n \end{bmatrix}, \tag{1}$$

where $\mathbf{x}_i \in \mathbf{R}^m (i = 1, 2, \ldots, n)$ is the m-dimensional row vector, and it indicates that a single thermal image with m pixel values.

Using the moving window strategy, the augmented matrix of the original observation $\mathbf{X}$ is constructed as follows:

$$\overline{\mathbf{X}} = \begin{bmatrix} \mathbf{x}_l & \mathbf{x}_{l-1} & \cdots & \mathbf{x}_1 \\ \mathbf{x}_{l+s} & \mathbf{x}_{l+s-1} & \cdots & \mathbf{x}_{s+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_n & \mathbf{x}_{n-1} & \cdots & \mathbf{x}_{n-l+1} \end{bmatrix}, \tag{2}$$

in which l is the moving window length and s is the moving window step. Usually, these two parameters make a trade-off between the model accuracy and calculation complexity. When the window size l is large and moving step s is small, the model accuracy will improve while the calculation complexity increases and vice versa. Based on the augmented matrix $\overline{\mathbf{X}}$, the first principal component (PC) is obtained by

$$\begin{aligned} &\max && \|\overline{\mathbf{X}}\mathbf{p}\|_2 \\ &\text{Subject to} && \|\mathbf{p}\|_2 \le 1, \end{aligned} \tag{3}$$

in which $\mathbf{p}$ is the principal eigenvector using the singular value decomposition, and it is also called the loading vector. $\|\cdot\|_2$ denotes the L_2 norm. The first PC can be calculated using $\mathbf{t} = \overline{\mathbf{X}}\mathbf{p}$, which can be treated as the linear combination of $\overline{\mathbf{X}}$. By calculating the first PC, the maximal correlated feature is obtained. Next, replacing $\overline{\mathbf{X}}$ using the model reconstruction error can obtain PCs orthogonal to each other in a similar manner [31]. Usually, the primary information of $\overline{\mathbf{X}}$ is mainly concentrated on several components. Therefore, few PCs will represent the original thermal images, and the defects can be visualized using these PCs. Assuming that the number of selected PCs is k, the size of the extracted feature matrix is $hwl \times k$. To visualize the final results, each PC can be reshaped into a two-dimensional matrix of $h \times w$. Hence, a total of $l \times k$ loading images can be generated, in order to achieve data compression and reduce the number of images to be observed. Due to the difference of features between defects and background, they are separated into different PCs and appear in different loading images. In addition, the measurement noise is retained in the residual subspace to achieve the purpose of noise suppression.

Even though MWPCT is able to extract both autocorrelations and cross-correlations in the 2-D expanded matrix, it still inherits some disadvantages of PCT. At first, each PC is linearly weighted by the original data, which means that all the elements in the loading vectors are usually nonzero, which brings out the difficulty and disturbance in final detection. To highlight the visual results and reduce the detection noise, a sparsity penalty can be added on the loadings to use few features and enhance the interpretability of the final results. Hence, the sparse moving window principal component thermography (SMWPCT) method is further developed next. The optimization problem of the first PC extracted by SMWPCT is usually expressed as follows:

$$\begin{aligned} &\max_{\widetilde{\mathbf{p}}} && \|\overline{\mathbf{X}}\widetilde{\mathbf{p}}\|_2^2 - \gamma\|\widetilde{\mathbf{p}}\|_0 \\ &\text{Subject to} && \|\widetilde{\mathbf{p}}\|_2 \le 1, \\ &\text{or } \max_{\widetilde{\mathbf{p}}} && \|\overline{\mathbf{X}}\widetilde{\mathbf{p}}\|_2 - \gamma\|\overline{\mathbf{X}}\widetilde{\mathbf{p}}\|_1 \\ &\text{Subject to} && \|\widetilde{\mathbf{p}}\|_2 \le 1, \end{aligned} \tag{4}$$

where γ is the tuning parameter that controls the sparsity of $\tilde{\mathbf{p}}$, $\|\cdot\|_0$ denotes the L_0 norm, which is calculated as the total number of nonzero elements in the vector and it represents the vector sparsity, and $\|\cdot\|_1$ denotes the L_1 norm, which represents the sum of the absolute values of the elements in the vector, and it is usually used to obtain the sparse result. Similar to MWPCT, the other PCs that are orthogonal to each other can be obtained repeatedly by performing the definition step and replacing $\overline{\mathbf{X}}$ with the current errors.

The sparse approach is able to select the most related PCs and restrict the feature variables that are not closely related to zero. In such manner, noise is further eliminated, and the dimension reduction becomes more condensed and easier to be interpreted. To make the model parameter estimation procedure more effective, the optimization problem can be transformed to be a Lasso regression problem; Zou et al. have introduced an L_2 norm regularization term in the objective function of the optimization function [32], which is given as

$$\begin{aligned} \min_{\mathbf{P},\boldsymbol{\Omega}} \quad & \left\|\overline{\mathbf{X}} - \overline{\mathbf{X}}\mathbf{P}\boldsymbol{\Psi}^T\right\|^2 + \gamma_1 \sum_{i=1}^{I} \left\|\boldsymbol{\psi}_i\right\|_2^2 + \gamma_2 \sum_{i=1}^{I} \left\|\boldsymbol{\psi}_i\right\|_1 \\ \text{Subject to} \quad & \mathbf{P}^{\mathbf{T}}\mathbf{P} = \mathrm{I}, \end{aligned} \tag{5}$$

where $\mathbf{P}$ is the loading matrix and $\boldsymbol{\Psi} = [\,\boldsymbol{\psi}_1 \;\; \boldsymbol{\psi}_2 \;\; \cdots \;\; \boldsymbol{\psi}_I\,]$ is a sparse approximation of $\mathbf{P}$ with I PCs. γ_1 and γ_2 are model tuning parameters. The combination of L_1 and L_2 constitutes an elastic net penalty and encouraging grouping which indicates that the variables of strong correlations will appear or disappear together [27].

Finally, the entire algorithms for the SMWPCT-based thermographic data analysis method are summarized as follows:

(1) Collect the thermal image data based on the pulsed thermography technology

(2) Rearrange the three-dimensional matrix to a two-dimensional form and normalize the measurement

(3) Select the appropriate moving window size and step size, and perform a moving window on the augmented matrix $\overline{\mathbf{X}}$

(4) Estimate the model parameter $\mathbf{P}$ and the sparse matrix $\mathbf{Q}$ by solving the optimization problem of equation (5)

(5) Reconstruct the thermal images based on $\mathbf{Q}$

(6) Observe the reconstructed thermal image to obtain the defect detection result

4. Case Study

In this section, a tested carbon fiber-reinforced polymer (CFRP) with subsurface defects is demonstrated to validate the feasibility of the proposed method. CFRP is obtained by pressing and drawing a plurality of continuous fibers and resin. It has been widely used in military, aerospace, racing, and other fields [9].

In this case, several Teflon strips were inserted into the fiberboard before the resin transfer molding to mimic the defective areas. In the tested CFRP board, there are totally ten defect areas, where the left bottom rectangle is a surface defect and the other nine rectangles are subsurface defects. The depth of these subsurface ones is different. Three defects on the left side were covered by one layer of fiber sheet while the middle column was beneath two layers and three on the right were covered by three layers. The thickness of each fiber sheet is about 0.26 cm. Besides, three kinds of defects are designed, where their sizes are $1.6\,\mathrm{cm} \times 1.6\,\mathrm{cm}$, $0.8\,\mathrm{cm} \times 0.8\,\mathrm{cm}$, and $0.4\,\mathrm{cm} \times 0.4\,\mathrm{cm}$, respectively (Figure 3).

To obtain the pulsed thermal imaging data, two 3000W flashes are used to heat the acquisition in the form of thermal pulse, and the heating time is about 3 ms. An infrared camera (TGS-G100EXD, NEC) is installed to collect thermal images in the reflection mode. The resolution of this camera is 320×240 pixels, and the sampling rate is 30 frames per second. A computer is connected to the camera for thermal imaging data processing. The process is cut to get the thermal images of the region of interest (ROI). Finally, there are 54 thermal images with 105×120 pixels which are collected. In Figure 4, the 1st, 10th, 20th, 30th, 40$^{\text{th}}$, and 50th thermal images of the thermal images are illustrated. The color bars indicate the pixel values. From the original thermal images, it can be seen that it is difficult to infer the defect locations owing to the nonuniform background and noises.

In order to evaluate the performance of the proposed SMWPCT method, several alternative methods PCT, SPCT, and MWPCT are also tested using the same thermal images. The experimental hardware platform is Window10, Intel i5-7500 CPU@3.40 GHz and 16 GBRAM. Also, the visibility of the defect area is measured by the signal-to-noise ratio (SNR) indicator, which refers to the ratio of signal to noise in an electronic device or system [33, 34]. In this paper, three kinds of SNR considered in [33, 34] are tested, which can be expressed as

$$\begin{aligned} \mathrm{SNR}_1 &= \frac{\left|\mathrm{M}_{\mathrm{def}} - \mathrm{M}_{\mathrm{n}}\right|}{\sigma_{\mathrm{n}}}, \\ \mathrm{SNR}_2 &= 10\log_{10}\frac{\left|\mathrm{M}_{\mathrm{def}} - \mathrm{M}_{\mathrm{n}}\right|^2}{\sigma_{\mathrm{n}}^2}, \\ \mathrm{SNR}_3 &= \frac{\left|\mathrm{M}_{\mathrm{def}} - \mathrm{M}_{\mathrm{n}}\right|}{\sqrt{\left(\sigma_{\mathrm{def}}^2 + \sigma_{\mathrm{n}}^2\right)/2}}, \end{aligned} \tag{6}$$

where $\mathrm{M}_{\mathrm{def}}$ is the average pixel value of the defect area and M_{n} is the average pixel value of the nondefect area. σ_{def} is the standard deviation of the pixel in the defect area and σ_{n} is the standard deviation of the pixel in the nondefect area. SNR is dimensionless, which reflects the contrast relationship between the defective area and the nondefective area. The larger the value of SNR, the more significant the defect.

The traditional PCT method is built based on all 54 thermal image samples. The detection results using the first six loading images of PCT are given in Figure 5(a). The loading values are represented by the red and blue color bars,

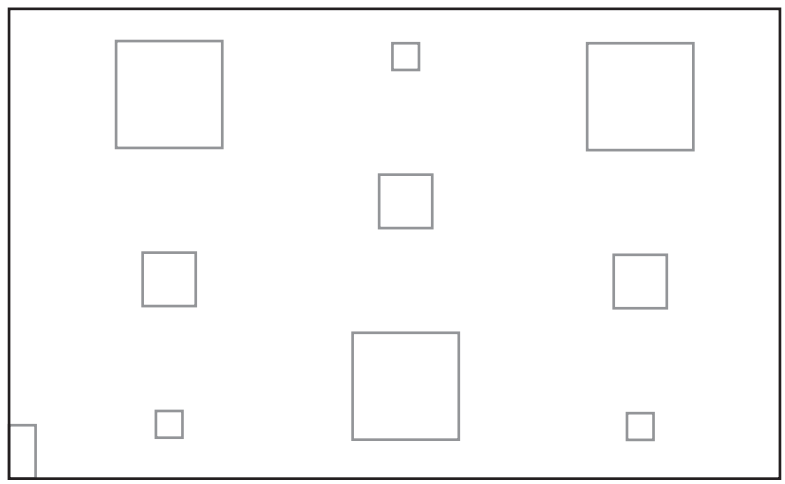


Figure 3: Illustration of defective regions.

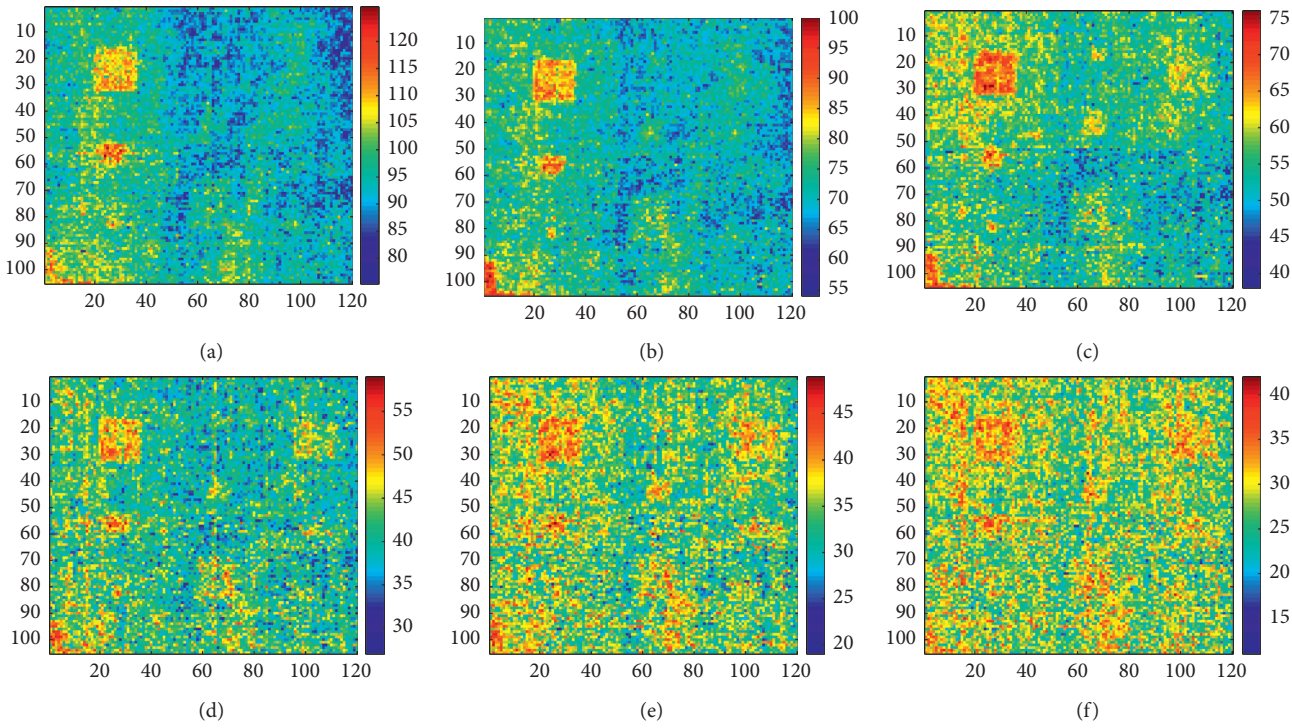
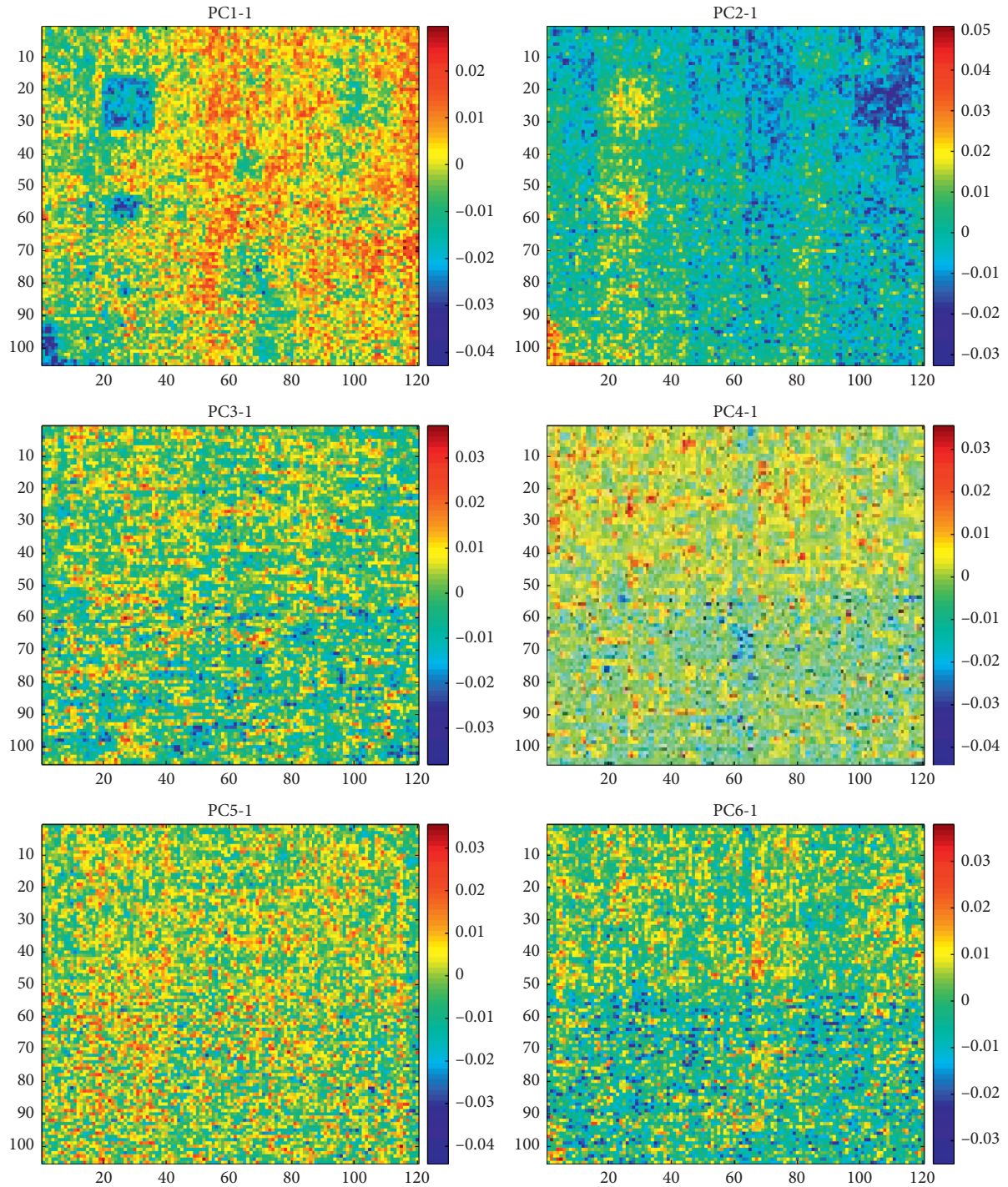


Figure 4: Illustration of defective regions. (a) Img-1. (b) Img-10. (c) Img-20. (d) Img-30. (e) Img-40. (f) Img-50.

in which the redder one indicates the larger value while the bluer one represents the lower value. From the final results, it can be seen that the PCT loading of the first several PCs can better reveal the position of the defect areas compared to the original thermal image. Since the heat transfer efficiency of the defect area is lower than that of the normal one, the loading value is smaller. And most of the defect information exists in the first loading image. However, PCT still has certain limitations. On the one hand, there are still more background or noise information in the first loading image, which cause some disturbances in defect detection and recognition. On the other hand, the defect information can still be observed more or less in the remaining loading images, which brings out difficulty on the selection of number of PCs.

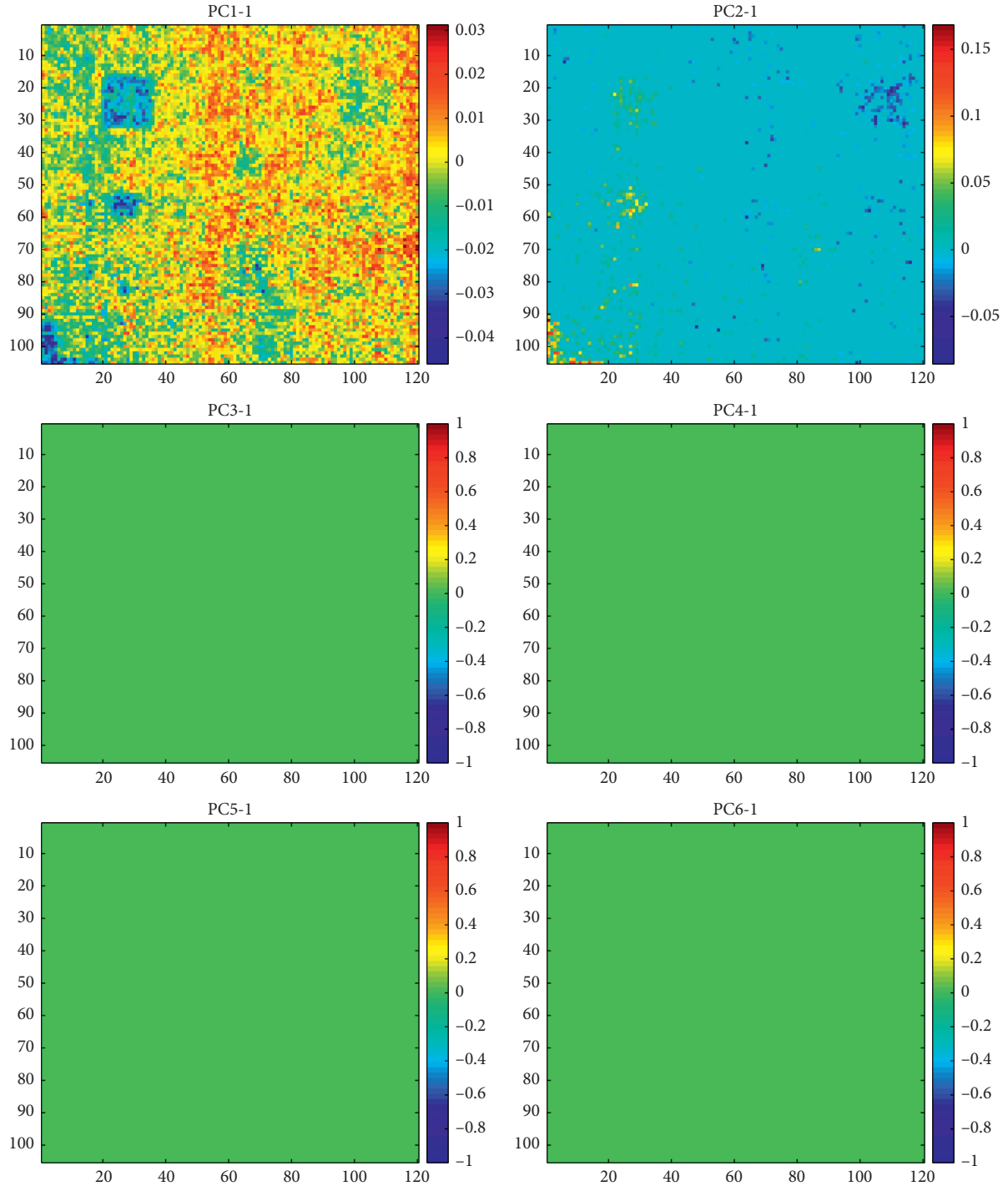
By applying sparsity penalties to the loadings, SPCT is able to reduce the background and noise interference of the result, which makes it more compact and interpretable.

In SPCT, the sparsity parameter is designed to penalize the loadings of different principal components, and it corresponds to an upper bound on the L1-norm of the BETA coefficients. Due to the nature of the L1 penalty, some coefficients will be shrunk to zero if the sparsity parameter is large enough. By contrast, SPCA results in a regular PCA when it equals to 0. Zou et al. [32] have specified the choice of penalty parameter, and it is set to 500 in this article. The



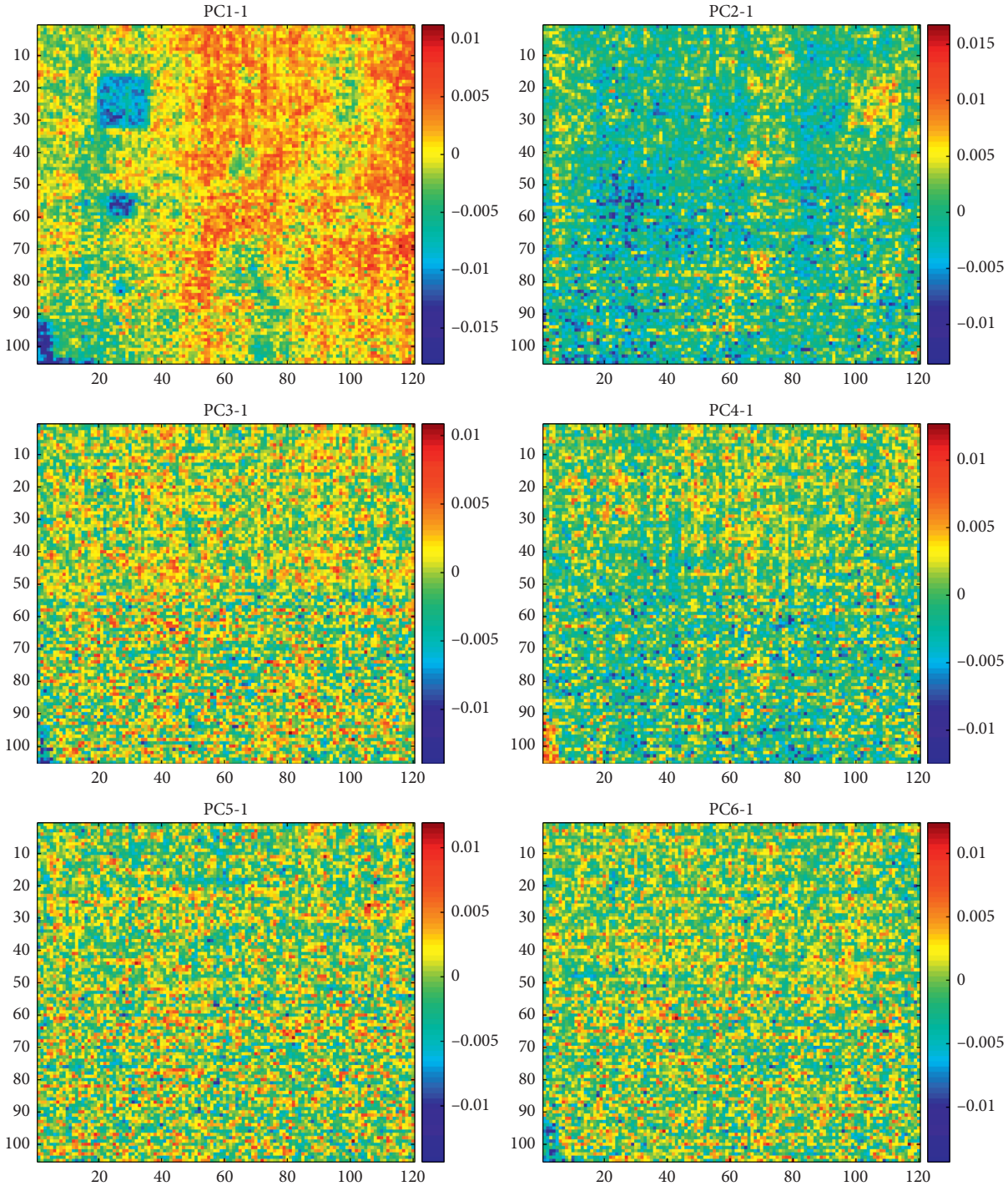
(a)

Figure 5: Continued.



(b)

Figure 5: Continued.



(c)

Figure 5: Continued.

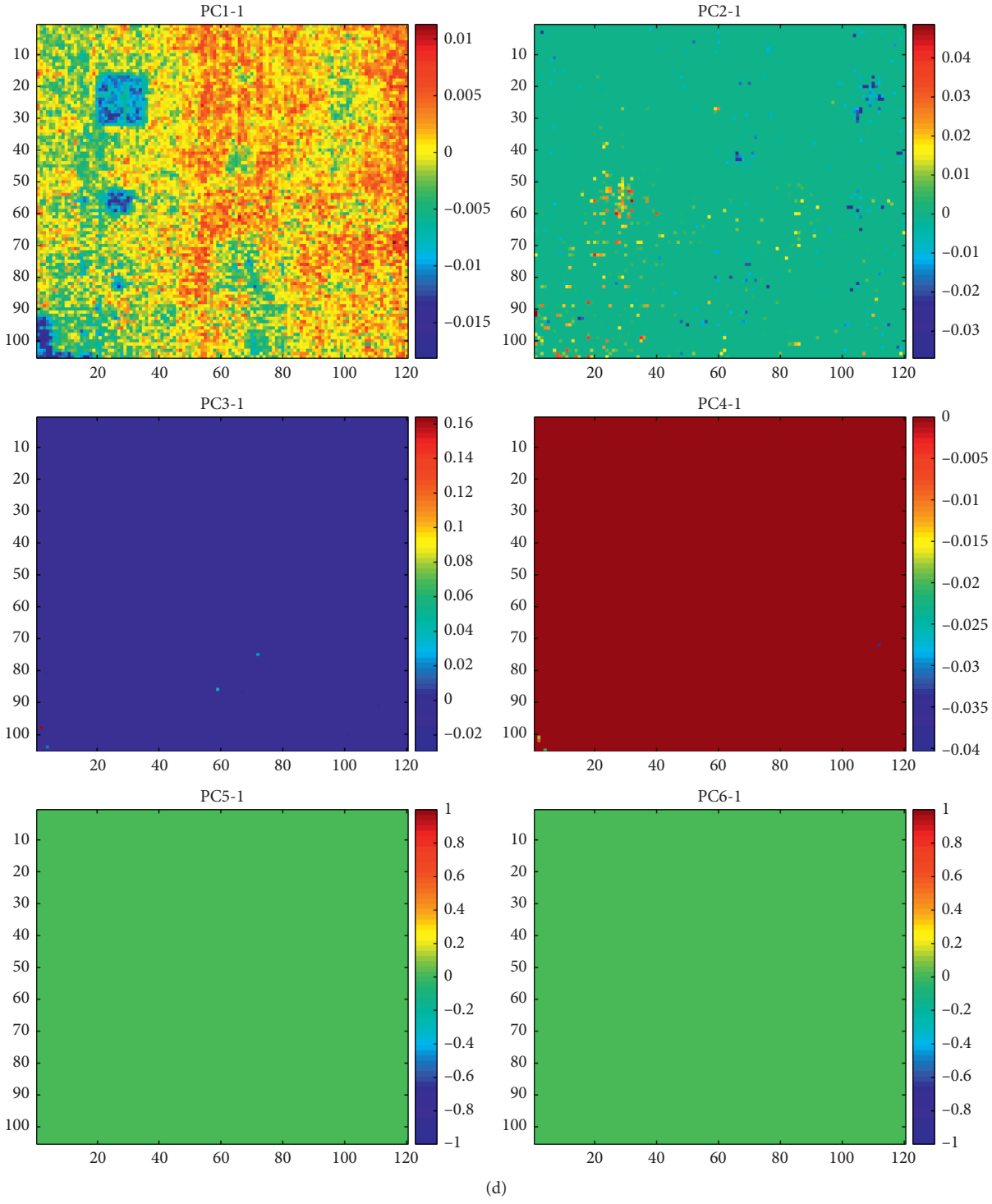


Figure 5: Results: (a) PCT, (b) SPCT, (c) MWPCT, and (d) SMWPCT.

results are shown in Figure 5(b). Compared with PCT, the loading image of SPCT is much sparse, in which the shallower defects are highlighted in the first loading image. It can be also seen that the deepest defect is better displayed in the second loading image. In the third and subsequent loading images, the loading image becomes very sparse due to much useful information retained. Therefore, the defect detection process for the thermal image can ignore these loading images, which also reduces the number of images that needs to be observed.

Next, MWPCT is compared with the moving window strategy. For the moving window technology, the window size and the step size are two model parameters to make a trade-off between the accuracy and computational load [35]. To select the proper values for these model parameters, several simulations have been made. Finally, the moving

Table 1: SNR_1 of different methods.

Methods/SNR_1/defect	Def_1	Def_2	Def_3	Def_4	Def_5	Def_6	Def_7	Sum
PCT	3.35	2.48	2.76	2.66	1.10	0.79	0.90	14.04
SPCT	3.67	2.70	3.01	2.89	1.21	0.88	0.98	15.32
MWPCT	4.32	2.94	3.19	3.11	1.14	0.77	0.94	16.42
SMWPCT	**4.45**	**3.21**	**3.62**	**3.37**	**1.39**	**0.99**	**1.11**	**18.15**

Table 2: SNR_2 of different methods.

Methods/SNR_2/defect	Def_1	Def_2	Def_3	Def_4	Def_5	Def_6	Def_7	Sum
PCT	10.50	7.88	8.81	8.48	0.85	−2.00	−0.87	39.39
SPCT	11.28	8.62	9.56	9.23	1.64	−1.15	−0.20	41.68
MWPCT	12.71	9.38	10.08	9.85	1.17	−2.27	−0.53	46.00
SMWPCT	**12.97**	**10.13**	**11.18**	**10.55**	**2.86**	**−0.05**	**0.90**	**48.65**

Table 3: SNR_3 of different methods.

Methods/SNR_3/defect	Def_1	Def_2	Def_3	Def_4	Def_5	Def_6	Def_7	Sum
PCT	1.84	1.36	1.52	1.46	0.61	0.44	0.50	7.73
SPCT	1.87	1.38	1.54	1.48	0.62	0.45	0.50	7.83
MWPCT	**2.08**	1.42	1.54	1.50	0.55	0.37	0.45	7.90
SMWPCT	2.00	**1.45**	**1.63**	**1.52**	**0.63**	**0.45**	**0.50**	**8.17**

window size is set to 7, and the step size is 5 in MWPCT in this case. The experiment result is shown in Figure 5(c). Compared to PCT, the defect areas in the first loading image of MWPCT are more prominent and are greatly reduced in the second loading image. It indicates that more accurate features have been captured using MWPCT. Besides, MWPCT concentrates the defect information in the first loading image so that we can only pay attention to the first loading image to identify the defect. Due to the inclusion of more defect information, the defect area is significantly improved, and the image appears that the defect area is darker in color.

Finally, the proposed SMWPCT method is tested, which combines the advantages of SPCT and MWPCT. It extracts the defect features based on the moving window strategy with sparsely loading images. The detection results are shown in Figure 5(d). It can be seen that both the concentration of defect features and the sparseness of irrelevant information are reflected in SMWPCT, and the saliency of defect areas has also been improved. The result is also quite reasonable since the proposed SMWPCT method extracts both cross-correlations from the temporal and spatial scales, and more accurate and compact features are obtained. Based on that, the defect detection procedure becomes straightforward and easy to be operated.

Moreover, the defect visibility is quantified by the signal-to-noise ratio indicator. Since the three defects in the last column are covered by three layers of fiberboard, when the heat conduction reaches the thickness of three-layer fiberboard, the loss of energy leads to poor visibility of the thermal image. Therefore, the calculation of the SNR index does not include the three deepest defects. Three kinds of different SNR values based on PCT, SPCT, MWPCT, and SMWPCT are listed in Tables 1–3. It can be seen that the performance of the proposed SMWPCT is superior to several alternatives in most cases even based on different SNR indicators. Also, the results have verified the effectiveness and reliability of SMWPCT.

5. Conclusion

Due to the existence of noises and nonuniform background, the artificial detection of the defects inside the composite products is difficult to achieve. Hence, it is necessary to apply the data analysis approaches based on the thermal imaging data. In this paper, a SMWPCT method is proposed, which combines SPCT with the moving window strategy. In SMWPCT, both dynamic and static information of the thermal imaging data can be captured under the sparse structure. The experimental results have illustrated the effectiveness of the proposed method. It reveals that not only the correlations between the pixels in different regions of the single thermal image but also the correlations of pixels in the same region at different sampling intervals are extracted in SMWPCT. Hence, the performance of the proposed method is superior to several alternatives.

Data Availability

The data used to support the findings of this study cannot be made freely available. Requests for access to these data should be made to Dr. Zhou (zhoule@zust.edu.cn).

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the NSFC-Zhejiang Joint Fund for the Integration of Industrialization and Informatization (U1609214), the National Natural Science Foundation of China (61603342), the Zhejiang Provincial Natural Science Foundation of China (LY19F030003), and Talent Project of Zhejiang Association for Science and Technology (no. 2018YCGC019).

References

[1] C. Ibarra-Castanedo, J. R. Tarpani, X. P. V. Maldague, and X. P. V. Maldague, "Nondestructive testing with thermography," *European Journal of Physics*, vol. 34, no. 6, pp. S91–S109, 2013.

[2] S. M. Shepard, "Introduction to active thermography for non-destructive evaluation," *Anti-Corrosion Methods and Materials*, vol. 44, no. 4, pp. 236–239, 1997.

[3] S. K. Lau, D. P. Almond, and J. M. Milne, "A quantitative analysis of pulsed video thermography," *NDT & E International*, vol. 24, no. 4, pp. 195–202, 1991.

[4] C. C. H. Guyott, P. Cawley, and R. D. Adams, "The non-destructive testing of adhesively bonded structure: a review," *The Journal of Adhesion*, vol. 20, no. 2, pp. 129–159, 1986.

[5] C. Garnier, M.-L. Pastor, F. Eyma, and B. Lorrain, "The detection of aeronautical defects in situ on composite structures using Non Destructive Testing," *Composite Structures*, vol. 93, no. 5, pp. 1328–1336, 2011.

[6] Y. Liu, K. Liu, Z. Gao et al., "Non-destructive defect evaluation of polymer composites via thermographic data analysis: a manifold learning method," *Infrared Physics & Technology*, vol. 97, pp. 300–308, 2019.

[7] S. M. Shepard, J. R. Lhota, B. A. Rubadeux et al., "Reconstruction and enhancement of active thermographic image sequences," *Optical Engineering*, vol. 42, no. 5, pp. 1337–1343, 2003.

[8] M. Pilla, M. Klein, X. Maldague et al., "New absolute contrast for pulsed thermography," *Proceedings of the QIRT*, vol. 5, pp. 53–58, 2002.

[9] K. Zheng, Y.-S. Chang, K.-H. Wang, and Y. Yao, "Improved non-destructive testing of carbon fiber reinforced polymer (CFRP) composites using pulsed thermograph," *Polymer Testing*, vol. 46, pp. 26–32, 2015.

[10] K. Zheng, Y.-S. Chang, and Y. Yao, "Defect detection in CFRP structures using pulsed thermographic data enhanced by penalized least squares methods," *Composites Part B: Engineering*, vol. 79, pp. 351–358, 2015.

[11] Y.-S. Chang, Z. Yan, K.-H. Wang, and Y. Yao, "Non-destructive testing of CFRP using pulsed thermography and multi-dimensional ensemble empirical mode decomposition," *Journal of the Taiwan Institute of Chemical Engineers*, vol. 61, pp. 54–63, 2016.

[12] L. Zhou, J. Zheng, Z. Ge, Z. Song, and S. Shan, "Multimode process monitoring based on switching autoregressive dynamic latent variable model," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 10, pp. 8184–8194, 2018.

[13] L. Yao and Z. Ge, "Deep learning of semi-supervised process data with hierarchical extreme learning machine and soft sensor application," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 2, pp. 1490–1498, 2017.

[14] X. Yuan, B. Huang, Y. Wang, C. Yang, and W. Gui, "Deep learning-based feature representation and its application for soft sensor modeling with variable-wise weighted SAE," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 7, pp. 3235–3243, 2018.

[15] L. Zhou, Y. Wang, Z. Ge, and Z. Song, "Multirate factor Analysis models for fault detection in multirate processes," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 7, pp. 4076–4085, 2019.

[16] Y. Liu, C. Yang, Z. Gao, and Y. Yao, "Ensemble deep kernel learning with application to quality prediction in industrial polymerization processes," *Chemometrics and Intelligent Laboratory Systems*, vol. 174, pp. 15–21, 2018.

[17] Q. Xuan, Z. Chen, Y. Liu et al., "Multiview generative adversarial network and its application in pearl classification," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 10, pp. 8244–8252, 2018.

[18] Z. Ge, Z. Song, S. X. Ding, and B. Huang, "Data mining and analytics in the process industry: the role of machine learning," *IEEE Access*, vol. 5, pp. 20590–20616, 2017.

[19] L. Zhou, G. Li, Z. Song et al., "Autoregressive dynamic latent variable models for process monitoring," *IEEE Transactions on Control Systems Technology*, vol. 25, no. 1, pp. 366–373, 2016.

[20] Y. Liu and Z. Gao, "Real-time property prediction for an industrial rubber-mixing process with probabilistic ensemble Gaussian process regression models," *Journal of Applied Polymer Science*, vol. 132, no. 6, 2015.

[21] F. J. Madruga, C. Ibarra-Castanedo, O. M. Conde, J. M. López-Higuera, and X. Maldague, "Infrared thermography processing based on higher-order statistics," *NDT & E International*, vol. 43, no. 8, pp. 661–666, 2010.

[22] X. Maldague, F. Galmiche, and A. Ziadi, "Advances in pulsed phase thermography," *Infrared Physics & Technology*, vol. 43, no. 3–5, pp. 175–181, 2002.

[23] N. Rajic, "Principal component thermography for flaw contrast enhancement and flaw depth characterisation in composite structures," *Composite Structures*, vol. 58, no. 4, pp. 521–528, 2002.

[24] I. T. Jolliffe, "Principal component analysis," *Journal of Marketing Research*, vol. 25, no. 4, p. 513, 2002.

[25] Z. Yan, C.-Y. Chen, L. Luo, and Y. Yao, "Stable principal component pursuit-based thermographic data analysis for defect detection in polymer composites," *Journal of Process Control*, vol. 49, pp. 36–44, 2017.

[26] Y. Yao, S. Sfarra, S. Lagüela et al., "Active thermography testing and data analysis for the state of conservation of panel paintings," *International Journal of Thermal Sciences*, vol. 126, pp. 143–151, 2018.

[27] J.-Y. Wu, S. Sfarra, and Y. Yao, "Sparse principal component thermography for subsurface defect detection in composite products," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 12, pp. 5594–5600, 2018.

[28] Y. Liu, J. Y. Wu, K. Liu et al., "Independent component thermography for non-destructive testing of defects in polymer composites," *Measurement Science and Technology*, vol. 30, no. 4, Article ID 044006, 2019.

[29] T. T. Cai, Z. Ma, and Y. Wu, "Sparse PCA: optimal rates and adaptive estimation," *The Annals of Statistics*, vol. 41, no. 6, pp. 3074–3110, 2013.

[30] B. Yousefi, S. Sfarra, C. Ibarra Castanedo, and X. P. V. Maldague, "Comparative analysis on thermal non-destructive testing imagery applying candid covariance-free incremental principal component thermography (CCIPCT)," *Infrared Physics & Technology*, vol. 85, pp. 163–169, 2017.

[31] L. W. Mackey, "Deflation methods for sparse PCA," *Advances in Neural Information Processing Systems*, vol. 10, pp. 1017–1024, 2009.

[32] H. Zou, T. Hastie, and R. Tibshirani, "Sparse principal component analysis," *Journal of Computational and Graphical Statistics*, vol. 15, no. 2, pp. 265–286, 2006.

[33] C. Ibarra-Castanedo, J.-M. Piau, S. Guilbert et al., "Comparative study of active thermography techniques for the nondestructive evaluation of honeycomb structures," *Research in Nondestructive Evaluation*, vol. 20, no. 1, pp. 1–31, 2009.

[34] R. Usamentiaga, C. Ibarra-Castanedo, and X. Maldague, "More than fifty shades of grey: quantitative characterization of defects and interpretation using snr and cnr," *Journal of Nondestructive Evaluation*, vol. 37, no. 2, p. 25, 2018.

[35] X. Wang, U. Kruger, and G. W. Irwin, "Process monitoring approach using fast moving window PCA," *Industrial & Engineering Chemistry Research*, vol. 44, no. 15, pp. 5691–5702, 2005.

Hindawi
Advances in Polymer Technology
Volume 2020, Article ID 6981302, 8 pages
https://doi.org/10.1155/2020/6981302

Research Article

Semi-Supervised Hybrid Local Kernel Regression for Soft Sensor Modelling of Rubber-Mixing Process

Haiqing Yu,[1] Jun Ji ,[2,3] Ping Li,[1] Fengjing Shao ,[2] Shunyao Wu,[2] Yi Sui,[2] Shujing Li,[2] Fengjiao He,[4] and Jinming Liu[2,3]

[1]*State Key Laboratory of Industrial Control Technology, College of Control Science and Engineering, Zhejiang University, Hangzhou, China*
[2]*College of Computer Science and Technology, Qingdao University, Qingdao, China*
[3]*Beijing Wanling Pangu Science and Technology Ltd., Beijing, China*
[4]*XiangYa School of Medicine, Central South University, Changsha, China*

Correspondence should be addressed to Jun Ji; junji@qdu.edu.cn

Received 11 July 2019; Accepted 26 September 2019; Published 31 January 2020

Guest Editor: Yuan Yao

Soft sensor techniques have been widely adopted in chemical industry to estimate important indices that cannot be online measured by hardware sensors. Unfortunately, due to the instinct time-variation, the small-sample condition and the uncertainty caused by the drifting of raw materials, it is exceedingly difficult to model the fed-batch processes, for instance, rubber internal mixing processing. Meanwhile, traditional global learning algorithms suffer from the outdated samples while online learning algorithms lack practicality since too many labelled samples of current batch are required to build the soft sensor. In this paper, semi-supervised hybrid local kernel regression (SHLKR) is presented to leverage both historical and online samples to semi-supervised model the soft sensor using proposed time-windows series. Moreover, the recursive formulas are deduced to improve its adaptability and feasibility. Additionally, the rubber Mooney soft sensor of internal mixing processing is implemented using real onsite data to validate proposed method. Compared with classical algorithms, the performance of SHLKR is evaluated and the contribution of unlabelled samples is discussed.

1. Introduction

Fed-batch processes play an important role in chemical and biochemical industry. They are widely adopted in the production of a vast range of fermentation-derived products such as fine-chemical industry, pharmaceuticals and food products. Rubber internal mixing [1] is a classical fed-batch process performed in an internal mixer to achieve an optimal Mooney viscosity for further processing. Since Mooney viscosity cannot be online measured while its laboratory assay is labour-intensive and time-consuming, soft-sensing approaches are investigated to establish a real-time evaluation of it. Furthermore, data-driven but not mechanism-modelling methods are commonly used for its soft sensor modelling because it is a complex nonlinear process without well-developed mechanism. Additionally, its instinctive time-variation, varying properties of natural rubber and additives accompanied with process drifting caused by field conditions. e.g., equipment aging, introduce a great amount of complexity to the process. Moreover, in order to avoid affecting the regular productions, small sample condition always occurred, which further reinforces the difficulty of rubber internal mixing modelling.

In the past decades, many data-driven techniques have been proposed. Extensive reviews can be found in work of Kadlec [2]. Among these methods, multivariate static techniques [3–6] have been widely used. However, these algorithms are relatively sensitive to measurement noise and commonly require a large number of samples to build the promising soft sensor as well. Meanwhile, various artificial neural network (ANN) algorithms [7] have been proposed and successfully applied to polymerization processes, but how to effectively construct the network topology is still an open question. To overcome these shortcomings, kernel-based methods, such as support vector regression [8], least squares support vector regression [9] are presented. These kernel

techniques can attain a better performance under small-sample condition owing to the structural risk minimization criterion.

Note that all the aforementioned algorithms are offline approaches, which can achieve a universal generalization performance but lack the mechanisms to leverage the time-variation characteristics such as drifting of the processes. So, kernel based online modelling algorithms [10–13] were presented. However, too many labelled samples of current batch are required to online build the model, while in most cases in industry field, those samples also have to be predicted instead of lab assay.

Therefore, both online and offline algorithms cannot effectively achieve the promising model [14–17]. On the other hand, taking advantage of the development of both information technology and industrial automation, there are lots of historical productive process data saved in the database of manufacturing execution system [18]. To leverage those data, local learning modelling algorithms [19, 20] were proposed. Nevertheless, those models are not stable owing to the outdated data, which would be used for training. Meanwhile, the unlabelled data are abundant, which contain the production data without indices to be predicted. According to the semi-supervised learning theory, those unlabelled data can be potentially used to improve the predictive model. Therefore, how to effectively leverage both existing historical and online productive process data to create the robust soft sensing model still need to be solved.

In our work, we explore the potential of the hybrid local semi-supervised mechanism to leverage both unlabelled and labelled data via the proposed time window mixed with both historical and online samples. To enhance its feasibility, corresponding recursive calculation formulas are deducted. Furthermore, the soft sensors using proposed and comparative algorithm are implemented to evaluate its performance. To the best of our knowledge, there is no such hybrid local semi-supervised algorithm presented in any article so far.

The remainder of this paper is organized as follows. In Section 2, the detail of proposed SHLKR method, including its recursive calculation derivation is presented. In Section 3, soft sensor modelling experiments of rubber internal mixing process using SHLKR method and comparative algorithms with real industrial field data are presented. Finally, in Section 4, the main contribution of this paper is summarized.

2. Materials and Methods

The thinking of local learning is to create the predictive model dedicated to the prediction of targeted unlabelled sample instead of building the global model using all samples. Since the model will only be created when the prediction is needed, it is also called "Just-in-time learning" or lazy learning [21]. Theoretically it can get more precise model under the condition that similar inputs lead to similar outputs.

Basically, there are three steps of the local learning modelling:

(1) Similar sample set selection: select similar samples from historical data based on one or some similarity calculation algorithms according to the features of the samples to be predicted.

(2) Local modelling: build the local learning model using selected samples with corresponding algorithm.

(3) Prediction: make the prediction and desert the predictive model.

Obviously, the key points of local learning are the algorithms to evaluate the similarity of samples and build the local model. Currently there are two categories that correlation based [19] and distance/angle based [10] similarity calculation algorithms. In this work, distance-based kernel is used because simply algorithm prone to be adopted under industrial application circumstances.

There are two major disadvantages of aforementioned local learning algorithm:

(1) In many cases the online time variation and drifting characteristics cannot be tracked since only similar historical data will be used for the modelling.

(2) Many unlabelled historical and online samples are orderly existed between labelled samples. Those time-series sequence data theoretically can be used to improve the model based on the manifold hypothesis [22] but currently leave unused.

In order to leverage those unused widely existed unlabelled data, we proposed recursive weighted kernel regression (RWKR) [23] before, which has already been validated in penicillin production process soft sensor modelling. But it behaves not promising for some other fed-batch processes, such as rubber internal mixing, since it behaves much more drifting and the time-based weighting mechanism does not work since the Mooney viscosity of rubber is not monotonic increased as the penicillin concentration in penicillin fermentation process. Therefore, in this paper, semi-supervised hybrid local kernel regression (SHLKR) is proposed to fully leverage both labelled and unlabelled data selected from historical and online data.

Different from traditional local kernel learning algorithms:

(1) Besides of labelled samples, combined with labelled samples, unlabelled samples are also used as time window during the training of SHLKR.

(2) Both historical data and online manufactural data are used during training. According to the current run's index of batch, hybrid training data set is formed by selecting corresponding historical samples joined with online manufactural samples, which can potentially improve the practicability and precision of the soft sensor.

2.1. SHLKR Flow. As is shown in Figure 1, the time window is defined as run's labelled sample $\left(x_{l_t}, y_{l_t}\right)$ with $\boldsymbol{X}_{u_t} = \left(x_{l_{t-1}}, y_{l_{t-u_t-1}}\right)$ which is the u_t unlabeled sequence samples between t and $t-1$ of current batch. In this way, each labeled sample associated

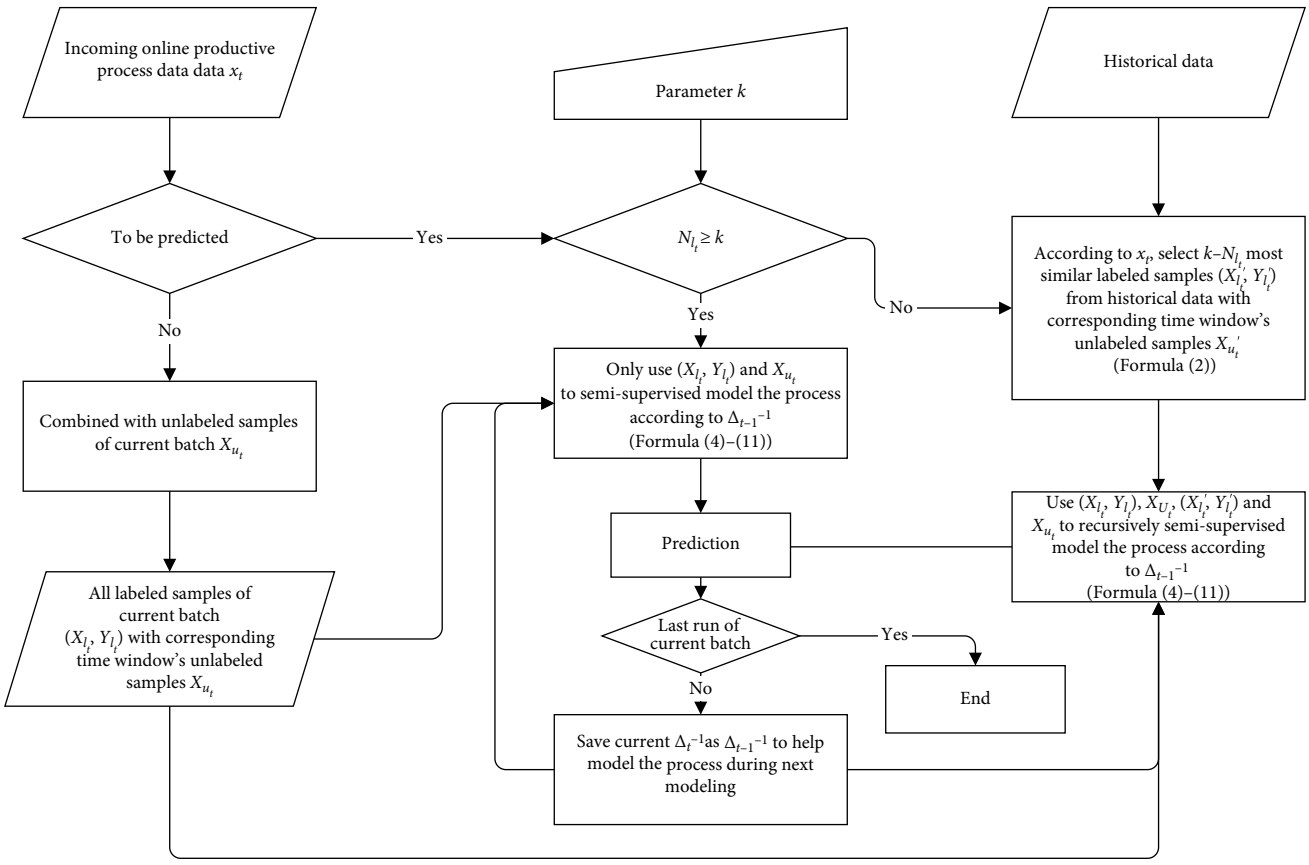


Figure 1: SHLKR flow chart.

with its u_t unlabeled samples is formed as an ordered sequence, which will be entirely used to semi-supervised model the soft sensor. According to the manifold hypothesis of semi-supervised learning theory [24–27], samples are trend to be similar within a small local space, unlabelled samples make the data space denser to more precisely describe the characteristic of data samples. So theoretically proposed semi-supervised data combination mechanism can more effectively model the soft sensor than only using labelled samples.

From the first run of first batch, the number of current labelled sample N_{l_t} is 0. If productive process data of current run x_t will only be collected for modelling in future, it will be added into the unlabelled sample set $\boldsymbol{X}_{u_t}$ of current batch, otherwise, since at this time only historical data can be used for modeling, evaluated by the similarity with x_t, k most similar historical labelled samples $\left(x'_{l_t}, y'_{l_t}\right)$ associated with the unlabeled samples $\boldsymbol{X}\prime_{u_t}$ within corresponding time windows are selected to semi-supervised train the model. On the other hand, if there are labelled samples $\left(\boldsymbol{X}_{l_t}, y_{l_t}\right)$ existing, they and associated unlabeled samples $\boldsymbol{X}_{u_t}$ will be both leveraged for training, in this case, if $N_{l_t} \geq k$, only online productive process data will be used, otherwise, $k - N_{l_t}$ most similar historical labeled samples and corresponding unlabeled samples will also be used to train the model.

2.2. SHLKR Recursive Calculation Derivation. Harmonic function is adapted to semi-supervised train the model. Its effectiveness and recursion have been validated before [23]. Although the historical data of training set cannot be recursively adopted since they depend on the x_t remaining online productive process data can be recursively added because all of them will be used for training. The larger t becomes, the more reduction it will have from following recursive calculation derivation.

Here we referred to the approach presented by Zhu et al. [28], in which the regularization framework is defined as follows:

$$\boldsymbol{Y}^* = \arg\min_{Y}\left\{\sum_{i,j=1}^{n} w_{ij}\left\|Y_i - Y_j\right\|^2 + \infty\sum_{i=1}^{l}\left\|Y_i - y_i\right\|^2\right\}, \quad (1)$$

where y_i is the real label of sample i, and ω_{ij} can be treated as the similarity between sample i and j, since Gaussian kernel is usually used to calculate the similarity, ω_{ij} is typically defined as

$$w_{ij} = K\left(x_i, x_j\right) = \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\sigma^2}\right). \quad (2)$$

Gram matrix $\boldsymbol{W}$ can be partitioned into 4 blocks for labelled samples L and unlabelled samples U:

$$\boldsymbol{W} = \begin{bmatrix} \boldsymbol{W}_{LL} & \boldsymbol{W}_{LU} \\ \boldsymbol{W}_{UL} & \boldsymbol{W}_{UU} \end{bmatrix}. \quad (3)$$

Then the solution of Equation (1) is formulated as:

$$\boldsymbol{Y}_U = \left(\boldsymbol{D}_{UU} - \boldsymbol{W}_{UU}\right)^{-1}\boldsymbol{W}_{UL}\boldsymbol{Y}_L, \tag{4}$$

$$\Delta_t^{-1} = \left(\boldsymbol{D}_{UU_t} - \boldsymbol{W}_{UU_t}\right)^{-1}, \tag{5}$$

$$\boldsymbol{D}_{UU_t} = \begin{bmatrix} \sum_j w_{(l'+1)j} & & \\ & \ddots & \\ & & \sum_j w_{(l'+n')(l'+n')} \end{bmatrix}, \tag{6}$$

$$\boldsymbol{W}_{UU_t} = \begin{bmatrix} w_{(l'+1)1} & \cdots & w_{(l'+1)(l'+n')} \\ \vdots & \ddots & \vdots \\ w_{(l'+1)(l'+n')} & \cdots & w_{(l'+n')(l'+n')} \end{bmatrix}, \tag{7}$$

here Δ_t^{-1} can also be divided into four parts:

$$\Delta_t^{-1} = \begin{bmatrix} \Delta_{OO_t} & \Delta_{OH_t} \\ \Delta_{HO_t} & \Delta_{HH_t} \end{bmatrix}^{-1}, \tag{8}$$

where Δ_{OH_t} is the kernel matrix between onlinemanufactural data and historical data of time t. Δ_{HO_t} is its transpose. Δ_{OO_t} and Δ_{HH_t} are the kernel matrixes of online manufactural data and historical data respectively. First the Δ_{OO_t} is considered as follows:

$$\Delta_{OO_t} = \begin{bmatrix} \Delta_{OO_{t-1}}{}' & \boldsymbol{p}^{\mathrm{T}} \\ \boldsymbol{p} & b \end{bmatrix}. \tag{9}$$

Here $\boldsymbol{p} = \left[-K\left(x_{u_1}, x_{u_t}\right), \ldots, -K\left(x_{u_{t-1}}, x_{u_t}\right)\right]$, $b = K\left(x_{u_1}, x_{u_1}\right) + \cdots + K\left(x_{u_1}, x_{u_{t-1}}\right)$ and:

$$\Delta_{OO_{t-1}}{}' = \Delta_{OO_{t-1}} + \begin{bmatrix} K\left(x_{u_1}, x_{u_{t-1}}\right) & & \\ & \ddots & \\ & & K\left(x_{u_{t-1}}, x_{u_{t-1}}\right) \end{bmatrix}. \tag{10}$$

Apply Sherman–Morrison–Woodbury to formula, then we get:

$$\Delta_{OO_t}{}^{-1} = \begin{bmatrix} \dfrac{\Delta_{OO_{t-1}}{}'^{-1} + \Delta_{OO_{t-1}}{}'^{-1}\boldsymbol{p}^T\boldsymbol{p}\Delta_{OO_{t-1}}{}'^{-1}}{b - \boldsymbol{p}\Delta_{OO_{t-1}}{}'^{-1}\boldsymbol{p}^T} & \dfrac{-\Delta_{OO_{t-1}}{}'^{-1}\boldsymbol{p}^T}{b - \boldsymbol{p}\Delta_{OO_{t-1}}{}'^{-1}\boldsymbol{p}^T} \\ -\dfrac{\boldsymbol{p}\Delta_{OO_{t-1}}{}'^{-1}}{b - \boldsymbol{p}\Delta_{OO_{t-1}}{}'^{-1}\boldsymbol{p}^T} & \dfrac{1}{b - \boldsymbol{p}\Delta_{OO_{t-1}}{}'^{-1}\boldsymbol{p}^T} \end{bmatrix}. \tag{11}$$

Then the $\Delta_{OO_t}{}^{-1}$ can be recursively calculated by $\Delta_{OO_{t-1}}{}'^{-1}$.

2.3. Application System. Smart Internal Mixing system is a product of MESNAC Co., Ltd., which is widely used in many rubber factories in China. It is mainly formed by four parts: internal mixing modelling, Mooney viscosity prediction, internal mixing process optimization and internal mixing expert system. As is shown in following Figure 2, Smart Internal Mixing system is embedded in the manufacturing execution system, which can monitor the online manufactural data and retrieve the historical manufactural data.

2.4. Experimental Data. Authorized by one rubber manufactory, 222 batches containing 19,148 runs historical samples were retrieved from the system. 2,140 of them were labelled and 17,008 runs are unlabelled which only contain manufactural information without Mooney viscosity value. All samples are from one rubber internal mixing formula to get rid of the formula variation impact. In the industrial application environment, to get the better performance, it also works to model the soft sensor respectively according to different rubber internal mixing formulas. Each sample includes:

(1) Index of current run.
(2) Density.
(3) Hardness.
(4) Minimum torque.
(5) Maximum torque.
(6) Elapsed time to reach 30% maximum torque.
(7) Elapsed time to reach 60% maximum torque.
(8) Elapsed time to increase 2 units after reaching minimum torque.

For labelled samples, all Mooney viscosity values were manually lab assayed. The Mooney viscosity values of first 10 batches are shown in Figure 3, the Mooney viscosity value of unlabelled samples are 0, the dash lines are used to separate different batches. Obviously, the run number of each batch changes a lot owing to its industrial manufactural requirement and the lab assay is performed generally every 8 runs. Besides of that, although the Mooney viscosity is required to be consisted, but the truth is it varies a lot within and between different batches under no obvious rules. It verified our hypothesis that data driven algorithms work in this situation to train the soft sensor.

3. Result and Discussion

To validate the performance of SHLKR, support vector machine (SVM) and Harmonic Functions based soft sensors are also implemented respectively to make the comparison, in which only labelled samples are used. To be faired, all these three algorithms are using the same labelled samples and only the unlabelled samples respective to those labelled samples are additionally used in SHLKR.

As is shown in Figure 4, the predictive results of all three different algorithms are plotted. The result is for last 27 of 222 batches as well as 1,777 of 19,148 runs including 1,577 unlabelled runs and 200 runs to be predicted. In order to predict those 200 samples, both 1,940 labelled and 15,431 unlabelled samples are used to train the soft sensor.

At the first step of training is to choose the parameter k. After the kernel width 1.1 is determined by leave-one-out cross validation [29], from 2 to 20, the results of using different k are shown in Figures 5(a)–5(c).

Because SVM cannot be resolved when $k < 8$, only SHLKR and Harmonic Functions have results shown in those figures. Obviously when $k = 5$, both of them have the best performance, when $k < 5$ they both behave unstably and when $k > 5$ they all trend to worse but stably. It means that: since k

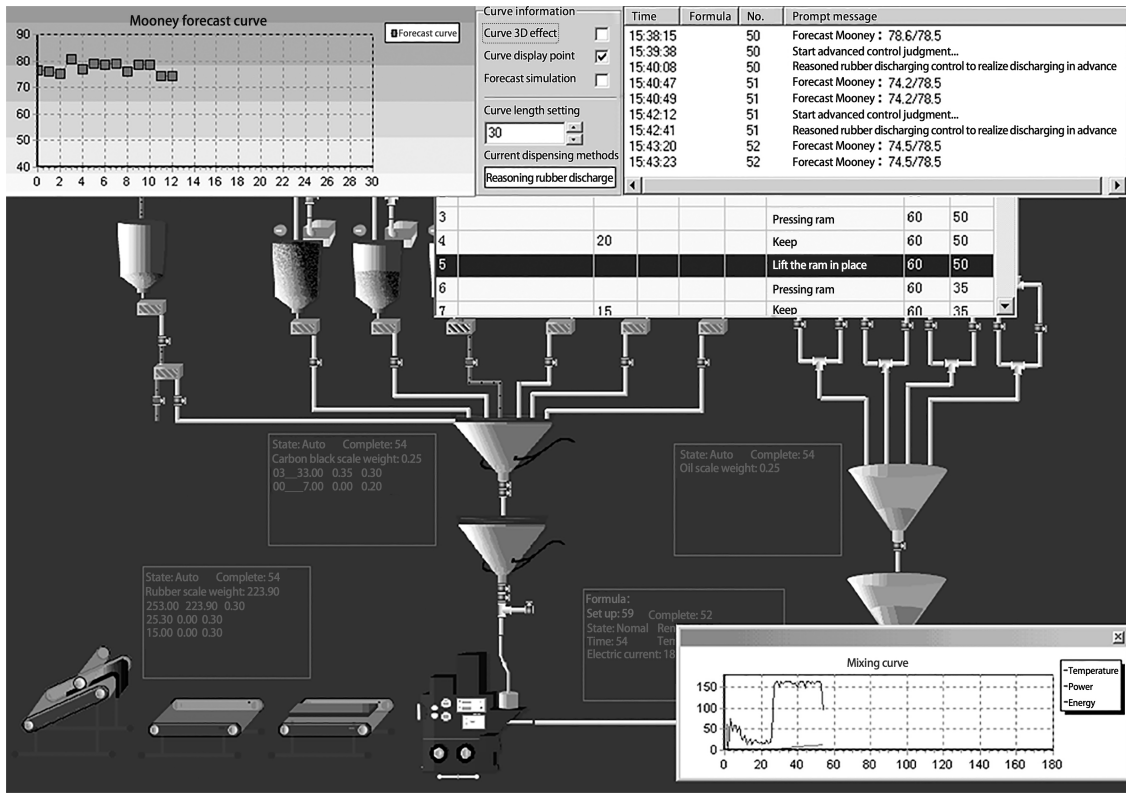


Figure 2: Smart internal mixing system.

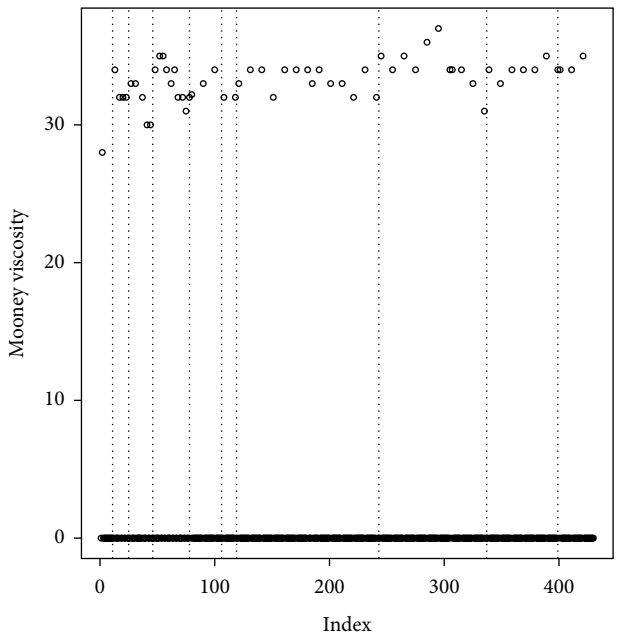


Figure 3: Mooney viscosity values of first 10 batches.

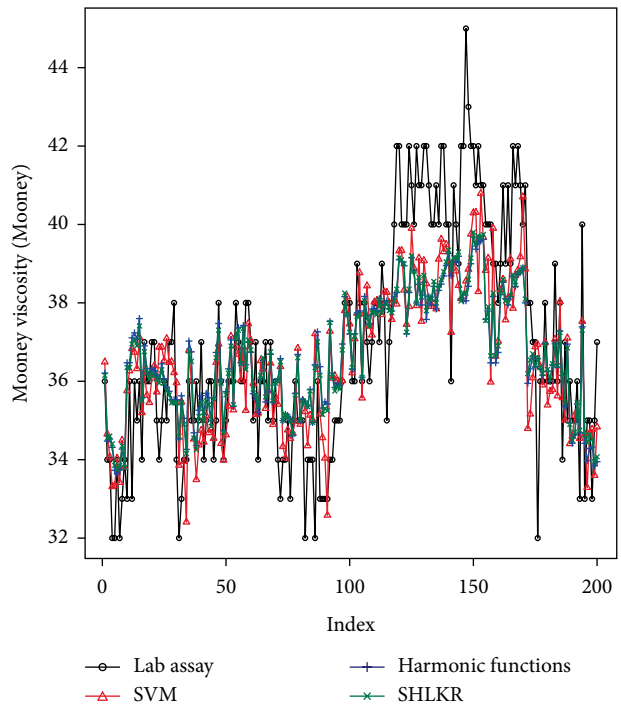


Figure 4: Predictive results of all three different algorithms.

onlycontrols the number of historical samples but not the online sample number, besides of too small sample size condition, the model suffers from too many historical samples, as well as that there will be an optimized k existing to trade-off between underfitting and overfitting. Because of that, theoretically k can be automatically selected by traversing from smaller k to larger ones. Besides of algorithms, also depends on the scale of the historical data and the varieties of noise and

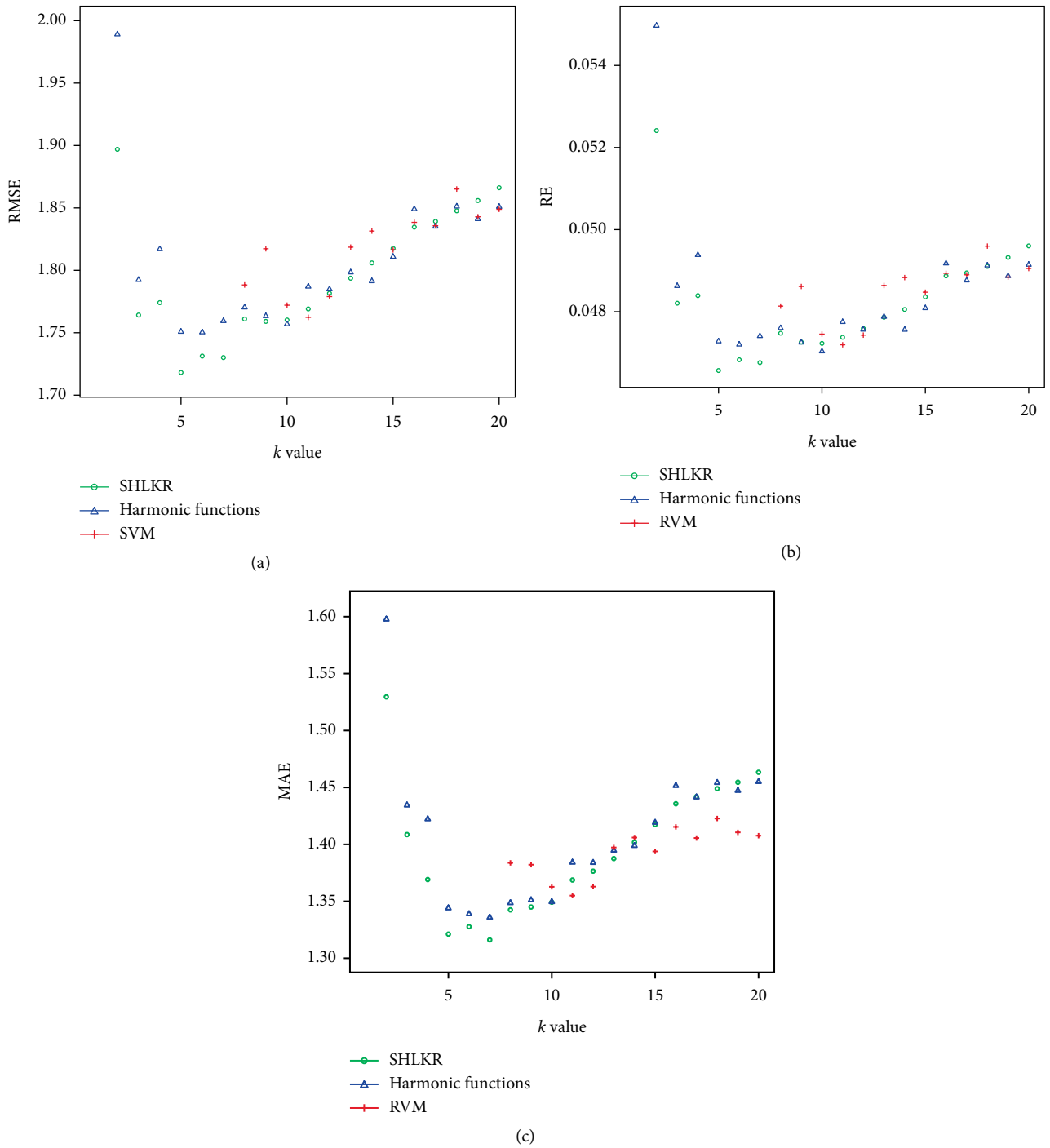


Figure 5: Error varying with k.

formula. Here the optimized values are $k = 5$ for SHLKR, $k = 6$ for Harmonic Functions and $k = 11$ for SVM, which are also determined by leave-one-out cross validation.

Some researches indicate that many indices have their own virtues to validate the soft-sensor model. In order to fully investigate the model performance, 3 commonly used criterions: Root-Mean-Square Error (RMSE), Relative root-mean-square Error (RE) and Mean Absolute Error (MAE) [30] are adopted. As is shown in Figures 5(a)–5(c) and Tables 1–4, N_h denotes batch number and N_p represents the run amount of corresponding batch. Among all algorithms, SVM behaves the worst since both Harmonic Functions and SHLKR algorithms

Table 1: Performance comparison between different algorithms.

	RMSE	RE	MAE
SVM	1.772	0.0479	1.363
HF	1.751	0.0473	1.344
SHLKR	1.718	0.0466	1.321

Table 2: RMSE comparison using different number of historical records.

N_h	0	1	2	3	4	5	6	7	8
N_p	43	12	13	17	19	20	24	25	27
SVM	1.621	1.408	1.156	1.286	1.174	2.244	2.251	1.831	1.961
HF	1.616	1.015	1.045	1.538	1.140	2.209	2.313	2.047	2.038
SHLKR	1.701	1.016	1.096	1.523	1.064	2.064	2.115	1.807	2.029

Table 3: RE comparison using different number of historical records.

N_h	0	1	2	3	4	5	6	7	8
N_p	43	12	13	17	19	20	24	25	27
SVM	0.046	0.040	0.033	0.043	0.032	0.063	0.062	0.053	0.054
HF	0.044	0.027	0.028	0.042	0.031	0.059	0.061	0.055	0.055
SHLKR	0.047	0.028	0.030	0.041	0.029	0.056	0.056	0.048	0.054

Table 4: MAE comparison using different number of historical records.

N_h	0	1	2	3	4	5	6	7	8
N_p	43	12	13	17	19	20	24	25	27
SVM	1.253	1.277	0.942	0.937	1.010	1.895	1.643	1.406	1.577
HF	1.299	0.850	0.770	1.126	0.927	1.752	1.592	1.647	1.697
SHLKR	1.360	0.857	0.858	1.093	0.882	1.644	1.453	1.408	1.666

are smooth hypothesis based. By leveraging unlabelled samples, SHLKR performs best, which has a 2.7% smaller RMSE than SVM, 1.9% smaller RMSE than Harmonic Functions, 1.5% smaller RE than the others, 3.9% smaller MAE than SVM and 1.7% smaller MAE than Harmonic Functions.

4. Conclusion

In this paper, we propose a new semi-supervised hybrid local kernel regression model for soft sensor modelling of internal rubber mixing processing. Distinguished from traditionally supervised models, it leverages unlabelled samples associated with labelled ones to benefit from widely existed supervised data. And the hybrid mechanism is proposed to effectively use both historical and online manufactural data to improve its practicability. Moreover the recursive formula is deduced to enhance its feasibility. With on-site data, soft sensors using proposed and comparative algorithms are implemented to make the evaluation. Experimental results demonstrate that it has a better performance than classical ones. In our future work, SHLKR will be applied to various rubber manufactories and more features will be added into your model, such as raw rubber information, energy cost of each rubber internal mixing phase etc., which will further increase the precision of proposed model.

Data Availability

The rubber mixing processing data used to support the findings of this study were supplied by Haiqing Yu under license and so cannot be made freely available. Requests for access to these data should be made to Haiqing Yu, 10130207@qq.com.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Haiqing Yu and Jun Ji are contributed equally to this work.

Funding

This work is partially supported by the National Natural Science of China (No. 61503208), the National Science Foundation of Shandong Province (No. ZR2015PF002) and the Ministry of Education of Humanities and Social Science Project (No. 15YJC860001).

References

[1] P. Freakley and S. Patel, "Internal mixing: a practical investigation of the flow and temperature profiles during a mixing cycle," *Rubber Chemistry and Technology*, vol. 58, no. 4, pp. 751–773, 1985.

[2] P. Kadlec, B. Gabrys, and S. Strandt, "Data-driven soft sensors in the process industry," *Computers & Chemical Engineering*, vol. 33, no. 4, pp. 795–814, 2009.

[3] 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," *Journal of Process Control*, vol. 19, no. 3, pp. 520–529, 2009.

[4] K. Kim, J.-M. Lee, and I.-B. Lee, "A novel multivariate regression approach based on kernel partial least squares with orthogonal signal correction," *Chemometrics and Intelligent Laboratory Systems*, vol. 79, no. 1-2, pp. 22–30, 2005.

[5] Q. Xuan, B. Fang, Y. Liu et al., "Automatic pearl classification machine based on a multistream convolutional neural network," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 8, pp. 6538–6547, 2018.

[6] Y. Yao and F. Gao, "A survey on multistage/multiphase statistical modeling methods for batch processes," *Annual Reviews in Control*, vol. 33, no. 2, pp. 172–183, 2009.

[7] J. C. B. Gonzaga, L. A. C. Meleiro, C. Kiang, and Filho R. Maciel, "ANN-based soft-sensor for real-time process monitoring and control of an industrial polymerization process," *Computers & Chemical Engineering*, vol. 33, no. 1, pp. 43–49, 2009.

[8] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer Science & Business Media, Berlin, 2013.

[9] J. A. K. Suykens, T. V. Gestel, J. D. Brabanter, B. D. Moor, and J. Vandewalle, "Least squares support vector machines," *International Journal of Circuit Theory & Applications*, vol. 27, no. 6, pp. 605–615, 2002.

[10] Y. Gao, J. Ji, H. Wang, and P. Li, "Adaptive least contribution elimination kernel learning approach for rubber mixing soft-sensing modeling," in *IEEE International Conference on Intelligent Computing & Intelligent Systems*, pp. 470–474, IEEE, Xiamen, China, 2010.

[11] Y. Liu and Z. Gao, "Real-time property prediction for an industrial rubber-mixing process with probabilistic ensemble Gaussian process regression models," *Journal of Applied Polymer Science*, vol. 132, no. 6, 2015.

[12] Y. Liu, N. Hu, H. Wang, and P. Li, "Soft chemical analyzer development using adaptive least-squares support vector regression with selective pruning and variable moving window size," *Industrial & Engineering Chemistry Research*, vol. 48, no. 12, pp. 5731–5741, 2009.

[13] H. Wang, P. Li, F. Gao, Z. Song, and S. X. Ding, "Kernel classifier with adaptive structure and fixed memory for process diagnosis," *AIChE Journal*, vol. 52, no. 10, pp. 3515–3531, 2006.

[14] Y. Liu, C. Yang, Z. Gao, and Y. Yao, "Ensemble deep kernel learning with application to quality prediction in industrial polymerization processes," *Chemometrics & Intelligent Laboratory Systems*, vol. 174, pp. 15–21, 2018.

[15] Q. Xuan, Z. Chen, Y. Liu, H. Huang, G. Bao, and D. Zhang, "Multiview generative adversarial network and its application in pearl classification," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 10, pp. 8244–8252, 2019.

[16] Y. Liu, Y. Fan, and J. Chen, "Flame images for oxygen content prediction of combustion systems using DBN," *Energy & Fuels*, vol. 31, no. 8, pp. 8776–8783, 2017.

[17] W. Zheng, X. Gao, Y. Liu, L. Wang, J. Yang, and Z. Gao, "Industrial Mooney viscosity prediction using fast semi-supervised empirical model," *Chemometrics and Intelligent Laboratory Systems*, vol. 171, pp. 86–92, 2017.

[18] Y. Liu, T. Chen, and J. Chen, "Auto-switch Gaussian process regression-based probabilistic soft sensors for industrial multigrade processes with transitions," *Industrial & Engineering Chemistry Research*, vol. 54, no. 18, pp. 5037–5047, 2015.

[19] K. Fujiwara, M. Kano, S. Hasebe, and A. Takinami, "Soft-sensor development using correlation-based just-in-time modeling," *AIChE Journal*, vol. 55, no. 7, pp. 1754–1765, 2010.

[20] Y. Liu, Z. Gao, P. Li, and H. Wang, "Just-in-time kernel learning with adaptive parameter selection for soft sensor modeling of batch processes," *Industrial & Engineering Chemistry Research*, vol. 51, no. 11, pp. 4313–4327, 2012.

[21] Z. Zheng and G. I. Webb, "Lazy learning of bayesian rules," *Machine Learning*, vol. 41, no. 1, pp. 53–84, 2000.

[22] M. Belkin, P. Niyogi, V. Sindhwani, and P. Bartlett, "Manifold regularization: a geometric framework for learning from examples," *Journal of Machine Learning Research*, vol. 7, no. 1, pp. 2399–2434, 2006.

[23] J. Ji, H. Wang, K. Chen, Y. Liu, N. Zhang, and J. Yan, "Recursive weighted kernel regression for semi-supervised soft-sensing modeling of fed-batch processes," *Journal of the Taiwan Institute of Chemical Engineers*, vol. 43, no. 1, pp. 67–76, 2012.

[24] X. Zhu and A. B. Goldberg, "Introduction to semi-supervised learning," *Synthesis Lectures on Artificial Intelligence and Machine Learning*, vol. 3, no. 1, 130 pages, 2009.

[25] W. Zheng, Y. Liu, Z. Gao, and J. Yang, "Just-in-time semi-supervised soft sensor for quality prediction in industrial rubber mixers," *Chemometrics and Intelligent Laboratory Systems*, vol. 180, pp. 36–41, 2018.

[26] Y. Liu and J. Chen, "Integrated soft sensor using just-in-time support vector regression and probabilistic analysis for quality prediction of multi-grade processes," *Journal of Process Control*, vol. 23, no. 6, pp. 793–804, 2013.

[27] Y. Liu, Q. Y. Wu, and J. Chen, "Active selection of informative data for sequential quality enhancement of soft sensor models with latent variables," *Industrial & Engineering Chemistry Research*, vol. 56, no. 16, pp. 4804–4817, 2017.

[28] X. Zhu, J. Lafferty, and R. Rosenfeld, *Semi-Supervised Learning with Graphs*, Carnegie Mellon University, Language Technologies Institute, School of Computer Science, Pittsburgh, PA, USA, 2005.

[29] K. Tuda, G. Rätsch, S. Mika, and K. R. Müller, "Learning to predict the leave-one-out error of kernel based classifiers," in *International Conference on Artificial Neural Networks*, pp. 331–338, Springer, Berlin, 2001.

[30] C. J. Willmott and K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," *Climate Research*, vol. 30, no. 1, pp. 79–82, 2005.