

# Semi-supervised selective ensemble learning based on distance to model for nonlinear soft sensor development



Weiming Shao, Xuemin Tian\*

College of Information and Control Engineering, China University of Petroleum, 66#, Changjiang West Road, Huangdao District, Qingdao 266580, China

## ARTICLE INFO

Communicated by Jingdong Wang

### Keywords:

Nonlinear soft sensor  
Semi-supervised ensemble learning  
Selective ensemble learning  
Distance to model  
Adaptive process state partition  
Particle swarm optimization

## ABSTRACT

Data-driven soft sensors have been widely used in process systems for delivering online estimations of hard-to-measure yet important quality-related variables. However, in many data-driven soft sensor applications, the process may be strongly nonlinear, and the number of labeled samples is limited, which are two major difficulties in developing high-accuracy soft sensors. To cope with this issue, a novel soft sensing method based on semi-supervised selective ensemble learning strategy is proposed. In the proposed method, the process is initially localized by an adaptive process state partition approach, where the information of unlabeled samples can be incorporated. Subsequently, a new distance to model (DM) criterion is defined for selective ensemble learning, which can overcome the drawback of the  $k$  nearest neighbor method. The newly defined DM criterion along with the incorporation of unlabeled samples can help to describe the relationships between query samples and local models more accurately, and therefore is able to provide higher estimation accuracy. The parameters of the proposed method are finalized automatically by the particle swarm optimization technique. The proposed method is investigated using three real-life benchmark datasets, and the simulation results demonstrate the effectiveness of the proposed method in dealing with nonlinear regression problems in the process system and in the urban pollutant monitoring area.

## 1. Introduction

In process industries, a large amount of hardware sensors are installed to provide data for the purpose of monitoring and control [1]. However, many quality-related key variables can not be measured online by these conventional sensors, such as melt index of polypropylene, concentration of butane, endpoint of crude oil, octane number of diesel oil, biomass concentration, etc. Measurements of those quality variables can be obtained through either online analyzer or offline laboratory analysis. Unfortunately, both of the two ways may result in high cost or hours of measurement delay, making them fail to satisfy modern industrial applications' ever increasing demands [2,3]. The infrequent or inaccurate measurements of those quality variables may lead to poor control performance, huge production losses and even safety hazards [4].

Soft sensors for prediction purpose have been increasingly adopted to cope with the above mentioned issue in recent years [1,5,6]. The essence of soft sensor is to construct a predictive mathematical model for describing the functional relationship between primary variables (i.e., those difficult-to-measure variables) and secondary variables (i.e., those easy-to-measure variables). Whenever the measurements of

primary variables are in need, such predictive model is invoked to infer or estimate their real values. In other word, the soft sensor is a kind of computer software but possesses the function of hardware sensors. Thereby, in contrast with those online analyzers and laboratory analyses, soft sensors have lower maintenance cost and their measurement delays are negligible.

Due to the advantages of data-driven soft sensors over model-driven soft sensors, and the popularization of the distributed control system, numerous modeling algorithms have been applied to develop data-driven soft sensors in many process industries, such as the distillation column process [7], the wastewater treatment process [8], the polymerization process [9], etc. The most commonly used ones include multivariate statistical regression algorithms such as Principal Component Regression (PCR) [10,11] and Partial Least Squares (PLS) [12,13], and machine learning algorithms such as Artificial Neural Networks (ANN) [14,15], Support Vector Machines (SVM) [9,16], Gaussian Process Regression (GPR) [17,18], etc. Normally the linear soft sensing methods such as the PCR and the PLS have more practical applications due to their simplicity revealed by a recent investigation among chemical processes in Japan [5,19], however, they can not model the process nonlinearity, while almost all industrial processes

\* Corresponding author.

E-mail addresses: [swmupc@163.com](mailto:swmupc@163.com) (W. Shao), [tianxm@upc.edu.cn](mailto:tianxm@upc.edu.cn) (X. Tian).

exhibit different degrees of nonlinearities. Those nonlinear soft sensors such as the ANN based ones and the SVM based ones can be adopted to alleviate such issue. Nevertheless, the nonlinear modeling algorithms have the risk of over-fitting especially in the scenario of small amount of training samples, and one single global model may not perform well for processes with strong nonlinearities in wide operating ranges. Ensemble learning based soft sensors, which constructs a series of sub-models and then combines them to give the final estimation of the primary variable, has proven to be theoretically and practically helpful to improve the prediction performance by reducing the estimation variance compared with the single-model based soft sensors [3,20].

Ensemble learning is composed of two steps, namely ensemble members generation and prediction combination [20]. In the first step, the entire dataset is divided into several sub-datasets, upon which sub-models are constructed. For the sake of dealing with the process nonlinearity, this paper focuses on the localization strategies that construct local models specialized on specific model regions, where the philosophy of ‘divide and conquer’ is adopted. Clustering-based methods, for example the fuzzy C-means (FCM) [21,22] and expectation maximization (EM) [18,23,24] are commonly used approaches to completing this mission. Recently, some adaptive localization schemes based on the moving window technique are reported [3,25–31]. The advantages of these methods over clustering-based approaches are that the number of local models is not predefined but adaptively determined, and new clusters can be easily identified and added to the local model set at the online operation stage. In addition, the correlation between process variables can be effectively taken into consideration. Therefore, this paper employs one of the moving window-based localization methods, i.e., the one developed in our previous work [3], as it not only accounts for both first order and second order information of the predicted errors, but also has the functionality of redundancy discrimination.

In the other step of ensemble learning, namely the prediction combination step, the key task is to assign the combination weight to each local model according to some criterion including fuzzy membership [32,33], Bayesian posterior probability [23,24,34], distance to subspaces [35,36], predicted errors [3,26,29,30,37], and so forth. Although these ensemble learning methods have been shown effective in developing high-performance soft sensors, they still have limitations. For example, due to the measurement noise, those criteria such as fuzzy membership, Bayesian posterior probability and distance to subspaces using only the query sample to measure the relevance between the query sample and certain local model might be unreliable. Although the  $k$  nearest neighbors ( $k$ NN) of the query sample are employed in some predicted errors-based criteria, further research indicates that in the case of some process drifts and small amount of labeled samples [3,30,37,38], the selected labeled  $k$ NN of the query sample may cover overlarge operating area and the predicted errors may fail to reflect the true relationship between the query sample and local models.

What's worse, in soft sensor applications, labeling samples is usually expensive and time-consuming, resulting in the fact that the number of labeled samples is limited. As a result, local model regions that rely on labeled sample only may not accurately capture the expertise area of the corresponding local model, which can lead to the problems of inaccurate estimation of the probability density function or the subspace. Making use of both labeled and unlabeled samples, which is termed as ‘semi-supervised learning’ in the machine learning field, is a potential way to deal with the limitation of using labeled samples only. For example, Xiang et al. realized the semi-supervised learning based on local spine regression for classification purpose [39]. Bao et al. and Ge and Song developed semi-supervised soft sensors through co-training [40] and Bayesian method [41], respectively, where the incorporation of unlabeled samples has been shown effective for accuracy enhancement. However, their works are limited to distance-based learning, not suitable for the moving-window

based localization focused in this paper.

Motivated by solving the issue discussed in the above paragraph, this paper proposes a novel ensemble learning method for nonlinear soft sensor development. In the proposed method, the process state is adaptively partitioned into local model regions, and the combination weights of local models are calculated based on a newly defined criterion namely the distance to model (DM). In addition, due to selective ensemble learning has been shown effective to improve the estimation performance compared with making an ensemble of all local models [3,42,43], an adaptive selective ensemble learning mechanism developed in our previous work [3] is incorporated. The soft sensing method developed in this paper is referred to as  $S^3$ EnL-DM, which represents ‘semi-supervised selective ensemble learning based on distance to model’. Our novel contributions in this paper are summarized as follows:

- (1) A DM criterion is proposed to measure the relevance between the query sample and certain local model. This DM criterion utilizes not only the query sample but also its neighbors selected by the time metric instead of the Euclidean distance metric, which can overcome the drawbacks of using only the query sample or selecting the query sample's neighbors by the  $k$ NN.
- (2) We propose to represent local model regions with not only labeled samples but also unlabeled samples. The reason why unlabeled samples are used is that sampling unlabeled samples is much easier and more frequent [44], and the utilization of unlabeled samples is expected helpful for describing the expertise area of the local model in the input space.
- (3) Based on the DM criterion and the local model region representation strategy, a semi-supervised ensemble learning method is developed, which can be used for soft sensor modeling of nonlinear process systems, or dealing with nonlinear regression problems in other areas.

The remaining of this paper is structured as follows. In Section 2, some preliminaries are presented, followed by the details of the proposed method, including the adaptive process state partition, the distance to model criterion, and the selective ensemble learning strategy in Section 3. In Section 4, case studies are conducted on benchmark datasets gathering from real-life processes, and simulation results and discussions are presented. Finally conclusions are put forward in Section 5.

## 2. Preliminaries

### 2.1. Least squares support vector regression

The least squares support vector regression (LSSVR) is a more computationally efficient nonlinear function estimator modified from the support vector regression (SVR). In this paper the LSSVR is applied to construct local models upon the partitioned local model regions because of its ability of handling process nonlinearity. The LSSVR model can be obtained by solving the following optimization problem [45]:

$$\begin{cases} \min J(\omega, \mathbf{e}) = \frac{1}{2} \omega^T \omega + \frac{\gamma}{2} \sum_{k=1}^N e_k^2 \\ s. t. \quad y_k = \omega^T \phi(\mathbf{x}_k) + b + e_k, \quad (k = 1, 2, \dots, N) \end{cases} \quad (1)$$

where  $\{\mathbf{x}_k, y_k\}$  is the  $k$ -th training sample,  $\phi(\cdot)$  is the nonlinear mapping function,  $\omega$  and  $b$  are the model coefficients in feature space,  $\gamma$  is the regularization parameter,  $\mathbf{e} = \{e_1, e_2, \dots, e_N\}^T$  is the training error vector, and  $N$  is the number of training samples. The solution to Eq. (1) is obtained by constructing the Lagrangian as

$$L(\omega, b, \mathbf{e}; \alpha) = J(\omega, \mathbf{e}) - \sum_{k=1}^N \alpha_k \{\omega^T \phi(\mathbf{x}_k) + b + e_k - y_k\} \quad (2)$$

where  $\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_N\}^T$  are Lagrange multiplier vector. Setting the derivatives of  $L(\omega, b, e; \alpha)$  with respect to  $\omega, b, e$  and  $\alpha$  to zeros leads to

$$\begin{cases} \frac{\partial L(\omega, b, e; \alpha)}{\partial \omega} = 0 \Rightarrow \omega = \sum_{k=1}^N \alpha_k \phi(\mathbf{x}_k) \\ \frac{\partial L(\omega, b, e; \alpha)}{\partial b} = 0 \Rightarrow \sum_{k=1}^N \alpha_k = 0 \\ \frac{\partial L(\omega, b, e; \alpha)}{\partial e_k} = 0 \Rightarrow \alpha_k = \gamma e_k, \quad k = 1, 2, \dots, N \\ \frac{\partial L(\omega, b, e; \alpha)}{\partial \alpha_k} = 0 \Rightarrow \omega^T \phi(\mathbf{x}_k) + b + e_k - y_k = 0, \quad k = 1, 2, \dots, N \end{cases} \quad (3)$$

After eliminating  $\omega$  and  $e$  one can get  $\alpha$  and  $b$  through solving the following system of linear equations:

$$\begin{bmatrix} \mathbf{K} + \frac{1}{\gamma} \mathbf{I} & \mathbf{1}_N \\ \mathbf{1}_N^T & 0 \end{bmatrix} \begin{bmatrix} \alpha \\ b \end{bmatrix} = \begin{bmatrix} \mathbf{Y} \\ 0 \end{bmatrix} \quad (4)$$

where  $\mathbf{I}$  represents unit matrix,  $\mathbf{1}_N$  is a column vector with all elements as 1,  $\mathbf{K} \in R^{N \times N}$  and  $\mathbf{K}(i, j) = K(\mathbf{x}_i, \mathbf{x}_j)$ . Here  $K(\cdot, \cdot)$  is called kernel function, and Gaussian kernel function is employed in this paper, which is defined as

$$\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right) \quad (5)$$

where  $\sigma > 0$  determines the kernel width. Finally the prediction of the LSSVR model,  $\hat{y}_t$ , given a query sample  $\mathbf{x}_t$  becomes

$$\hat{y}_t = \sum_{k=1}^N \alpha_k K(\mathbf{x}_t, \mathbf{x}_k) + b \quad (6)$$

## 2.2. Particle swarm optimization

The particle swarm optimization (PSO) technique seeks the solution to an optimization problem by mimicking birds' hunting behaviors. In this paper, it is employed to finalize model parameters for the proposed method due to its powerful search ability and easy implementation. In the PSO, each bird is named as a 'particle', and the solution to an optimization problem corresponds to the position of one bird in the search space. Each particle has its position, speed and fitness value, and tracks the best particle to seek the solution. The PSO starts to work by randomizing a group of particles, and ends the search procedure after several times of iterations. At each iteration, one particle updates its position according to two best values. The first one is called 'personally best value (**pbest**)' which is the best position for a particle itself, and the other one is called 'globally best value (**gbest**)' which is the best position for all particles. Formulas for calculating the position and the speed of one particle in the standard PSO are given as [46]:

$$\begin{cases} \mathbf{v}_i(k+1) = w \cdot \mathbf{v}_i(k) + c_1 \cdot \text{rand} \cdot (\mathbf{pbest}_i - \mathbf{p}_i(k)) + c_2 \cdot \text{rand} \cdot (\mathbf{gbest} - \mathbf{p}_i(k)) \\ \mathbf{p}_i(k+1) = \mathbf{p}_i(k) + \mathbf{v}_i(k) \end{cases} \quad (7)$$

where  $\mathbf{v}_i(k)$ ,  $\mathbf{p}_i(k)$  and  $\mathbf{pbest}_i$  represent the speed, position and personally best position of the  $i$ -th particle at the  $k$ -th iteration,  $\text{rand}$  means random value between 0 and 1,  $c_1$  and  $c_2$  are accelerating constants, and  $w$  is called inertial weight, which is an important parameter for the PSO. Here a linearly decreasing inertial weight which has both good global and local search ability is adopted to calculate  $w$  [46]:

$$w = w_{\max} - \frac{w_{\max} - w_{\min}}{K} \cdot k \quad (8)$$

where  $w_{\max}$ ,  $w_{\min}$ ,  $K$ ,  $k$  represent the maximum weight, the minimum weight, the maximum iteration number and the iteration index, respectively.

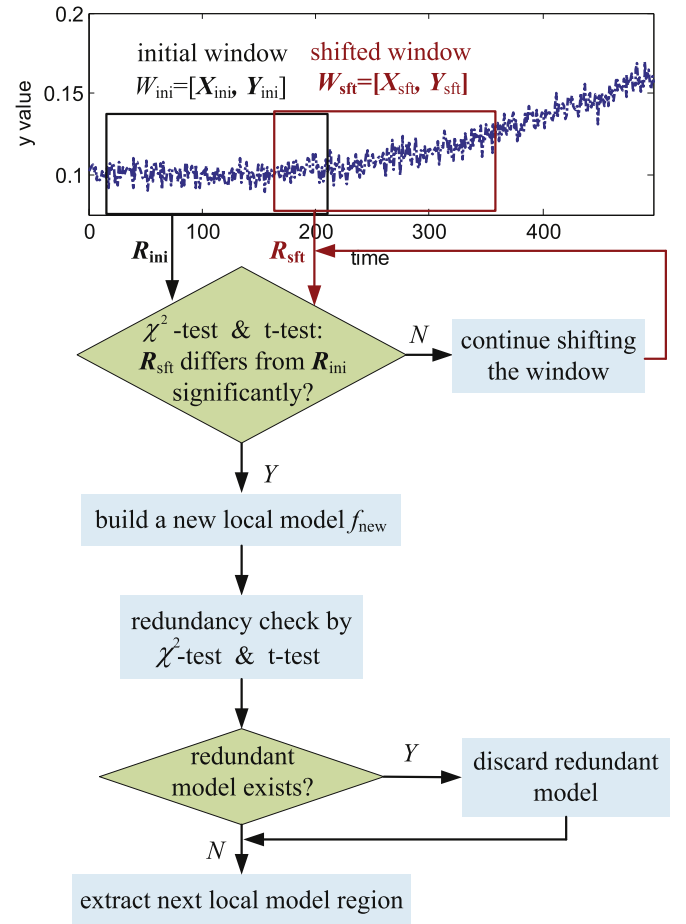


Fig. 1. Adaptive process state partition way using a moving window.

## 3. Selective semi-supervised ensemble learning based on distance to model

The proposed soft sensing method, i.e., the S<sup>3</sup>EnL-DM, is composed of the offline operation stage and the online operation stage. At the offline stage, the process state is divided into local model regions adaptively, upon which independent local models are constructed. The parameter selection for S<sup>3</sup>EnL-DM is also completed at this stage with the aid of PSO. At the online operation stage, whenever the estimation need for a query sample arises, the relationship between the query sample and each local model will be quantified using the DM based criterion. Only those local models which are expected to have higher accuracy for the query sample are selected and combined via the Bayesian inference to give the final estimation for the query sample. The mechanisms in both of the two stages are detailed in this section.

### 3.1. Adaptive process state partition

The adaptive way for process state partition proposed in Ref. [3] is used to prepare concise local model set, which is illustrated in Fig. 1.

Initially, a data window  $W_{ini} = [\mathbf{X}_{ini}, \mathbf{Y}_{ini}]$  with  $W$  consecutive-time samples is set, upon which a local model  $f_{ini}$  is constructed. Subsequently the window is moved one sample step ahead, and a shifted window  $W_{sft} = [\mathbf{X}_{sft}, \mathbf{Y}_{sft}]$  can be obtained. Then the predicted errors for  $\mathbf{Y}_{ini}$  and  $\mathbf{Y}_{sft}$  based on  $f_{ini}$ , denoted as  $\mathbf{R}_{ini}$  and  $\mathbf{R}_{sft}$ , respectively, are calculated. If  $\mathbf{R}_{ini}$  and  $\mathbf{R}_{sft}$  are not significantly different, the performance of  $f_{ini}$  is considered not deteriorated on  $W_{sft}$ , then  $W_{sft}$  doesn't represent a new local model region different from that represented by  $W_{ini}$ . Therefore, the model region represented by  $W_{sft}$  will not be stored, and the window continues to move and new

$\mathbf{R}_{\text{sft}}$  will be calculated by  $f_{\text{ini}}$ . Once  $\mathbf{R}_{\text{sft}}$  significantly deviates from  $\mathbf{R}_{\text{ini}}$ , it indicates a new local model region is identified, upon which a new local model  $f_{\text{new}}$  is built. Assume at this point  $L$  ( $L \geq 1$ ) local models  $\{f_l\}_{l=1}^L$  have been constructed and stored. Even though  $f_{\text{new}}$  differs from  $f_{\text{ini}}$ , whether it is independent of other stored local models  $\{f_l\}_{l=1}^{L-1}$  has not been examined. If there exists one local model  $f_l$  ( $1 \leq l \leq L-1$ ) similar to  $f_{\text{new}}$ , then  $f_{\text{new}}$  will not be preserved. That is, redundancy check needs to be performed. The tasks of both new local process state identification and redundant local model detection are completed through the  $t$ -test and  $\chi^2$ -test, which accounts for the mean and variance information, respectively. Finally, such localization method generates a group of local model regions  $\{W_l = [\mathbf{X}_l, \mathbf{Y}_l]\}_{l=1}^L$  and the corresponding local models  $\{f_l\}_{l=1}^L$ .

The detailed procedures for implementing such adaptive process state partition are given in Ref. [3]. In this paper, the local experts are constructed by the LSSVR on the identified local model regions  $\{W_l = [\mathbf{X}_l, \mathbf{Y}_l]\}_{l=1}^L$ , and a group of local LSSVR models are obtained, which are denoted as  $\{f_l^{\text{LSSVR}}\}_{l=1}^L$ . However, it needs to be pointed out that in some applications, due to the window size is too small required for handling the process nonlinearity, the initial model  $f_{\text{ini}}$  constructed by LSSVR may over-fit the training samples, making the variance of  $\mathbf{R}_{\text{ini}}$  quite smaller, which will generate unnecessarily large amount of local models. In this scenario, the  $f_{\text{ini}}$  can be constructed by linear modeling algorithms such as PLS, which can also generate a group of local model regions  $\{W_l = [\mathbf{X}_l, \mathbf{Y}_l]\}_{l=1}^L$ . Then local LSSVR models  $\{f_l^{\text{LSSVR}}\}_{l=1}^L$  are constructed upon the corresponding local model regions and stored.

In addition, within each local model region, not only labeled samples but also unlabeled samples are contained, which is illustrated in Fig. 2.

Normally the unlabeled samples consisting of  $\mathbf{X}$ -Variable (i.e., those process variables) are sampled much faster than the labeled samples consisting of both  $\mathbf{X}$ -Variable and  $\mathbf{Y}$ -Variable (i.e., those

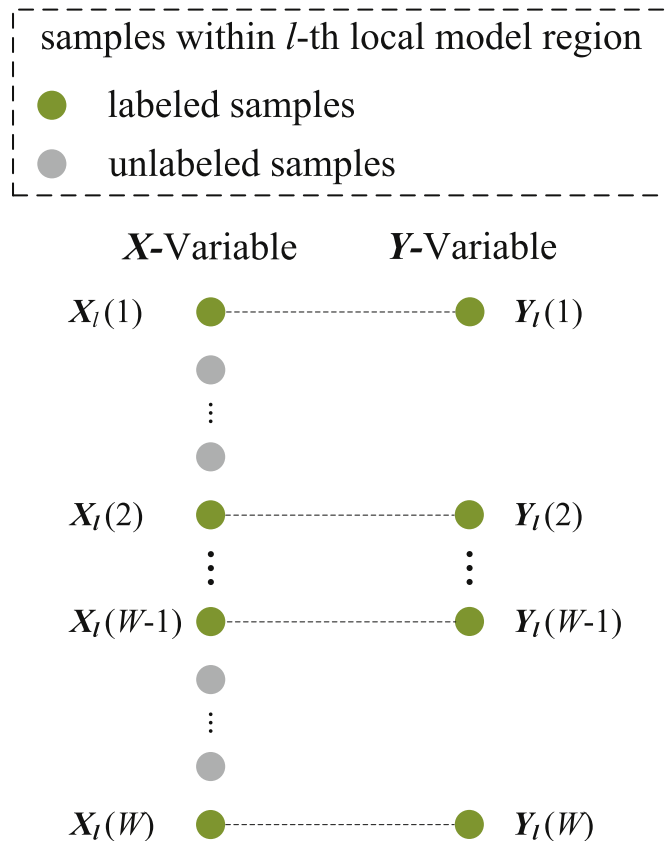


Fig. 2. Illustration of samples within the  $l$ -th local model region.

quality variables), however, in most cases, those unlabeled samples are discarded. In this paper, unlabeled samples are utilized for more accurately describing the receptive fields of local models, which is expected to more reasonably assign combination weights to local models given the query sample. Let  $\mathbf{X}_l^u$  ( $1 \leq l \leq L$ ) represent the unlabeled samples within the  $l$ -th local model region, and all samples of process variables within the  $l$ -th local model region can be denoted as

$$\bar{\mathbf{X}}_l = \begin{bmatrix} \mathbf{X}_l^l \\ \mathbf{X}_l^u \end{bmatrix} \quad (9)$$

Therefore, the local model regions can be represented in terms of only process variables as  $\{\bar{W}_l = \bar{\mathbf{X}}_l\}_{l=1}^L$ .

### 3.2. Distance to model

It has been shown that the DM criterion can reflect the estimation accuracy of a soft sensor model [47], which defines two ways of calculating DM, i.e., the distance to the average of training data and the distance to the nearest neighbor in the training data. In general, the smaller the DMs between the query sample and the soft sensor model are, the higher the estimation accuracy of the soft sensor model for the query sample is. According to the distance to the nearest neighbor, given a query sample  $\mathbf{x}_t$  at time instant  $t$  and the  $l$ -th local model region represented by  $\bar{\mathbf{X}}_l$  ( $1 \leq l \leq L$ ), the DM between  $\mathbf{x}_t$  and the  $l$ -th local model,  $\text{DM}_l(\mathbf{x}_t, \bar{\mathbf{X}}_l)$  ( $1 \leq l \leq L$ ), is defined as

$$\text{DM}_l(\mathbf{x}_t, \bar{\mathbf{X}}_l) = \min \{\|\mathbf{x}_t - \bar{\mathbf{X}}_l(n_i)\|, n_i = 1, 2, \dots, N_l\} \quad (10)$$

where  $\bar{\mathbf{X}}_l(n_i)$  represents the  $n_i$ -th sample within  $\bar{W}_l$  and  $N_l$  is the number of samples within  $\bar{W}_l$ . Note that using only  $\mathbf{x}_t$  may not suffice to measure the true DM, therefore, we augment  $\mathbf{x}_t$  with its neighbors defined as  $2T$  consecutive-time samples before and after time instant  $t$ , which forms  $\mathbf{X}_t$  as

$$\mathbf{X}_t = [\mathbf{x}_{t-T}, \mathbf{x}_{t-T+1}, \dots, \mathbf{x}_{t-1}, \mathbf{x}_t, \mathbf{x}_{t+1}, \dots, \mathbf{x}_{t+T-1}, \mathbf{x}_{t+T}] \quad (11)$$

And the DM between  $\mathbf{X}_t$  and the  $l$ -th local model represented by  $\bar{\mathbf{X}}_l$ ,  $\text{DM}_l^*(\mathbf{X}_t, \bar{\mathbf{X}}_l)$  ( $1 \leq l \leq L$ ), is defined as

$$\text{DM}_l^*(\mathbf{X}_t, \bar{\mathbf{X}}_l) = \frac{\sum_{j=-T}^T s_j \text{DM}_l(\mathbf{x}_{t+j}, \bar{\mathbf{X}}_l)}{\sum_{j=-T}^T s_j} \quad (12)$$

where  $s_j$  ( $j = -T, -T+1, \dots, T$ ) representing the similarity between  $\mathbf{x}_{t+j}$  and  $\mathbf{x}_t$ , which is calculated as

$$s_j = \exp\left(-\frac{\|\mathbf{x}_t - \mathbf{x}_{t+j}\|}{\text{std}_x}\right) \quad (13)$$

where  $\text{std}_x$  is standard deviation of the distances calculated in input space among both labeled and unlabeled samples. When  $T$  is set to 0,  $\text{DM}_l^*(\mathbf{X}_t, \bar{\mathbf{X}}_l)$  is identical to  $\text{DM}_l(\mathbf{x}_t, \bar{\mathbf{X}}_l)$ , and the proposed DM criterion would degenerate to the second DM definition used in Ref. [47].

It should be pointed out that it is  $\bar{\mathbf{X}}_l$  including the unlabeled samples that being employed to calculate the DM rather than  $\mathbf{X}_l$ , so as to better reflect the true DM. In addition,  $\mathbf{x}_t$  is augmented by samples selected according to the time sequence instead of the Euclidean distance, such that the correlation among process variables can be taken into account. Meanwhile, the proposed DM according to Eq. (12) can not be calculated immediately when  $\mathbf{x}_t$  is measured, while requires to wait for  $T$  sampling instants in order to obtain the subsequent  $T$  samples. This is a drawback of the proposed DM criterion, however, since the sampling rate for the unlabeled samples is usually very fast, the waiting time is very small (usually only a few seconds) and can be negligible compared with the sampling interval of labeled samples which may take hours. Consequently, using the proposed DM criterion at the online operation stage is not a problem.



### 3.3. Selective ensemble learning

In the ensemble learning, the estimated  $y$ -value,  $\hat{y}_t$ , for a query sample  $\mathbf{x}_t$  is obtained as the combination of the estimations of the off-line prepared local LSSVR models. According to the Bayesian inference framework [23],  $\hat{y}_t$  can be calculated as

$$\hat{y}_t = \sum_{l=1}^L P(f_l^{LSSVR} | \mathbf{x}_t) f_l^{LSSVR}(\mathbf{x}_t) \quad (14)$$

where  $P(f_l^{LSSVR} | \mathbf{x}_t)$  ( $1 \leq l \leq L$ ) and  $f_l^{LSSVR}(\mathbf{x}_t)$  ( $1 \leq l \leq L$ ) are the posterior probability and the prediction of the  $l$ -th local LSSVR model given  $\mathbf{x}_t$ , respectively. Based on the formulas for conditional probability and total probability,  $P(f_l^{LSSVR} | \mathbf{x}_t)$  is computed as

$$P(f_l^{LSSVR} | \mathbf{x}_t) = \frac{P(f_l^{LSSVR}) P(\mathbf{x}_t | f_l^{LSSVR})}{\sum_{l=1}^L P(f_l^{LSSVR}) P(\mathbf{x}_t | f_l^{LSSVR})} \quad (15)$$

where  $P(f_l^{LSSVR})$  ( $1 \leq l \leq L$ ) represents the prior probability of  $f_l^{LSSVR}$  that the current process property can be described by  $f_l^{LSSVR}$  and  $P(\mathbf{x}_t | f_l^{LSSVR})$  ( $1 \leq l \leq L$ ) means the likelihood that  $\mathbf{x}_t$  comes from the  $l$ -th local model region. Since the DM can reflect the prediction accuracy as demonstrated in Ref. [47], the likelihood of the  $l$ -th local model is computed based on  $DM_l^*(\mathbf{X}_t, \bar{\mathbf{X}}_t)$ . In what follows, without causing any confusion,  $DM_l^*(\mathbf{X}_t, \bar{\mathbf{X}}_t)$  is simplified as  $DM_l^*$  for simplicity. In this paper, the Bayesian assumption is adopted, meaning that all local models have the same prior probability, that is,  $P(f_l^{LSSVR}) = 1/L$ .

Because smaller  $DM^*$  implies higher estimation accuracy and the corresponding local model should be assigned to larger combination weight,  $P(\mathbf{x}_t | f_l^{LSSVR})$  is defined as

$$P(\mathbf{x}_t | f_l^{LSSVR}) = J_l^* / \sum_{l=1}^L J_l^* \quad (16)$$

where

$$J_l^* = \frac{1}{DM_l^*} / \max \left\{ \frac{1}{DM_1^*}, \frac{1}{DM_2^*}, \dots, \frac{1}{DM_L^*} \right\} \quad (17)$$

In Eq. (17), larger  $J_l^*$  ( $l = 1, 2, \dots, L$ ) represents higher estimation accuracy, and the normalization term in the denominator will not influence the combination results but it can scale  $J_l^*$  into the range of (0,1], which can facilitate the implementation of selective ensemble learning. Up to now, the posterior probability of the  $l$ -th local model can be obtained and  $\hat{y}_t$  can be computed through making ensemble of all local models' predictions with Eq. (14). However, the selective ensemble learning which combines part of local models can achieve better generalization performance, as it can reach a good equilibrium between the prediction bias and variance [3]. In order to perform the selective ensemble learning, which models should be selected at each prediction round must be determined. Intuitively, only those ensemble members whose estimation accuracies given  $\mathbf{x}_t$  are higher than a certain satisfaction degree are fused and the rest ones will not be involved. In this paper, such satisfaction degree is determined by setting a pre-defined threshold value  $J_{thr}$  on  $J_l^*$ . Since  $J_l^* \in (0,1]$ , the range of  $J_{thr}$  is limited to (0,1]. Thereby the index set of the selected local models, which is denoted as  $L_S = \{l_1, l_2, \dots, l_S\}$ , can be determined as

$$L_S = \{l | J_l^* \geq J_{thr}\} \quad (18)$$

where  $S$  represents the number of selected local models. Note that  $L_S$  is usually not fixed but adaptively determined depending on the query sample  $\mathbf{x}_t$  at each prediction round, and  $L_S$  will never be empty, that is, at least one local model will be selected due to the normalization operation in Eq. (17). Subsequently,  $\hat{y}_t$  is computed by building an ensemble of the selected local models, which is shown as

$$\hat{y}_t = \sum_{j=1}^S P(f_{l_j}^{LSSVR} | \mathbf{x}_t) f_{l_j}^{LSSVR}(\mathbf{x}_t) \quad (19)$$

and

$$P(f_{l_j}^{LSSVR} | \mathbf{x}_t) = \frac{P(f_{l_j}^{LSSVR}) P(\mathbf{x}_t | f_{l_j}^{LSSVR})}{\sum_{j=1}^S P(f_{l_j}^{LSSVR}) P(\mathbf{x}_t | f_{l_j}^{LSSVR})} \quad (20)$$

where  $j=1, 2, \dots, S$ . Since the Bayesian assumption is adopted, the prior probability  $P(f_{l_j}^{LSSVR}) = 1/S$ , and according to Eq. (16) the likelihood  $P(\mathbf{x}_t | f_{l_j}^{LSSVR}) = J_{l_j}^* / \sum_{j=1}^S J_{l_j}^*$ . Finally, we can get the formula of computing  $\hat{y}_t$  with those selected local models as

$$\hat{y}_t = \sum_{j=1}^S J_{l_j}^* f_{l_j}^{LSSVR}(\mathbf{x}_t) / \sum_{j=1}^S J_{l_j}^* \quad (21)$$

Detailed procedures for implementing the above selective ensemble learning for estimating the  $y$ -value given a query sample  $\mathbf{x}_t$  are presented as follows.

**Step (1):** Collect 2  $T$  consecutive-time samples before and after  $\mathbf{x}_t$  to form  $\mathbf{X}_t$  according to Eq. (11).

**Step (2):** Calculate similarities  $s_j$  ( $j = -T, -T+1, \dots, T$ ) between  $\mathbf{x}_t$  and  $\mathbf{x}_{t+j}$  using Eq. (13).

**Step (3):** Get  $DM_l$  ( $l = 1, 2, \dots, L$ ) with Eq. (10) and get  $DM_l^*$  ( $l = 1, 2, \dots, L$ ) with Eq. (12).

**Step (4):** Compute  $J_l^*$  ( $l = 1, 2, \dots, L$ ) via Eq. (17).

**Step (5):** Obtain the index set of the selected local models,  $L_S$ , through Eq. (18).

**Step (6):** Estimate the real  $y$ -value for  $\mathbf{x}_t$  based on Eq. (21).

### 3.4. PSO aided model parameter determination

The proposed soft sensing method involves several parameters that need to be determined, namely the window size  $W$ , the kernel parameter  $\sigma$  and the regularization parameter  $\gamma$  in the LSSVR algorithm, the time span  $T$  and the threshold value  $J_{thr}$  for selective ensemble learning. The significance levels for hypothesis testing can be pre-defined as small positive values such as 0.05 according to Refs. [3,26]. In general, smaller  $W$  can divide the process state more detailedly, and hence is able to enhance the ability of handling the process nonlinearity. Nevertheless, too small window size has the risk of causing model instability and un-reliable estimations of the predicted residuals' mean and variance [3]. The threshold value  $J_{thr}$  is an important value for the proposed method, as it decides the degree of selective ensemble learning. Specifically,  $J_{thr} = 0$  means all local models are involved for ensemble learning, while when  $J_{thr}$  is set to 1, only one local model with the smallest DM is chosen to predict the  $y$ -value for the query sample. Improper selection of  $J_{thr}$  could not obtain a satisfactory equilibrium between the prediction variance and bias, which may result in unacceptable estimation performance. It can also be expected that the time span  $T$  should be neither too small nor too large. In addition, two parameters involved in the LSSVR algorithm are set identical among local LSSVR models, although they can be automatically determined through the leave-one-out cross-validation in some applications [48].

Generally, multi-tuning of these parameters may be intractable, which takes too much effort and time if it is done manually. In this paper, the parameter tuning is completed by the PSO automatically. The fitness function in the PSO is set as the predicted root mean squares error on the validation dataset ( $RMSE^{val}$ ) that reflects the estimation accuracy. The  $RMSE^{val}$  is defined as

$$\text{RMSE}^{\text{val}} = \sqrt{\sum_{t=1}^{N_{\text{val}}} (\hat{y}_t^{\text{val}} - y_t^{\text{val}})^2 / N_{\text{val}}} \quad (22)$$

where  $y_t^{\text{val}}$  and  $\hat{y}_t^{\text{val}}$  represent the real  $y$ -value and its estimation for the  $t$ th sample in the validation dataset, respectively, and  $N_{\text{val}}$  is the number of samples in the validation dataset. According to the definition of  $\text{RMSE}^{\text{val}}$ , the model parameters of the proposed method is determined as

$$(W^*, \sigma^*, \gamma^*, T^*, J_{\text{thr}}^*) = \arg \min_{W, \sigma, \gamma, T, J_{\text{thr}}} \text{RMSE}^{\text{val}}(W, \sigma, \gamma, T, J_{\text{thr}}) \quad (23)$$

In the PSO, the iteration procedure can be terminated by either limiting the maximum iteration number, or setting a threshold value for **gbest** differential between two iteration rounds. Here we adopt the former one, and the maximum iteration number is set as 50. In addition, the number of particles is set as ten times the number of parameters that need to be determined.

#### 4. Case studies

In this section, the performance of our proposed soft sensing method, i.e., the  $\text{S}^3\text{EnL-DM}$ , is first tested on two benchmark datasets, which are obtained from the industrial debutanizer column process (DCP) and the sulfur recover unit (SRU). These datasets can be publicly downloaded from the website provided by Ref. [49], and have been widely employed for evaluating the performance of various soft sensors [3,34,44,49–51]. Subsequently, the application of the  $\text{S}^3\text{EnL-DM}$  in the urban pollutant monitoring field for developing calibration model for the electronic nose is reported. The performance of some benchmark soft sensing methods is also presented for comparison purpose. These benchmark methods include the PLS [52], the LSSVR [45], the just-in-time learning LSSVR (JITL-LSSVR) [53], the ensemble of LSSVR (EnLSSVR) [37] that employs LSSVR to construct local experts. In addition, in order to verify the effectiveness of those unlabeled samples employed by the  $\text{S}^3\text{EnL-DM}$ , one more soft sensor model is developed denoted as  $\text{SEnL-DM}$ . The only difference between the  $\text{SEnL-DM}$  and the  $\text{S}^3\text{EnL-DM}$  is that in the  $\text{SEnL-DM}$ , the unlabeled samples in the training dataset are discarded. In other word, in the  $\text{SEnL-DM}$ , the stored local model regions are represented by labeled samples only instead of by both labeled and unlabeled sample. The performance of those soft sensors is quantified using the RMSE index on the testing dataset which is denoted as  $\text{RMSE}^{\text{tst}}$ .

##### 4.1. Application to the debutanizer column process

The debutanizer column process (DCP) is one part of the naphtha splitter plant, and is mainly used to remove butane and propane as overheads from the naphtha stream [49], which is shown in Fig. 3. At the bottom of DCP, the content of butane is required to be minimized. To realize this objective, the measurement of butane content should be obtained in real-time. However, the butane content is usually analyzed by a gas chromatograph installed at the top of the DCP, which introduces a large measurement delay. Therefore, it is desirable to deliver real-time estimation of the butane content through developing a soft sensor. Seven hardware sensors which are marked with yellow circles in Fig. 3, are installed to obtain secondary variables, the explanations of which are presented in Table 1.

The DCP dataset contains 2394 samples, and the time alignment matching has been completed. They are evenly divided into three groups: the first one containing half of the dataset serves as the training dataset, and both of the rest two contain a quarter of the dataset which serves as the validation dataset and the testing dataset, respectively. In the training dataset, we assume only a part of samples are labeled and the rest ones are unlabeled. The proportion of the labeled samples in the training dataset is denoted as  $p$ , and the performance of all mentioned soft sensors are tested under different proportions of labeled samples. In this case study, the proposed  $\text{S}^3\text{EnL-DM}$  based

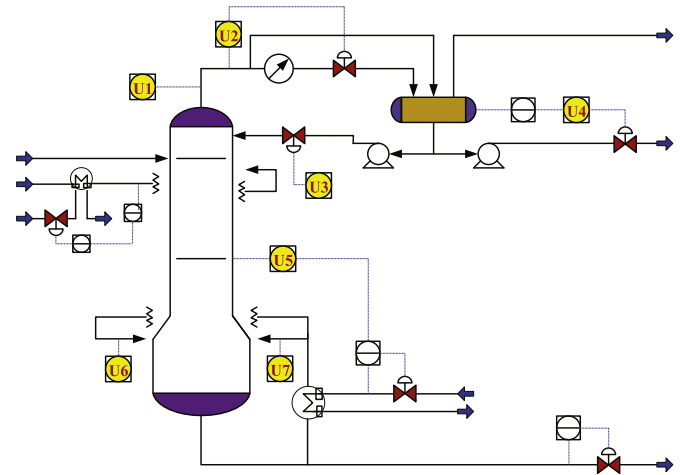


Fig. 3. Block diagram of the DCP. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

Table 1

Descriptions of secondary variables for developing soft sensor for the DCP.

Secondary variable	Explanation
$u_1$	Top temperature
$u_2$	Top pressure
$u_3$	Reflux flow
$u_4$	Flow to next process
$u_5$	6th tray temperature
$u_6$	Bottom temperature
$u_7$	Bottom temperature

soft sensor employs the PLS to assist process state partition due to the reason explained in Section 3.1, and the latent variable number of PLS is determined by 5-fold cross validation (CV), such that the model complexity of the  $\text{S}^3\text{EnL-DM}$  can be reduced, the computational burden of the PSO can be decreased and the difficulty of the optimization can be lowered. Note that the 5-fold CV is implemented on the  $W$  samples within each local model region which is a part of the training dataset. In the 5-fold CV, the  $W$  samples are first divided into five disjoint subsets. In each iteration a temporary PLS model is trained on a different combination of four subsets, and the unused one is used to calculate the estimation errors based on the temporary PLS model. After five iterations, the latent variable number which can minimize the total estimation errors is selected to construct the final local PLS model using all the  $W$  samples. The significance levels for  $t$ -test and  $\chi^2$ -test are both set as 0.05, the scaling parameter  $\text{std}_x$  is calculated as 0.218 according to its definition, and window size  $W$  is fixed as 10 in advance for simplicity. The remaining parameters are optimized by the PSO, and the candidate set for each parameter are determined by either its physical meaning or trial-and-error [3], which are summarized in Table 2. The optimized values of the parameters of the  $\text{S}^3\text{EnL-DM}$  are tabulated in Table 3.

The time trend plot comparisons of the prediction results of those soft sensors under the condition that 20% samples in the training set are labeled are illustrated in Fig. 4(a–f), and the prediction accuracies of those soft sensors under different sizes of labeled dataset are quantified

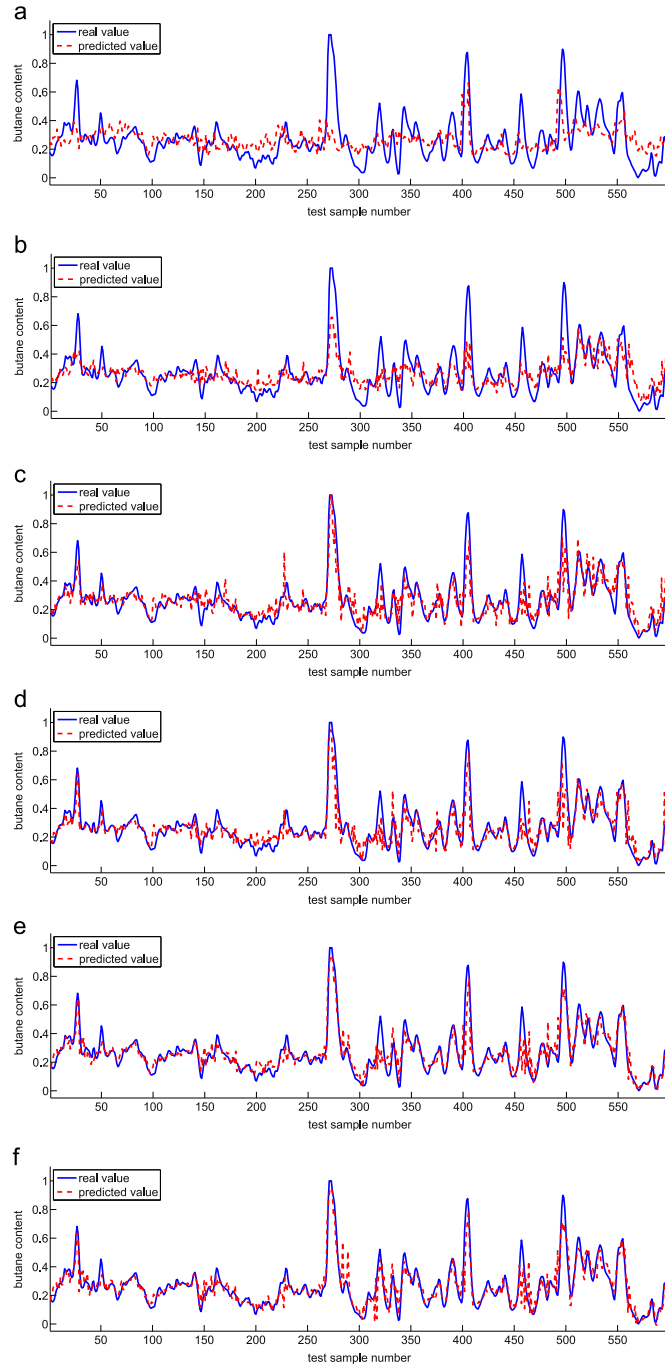
Table 2

Candidate sets for parameters of the  $\text{S}^3\text{EnL-DM}$  that need to be optimized in the case of the DCP.

Ratio of labeled samples	$\gamma$	$\sigma$	$J_{\text{thr}}$	$T$
$1/10 \leq p \leq 1/4$	[1,50]	[0.05, 0.5]	[0,1]	[0,15]
$p = 1/3$	[1,100]			
$p = 1/2$	[1,200]			

**Table 3**Optimized parameters of the S<sup>3</sup>EnL-DM based soft sensor under different proportions of labeled samples in the training data set.

Parameter	$p = 1/10$	$p = 1/9$	$p = 1/8$	$p = 1/7$	$p = 1/6$	$p = 1/5$	$p = 1/4$	$p = 1/3$	$p = 1/2$
$\gamma$	10.37	14.71	15.95	27.05	26.23	25.35	22.28	73.14	100.00
$\sigma$	0.175	0.187	0.245	0.219	0.284	0.331	0.343	0.228	0.327
$J_{thr}$	0.706	0.729	0.772	0.645	0.735	0.823	0.799	0.745	0.853
$T$	9	7	7	9	5	6	6	7	5

**Fig. 4.** Time trend plot comparisons of predicted butane content between various soft sensors with  $p=1/5$ : (a) predictions of PLS, (b) predictions of LSSVR, (c) predictions of JITL-LSSVR, (d) predictions of EnLSSVR, and (e) predictions of SEnL-DM; and (f) predictions of S<sup>3</sup>EnL-DM.

in Table 4 in terms of RMSE<sup>tst</sup>. Here the latent variable number of the PLS model is selected on the validation dataset, and the LSSVR model with Gaussian kernel is realized by using the LSSVM toolbox down-

loaded from <http://www.esat.kuleuven.be/sista/lssvmlab/>. In addition, the EnLSSVR, the SEnL-DM and the S<sup>3</sup>EnL-DM share the same local model set, which are identified in the S<sup>3</sup>EnL-DM, but other parameters required by the EnLSSVR and the SEnL-DM are optimized by the PSO.

It is observed from Fig. 4(a) that the performance of the PLS based soft sensor fail to model this process especially in predicting the peak values of the butane concentration, due to its limitation in handling nonlinear process. Meanwhile, Fig. 4(b) shows the LSSVR outperforms the PLS to some extent, however, its estimation performance is still disappointing, indicating that a single nonlinear function estimator may have difficulty in modeling the process with strong nonlinearity. In contrast, the rest four soft sensors which employ the ‘divide and conquer’ strategy could achieve much better estimation performance as shown in Fig. 4(c–f). In particular, the proposed two methods, i.e., the SEnL-DM and S<sup>3</sup>EnL-DM, can further enhance the estimation accuracy compared with the JITL-LSSVR and the EnLSSVR, demonstrating the effectiveness of the proposed ensemble learning strategies. The quantified prediction accuracy listed in Table 4 confirms the above analyses. In addition, the predicted RMSE<sup>tst</sup> tabulated in last two rows of Table 4 demonstrate that the S<sup>3</sup>EnL-DM outperforms the SEnL-DM, which can not be distinguished clearly in Fig. 4(e–f).

It should be pointed out that satisfactory performance of the S<sup>3</sup>EnL-DM demonstrated in Fig. 4 and Table 4 relies on properly tuning model parameters, which is realized through proper selection of validation samples. If we randomly allocate samples to the validation dataset and the testing dataset, where the training dataset keeps invariant required by the localization method, the average predicted RMSE<sup>tst</sup> and the corresponding standard deviation under different proportions of validation samples would be as presented in Table 5. Here  $\zeta$  represents the proportion of validation samples, and the average performance is obtained via 10 times independent simulations. In each simulation, model parameters are optimized by the PSO on the validation dataset with specific proportion of labeled samples in the training dataset and that of the validation samples.

It is indicated by Table 5 that, generally the larger the proportion of validation samples (i.e.,  $\zeta$ ), the better estimation performance the S<sup>3</sup>EnL-DM achieves. However, the standard deviation of the average prediction accuracy is likely to increase as  $\zeta$  becomes bigger. Moreover, comparing the results reported in Table 5, and the predicted RMSE<sup>tst</sup> of the S<sup>3</sup>EnL-DM listed in Table 4, one can find that under many values of ‘ $p$ ’ when  $\zeta$  exceeds 80%, the prediction accuracies of the S<sup>3</sup>EnL-DM with random selection of validation samples are inferior to that of the S<sup>3</sup>EnL-DM with deterministic selection of validation samples where the proportion of validation samples are 50%. That is, evenly selecting validation samples can help to enhance the prediction accuracy of the S<sup>3</sup>EnL-DM. Therefore, we can basically infer that making the validation samples as disperse as possible over the entire operation area, and as many as possible, is beneficial to properly tuning the algorithm parameters of the S<sup>3</sup>EnL-DM, which is suggested and adopted in the rest two case studies.

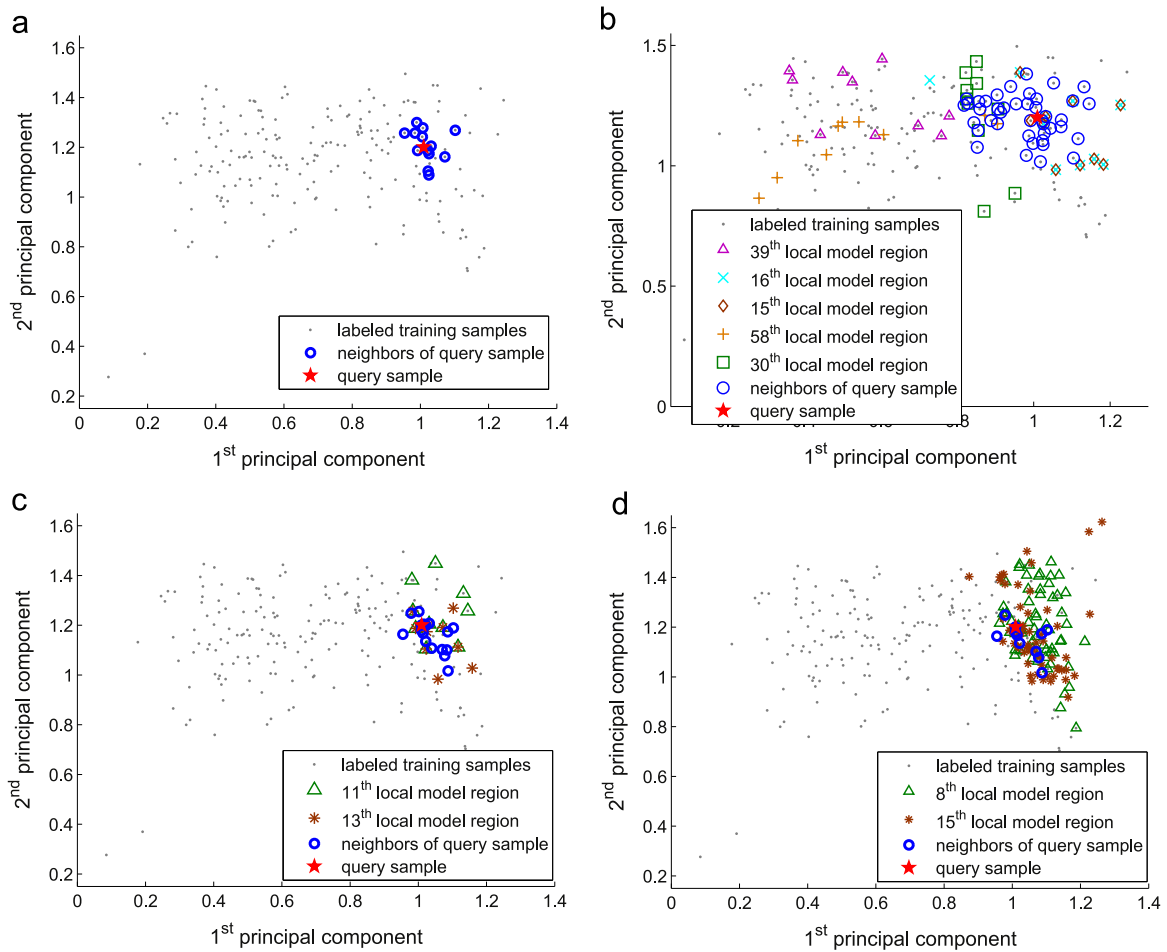
With the aid of visualization using the principal component analysis, the mechanisms of the JITL-LSSVR, the EnLSSVR, the SEnL-DM and the S<sup>3</sup>EnL-DM can be revealed by taking the 110<sup>th</sup> testing sample as example in the case of  $p=1/6$  as shown in Fig. 5, which is helpful for understanding the reason why the proposed

**Table 4**Predicted RMSE<sup>1st</sup> of various soft sensors under different proportion of labeled samples.

Soft sensing method	$p = 1/10$	$p = 1/9$	$p = 1/8$	$p = 1/7$	$p = 1/6$	$p = 1/5$	$p = 1/4$	$p = 1/3$	$p = 1/2$
PLS	0.1416	0.1410	0.1405	0.1409	0.1408	0.1405	0.1403	0.1404	0.1404
LSSVR	0.1243	0.1234	0.1221	0.1218	0.1130	0.1099	0.0958	0.0931	0.0747
JITL-LSSVR	0.1154	0.1082	0.1093	0.1093	0.1006	0.0927	0.0906	0.0822	0.0666
EnLSSVR	0.1127	0.1063	0.1042	0.1122	0.0999	0.0862	0.0828	0.0785	0.0589
SEnL-DM	0.0993	0.0925	0.0913	0.0872	0.0765	0.0717	0.0644	0.0467	0.0322
S <sup>3</sup> EnL-DM	0.0952	0.0892	0.0870	0.0801	0.0711	0.0685	0.0596	0.0439	0.0311

**Table 5**Average predicted RMSE<sup>1st</sup> of the S<sup>3</sup>EnL-DM under different proportion of validation samples with random partition.

	$\zeta=20\%$	$\zeta=40\%$	$\zeta=60\%$	$\zeta=80\%$
$p = 1/10$	$0.0985 \pm 0.0017$	$0.0960 \pm 0.0039$	$0.0957 \pm 0.0037$	$0.0928 \pm 0.0062$
$p = 1/9$	$0.0920 \pm 0.0022$	$0.0908 \pm 0.0030$	$0.0901 \pm 0.0031$	$0.0876 \pm 0.0044$
$p = 1/8$	$0.0884 \pm 0.0026$	$0.0873 \pm 0.0031$	$0.0857 \pm 0.0034$	$0.0838 \pm 0.0041$
$p = 1/7$	$0.0833 \pm 0.0031$	$0.0817 \pm 0.0032$	$0.0812 \pm 0.0031$	$0.0793 \pm 0.0052$
$p = 1/6$	$0.0754 \pm 0.0023$	$0.0733 \pm 0.0024$	$0.0718 \pm 0.0027$	$0.0706 \pm 0.0032$
$p = 1/5$	$0.0732 \pm 0.0018$	$0.0715 \pm 0.0020$	$0.0712 \pm 0.0031$	$0.0698 \pm 0.0019$
$p = 1/4$	$0.0611 \pm 0.0011$	$0.0599 \pm 0.0013$	$0.0597 \pm 0.0022$	$0.0570 \pm 0.0033$
$p = 1/3$	$0.0455 \pm 0.0013$	$0.0450 \pm 0.0013$	$0.0446 \pm 0.0019$	$0.0440 \pm 0.0051$
$p = 1/2$	$0.0342 \pm 0.0029$	$0.0337 \pm 0.0024$	$0.0323 \pm 0.0031$	$0.0311 \pm 0.0025$

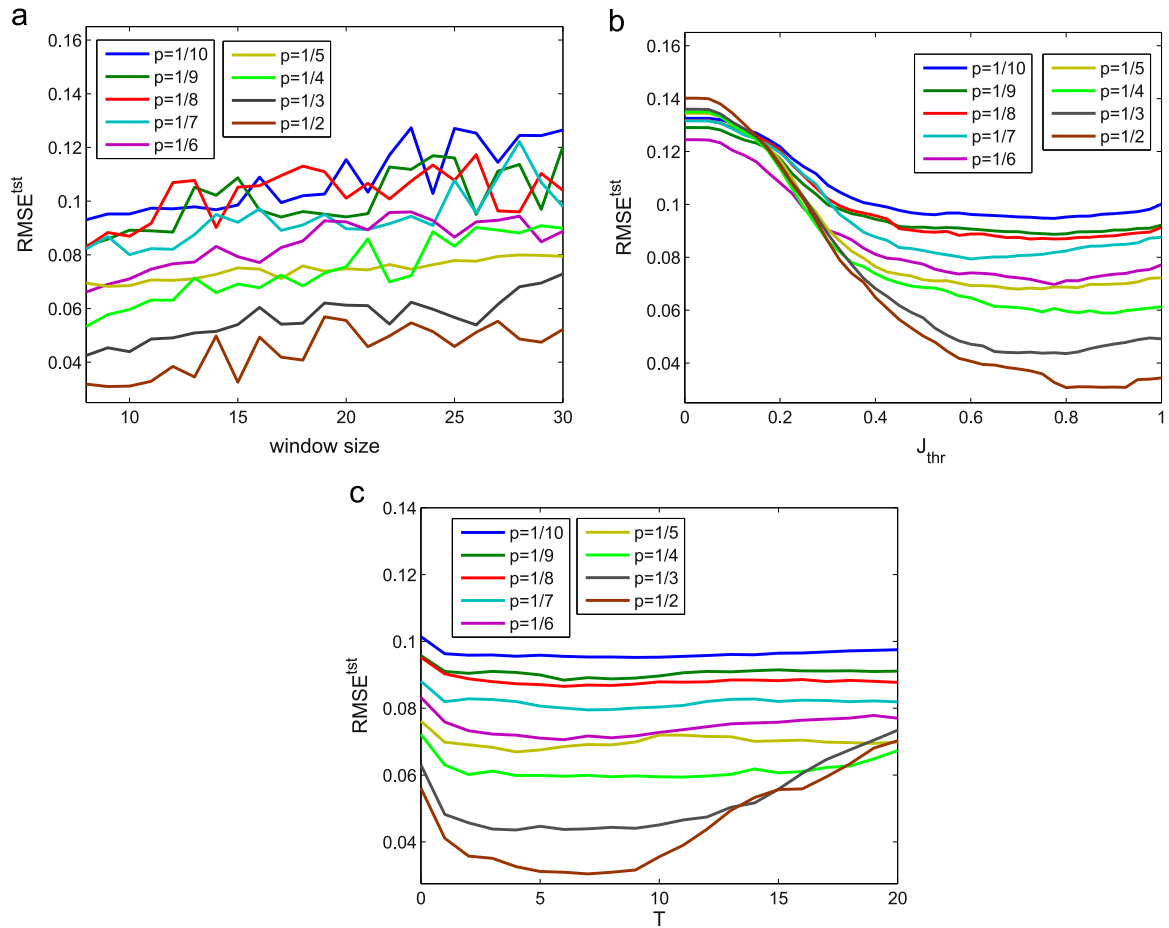
**Fig. 5.** Mechanisms of various soft sensors for the 110<sup>th</sup> testing sample under the condition that  $p=1/6$ : (a) JITL-LSSVR, (b) EnLSSVR, (c) SEnL-DM, and (d) S<sup>3</sup>EnL-DM.

methods show superiorities over the other two methods. In Fig. 5(a), the JITL-LSSVR determines the  $k$ NN of the query sample according to the Euclidean distance criterion and constructs a local model on those selected neighbors. It ignores the correlation among process variables and the samples that are spatially close in terms of the Euclidean

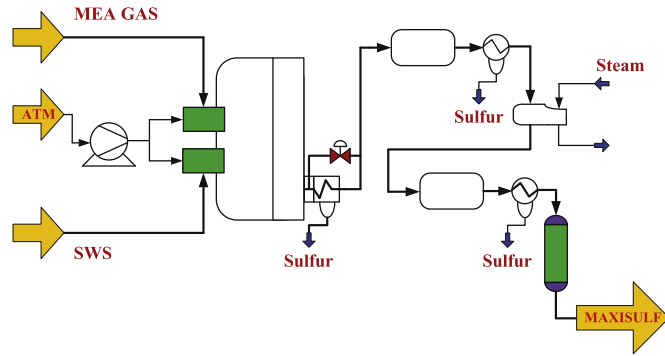
distance criterion may not have the same correlation [25], which makes the mapping relationship between the secondary variables and the primary variable among those selected neighbors still nonlinear even after the nonlinear mapping.

This issue can be tackled by the adaptive process state partition





**Fig. 6.** Estimation accuracy as a function of parameters of the proposed soft sensor: (a)  $W$ ; (b)  $J_{thr}$ ; and (c)  $T$ . In each case, the rest parameters remain the same as their values optimized by the PSO on the validation dataset.



**Fig. 7.** Flowchart of the SRU.

**Table 6**

Secondary and primary variables of the soft sensors for SRU.

Variable	Explanation
Secondary variable	MEA gas flow First air flow Second air flow Gas flow in SWS zone Air flow in SWS zone
Primary variable	Concentration of H <sub>2</sub> S Concentration of SO <sub>2</sub>

**Table 7**

Candidate sets and optimized values for parameters of the S<sup>3</sup>EnL-DM in the case of SRU.

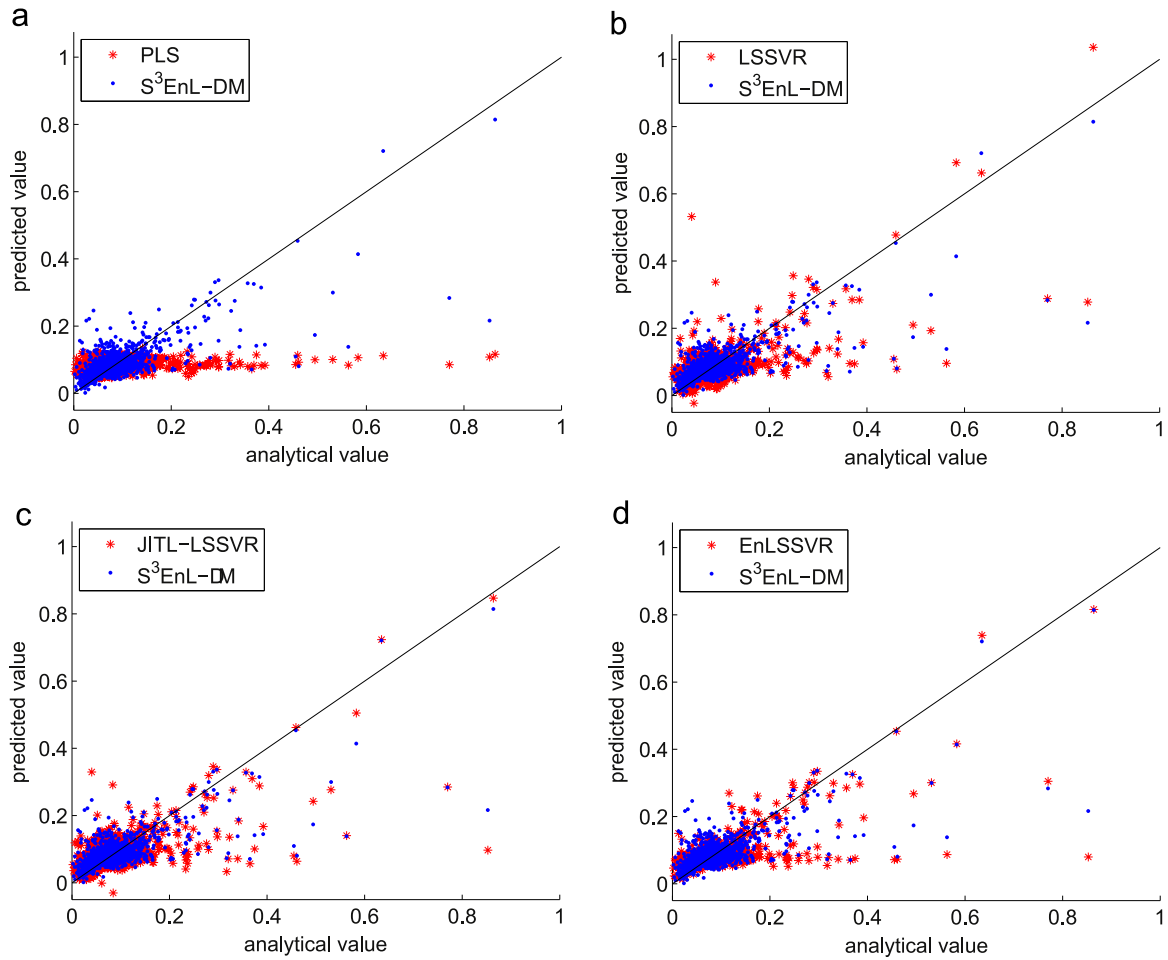
	$\gamma$	$\sigma$	$J_{thr}$	$T$
Candidate set	[1,50]	[0.01, 0.5]	[0,1]	[0,15]
Optimized value	for H <sub>2</sub> S for SO <sub>2</sub>	11.57 14.74	0.100 0.097	0.806 0.708
				3 5

**Table 8**

Predicted RMSE<sup>test</sup> of various soft sensors for estimating concentrations of H<sub>2</sub>S and SO<sub>2</sub>.

Concentration	PLS	LSSVR	JITL-LSSVR	EnLSSVR	SEnL-DM	S <sup>3</sup> EnL-DM
H <sub>2</sub> S	0.0519	0.0415	0.0401	0.0382	0.0376	0.0367
SO <sub>2</sub>	0.0573	0.0517	0.0475	0.0477	0.0421	0.0408

based on a moving window employed by the EnLSSVR, the SEnL-DM and the S<sup>3</sup>EnL-DM, which gathers correlation-related samples to form one local region as embodied in Fig. 5(b-d). However, the EnLSSVR still utilizes the  $k$ NN of the query sample to assign weights to local models, and it may suffer from the difficulty of reasonably assigning combination weights to local models due to the above analyzed limitation of the  $k$ NN method. For example, in this application with  $p=1/6$  for the 110<sup>th</sup> testing sample, the EnLSSVR assigns five largest combination weights to the 39<sup>th</sup>, the 16<sup>th</sup>, the 15<sup>th</sup>, the 58<sup>th</sup> and the 30<sup>th</sup> local model, and the corresponding local model regions are illustrated in Fig. 5(b). As can be seen, the query sample is located beyond the ‘expertise area’ of the 39<sup>th</sup> local model that has the largest



**Fig. 8.** Scatter plot comparisons for estimating the concentration of  $H_2S$  between the  $S^3EnL-DM$  and the PLS, the LSSVR, the JITL-LSSVR and the EnLSSVR.

contribution to the final estimation for the query sample. In this scenario, high estimation accuracy can not be guaranteed.

In contrast, in both the  $SEnL-DM$  and the  $S^3EnL-DM$  the neighbors of the query sample are selected according to the time sequence rather than the  $kNN$  method, which are marked with blue cycles in Fig. 5(c-d). For the 110<sup>th</sup> testing sample, the  $SEnL-DM$  selects the 11<sup>th</sup>–13<sup>th</sup> local models, while the  $S^3EnL-DM$  selects the 8<sup>th</sup>–15<sup>th</sup> local models. The 11<sup>th</sup> and the 13<sup>th</sup> local model regions are visualized with labeled samples in Fig. 5(c), and the 8<sup>th</sup> and the 15<sup>th</sup> local model regions are visualized with both labeled and unlabeled samples in Fig. 5(d). It can be inferred from Fig. 5(c-d) that the query sample is located within the ‘expertise area’ of the selected local models, implying that the proposed DM criterion can effectively select proper local models given a query sample and high estimation accuracy can be expected. It is notable that as the  $S^3EnL-DM$  employs both labeled and unlabeled samples to describe the ‘expertise area’ of one local model, the DM between the query sample and the local model can be calculated more accurately. In other word, it is the unlabeled samples in the training dataset that help the  $S^3EnL-DM$  to outperform the  $SEnL-DM$ .

The influences of model parameters of the proposed  $S^3EnL-DM$  method on the estimation performance under different proportions of labeled samples in the training dataset are demonstrated in Fig. 6(a-c). Such sensitivity analysis of the estimation accuracy with respect to model parameters is helpful to understand the proposed method, and then facilitate the determination of model parameters. As can be seen from Fig. 6(a), smaller window sizes can enhance the estimation accuracy because they are helpful to deal with the process nonlinearity, while larger window size not only increases the prediction error but also causes large fluctuation of the estimation performance.

Nevertheless, too small window sizes should be resisted, as they may result in the problem of over-fitting among local models.

Fig. 6(b) proves the functionality of the selective ensemble learning. When  $J_{thr} = 1$ , only one local model is selected given a query sample, which is referred to as the ‘best model selection’ strategy; while  $J_{thr} = 0$  meaning all local models are combined. Apparently, neither of the two strategies is the best choice. In particular, when all ensemble members are combined, the estimation error is rather high compared with the ‘best model selection’ strategy. This is due to the bias term in the bias/variance decomposition of the estimation error [3]. It is worth to point out that the proposed selective ensemble learning strategy has two delightful properties. The first one is that the threshold value  $J_{thr}$  is limited to the range of [0 1], and the other one is that there is wide range of  $J_{thr}$  for achieving high estimation performance. Such two properties enable us to maximize the benefit of the selective ensemble learning strategy with few difficulties.

It is observed in Fig. 6(c) that augmenting the query sample with its time sequence neighbors is necessary, since setting  $T$  to zero always generates larger  $RMSE^{test}$  than setting  $T$  properly. Note that there is also wide space of  $T$  for getting low estimation error. In addition, it is interesting to notice that in the case of small proportion of labeled samples such as  $p=1/10$ , increasing  $T$  has little influence on the estimation accuracy, however, when the proportion of labeled samples is large, for example  $p=1/2$ , increasing  $T$  from 10 to 20 can remarkably deteriorate the estimation performance. Actually, as the size of the labeled samples increases, the coverage area of one local model becomes narrower, since the window size under each value of  $p$  is set to the same value, i.e., 10. Consequently, when  $p=1/10$ , increasing  $T$  in small range will not significantly change the DM between the query

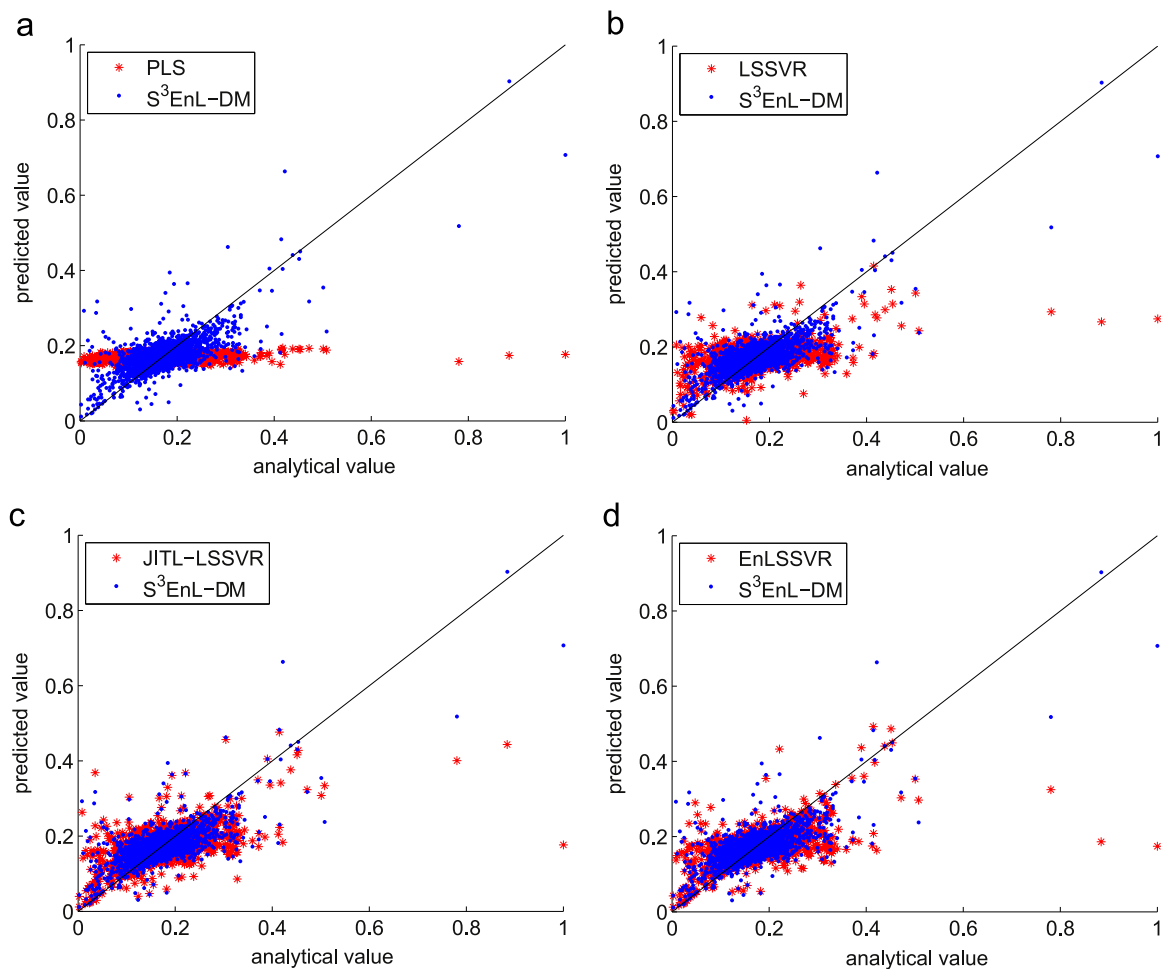


Fig. 9. Scatter plot comparisons for estimating the concentration of  $\text{SO}_2$  between the  $\text{S}^3\text{EnL-DM}$  and the PLS, the LSSVR, the JITL-LSSVR and the EnLSSVR.

Table 9

Predicted  $\text{RMSE}^{\text{test}}$  of various calibration models for estimating the concentration of benzene.

PLS	LSSVR	JITL-LSSVR	EnLSSVR	$\text{S}^3\text{EnL-DM}$
2.3782	1.3919	1.3656	1.3436	0.7947

sample and one local model because the local model region is very 'wide'. While in the case that  $p=1/2$ , when  $T > 10$ , the coverage area of the neighbors of the query sample may be wider than that of one local model region. As a result, the DM between the query sample and the local model may be changed improperly.

#### 4.2. Application to the sulfur recovery unit

In this subsection, the performance of those soft sensing methods conducted on the debutanizer column process are further tested on a sulfur recovery unit (SRU), whose flowchart is depicted in Fig. 7 [49]. The SRU process is utilized to recycle the pure sulfur and remove the pollutant that is harmful to human body and atmosphere. The inputs of the SRU process are the monoethanolamine (MEA) gas that is abundant with hydrogen sulfide ( $\text{H}_2\text{S}$ ) and the sour water stripper (SWS) gas abounding with  $\text{H}_2\text{S}$  and ammonia ( $\text{NH}_3$ ). The  $\text{H}_2\text{S}$  is transformed into pure sulfur with sulfur dioxide ( $\text{SO}_2$ ) generated. Those unconverted residual gas is the feed of a downstream process, i.e., the Maxisulfur plant for completing the final transformation. The concentrations of  $\text{H}_2\text{S}$  and  $\text{SO}_2$  contained in the exhaust gas should be monitored. In this case study, soft sensors are developed to estimate

the concentrations of  $\text{H}_2\text{S}$  and  $\text{SO}_2$ , where the secondary and primary variables are presented in Table 6.

There are 10,081 samples in the SRU dataset, which are also partitioned evenly into three groups, namely the training dataset, the validation dataset and the testing dataset. The training dataset contains half of the available samples, while both the validation dataset and the testing dataset contain a quarter of all samples. Meanwhile, in the training dataset, only 20% samples are regarded as labeled, and the rest ones are treated as unlabeled samples. The required model parameters of those employed soft sensing methods are finalized in the same ways as those used in the debutanizer column process. In the  $\text{S}^3\text{EnL-DM}$ , the window size is chosen as 20, and the confidence levels are set as 0.05. The rest parameters, namely  $J_{\text{thr}}$ ,  $\sigma$ ,  $\gamma$ , and  $T$ , are optimized by the PSO on the validation dataset. The candidate sets for  $J_{\text{thr}}$ ,  $\sigma$ ,  $\gamma$ , and  $T$  are again determined by their physical interpretation or trial-and-error, and they are presented in Table 7, while the optimized values of those parameters are included.

Estimation accuracies of various soft sensing methods in terms of  $\text{RMSE}^{\text{test}}$  are presented in Table 8. Note that in this case study, the process state partition is completed by LSSVR, requiring no assistance of PLS. Through Table 8, the advantage of the proposed methods over the other four can be confirmed, and the functionality of the utilization of unlabeled samples can be demonstrated by comparing the performance of the  $\text{SEnL-DM}$  with that of the  $\text{S}^3\text{EnL-DM}$ .

Scatter plot comparisons between the  $\text{S}^3\text{EnL-DM}$  and the PLS, the LSSVR, the JITL-LSSVR and the EnLSSVR are provided in Fig. 8 and Fig. 9, respectively, where the predictions of the PLS based soft sensor for the concentrations of both  $\text{H}_2\text{S}$  and  $\text{SO}_2$  stay around the mean of the target variable because of the strong nonlinearity in this SRU process.

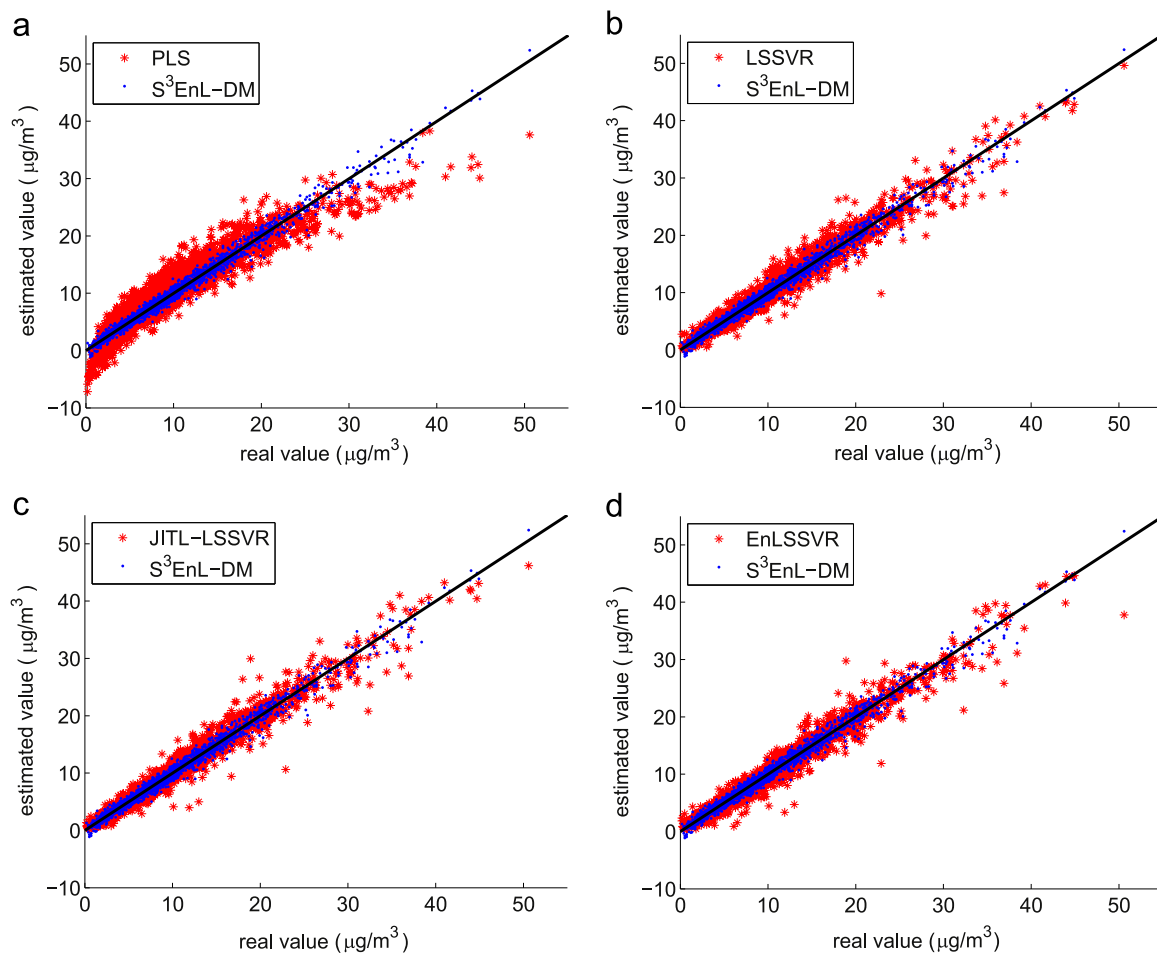


Fig. 10. Scatter plot comparisons for estimating the concentration of benzene between the  $S^3\text{EnL-DM}$  and the PLS, the LSSVR, the JITL-LSSVR and the EnLSSVR.

In addition, the scatters obtained by the proposed  $S^3\text{EnL-DM}$  method are clustered more tightly along the diagonal line for the concentrations of both  $\text{H}_2\text{S}$  and  $\text{SO}_2$ , showing that the  $S^3\text{EnL-DM}$  has better ability in dealing with this nonlinear process. In particular, large values of the concentrations of both  $\text{H}_2\text{S}$  and  $\text{SO}_2$ , for example the peak value, should be paid more attention to, as they may result in dangerous situations. In this respect, the  $S^3\text{EnL-DM}$  is more preferable, because its predictions for large values of the primary variables are more accurate compared with those of other soft sensors.

#### 4.3. Application to the urban pollution monitoring

In addition to being used for developing soft sensors in process systems, the proposed  $S^3\text{EnL-DM}$  is applicable to dealing with other nonlinear regression problems, such as the calibration of the electronic nose in the urban pollution monitoring scenario. As is known to all, pollutants in urban atmosphere could cause severe damages, and some of them such as the benzene may induce cancers when human body is exposed in it for long time. Therefore, the concentrations of these urban atmospheric pollutants need to be precisely measured. Currently, this task is primarily completed by industrial spectrometers, which are installed in spatially distributed stations [54]. Due to some reasons such as the sizes and the costs, these equipments are distributed very sparsely, preventing accurate evaluation of the distribution of gases and particles concentrations in a city. Using gas multi-sensor devices, which is referred to as ‘electronic nose’, is an effective and economic way to densify the monitoring network. However, without accurate calibration, the electronic nose usually fails to provide satisfactory estimation ability.

The calibration problem is in essence a regression problem, and the most frequently used approaches for developing the calibration model include PLS, ANN, SVR, etc. In this case study, we employ the  $S^3\text{EnL-DM}$  to construct a calibration model to perform calibration of the electronic nose for the dangerous pollutant benzene. In the calibration model, six variables, namely CO concentration,  $\text{NO}_x$  concentration,  $\text{NO}_2$  concentration,  $\text{O}_3$  concentration, temperature and humidity are selected as the input variables. Note that the concentration of non-metanic hydrocarbons (NMHC) is not included, since its analyzer failed to work 8 days after the measurement campaign started [55]. The reference data were collected at one main street located in the center of an Italian city, and is available at <http://archive.ics.uci.edu/ml/datasets.html>. The collected data covered 13 months, lasting from March 2004 to April 2005, and samples are stored hourly. After removing some missing data, 8991 samples are available.

Besides the  $S^3\text{EnL-DM}$ , the PLS, the LSSVR, the JITL-LSSVR, and the EnLSSVR are also employed to develop calibration models for comparison purpose. The entire samples are divided into training dataset, validation dataset and testing dataset in the same way as those used in the previous two case studies. The parameters of the five calibration models for estimating the concentration of benzene studied in this section are again optimized by the PSO on the validation dataset. In the  $S^3\text{EnL-DM}$ , before optimization, the significance levels are set as 0.05, the  $\text{std}_x$  is calculated as 0.242, and the remaining model parameters are optimized as follows:  $W=50$ ,  $\gamma=10617$ ,  $\sigma=1.517$ ,  $J_{\text{thr}}=0.678$ , and  $T=94$ . The search candidates for those parameters that need to be manually selected are also determined by their physical meanings or the trial-and-error manner.

The estimated RMSE of various calibration models on the testing



dataset are tabulated in Table 9, and the estimation performance comparisons between the  $S^3$ EnL-DM and other methods in terms of scatter plot are presented as Fig. 10. By comparing the estimation errors for the benzene concentration achieved by the conventional linear and nonlinear calibration models constructed by the PLS and the LSSVR, respectively, one can find that the performance of the PLS-based model is much worse than that of the LSSVR based model, indicating the strong nonlinear mathematical relationship between the input variables and the benzene concentration. In addition, Table 9 demonstrates that the calibration models constructed by those local learning based methods, namely the JITL-LSSVR, the EnLSSVR, and the  $S^3$ EnL-DM could further improve the estimation accuracy compared with the LSSVR. Furthermore, it can be seen that the predicted RMSE<sup>test</sup> obtained by the  $S^3$ EnL-DM is much smaller than those obtained by the JITL-LSSVR and the EnLSSVR, and Fig. 10 demonstrates the prediction scatters of the  $S^3$ EnL-DM are distributed much tightly and evenly along the diagonal line. These facts and comparisons show that the accuracy improvement achieved by the  $S^3$ EnL-DM based calibration model for estimating the benzene concentration is significant, which verifies the effectiveness of the  $S^3$ EnL-DM in dealing with nonlinear regression problems.

## 5. Conclusion

In this paper, a soft sensing method is proposed for nonlinear processes with limited labeled samples, which is referred to as the  $S^3$ EnL-DM. The novel contributions of this paper can be twofold. On the one hand, the unlabeled samples collected in the training dataset are incorporated, which can better describe the expertise area of the local model. On the other hand, a novel DM criterion is proposed for implementing the selective ensemble learning. These two contributions have been proven helpful to improve the performance of soft sensor using datasets coming from two real-life industrial processes. Moreover, the mechanism why the proposed soft sensor can work better is revealed and parameter sensitivity analysis of the proposed method is conducted. In addition to the process system, the proposed method is demonstrated to have effectiveness in dealing with nonlinear regression problems in the urban pollutant monitoring scenario, which constructs calibration model for the electronic nose.

It should be pointed out that although the proposed method outperforms some methods such as the PLS and the LSSVR, the improved performance is achieved at the cost of higher model complexity and more optimization time. Fortunately, the parameter determination of the proposed method can be automatically completed by the PSO, and the optimization time is consumed at the offline stage which would not impose computational burden at the online stage.

## Acknowledgement

This work was supported by the National Science Foundation of China (Grant no. 61273160) and the Fundamental Research Funds for the Central Universities (Grant no. 14CX06067A).

## References

- [1] P. Kadlec, B. Gabrys, S. Strandt, Data-driven soft sensors in the process industry, *Comput. Chem. Eng.* 33 (4) (2009) 795–814.
- [2] Z.Q. Ge, Z.Q. Song, Ensemble independent component regression models and soft sensing application, *Chemom. Intell. Lab. Syst.* 130 (15) (2014) 115–122.
- [3] W.M. Shao, X.M. Tian, Adaptive soft sensor for quality prediction of chemical processes based on selective ensemble of local partial least squares models, *Chem. Eng. Res. Des.* 95 (2015) 113–132.
- [4] A.K. Pani, H.K. Mohanta, A survey of data treatment techniques for soft sensor design, *Chem. Prod. Process Model.* 6 (1) (2011) 1–21.
- [5] M. Kano, M. Ogawa, The state of the art in chemical process control in Japan: good practice and questionnaire survey, *J. Process Control* 20 (9) (2010) 969–982.
- [6] M. Kano, K. Fujiwara, Virtual sensing technology in process industries: trends and challenges revealed by recent industrial applications, *J. Chem. Eng. Jpn.* 46 (1) (2013) 1–17.
- [7] C. Zhou, Q.Y. Liu, D.X. Huang, J. Zhang, Inferential estimation of kerosene dry point in refineries with varying crudes, *J. Process Control* 22 (6) (2011) 1122–1126.
- [8] A.J. Yan, H.S. Shao, P. Wang, A soft-sensing method of dissolved oxygen concentration by group genetic case-based reasoning with integrating group decision making, *Neurocomputing* 169 (2015) 422–429.
- [9] Z. Cheng, X.G. Liu, Optimal online soft sensor for product quality monitoring in propylene polymerization process, *Neurocomputing* 149 (2015) 1216–1224.
- [10] J. Tang, W. Yu, T.Y. Chai, L.J. Zhao, On-line principal component analysis with application to process modeling, *Neurocomputing* 82 (1) (2012) 167–178.
- [11] Z.Q. Ge, Quality prediction and analysis for large-scale processes based on multi-level principal component modeling strategy, *Control Eng. Pract.* 31 (2014) 9–23.
- [12] J. Tang, T.Y. Chai, L.J. Zhao, W. Yu, H. Yue, Soft sensors for parameters of mill load based on multi-spectral segments PLS sub-models and on-line adaptive weighted fusion algorithm, *Neurocomputing* 78 (1) (2012) 38–47.
- [13] Y.Q. Liu, Y.P. Pan, D.P. Huang, Development of a novel adaptive soft-sensor using Variational Bayesian PLS with accounting for online identification of key variables, *Ind. Eng. Chem. Res.* 54 (1) (2015) 338–350.
- [14] Z.Y. Zhang, T. Wang, X.G. Liu, Melt index prediction by aggregated RBF neural networks trained with chaotic theory, *Neurocomputing* 131 (5) (2014) 368–376.
- [15] C. Shang, F. Yang, D.X. Huang, W.X. Lyu, Data-driven soft sensor development based on deep learning technique, *J. Process Control* 24 (3) (2014) 223–233.
- [16] C. Shang, X.Q. Gao, F. Yang, D.X. Huang, Novel Bayesian framework for dynamic soft sensor based on support vector machine with finite impulse response, *IEEE Trans. Control Syst. Technol.* 22 (4) (2014) 1550–1557.
- [17] F. Sciascio, A.N. Amicarelli, Biomass estimation in batch biotechnology process by Bayesian Gaussian process regression, *Comput. Chem. Eng.* 32 (2) (2008) 3264–3273.
- [18] R. Grbić, D. Slišković, P. Kadlec, Adaptive soft sensor for online prediction and process monitoring based on a mixture of Gaussian process models, *Comput. Chem. Eng.* 58 (11) (2013) 84–97.
- [19] S. Kim, M. Kano, S. Hasebe, A. Takinami, T. Seki, Long-term industrial applications of inferential control based on just-in-time soft sensors: economic impact and challenges, *Ind. Eng. Chem. Res.* 52 (35) (2013) 12346–12356.
- [20] P. Kadlec, On Robust and Adaptive Soft Sensors (Ph.D Thesis), Bournemouth University, Poole, 2009.
- [21] Y. Liu, C.L. Li, Z.L. Gao, A novel unified correlation model using ensemble support vector regression for prediction of flooding velocity in randomly packed towers, *J. Ind. Eng. Chem.* 22 (3) (2014) 1109–1118.
- [22] A.R. Gholami, M. Shahbazian, Soft sensor design based on fuzzy C-Means and RFN\_SVR for a stripper column, *J. Nat. Gas. Sci. Eng.* 25 (2015) 23–29.
- [23] S. Khatibisepehr, B. Huang, F.W. Xu, A. Espejo, A. Bayesian, approach to design of adaptive multi-model inferential sensors with application in oil sand industry, *J. Process Control* 22 (10) (2012) 1913–1929.
- [24] J. Yu, Online quality prediction of nonlinear and non-Gaussian chemical processes with shifting dynamics using finite mixture model based Gaussian process regression approach, *Chem. Eng. Sci.* 82 (12) (2012) 22–30.
- [25] K. Fujiwara, M. Kano, S. Hasebe, A. Takinami, Soft-sensor development using correlation-based just-in-time modeling, *AIChE J.* 55 (7) (2009) 1754–1764.
- [26] P. Kadlec, B. Gabrys, Local learning based adaptive soft sensor for catalyst activation prediction, *AIChE J.* 57 (5) (2011) 1288–1301.
- [27] W.D. Ni, S.K. Tan, W.J. Ng, S.D. Brown, Localized, adaptive recursive partial least squares regression for dynamic system modeling, *Ind. Eng. Chem.* 55 (23) (2012) 8025–8039.
- [28] W.M. Shao, X.M. Tian, H.L. Chen, Adaptive anti-over-fitting soft sensing method based on local learning, in: *Preprints of 10th IFAC International Symposium Dynamics and Control of Process Systems (Mumbai, India), Dec. 18–20, 2013*, pp. 415–420.
- [29] H. Kaneko, K. Funatsu, Adaptive soft sensor based on online support vector regression and Bayesian ensemble learning for various states in chemical plants, *Chemom. Intell. Lab. Syst.* 137 (15) (2014) 57–66.
- [30] W.M. Shao, X.M. Tian, P. Wang, Local partial least squares based online soft sensing method for multi-output processes with adaptive process states division, *Chin. J. Chem. Eng.* 22 (7) (2014) 828–836.
- [31] H.P. Jin, X.G. Chen, Y.W. Yang, H. Zhang, L. Wang, L. Wu, Multi-model adaptive soft sensor modeling method using local learning and online support vector regression for nonlinear time-variant batch processes, *Chem. Eng. Sci.* 131 (2015) 282–303.
- [32] J.L. Liu, On-line soft sensor for polyethylene process with multiple production grades, *Control Eng. Pract.* 15 (7) (2007) 769–778.
- [33] Y. Lv, J.Z. Liu, T.T. Yang, D.L. Zeng, A novel least squares support vector machine ensemble model for NO<sub>x</sub> emission prediction of a coal-fired boiler, *Energy* 55 (15) (2013) 319–329.
- [34] Z.Q. Ge, B. Huang, Z.H. Song, Nonlinear semisupervised principal component regression for soft sensor modeling and its mixture form, *J. Chemom.* 28 (11) (2014) 793–804.
- [35] Z.Q. Ge, T. Chen, Z.Q. Song, Quality prediction for polypropylene production process based on CLGPR model, *Control Eng. Pract.* 19 (5) (2011) 423–432.
- [36] Z.Q. Ge, Z.H. Song, P.L. Wang, Probabilistic combination of local independent component regression model for multimode quality prediction in chemical processes, *Chem. Eng. Res. Des.* 92 (3) (2014) 501–512.
- [37] W.M. Shao, X.M. Tian, Soft sensor for nonlinear processes based on ensemble of partial least squares with adaptive localization, in: *Proceedings of the 27th Chinese Control and Decision Conference*, 743–748, May 23–25, 2015, Qingdao, China.
- [38] H. Kaneko, K. Funatsu, Classification of the degradation of soft sensor models and discussion on adaptive models, *AIChE J.* 59 (7) (2013) 2339–2347.

- [39] S. Xiang, F.P. Nie, C.S. Zhang, Semi-supervised classification via local spine regression, *IEEE Trans. Pattern Anal. Mach. Intell.* 32 (11) (2010) 2039–2053.
- [40] L. Bao, X.F. Yuan, Z.Q. Ge, Co-training partial least squares model for semi-supervised soft sensor development, *Chemom. Intell. Lab. Syst.* 147 (2015) 75–85.
- [41] Z.Q. Ge, Z.H. Song, Semisupervised Bayesian method for soft sensor modeling with unlabeled data samples, *AIChE J.* 57 (8) (2011) 2109–2119.
- [42] Z.H. Zhou, J.X. Wu, W. Tang, Ensembling neural networks: many could be better than all, *Artif. Intell.* 137 (1–2) (2002) 239–263.
- [43] B. Han, B. He, R. Nian, M.M. Ma, S.J. Zhang, M.H. Li, A. Lendasse, LARSEN-ELM: selective ensemble of extreme learning machines using LARS for blended data, *Neurocomputing* 149 (3) (2015) 285–294.
- [44] Z.Q. Ge, Active learning strategy for smart soft sensor development under a small number of labeled data samples, *J. Process Control* 24 (9) (2014) 1454–1461.
- [45] J.A.K. Suykens, T.V. Gestel, J.D. Brabanter, D.B. Moor, J. Vandewalle, *Least Squares Support Vector Machines*, World Scientific, Singapore, 2002.
- [46] Y. Shi, R.C. Eberhart, Empirical study of particle swarm optimization, in: *Proceedings of the 1999 Congress on Evolutionary Computation*, Washington: IEEE, 1999, 3, 1945–1950.
- [47] H. Kaneko, M. Arakawa, K. Funatsu, Applicability domains and accuracy of prediction of soft sensor models, *AIChE J.* 57 (6) (2011) 1506–1513.
- [48] Y. Liu, N.P. Hu, H.Q. Wang, P. Li, Soft chemical analyzer development using adaptive least squares support vector regression with selective pruning and variable moving window size, *Ind. Eng. Chem. Res.* 48 (12) (2009) 5731–5741.
- [49] L. Fortuna, S. Graziani, A. Rizzo, M.G. Xibilia, *Soft sensors for monitoring and control of industrial processes*, Springer-Verlag, London, 2007.
- [50] P. Kadlec, R. Grbić, B. Gabrys, Review of adaptation mechanisms for data-driven soft sensors, *Comput. Chem. Eng.* 35 (1) (2011) 1–24.
- [51] J.L. Zhu, Z.Q. Ge, Z.H. Song, Robust supervised probabilistic principal component analysis model for soft sensing of key process variables, *Chem. Eng. Sci.* 122 (2015) 573–584.
- [52] B.S. Dayal, J.F. MacGregor, P.L.S. Improved, algorithm, *J. Chemom.* 11 (1) (1997) 73–85.
- [53] Z.Q. Ge, Z.H. Song, A comparative study of just-in-time learning based methods for online soft sensor modeling, *Chemom. Intell. Lab. Syst.* 104 (2) (2010) 306–317.
- [54] De.S. Vito, E. Massera, M. Piga, L. Martinotto, Di.G. Francia, On field calibration of an electronic nose for benzene estimation in an urban pollution monitoring scenario, *Sens. Actuators* 129 (2) (2008) 750–757.
- [55] De.S. Vito, M. Piga, L. Martinotto, Di.G. Francia, CO, NO<sub>2</sub> and NO<sub>x</sub> urban pollution monitoring with on-field calibrated electronic nose by automatic bayesian regularization, *Sens. Actuators* 143 (1) (2009) 182–191.



**Weiming Shao** was born in Shandong, China, on September 22, 1986. He received his Bachelor degree and Doctor degree in 2009 and 2016, respectively, both from China University of Petroleum. His current research interests focus on statistical learning and machine learning based adaptive modeling of industrial processes.



**Xuemin Tian** received his Bachelor of Engineering from Huadong Petroleum Institute, Dongying, China, in January 1982, and his M.S. degree from Beijing University of Petroleum, Beijing, China, in June 1994. From September 2001 to June 2002, he served as a visiting professor at Center of Process Control, University of California in Santa Barbara. He is a professor of Process Control at China University of Petroleum (Hua Dong). Professor Tian's research interests are in modeling, advanced process control and optimization for petrol-chemical processes as well as fault detection and diagnosis, and process monitoring.