IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 3, 2018, accepted September 17, 2018, date of publication October 1, 2018, date of current version October 19, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2872752

Adaptive Soft Sensor Development for Multi-Output Industrial Processes Based on Selective Ensemble Learning

WEIMING SHAO[1], SHENG CHEN[2,3], (Fellow, IEEE), AND CHRIS J. HARRIS[2]

[1]College of Control Science and Engineering, Zhejiang University, Hangzhou 310007, China
[2]School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K.
[3]King Abdulaziz University, Jeddah 21589, Saudi Arabia

Corresponding author: Sheng Chen (sqc@ecs.soton.ac.uk)

This work was supported in part by the National Natural Science Foundation of China under Grant 61703367 and in part by the China Postdoctoral Science Foundation.

ABSTRACT Soft sensors are vital for online predictions of quality-related yet difficult-to-measure variables in process industry. In this paper, an adaptive soft sensing approach based on selective ensemble learning is proposed for multi-output nonlinear and time-varying industrial processes, which we refer to as the selective ensemble learning for multi-outputs (SEL-MO). Specifically, an adaptive localization approach is developed for dealing with the process nonlinearity based on the statistical hypothesis testing theory, which can construct redundancy-free local model set. At the online operation stage, these constructed local models are partially combined under an adaptive selective ensemble learning framework, where the weightings of local models are query-sample-oriented such that both gradual and abrupt changes in the process characteristics can be handled. In addition, an insensitivity strategy is proposed to enhance the online computational efficiency of the SEL-MO by avoiding the unnecessary search of the historical data set. Case studies are carried out on a simulated fed-batch penicillin process and a real-life industrial primary reformer, and the results obtained demonstrate the effectiveness of the proposed method.

INDEX TERMS Adaptive soft sensor, multi-output process, adaptive localization, statistical hypothesis testing, selective ensemble learning.

I. INTRODUCTION

In many modern industrial processes, a class of variables closely related to product quality, known as the primary variables, must be monitored and controlled online in real-time, in order to increase profit as well as to enhance process safety and environmental protection [1]. Measurements of these primary variables often suffer from large measurement delay, high maintenance cost or low accuracy. Alternatively, they can be estimated by soft sensors, which are free from measurement delay, easy to maintain, and able to provide accurate and reliable estimations of the primary variables. Therefore, in recent years, the soft sensing technology has gained fast-increasing applications in many industrial processes, such as distillation column process, polymer production process, chemical reactor, etc [2], [3].

Data-driven soft sensors have good versatility and do not require in-depth knowledge of the target process that is usually difficult to master. Moreover, they are capable of describing the true process conditions well because they are based on the real process data directly collected from plants [4]. With the wide application of computer technology based distributed control systems, abundant process data have been gathered [5]. In the past few decades, various statistical and machine learning methods have become powerful tools to develop predictive data-driven soft sensors. Extensive reviews of these soft sensing approaches and their applications to industrial processes can be found in [4] and [6]. Linear modeling approaches, in particular the partial least squares (PLS), seem to be dominant methods for applications to industrial plants, according to a questionnaire survey for process control of chemical plants in Japan [7]. However, linear soft sensors cannot perform well in nonlinear processes. An effective solution is to embed linear modeling within the local learning framework, such as the ensemble learning [1].

Most industrial processes also exhibit time-varying characteristics due to the process drifts resulting from changes

of operating conditions and raw materials, catalyst deactivation, mechanical abrasions, or external climatic variations, etc [3], [7]. The adverse consequence of process drifts, which can be either gradual or abrupt, is causing the performance deterioration of soft sensor, in particular, failure to maintain long-term high-accuracy. This is recognized as the most crucial problem encountered in applications [7], [8]. Various adaptive mechanisms are employed, which can realize online self-updating with newly measured labeled samples in order for soft sensors to maintain satisfactory performance over a long operating period. Commonly adopted adaptive strategies include recursive strategy [9]–[11], moving-window (MW) strategy [12]–[14], just-in-time learning (JITL) strategy [15]–[17], ensemble learning strategy [18], etc.

Recursive methods generally handle gradual changes well but are unable to give accurate predictions in time after abrupt changes. Moreover, recursive methods model the entire set of process states with a single global model, and may not handle well the situations of process-state dependent nonlinearities. MW models also provide high-performance for gradually changing processes and can deal effectively with the degradations caused by drifts of secondary variables, but they share the same drawback with recursive models in handling abrupt changes. JITL-based models, which belong to the local learning framework, can tackle well the process nonlinearity as well as both gradual and abrupt changes simultaneously. However, failures in accurately selecting relevant samples owing to the reasons such as measurement noise or drifts of secondary variables often prevent JITL-based soft sensors to perform satisfactorily.

Unlike the above-discussed single model-based adaptive strategies, ensemble learning-based soft sensors, which first construct a group of local models and then combine their outputs together for prediction, are able to handle process nonlinearities and time-varying changes simultaneously. Ensemble methods have recently become focal point in soft sensor development [1], [5], [19]–[22]. To effectively deal with process nonlinearity, ensemble learning performs localization to partition the process states into local model regions upon which corresponding local models are constructed. Localization methods can be categorized into two groups, the distance-based ones and the MW-based ones. The former, which gathers spatially closed samples together, typically includes clustering methods, Gaussian mixture models [23] and JITL methods [3], [24]. However, there exist some issues associated with this group. For example, clustering methods cannot perform online inclusion of new cluster members without retraining from scratch. Although JITL methods can overcome this drawback, they may have difficulty in properly building the probabilistic data descriptor models, due to high-dimensionality of secondary variables and co-linearity as a result of ill-conditioned covariance matrices.

MW-based methods, which gather time-relevant samples to construct local model regions by exploiting the fact that 'data measured close in time have strong relationships and correlations' [25], are able to solve the problems related to clustering-based and JITL-based localization methods. The key task in MW-based methods is to judge whether two local model regions are significantly different. To fulfill this task, the studies [3], [19], [24] employed the statistical hypothesis testing, t-test, to detect the significant difference between two predicted residuals' means. The t-test is a versatile supervised method. However, it neglects the influence of the second-order information of the predicted residuals, which may lead to improper judgement. To deal with this issue, employing χ^2-test in addition to t-test was proposed for performing hypothesis testing [26], where the effectiveness of considering the variance information was confirmed.

To prevent performance degradation, ensemble learning needs to provide two levels of online adaptation. The first one is the adaptation of local model set, including updating local models recursively [23], [24] and online inclusion/exclusion of local models [26], [27]. The later is advantageous in following re-occurred process states. The second adaptation level is the adaptation of the combining weights for local models. Commonly used adaptive weighting criteria include Bayesian posterior probability [23], [24], monitoring statistics (such as T^2, I^2 and Q) [5], estimated prediction variance [28], predicted errors for the newest samples [22], [29]. In addition, weighing local models by mining the information contained in both labeled and unlabeled samples was proposed to handle both abrupt and gradual changes in process characteristics [1], [27]. Recently, it has been recognized that combing part of ensemble members rather than all of them may achieve better performance, and several selective ensemble learning (SEL) strategies were proposed. For example, Zhou *et al.* [30] and Zhou *et al.* [31] first proved that partial ensembling of neural networks could perform better, and subsequently Zhao and Liu used this philosophy to model the waste water treatment plant with extreme learning machines where good performance was achieved. Further, Kaneko and Funatsu combined part of local support vector regression models for soft sensor development, and their experimental results demonstrated high-accuracy of SEL [29]. An adaptive framework was also developed for SEL, which is capable of balancing prediction bias and variance, and illustrated the superiority of the SEL over the traditional ensemble learning strategy [27].

Similar to most of the data-based soft sensors discussed above, the SEL-based adaptive soft sensors developed in our previous works [1], [26], [27] were for single-output soft sensors, i.e., soft sensors for predicting single primary variable, which is referred to as the SEL-SO. However, a great deal of industrial processes have multiple primary variables to predict, and a naive approach would be to construct multiple single-output soft sensors, one for each primary variable, by directly using a single-output soft sensor design. But this is inappropriate because it will not only dramatically increase the cost but also ignore the inherent relationships between the primary variables. Luckily, many data-based soft sensor designs originally developed for the single-primary-variable

case, such as the PLS [32], the locally weighted PLS (LW-PLS) [33], the MWPLS [13] and the JITL-PLS [34], can easily be extended to design a multi-output soft sensor for predicting multiple primary variables. However, it is much more challenging to extend the SEL-SO to multi-output soft sensor design. This is because the key components of the SEL-SO, including its adaptive localization scheme, adaptive local model weighting strategy and adaptive selective ensemble learning framework, are restricted to single-output modeling and they are not applicable to multi-output modeling. Extending these components to be applicable to the multi-output case is complicated and certainly nontrivial. Against this background, this paper develops a new SEL-based adaptive soft sensor for multi-output modeling, referred to as the SEL-MO. Our main contributions are as follows.

- Because the adaptive localization scheme of the SEL-SO cannot be applied to the multi-output process, we develop a new adaptive localization approach with the appropriate statistics for multi-output model hypothesis testing in order to construct an accurate and redundancy-free multi-output local model set.
- During online operation, the local models' weights may be updated too frequently, which imposes heavy online computational burden, particularly, for multi-output processes. We propose a new insensitivity strategy for adaptive weighting, which significantly reduces online complexity while maintaining satisfactory performance.
- Because the adaptive selective ensemble learning (SEL) framework developed in our previous works cannot be applied directly to the multi-output case, we develop the theoretical framework necessary for adaptive multi-output soft sensors and, therefore, complete the design of the effective SEL-MO.

Two case studies, a simulated fed-batch penicillin process and a real-life industrial primary reformer, are included to evaluate the effectiveness of our proposed SEL-MO design as well as to demonstrate its superior performance over several state-of-the-art adaptive multi-output soft sensor designs.

II. PARTIAL LEAST SQUARES

Since it is a basic regression algorithm in our SEL-MO, the PLS algorithm is first briefly reviewed.

Let the m- and p-dimensional input and output vectors of the ith sample be $\boldsymbol{x}_i = \left[x_{1,i}\ x_{2,i} \cdots x_{m,i}\right]^{\mathrm{T}} \in \mathbb{R}^m$ and $\boldsymbol{y}_i = \left[y_{1,i}\ y_{2,i} \cdots y_{p,i}\right]^{\mathrm{T}} \in \mathbb{R}^p$, respectively, and assume that there are N data samples for model construction. Then the input and output data matrices are denoted by $\boldsymbol{X} = [\boldsymbol{x}_1\, \boldsymbol{x}_2 \cdots \boldsymbol{x}_N]^{\mathrm{T}} \in \mathbb{R}^{N\times m}$ and $\boldsymbol{Y} = \left[\boldsymbol{y}_1\, \boldsymbol{y}_2 \cdots \boldsymbol{y}_N\right]^{\mathrm{T}} \in \mathbb{R}^{N\times p}$, respectively. Assuming that the modeling data have been mean-centered and appropriately scaled, the PLS algorithm models the mapping relationship between $\boldsymbol{X}$ and $\boldsymbol{Y}$ as

$$\boldsymbol{Y} \approx \boldsymbol{X}\boldsymbol{C}^{PLS}, \tag{1}$$

where $\boldsymbol{C}^{PLS} = \left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\right)^{+}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y} \in \mathbb{R}^{m\times p}$ is the regression coefficient matrix, and $(\bullet)^{+}$ denotes the generalized inverse operator. In the PLS algorithm, data matrices $\boldsymbol{X}$ and $\boldsymbol{Y}$ are first decomposed respectively as

$$\boldsymbol{X} = \sum\nolimits_{i=1}^{A} \boldsymbol{t}_i\boldsymbol{p}_i^{\mathrm{T}} + \boldsymbol{E}_{\boldsymbol{X}} = \boldsymbol{T}\boldsymbol{P}^{\mathrm{T}} + \boldsymbol{E}_{\boldsymbol{X}}, \tag{2}$$

$$\boldsymbol{Y} = \sum\nolimits_{i=1}^{A} \boldsymbol{u}_i\boldsymbol{q}_i^{\mathrm{T}} + \boldsymbol{E}_{\boldsymbol{Y}} = \boldsymbol{U}\boldsymbol{Q}^{\mathrm{T}} + \boldsymbol{E}_{\boldsymbol{Y}}, \tag{3}$$

where A denotes the number of latent variables, $\boldsymbol{T} = \left[\boldsymbol{t}_1 \cdots \boldsymbol{t}_A\right]^{\mathrm{T}} \in \mathbb{R}^{N\times A}$ and $\boldsymbol{U} = \left[\boldsymbol{u}_1 \cdots \boldsymbol{u}_A\right]^{\mathrm{T}} \in \mathbb{R}^{N\times A}$ denote the score matrices, and $\boldsymbol{P} = \left[\boldsymbol{p}_1 \cdots \boldsymbol{p}_A\right]^{\mathrm{T}} \in \mathbb{R}^{m\times A}$ and $\boldsymbol{Q} = \left[\boldsymbol{q}_1 \cdots \boldsymbol{q}_A\right]^{\mathrm{T}} \in \mathbb{R}^{p\times A}$ represent the loading matrices of $\boldsymbol{X}$ and $\boldsymbol{Y}$, respectively, while $\boldsymbol{E}_{\boldsymbol{X}} \in \mathbb{R}^{N\times m}$ and $\boldsymbol{E}_{\boldsymbol{Y}} \in \mathbb{R}^{N\times m}$ denote the respective residual matrices.

Realizing (2) and (3) involves iterative operations, where in the ith iteration it is required that $\boldsymbol{t}_i$ and $\boldsymbol{u}_i$ can maximally represent the variations of $\boldsymbol{X}$ and $\boldsymbol{Y}$, respectively, while $\boldsymbol{t}_i$ can best explain $\boldsymbol{u}_i$. Thus in the ith iteration, the PLS algorithm needs to solve the following optimization problem:

$$\max\ cov\left(\boldsymbol{t}_i, \boldsymbol{u}_i\right) = \max \sqrt{var\left(\boldsymbol{t}_i\right) var\left(\boldsymbol{u}_i\right)}\, r\left(\boldsymbol{t}_i, \boldsymbol{u}_i\right), \tag{4}$$

where $cov\,(\bullet, \bullet)$, $var\,(\bullet)$ and $r\,(\bullet, \bullet)$ represent the operators of calculating covariance, variance and correlation coefficient, respectively. The optimization (4) can be transformed into an eigenvector decomposition problem, and the linear relationship between $\boldsymbol{t}_i$ and $\boldsymbol{u}_i$ can be obtained via the least squares regression as $\boldsymbol{u}_i = b_i\boldsymbol{t}_i$ with $b_i = \boldsymbol{t}_i^{\mathrm{T}}\boldsymbol{u}_i / \boldsymbol{t}_i^{\mathrm{T}}\boldsymbol{t}_i$. Pseudo-codes and computer programs for the PLS algorithm can readily be found in [9] and [32].

III. PROPOSED SEL-MO ADAPTIVE SOFT SENSOR

In order to deal with process nonlinearities, the SEL-MO performs adaptive localization partitioning of the process states into local model regions, on which locally valid PLS models are constructed. These operations are implemented at the offline stage. At the online operation stage, query sample-oriented SEL strategy is designed, which adaptively determines the combining weights of ensemble members in order to provide the soft sensor with adaptive capability.

A. ADAPTIVE LOCALIZATION VIA STATISTICAL HYPOTHESIS TESTING

The task of the localization is to establish the local experts $\{\boldsymbol{f}_l\}_{l=1}^{L}$ that are valid in their corresponding local model regions represented by the sub-datasets $\{\boldsymbol{X}_l, \boldsymbol{Y}_l\}_{l=1}^{L}$. The schematic of the proposed localization approach is illustrated in Fig. 1. The basic idea is as follows. A data window $\boldsymbol{W}_{ini} = \{\boldsymbol{X}_{ini} \in \mathbb{R}^{W\times m}, \boldsymbol{Y}_{ini} \in \mathbb{R}^{W\times p}\}$ with W consecutive-time samples is initially set, and an initial local model $\boldsymbol{f}_{ini}$ is built on it by the PLS algorithm. Subsequently, a shifted window $\boldsymbol{W}_{sft} = \{\boldsymbol{X}_{sft} \in \mathbb{R}^{W\times m}, \boldsymbol{Y}_{sft} \in \mathbb{R}^{W\times p}\}$ is obtained by moving the window one sample step ahead. If the two local regions, $\boldsymbol{W}_{ini}$ and $\boldsymbol{W}_{sft}$, are not significantly different, it is considered that the samples within $\boldsymbol{W}_{sft}$ come from the same process state as those within $\boldsymbol{W}_{ini}$, and the window is continued to be shifted forward. Otherwise, $\boldsymbol{W}_{sft}$ is considered to represent a new state different from the one represented by $\boldsymbol{W}_{ini}$, and a

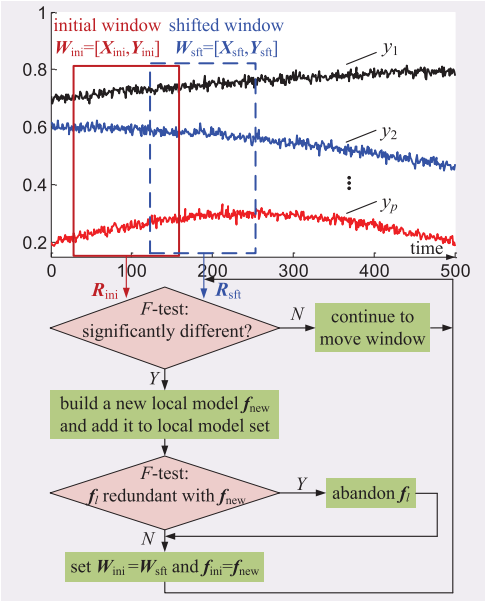


FIGURE 1. Schematic of the adaptive localization for multi-output soft sensors.

new local model $\boldsymbol{f}_{new}$ should be constructed on $\boldsymbol{W}_{sft}$ and it is added to the local model set.

Denote the predicted residual matrices of $\boldsymbol{f}_{ini}$ for the data samples within $\boldsymbol{W}_{ini}$ and $\boldsymbol{W}_{sft}$ as

$$\boldsymbol{R}_{ini} = \boldsymbol{f}_{ini}(\boldsymbol{X}_{ini}) - \boldsymbol{Y}_{ini}, \tag{5}$$

$$\boldsymbol{R}_{sft} = \boldsymbol{f}_{ini}(\boldsymbol{X}_{sft}) - \boldsymbol{Y}_{sft}, \tag{6}$$

respectively, where $\boldsymbol{R}_{ini} \in \mathbb{R}^{W\times p}$ and $\boldsymbol{R}_{sft} \in \mathbb{R}^{W\times p}$. If $\boldsymbol{R}_{ini}$ and $\boldsymbol{R}_{sft}$ are not significantly different, the performance of $\boldsymbol{f}_{ini}$ on $\boldsymbol{W}_{ini}$ and $\boldsymbol{W}_{sft}$ can be considered to be similar, implying that the two data windows are essentially the same. As $\boldsymbol{f}_{ini}$ is a multi-output linear model, $\boldsymbol{R}_{ini}$ and $\boldsymbol{R}_{sft}$ are considered not significantly different when both their mean vectors and covariance matrices are the same. Accordingly, the null hypothesis, H_0^{μ} and H_0^{Σ}, are set as

$$H_0^{\mu} : \boldsymbol{\mu}_{sft} = \boldsymbol{\mu}_{ini}, \tag{7}$$

$$H_0^{\Sigma} : \boldsymbol{\Sigma}_{sft} = \boldsymbol{\Sigma}_{ini}, \tag{8}$$

in which $\boldsymbol{\mu}_{sft}$ and $\boldsymbol{\mu}_{ini}$ are the mean vectors of the populations where the samples of $\boldsymbol{R}_{sft}$ and $\boldsymbol{R}_{ini}$ come from, respectively, while $\boldsymbol{\Sigma}_{sft}$ and $\boldsymbol{\Sigma}_{ini}$ are the covariance matrices of the populations where $\boldsymbol{R}_{sft}$ and $\boldsymbol{R}_{ini}$ come from, respectively. Moreover, $\boldsymbol{\mu}_{ini}$ and $\boldsymbol{\Sigma}_{ini}$ are estimated directly based on $\boldsymbol{R}_{ini}$, yielding $\boldsymbol{\mu}_{ini} = \boldsymbol{0}$ and $\boldsymbol{\Sigma}_{ini} \approx \frac{1}{W-1}\boldsymbol{R}_{ini}^{\mathrm{T}}\boldsymbol{R}_{ini}$. Note that in [35], only the null hypothesis test H_0^{μ} was employed to judge whether $\boldsymbol{R}_{ini}$ and $\boldsymbol{R}_{sft}$ are the same, which is insufficient. In order to determine whether or not to accept H_0^{μ}, the statistic F_{μ} is constructed as

$$F_{\mu} = \frac{(W-p)W}{(W-1)p}\left(\widehat{\boldsymbol{\mu}}_{sft} - \boldsymbol{\mu}_{ini}\right)\widehat{\boldsymbol{\Sigma}}_{sft}^{-1}\left(\widehat{\boldsymbol{\mu}}_{sft} - \boldsymbol{\mu}_{ini}\right)^{\mathrm{T}}, \tag{9}$$

where $\widehat{\boldsymbol{\mu}}_{sft}$ and $\widehat{\boldsymbol{\Sigma}}_{sft}$ are the mean vector and covariance matrix of $\boldsymbol{R}_{sft}$, respectively, estimated with $\widehat{\boldsymbol{\Sigma}}_{sft} = \frac{1}{W-1}\sum_{i=1}^{W}\left(\boldsymbol{R}_{sft}(i,:) - \widehat{\boldsymbol{\mu}}_{sft}\right)^{\mathrm{T}}\left(\boldsymbol{R}_{sft}(i,:) - \widehat{\boldsymbol{\mu}}_{sft}\right)$ and $\widehat{\boldsymbol{\mu}}_{sft} = \frac{1}{W}\sum_{i=1}^{W}\boldsymbol{R}_{sft}(i,:)$ in which $\boldsymbol{R}_{sft}(i,:)$ denotes the ith row of $\boldsymbol{R}_{sft}$. Assuming that both $\boldsymbol{R}_{ini}$ and $\boldsymbol{R}_{sft}$ follow the multivariate normal distributions, then F_{μ} follows the F distribution with the degrees of freedom p and $W-p$, denoted as $F_{\mu} \sim \mathcal{F}(p, W-p)$, when H_0^{μ} holds. Therefore, the condition of accepting H_0^{μ} is

$$F_{\mu} < \lambda_{\mu}, \tag{10}$$

where λ_{μ} is the threshold value given the significance level α_{μ} which satisfies Pro $\{F_{\mu} < \lambda_{\mu}\} = 1-\alpha_{\mu}$.

In order to test whether to accept H_0^{Σ} or not, we consider the following statistic

$$F_{\Sigma} = \frac{W-1}{b}\left(\ln|\boldsymbol{\Sigma}_{ini}| - p - \ln\left|\frac{\widehat{\boldsymbol{\Sigma}}_{sft}}{W-1}\right| + \mathrm{tr}\left(\frac{\widehat{\boldsymbol{\Sigma}}_{sft}\boldsymbol{\Sigma}_{ini}^{-1}}{W-1}\right)\right), \tag{11}$$

where $|\bullet|$ and tr $(\bullet)$ denote the determinant and trace operators, respectively, while $b = \frac{z_1}{1-D_1-z_1/z_2}$, $z_1 = \frac{p(p+1)}{2}$, $z_2 = \frac{z_1+2}{D_2-D_1^2}$, $D_1 = \frac{2p+1-2/(p+1)}{6(W-1)}$, and $D_2 = \frac{(p-1)(p+2)}{6(W-1)^2}$. When W or p is sufficiently large and H_0^{Σ} holds, F_{Σ} approximately follows the F distribution with the degrees of freedom z_1 and z_2, and the condition of accepting H_0^{Σ} is

$$F_{\Sigma} < \lambda_{\Sigma}, \tag{12}$$

where λ_{Σ} is the threshold value given the significance level α_{Σ} which satisfies Pro $\{F_{\Sigma} < \lambda_{\Sigma}\} = 1-\alpha_{\Sigma}$.

In summary, $\boldsymbol{W}_{ini}$ and $\boldsymbol{W}_{sft}$ are not significantly different only when both (10) and (12) are satisfied.

Assume that the local model set consists of $L > 1$ independent local models $\{\boldsymbol{f}_l\}_{l=1}^{L}$, and $\boldsymbol{f}_{ini} = \boldsymbol{f}_L$ is used. When $\boldsymbol{W}_{ini}$ and $\boldsymbol{W}_{sft}$ are tested to be different, the new local model $\boldsymbol{f}_{new}$ is different from $\boldsymbol{f}_L$. However, it cannot be asserted that $\boldsymbol{f}_{new}$ differs from the rest $L-1$ models in $\{\boldsymbol{f}_l\}_{l=1}^{L-1}$, as some process states may re-occur. Thus, redundancy check after the detection of the new local model is necessary to delete redundant model. This task is also fulfilled based on the statistical hypothesis testing. Specifically, let the predicted residual matrices $\boldsymbol{R}_{new}$ and $\boldsymbol{R}_l$, calculated on the samples of $\boldsymbol{W}_{sft}$ by $\boldsymbol{f}_{new}$ and the lth local model $\boldsymbol{f}_l$ for $1 \le l \le L-1$, be denoted respectively as

$$\boldsymbol{R}_{new} = \boldsymbol{f}_{new}\left(\boldsymbol{X}_{sft}\right) - \boldsymbol{Y}_{sft}, \tag{13}$$

$$\boldsymbol{R}_l = \boldsymbol{f}_l\left(\boldsymbol{X}_{sft}\right) - \boldsymbol{Y}_{sft}. \tag{14}$$

To test whether $\boldsymbol{R}_{new}$ and $\boldsymbol{R}_l$ are significantly different or not, the two null hypotheses, $H_0^{\mu_l}$ and $H_0^{\Sigma_l}$, are set as

$$H_0^{\mu_l} : \boldsymbol{\mu}_l = \boldsymbol{\mu}_{new}, \tag{15}$$

$$H_0^{\Sigma_l} : \boldsymbol{\Sigma}_l = \boldsymbol{\Sigma}_{new}, \tag{16}$$

in which $\boldsymbol{\mu}_l$ and $\boldsymbol{\mu}_{new}$ are the mean vectors of the populations where the samples of $\boldsymbol{R}_l$ and $\boldsymbol{R}_{new}$ come from, respectively, while $\boldsymbol{\Sigma}_l$ and $\boldsymbol{\Sigma}_{new}$ are the covariance matrices of the populations where the samples of $\boldsymbol{R}_l$ and $\boldsymbol{R}_{new}$ come from, respectively. Here $\boldsymbol{\mu}_{new}$ and $\boldsymbol{\Sigma}_{new}$ are estimated based on $\boldsymbol{R}_{new}$, respectively, as $\boldsymbol{\mu}_{new} = \boldsymbol{0}$ and $\boldsymbol{\Sigma}_{new} \approx \frac{1}{W-1}\boldsymbol{R}_{new}^{\mathrm{T}}\boldsymbol{R}_{new}$.

Algorithm 1 Adaptive Localization Procedure for Multi-Output Soft Sensors

Step 1) Initialization.
Set $\boldsymbol{W}_{ini}$ with W consecutive-time samples from historical dataset, and construct local PLS model $\boldsymbol{f}_{ini}$ on $\boldsymbol{W}_{ini}$;
Calculate $\boldsymbol{R}_{ini}$ using (5), and estimate $\boldsymbol{\mu}_{ini}$ and $\boldsymbol{\Sigma}_{ini}$;
Set $L = 1, \boldsymbol{f}_L = \boldsymbol{f}_{ini}$, and $\boldsymbol{W}_{sft} = \boldsymbol{W}_{ini}$;
Step 2) New local model detection.
Step 2a) Shift $\boldsymbol{W}_{sft}$ one sample step ahead to get new $\boldsymbol{W}_{sft}$, calculate $\boldsymbol{R}_{sft}$ using (6), and estimate $\widehat{\boldsymbol{\mu}}_{sft}$ and $\widehat{\boldsymbol{\Sigma}}_{sft}$;
Construct F_{μ} and F_{Σ} using (9) and (11), respectively;
if both conditions (10) and (12) are satisfied **then**
  Go to *Step 2a)*;
end if
Build a new model $\boldsymbol{f}_{new}$ on $\boldsymbol{W}_{sft}$, calculate $\boldsymbol{R}_{new}$ using (13), and estimate $\boldsymbol{\mu}_{new}$ and $\boldsymbol{\Sigma}_{new}$;
Step 3) Redundant local model check.
for $l = 1, 2, \cdots, L-1$ **do**
  Calculate $\boldsymbol{R}_l$ according to (14), and estimate corresponding $\widehat{\boldsymbol{\mu}}_l$ and $\widehat{\boldsymbol{\Sigma}}_l$;
  Construct $F_{\boldsymbol{\mu}}^{(l)}$ and $F_{\boldsymbol{\Sigma}}^{(l)}$ according to (17) and (18);
  if condition (19) is satisfied **then**
    Delete $\boldsymbol{f}_l$, set $\boldsymbol{f}_i = \boldsymbol{f}_{i+1}$ for $i = l, l+1, \cdots, L-1$, set $L = L-1$, and go to *Step 3a)*;
  end if
end for
Step 3a) Set $L = L+1$, add $\boldsymbol{f}_{new}$ to local model set as $\boldsymbol{f}_L$;
Set $\boldsymbol{W}_{ini} = \boldsymbol{W}_{sft}$ and $\boldsymbol{f}_{ini} = \boldsymbol{f}_{new}$, and return to **Step 2**);

Subsequently, the following two statistics are constructed

$$F_{\boldsymbol{\mu}}^{(l)} = \frac{(W-p)W}{(W-1)p}\left(\widehat{\boldsymbol{\mu}}_l - \boldsymbol{\mu}_{new}\right)\widehat{\boldsymbol{\Sigma}}_l^{-1}\left(\widehat{\boldsymbol{\mu}}_l - \boldsymbol{\mu}_{new}\right)^{\mathrm{T}}, \tag{17}$$

$$F_{\Sigma}^{(l)} = \frac{W-1}{b}\left(\ln|\boldsymbol{\Sigma}_{new}| - p - \ln\left|\frac{\widehat{\boldsymbol{\Sigma}}_l}{W-1}\right| + \mathrm{tr}\left(\frac{\widehat{\boldsymbol{\Sigma}}_l \boldsymbol{\Sigma}_{new}^{-1}}{W-1}\right)\right), \tag{18}$$

where $\widehat{\boldsymbol{\mu}}_l$ and $\widehat{\boldsymbol{\Sigma}}_l$ are the mean vector and covariance matrix of $\boldsymbol{R}_l$, with $\widehat{\boldsymbol{\mu}}_l = \frac{1}{W}\sum_{i=1}^{W}\boldsymbol{R}_l(i,:)$ and $\widehat{\boldsymbol{\Sigma}}_l = \frac{1}{W-1}\sum_{i=1}^{W}\left(\boldsymbol{R}_l(i,:) - \widehat{\boldsymbol{\mu}}_l\right)^{\mathrm{T}}\left(\boldsymbol{R}_l(i,:) - \widehat{\boldsymbol{\mu}}_l\right)$. Under the assumption that $\boldsymbol{R}_l$ and $\boldsymbol{R}_{new}$ follow the multivariate normal distributions, $F_{\boldsymbol{\mu}}^{(l)} \sim \mathcal{F}(p, W-p)$ when $H_0^{\mu_l}$ holds, and $F_{\Sigma}^{(l)} \sim \mathcal{F}(z_1, z_2)$ when $H_0^{\Sigma_l}$ holds. Hence, $\boldsymbol{f}_l$ and $\boldsymbol{f}_{new}$ are regarded to be identical if the following condition is met

$$F_{\boldsymbol{\mu}}^{(l)} < \lambda_{\mu} \;\&\; F_{\Sigma}^{(l)} < \lambda_{\Sigma}. \tag{19}$$

Under this circumstance, either $\boldsymbol{f}_l$ or $\boldsymbol{f}_{new}$ is redundant and one of them should be removed. Since $\boldsymbol{f}_l$ is relatively 'older' than $\boldsymbol{f}_{new}$, $\boldsymbol{f}_{new}$ is kept and $\boldsymbol{f}_l$ is removed.

The proposed adaptive localization procedure is summarized in Algorithm 1.

Remark 1: At the online stage, the process will record the new secondary-variable measurements or samples $\{\boldsymbol{x}_q\}$.

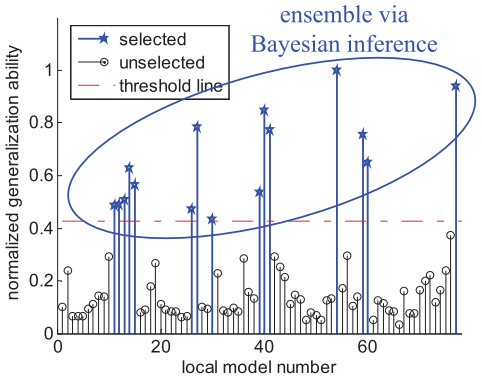


FIGURE 2. Schematic diagram of the selective ensemble learning for multi-output soft sensors.

After the labels or the true outputs $\boldsymbol{y}_q$ associated with these secondary-variable samples $\boldsymbol{x}_q$ are eventually known, for example, provided by online analyzer in a hour later, the newly acquired 'online' labeled data $\{\boldsymbol{x}_q, \boldsymbol{y}_q\}$ can be added to the historical dataset, and the above adaptive localization procedure can continue by moving the window forward to identify new local models and discard redundant models from the local model set. In this way, any newly emerging process state during the (just finished) online operation period can be captured and added to the local model set.

Remark 2: The significance levels, α_{μ} and α_{Σ}, are usually set to small values, e.g., 0.05. The selection of the window size W is a trade-off between the adaptation capability for capturing fast time-varying process states and the accuracy of the statistics. How to determine appropriate values for the algorithmic parameters of the SEL-MO, including W, is addressed in Section III-C.

B. SELECTIVE ENSEMBLE LEARNING WITH INSENSITIVITY STRATEGY

Fig. 2 depicts the schematic diagram of the proposed SEL-MO for adaptive soft sensor, which makes an ensemble of partial members of the local models generated by the localization scheme to estimate the output of a query sample $\boldsymbol{x}_q$. There are three main steps in forming the SEL-MO based soft sensor: 1) the generalization ability of each local model for the given $\boldsymbol{x}_q$ is measured; 2) some models with good generalization abilities are selected; and 3) those selected local models are combined together appropriately. Accordingly, there are three questions that must be answered:

- Q 1: how to quantify the generalization ability of each local model given query sample $\boldsymbol{x}_q$?
- Q 2: how to determine which models should be selected and which models should be filtered out?
- Q 3: how to make the ensemble of those selected models appropriately?

To answer Q 1, the generalization ability of the lth local model $\boldsymbol{f}_l$, denoted as $\zeta(l)$, for $1 \le l \le L$, must be quantified. It has been shown that the information contained in the newest labeled sample and the neighbors of the query sample can

be utilized to measure the generalization ability of a single-output local model [26], [27], [35]. Let the newest labeled sample be denoted as $\{\boldsymbol{x}_0, \boldsymbol{y}_0\}$. Assume that K nearest neighbors of $\boldsymbol{x}_q$ from the historical dataset are utilized, which are denoted by $\{\boldsymbol{x}_k^{(q)}, \boldsymbol{y}_k^{(q)}\}_{k=1}^K$. Further define the absolute error vector $\boldsymbol{e}_0^{(l)} \in \mathbb{R}^p$ of the lth local model for $\{\boldsymbol{x}_0, \boldsymbol{y}_0\}$ as

$$\boldsymbol{e}_0^{(l)}[j] = \left|\boldsymbol{f}_l(\boldsymbol{x}_0)[j] - \boldsymbol{y}_0[j]\right|, \tag{20}$$

and the absolute error vectors $\boldsymbol{e}_k^{(l)} \in \mathbb{R}^p$, $1 \le k \le K$, of the lth local model for $\{\boldsymbol{x}_k^{(q)}, \boldsymbol{y}_k^{(q)}\}_{k=1}^K$ as

$$\boldsymbol{e}_k^{(l)}[j] = \left|\boldsymbol{f}_l(\boldsymbol{x}_k^{(q)})[j] - \boldsymbol{y}_k^{(q)}[j]\right|, \tag{21}$$

where $[j]$ denotes the jth element of the corresponding vector. Then the generalization ability of the lth multi-output local model can be quantified by

$$\begin{aligned}\zeta(l) &= \frac{1}{\beta\left(\boldsymbol{e}_0^{(l)}\right)^{\mathrm{T}}\boldsymbol{\theta} + (1-\beta)\sum_{k=1}^{K} w_k\left(\boldsymbol{e}_k^{(l)}\right)^{\mathrm{T}}\boldsymbol{\theta} \Big/ \sum_{k=1}^{K} w_k} \\ &= \frac{1}{\sum_{j=1}^{p}\theta_j\left(\beta\boldsymbol{e}_0^{(l)}[j] + (1-\beta)\left(\sum_{k=1}^{K} w_k\boldsymbol{e}_k^{(l)}[j] \Big/ \sum_{k=1}^{K} w_k\right)\right)},\end{aligned} \tag{22}$$

where $\boldsymbol{\theta} = \left[\theta_1\ \theta_2 \cdots \theta_p\right]^{\mathrm{T}} \in \mathbb{R}^p$ with θ_j representing the importance of the jth primary variable, and w_k is the weight for the kth neighbor of $\boldsymbol{x}_q$, while β is the linking parameter that connects the newest sample $\boldsymbol{x}_0$ and the query sample $\boldsymbol{x}_q$.

The importance parameters θ_j, $1 \le j \le p$, are determined by the prior knowledge. For those primary variables that are more important, such as having higher price, higher values may be assigned to their corresponding importance parameters. In the case of no prior knowledge, all the primary variables may be treated with equal importance by assigning $\theta_j = 1$ for $1 \le j \le p$. The weighting w_k is defined as

$$w_k = \exp\left(-d\left(\boldsymbol{x}_q, \boldsymbol{x}_k^{(q)}\right)/\sigma_d\right), \tag{23}$$

where $d(\bullet, \bullet)$ stands for the Euclidean distance between two vectors and σ_d denotes the standard deviation of the distances between every two samples in the historical dataset. The linking parameter β is adaptively determined according to

$$\beta = \exp\left(-\varphi d\left(\boldsymbol{x}_q, \boldsymbol{x}_0\right)\right), \tag{24}$$

with φ is the scaling factor. It can be inferred that using (23) reduces the influence of faraway neighbors of $\boldsymbol{x}_q$, while by using (24) the influence of the newest labeled sample $\{\boldsymbol{x}_0, \boldsymbol{y}_0\}$ on the query sample $\boldsymbol{x}_q$ can be adaptively adjusted. These two rules are helpful to deal with both gradual and abrupt drifts in the process characteristics. Since the denominator of $\zeta(l)$ is composed of the relevant predicted errors, the larger $\zeta(l)$ is, the better generalization ability the lth local model has. For the convenience, the normalized $\zeta(l)$, denoted as $\widetilde{\zeta}(l)$, is used, which is given by

$$\widetilde{\zeta}(l) = \frac{\zeta(l)}{\max\{\zeta(1), \zeta(2), \cdots, \zeta(L)\}} \in (0,\ 1]. \tag{25}$$

Q 2 is answered by selecting those local models whose generalization abilities are above a threshold level δ. Hence, the indexes of the selected local models, represented by the index set $\mathcal{L}_M = \{l_1, \cdots, l_M\}$, are determined according to

$$\mathcal{L}_M = \left\{l \middle| 1 \le l \le L \text{ and } \widetilde{\zeta}(l) \ge \delta\right\}, \tag{26}$$

where M indicates the number of the selected local models which obviously varies in each prediction round. It can be seen that due to the normalization in (25), $\delta \in [0,\ 1]$ should be used. With $\delta = 0$, every local model is selected, while with $\delta = 1$, only one local model is selected. In general, the smaller δ is set, the more local models are combined, and vice versa. Thus, the threshold level δ is an important parameter to facilitate an appropriate implementation of the SEL-MO.

Q 3 is answered by building an ensemble of those selected local models under the Bayesian inference framework. Specifically, assume that the prior probabilities of the local models in the selected model set $\mathcal{L}_M$ are equal, i.e.,

$$\Pr\left(\boldsymbol{f}_{l_j}\right) = \frac{1}{M}, \quad l_j \in \mathcal{L}_M. \tag{27}$$

Further define the likelihood that $\boldsymbol{x}_q$ can be described by the l_jth local model to be $\widetilde{\zeta}(l_j)$, i.e.,

$$\Pr\left(\boldsymbol{x}_q | \boldsymbol{f}_{l_j}\right) = \widetilde{\zeta}(l_j). \tag{28}$$

According to the Bayesian inference, the estimated output $\widehat{\boldsymbol{y}}_q$ for the query sample $\boldsymbol{x}_q$ is calculated as

$$\widehat{\boldsymbol{y}}_q = \sum_{j=1}^{M} \widetilde{\zeta}(l_j)\boldsymbol{f}_{l_j}(\boldsymbol{x}_q) \Big/ \sum_{j=1}^{M} \widetilde{\zeta}(l_j). \tag{29}$$

It is worth pointing out that since at the online operation stage, the weight of each ensemble member is adaptively determined according to the query sample and the localization scheme may continue to adaptively update the local model set, our proposed SEL-MO based soft sensor is self-adaptive.

In each prediction round, searching the historical database to find the K neighbors of the query sample $\boldsymbol{x}_q$ is required to calculate the generalization measure (22). This is the main computational burden of the SEL-MO at the online operation stage, particularly when the size of the historical database is large and the output dimension p is high. Intuitively, when there is no abrupt drift in the process, by using only the newest labeled sample $\{\boldsymbol{x}_0, \boldsymbol{y}_0\}$ without involving the neighbors of $\boldsymbol{x}_q$, the SEL is still capable of providing good estimation performance. By contrast, when the process is in an abrupt change, using $\{\boldsymbol{x}_0, \boldsymbol{y}_0\}$ alone is insufficient and it is necessary to search for the neighbors of $\boldsymbol{x}_q$, in order to provide an accurate estimate. Based on this discussion, an insensitivity strategy is proposed which limits the search frequency of the historical database when appropriate, in order for the SEL-MO to achieve high online computational efficiency, while maintaining an accurate estimation performance. Note that this insensitivity strategy has not been discussed in the existing SEL-related literature.

Algorithm 2 Online Operations of the SEL-MO Based Adaptive Soft Sensor

Calculate β using (24) and $\boldsymbol{e}_0^{(l)}$ for $1 \le l \le L$ using (20);
if condition (30) holds **then**
  Calculate $\zeta(l)$ for $1 \le l \le L$ using (31);
else
  Search for $\boldsymbol{x}_q$'s K nearest neighbors based on Euclidean distance;
  Calculate w_k for $1 \le k \le K$ using (23), and calculate $\boldsymbol{e}_k^{(l)}$ for $1 \le l \le L$ and $1 \le k \le K$ using (21);
  Calculate $\zeta(l)$ for $1 \le l \le L$ using (22);
end if
Normalize $\zeta(l)$ for $1 \le l \le L$ according to (25);
Determine indexes of selected local models using (26);
Make an ensemble of selected local models to estimate primary variables using (29);

When the process characteristics are varying slowly, $\boldsymbol{x}_q$ is located near to $\{\boldsymbol{x}_0, \boldsymbol{y}_0\}$, and according to (24), β is large. By contrast, when abrupt changes in the process occur, $\boldsymbol{x}_q$ is located faraway from $\{\boldsymbol{x}_0, \boldsymbol{y}_0\}$, and β becomes small. Therefore, the value of β embodies the varying speed of the process characteristics, and the process is considered to change slowly if

$$\beta > \varepsilon, \tag{30}$$

where $\varepsilon \in [0,\ 1]$ is a pre-set threshold. When (30) holds, $1 - \beta$ is small and the second term in the denominator of $\zeta(l)$ becomes small. Thus the SEL-MO is insensitive to the neighbours of $\boldsymbol{x}_q$ when (30) holds, and the generalization measure of the lth local model can be simplified to

$$\zeta(l) = \frac{1}{\left(\boldsymbol{e}_0^{(l)}\right)^{\mathrm{T}}\boldsymbol{\theta}} = \frac{1}{\sum_{j=1}^{p}\theta_j \boldsymbol{e}_0^{(l)}[j]}. \tag{31}$$

From (24), $\beta \in [0,\ 1]$. Therefore, if ε is set to 0, (30) always holds and searching for the neighbors of $\boldsymbol{x}_q$ never takes place. On the other hand, if $\varepsilon = 1$ is chosen, searching for the neighbors of $\boldsymbol{x}_q$ always takes place for every $\boldsymbol{x}_q$.

Algorithm 2 summarizes the online operations of the SEL-MO based adaptive soft sensor given a query sample $\boldsymbol{x}_q$.

C. PSO BASED ALGORITHMIC PARAMETER OPTIMIZATION

The key algorithmic parameters of the SEL-MO include the window size W, the number of latent variables in the PLS algorithm A, the scaling factor φ, the query sample's neighborhood size K, and the degree of the SEL δ. Tuning of these parameters manually is intractable. Therefore, the particle swarm optimization (PSO) technique [36]–[38] is employed to realize the automatic parameter optimization for the SEL-MO.

The determination of the algorithmic parameters for the SEL-MO based soft sensor can be casted as the following optimization problem

$$\boldsymbol{ssap}^{\star} = \arg\min_{\boldsymbol{ssap}} \mathrm{cost}(\boldsymbol{ssap}), \tag{32}$$

where the algorithmic parameter vector of the SEL-MO is

$$\boldsymbol{ssap} = [W\ A\ K\ \varphi\ \delta]^{\mathrm{T}}, \tag{33}$$

and the cost function calculated on the validation dataset is given by

$$\mathrm{cost}(\boldsymbol{ssap}) = \sum_{j=1}^{p}\theta_j \mathrm{RMSE}_j^{val}(\boldsymbol{ssap}). \tag{34}$$

in which $\mathrm{RMSE}_j^{val}(\boldsymbol{ssap})$ is the predicted root mean squares error (RMSE) by the SEL-MO with $\boldsymbol{ssap}$ for the jth primary variable obtained on the validation dataset.

The PSO algorithm is used to solve this optimization problem and, therefore, to obtain the optimal $W^{\star}$, $A^{\star}$, $K^{\star}$, $\varphi^{\star}$ and $\delta^{\star}$ for the SEL-MO based soft sensor. Implementation of PSO can readily be found in [36]–[38]. In this PSO based optimization, ε is set to 1, as this is an offline optimization and the aim is to achieve the best performance. The influence of ε on the achievable performance at the online operation stage will be further investigated in the next section.

IV. CASE STUDIES

The performance of the proposed soft sensor is investigated using two chemical processes, a simulated penicillin fermentation process with three primary variables and a real-life industrial primary reformer with four primary variables. Three PLS based soft sensors, the LW-PLS [33], the MWPLS [13] and the JITL-PLS [34], as well as the least squares support vector regression (LSSVR) [39] with Gaussian kernel are used as the benchmarks. We also implement the p multiple single-output soft sensors using our previous SEL-SO [26] as a comparison. To quantitatively evaluate the performance of a soft sensor, the RMSE is adopted, which is defined as

$$\mathrm{RMSE}_j = \sqrt{\sum_{t=1}^{N_t}\left(\widehat{y}_{j,t} - y_{j,t}\right)^2 / N_t}, \tag{35}$$

for $1 \le j \le p$, where $y_{j,t}$ is the jth primary variable of the tth query sample, and $\widehat{y}_{j,t}$ denotes the predicted value of $y_{j,t}$, while N_t is the number of the test samples.

A. FED-BATCH PENICILLIN FERMENTATION PROCESS

The penicillin fermentation process is a biochemical fed-batch process with nonlinear dynamics and multi-phase characteristics, which has been widely adopted for performance assessment of adaptive soft sensors [2], [3]. A simulator for simulating this fermentation process under a variety of operating conditions, referred to as PenSim, is available at [40]. For our soft sensor modeling of this fermentation process, the penicillin concentration, biomass concentration and substrate concentration are selected as the primary variables, while the other 10 process variables are used as the secondary variables, as tabulated in Table 1.

TABLE 1. Secondary and primary variables of soft sensors for the fermentation process [2].

No.	Secondary variable	No.	Primary variable
1	Aeration rate	1	Penicillin concentration
2	Agitator power	2	Biomass concentration
3	Substrate feed rate	3	Substrate concentration
4	Substrate feed temperature		
5	Dissolved oxygen concentration		
6	Culture volume		
7	Carbon dioxide concentration		
8	pH		
9	Fermentor temperature		
10	Generated heat		

The measurement noise is added to each of the secondary variables, and the noise level is specified by

$$\text{noise level} = (\sigma_n/\sigma_s) \times 100\%, \tag{36}$$

where σ_s and σ_n represent the standard deviation of the secondary variable considered and that of the noise, respectively. Six batches of data were generated using the PenSim tool with default simulation conditions [40], where the simulation duration was 400 h and the sampling interval was set to 0.5 h. Therefore, each batch is composed of 800 samples. The entire dataset is divided into three parts: the first two batches of samples are used as the training dataset, the middle two batches of samples are used as the validation dataset, and the last two batches of samples are used as the testing dataset.

When optimizing the five algorithmic parameters of the SEL-MO using the PSO on the validation dataset, ε was set to 1, i.e., the insensitivity strategy was not invoked, while the noise level was set to 10%. In addition, the three primary variables were assigned to the same importance, i.e., $\theta_1 = \theta_2 = \theta_3 = 1$. The parameters of the five benchmark soft sensing approaches were also optimized on the same validation dataset using the PSO. The optimized parameters of all the six soft sensing methods for the penicillin fermentation process are summarized as follows.

- LSSVR: The kernel width is 2.99 and the regularization parameter is 98061.
- LW-PLS: The neighborhood size of query sample is 15, $A = 1$, and the scaling parameter is 5.26.
- JITL-PLS: The neighborhood size of query sample is 15, and $A = 3$.
- MWPLS: The window size is 15, and $A = 2$.
- Three SEL-SOs: for penicillin, $W = 16$, $A = 2$, $K = 3$, $\varphi = 1.469$, and $\delta = 0.383$; for biomass, $W = 16$, $A = 3$, $K = 41$, $\varphi = 0.545$, and $\delta = 0.360$; for substrate, $W = 39$, $A = 8$, $K = 5$, $\varphi = 2.742$, and $\delta = 0.433$.
- SEL-MO: $W = 15$, $A = 2$, $K = 5$, $\varphi = 0.5$, and $\delta = 0.47$.

Note that for the LW-PLS, JITL-PLS, MWPLS, three SEL-SOs and SEL-MO, the lower limit of the searching range for the neighborhood size or the window size was set to 15, because too small a window size or neighborhood size should be avoided, in terms of model stability and noise resistance.

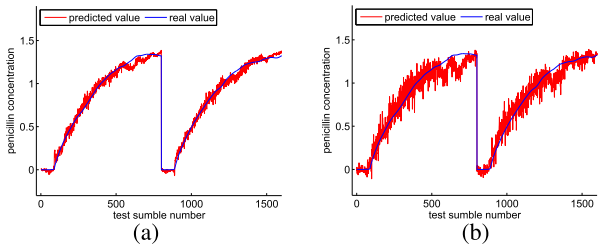


FIGURE 3. Predictions of the penicillin concentration by the LSSVR with (a) noise level 10%, and (b) noise level 30%.

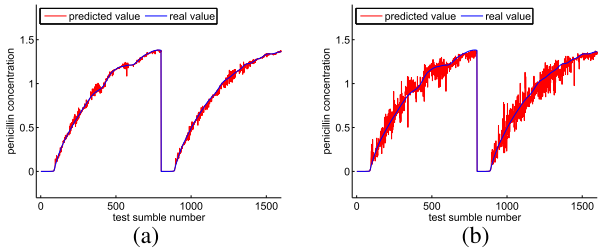


FIGURE 4. Predictions of the penicillin concentration by the LW-PLS with (a) noise level 10%, and (b) noise level 30%.

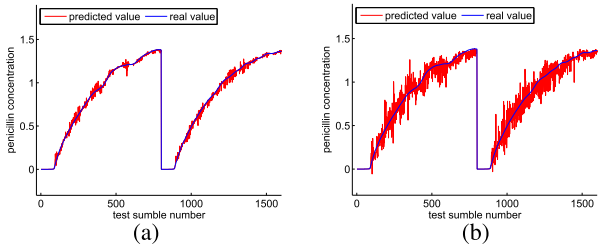


FIGURE 5. Predictions of the penicillin concentration by the JITL-PLS with (a) noise level 10%, and (b) noise level 30%.

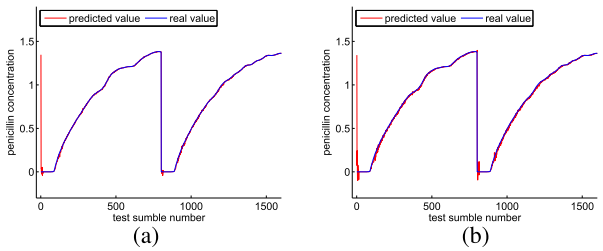


FIGURE 6. Predictions of the penicillin concentration by the MWPLS with (a) noise level 10%, and (b) noise level 30%.

The predictions of the three primary variables of the fermentation process by the six soft sensors on the two testing batches consisting of 1600 samples are plotted in Figs. 3 to 8, Figs. 9 to 14, and Figs. 15 to 20, respectively, for the two given noise levels of 10% and 30%. The quantitative prediction accuracies of the six soft sensors, in terms of RMSE, are further tabulated in Table 2 for various noise levels, where the results of the proposed SEL-MO were obtained without the insensitivity strategy, i.e., with $\varepsilon = 1$.

It can be seen that the nonlinear LSSVR is inferior to the two adaptive nonlinear local learning methods, the

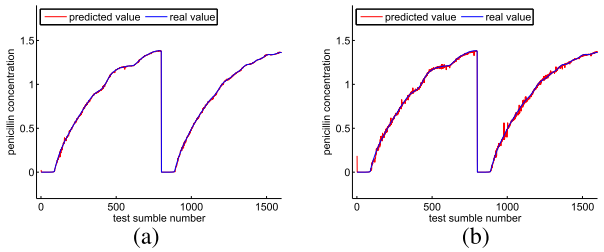


FIGURE 7. Predictions of the penicillin concentration by the SEL-SO with (a) noise level 10%, and (b) noise level 30%.

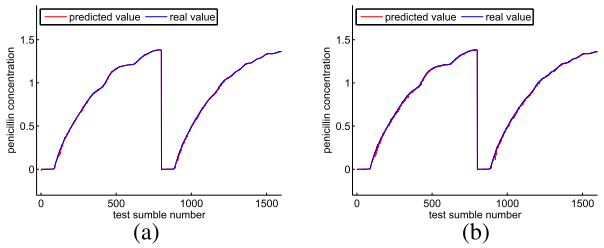


FIGURE 8. Predictions of the penicillin concentration by the SEL-MO with (a) noise level 10%, and (b) noise level 30%.

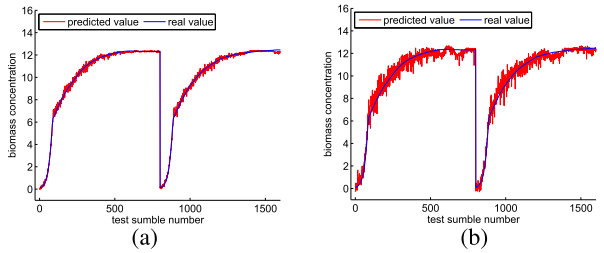


FIGURE 9. Predictions of the biomass concentration by the LSSVR with (a) noise level 10%, and (b) noise level 30%.

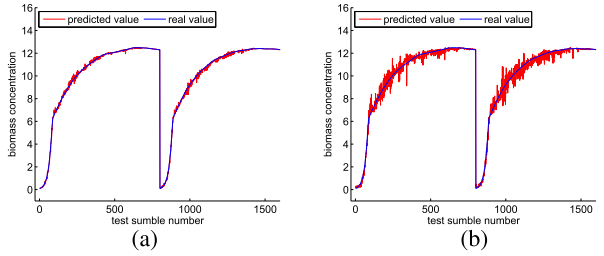


FIGURE 10. Predictions of the biomass concentration by the LW-PLS with (a) noise level 10%, and (b) noise level 30%.

JIT-PLS and MWPLS, and the LSSVR based soft sensor is prone to the measurement noise. However, the performance of the JIT-PLS soft sensor is also prone to the measurement noise, while the prediction accuracy of the MWPLS soft sensor deteriorates sharply when the process characteristics are changing rapidly. The results of Figs. 3 to 20 and Table 2 confirm that the proposed SEL-MO is much more superior to the LSSVR, LW-PLS, JIT-PLS and MWPLS soft sensing methods, as it can deal with the process nonlinearity and time-varying characteristics as well as the measurement noise much more effectively.

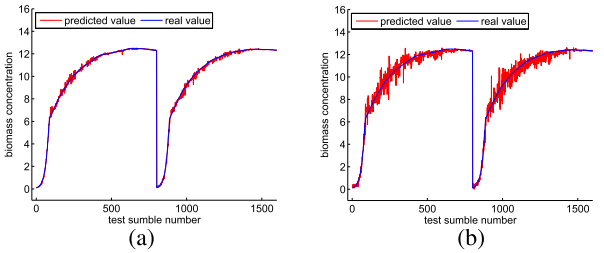


FIGURE 11. Predictions of the biomass concentration by the JITL-PLS with (a) noise level 10%, and (b) noise level 30%.

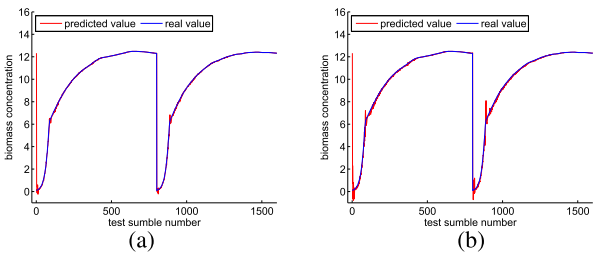


FIGURE 12. Predictions of the biomass concentration by the MWPLS with (a) noise level 10%, and (b) noise level 30%.

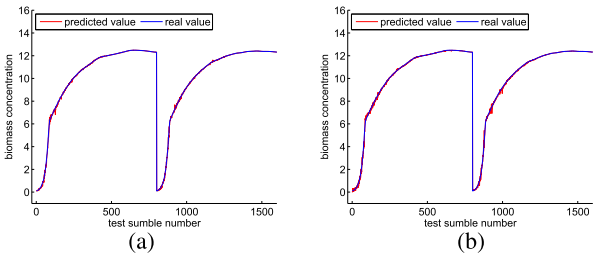


FIGURE 13. Predictions of the biomass concentration by the SEL-SO with (a) noise level 10%, and (b) noise level 30%.

It should be pointed out that with noise levels set as 5%, 10% and 25%, the generalized RMSE for predicting the concentration of substrate by the SEL-MO based soft sensor is larger than that by the SEL-SO based one. This may be explained from the objective function for parameter optimization defined in (34). In the SEL-MO, the PSO makes trade-off among various primary variables; in contrast, the SEL-SO just needs to consider one single primary variable. On the other hand, as can be seen from Figs. 15 to 20, for quite a long period, the substrate concentration is basically constant, where no correlation with the other two primary variables can be extracted. Therefore, the SEL-SO is able to achieve slight advantage over the SEL-MO. However, we should notice that the accuracy enhancement for the penicillin and biomass concentrations made by the SEL-MO is more remarkable compared with that for the substrate concentration made by the SEL-SO.

The reason that the JITL-PLS is prone to the effect of noise is because it is a spatial method [41], where the relevant samples for local model construction are selected according to the Euclidean distance metric. In the case of strong process

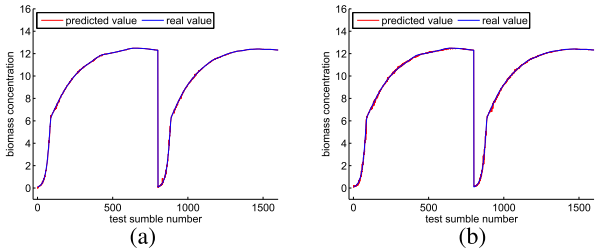


FIGURE 14. Predictions of the biomass concentration by the SEL-MO with (a) noise level 10%, and (b) noise level 30%.

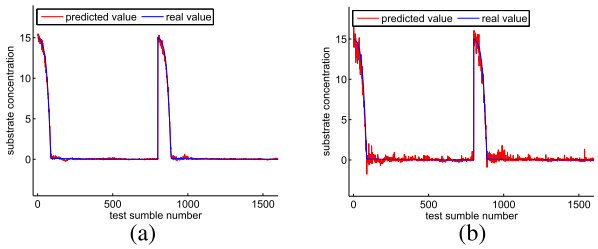


FIGURE 15. Predictions of the substrate concentration by the LSSVR with (a) noise level 10%, and (b) noise level 30%.

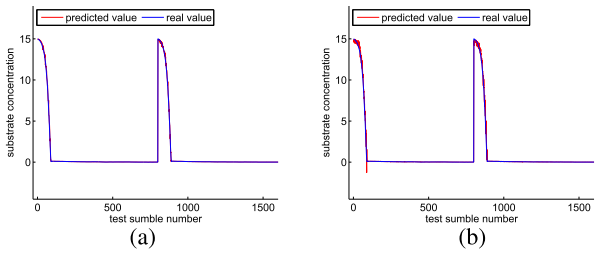


FIGURE 16. Predictions of the substrate concentration by the LW-PLS with (a) noise level 10%, and (b) noise level 30%.

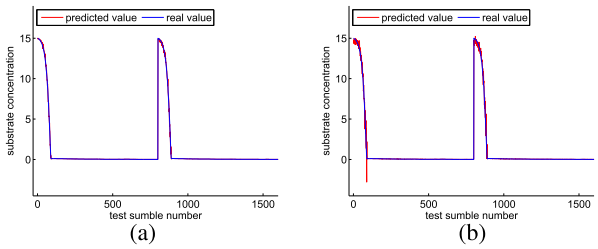


FIGURE 17. Predictions of the substrate concentration by the JITL-PLS with (a) noise level 10%, and (b) noise level 30%.

nonlinearity, the selected samples become less relevant with the query sample as the noise level increases, which contributes to the performance deterioration of the JITL-PLS. The temporal MWPLS and the SEL-MO on the other hand construct the local sample set based on the time-relevance criterion instead of the distance-relevance criterion, which is much less affected by the measurement noise. Therefore, the MWPLS and SEL-MO are more robust against the noise than the JITL-PLS. Furthermore, in the SEL-MO, the newest labeled samples that dominate the online weighting

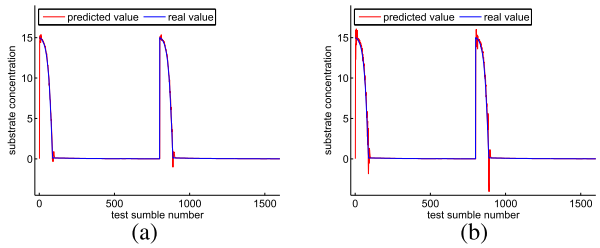


FIGURE 18. Predictions of the substrate concentration by the MWPLS with (a) noise level 10%, and (b) noise level 30%.

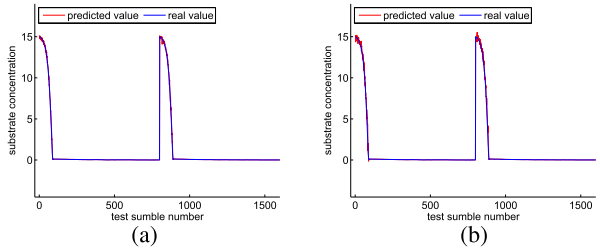


FIGURE 19. Predictions of the substrate concentration by the SEL-SO with (a) noise level 10%, and (b) noise level 30%.

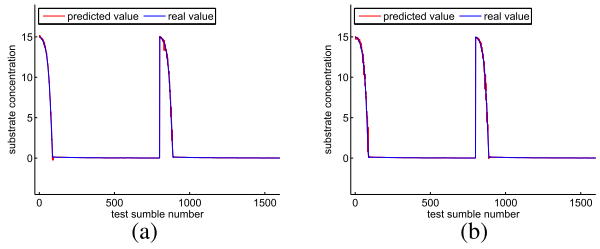


FIGURE 20. Predictions of the substrate concentration by the SEL-MO with (a) noise level 10%, and (b) noise level 30%.

of local models in most test samples are also free of noise. Thus, the SEL-MO is more robust against the noise than the MWPLS.

The performance of the MWPLS is poor when the process undergoes abrupt changes. This issue is effectively tackled by the SEL-MO, because the information of the query sample is taken into consideration using (22). Fig. 21 compares the performance of the MWPLS and SEL-MO on a segment of the testing samples. As can be seen from Fig. 21, unlike the MWPLS which produces large errors when the abrupt changes occur at the beginning of the new batch and at the beginning of the exponential phase within one batch, the SEL-MO can adapt to a new state quickly and accurately. This property of the SEL-MO is highly desirable for modeling systems with transition process, because the completion of the transition can be detected in time with the SEL-MO, which is extremely useful for reducing the amount of off-grade products [42].

Compared to the SEL-MO soft sensor, using three SEL-SO soft sensors to predict the three primary variables imposes three times more computational complexity in the off-line model construction and more critically requires three times

TABLE 2. RMSE prediction performance of various soft sensors for the penicillin fermentation process.

noise (%)	method	penicillin	biomass	substrate
	LSSVR	0.0309	0.1023	0.1290
	LW-PLS	0.0112	0.0627	0.0749
	MWPLS	0.0484	0.4354	0.5400
5	JITL-PLS	0.0121	0.0673	0.0731
	Three SEL-SOs	0.0059	0.0536	**0.0718**
	SEL-MO	**0.0037**	**0.0357**	0.0728
	LSSVR	0.0457	0.1613	0.1761
	LW-PLS	0.0219	0.1074	0.0901
	MWPLS	0.0488	0.4376	0.5445
10	JITL-PLS	0.0246	0.1210	0.0875
	Three SEL-SOs	0.0073	0.0665	**0.0860**
	SEL-MO	**0.0057**	**0.0483**	0.0890
	LSSVR	0.0563	0.2297	0.2311
	LW-PLS	0.0342	0.1648	0.1254
	MWPLS	0.0499	0.4433	0.5523
15	JITL-PLS	0.0385	0.1822	0.1070
	Three SEL-SOs	0.0087	0.0732	0.1123
	SEL-MO	**0.0067**	**0.0563**	**0.1035**
	LSSVR	0.0693	0.2874	0.2608
	LW-PLS	0.0450	0.2111	0.1496
	MWPLS	0.0493	0.4452	0.5647
20	JITL-PLS	0.0508	0.2338	0.1323
	Three SEL-SOs	0.0107	0.0756	0.1194
	SEL-MO	**0.0077**	**0.0635**	**0.1176**
	LSSVR	0.0857	0.3602	0.3254
	LW-PLS	0.0547	0.2592	0.1772
	MWPLS	0.0494	0.4479	0.5748
25	JITL-PLS	0.0628	0.2940	0.1648
	Three SEL-SOs	0.0129	0.0854	**0.1438**
	SEL-MO	**0.0086**	**0.0758**	0.1461
	LSSVR	0.0998	0.4327	0.3907
	LW-PLS	0.0733	0.3463	0.2141
	MWPLS	0.0510	0.4607	0.5997
30	JITL-PLS	0.0820	0.3869	0.2336
	Three SEL-SOs	0.0159	0.0917	0.1624
	SEL-MO	**0.0093**	**0.0837**	**0.1619**

more online computational costs. Moreover, the results of Figs. 7, 8, 13, 14, 19 and 20 as well as Table 2 also confirm that the proposed SEL-MO soft sensor provides more accurate predictions for the penicillin, biomass and substrate concentrations than the three SEL-SO soft sensors.

The above results for the SEL-MO are obtained without the insensitivity strategy, i.e., with $\varepsilon = 1$. Fig. 22 plots the variations of the connecting factor β given the noise level 15%. Observe from Fig. 22 that for most of the testing samples, β is close to 1, and it drops significantly only when abrupt changes occur, e.g., the start of a new batch. This fact implies that when the process characteristics are varying slowly, the information of the newest labeled sample is dominant in quantifying the generalization ability of a local model using (22) and the neighbors of the query sample may be dispensable.

When ε is set to be smaller than 1, the insensitivity strategy is applied. The influence of ε on the achievable performance and computational complexity of the SEL-MO is now investigated. Specifically, its prediction RMSE, the search times of database to find the neighbors of query sample, and the online consumed physical CPU time (CPT^{online}, in seconds), are calculated. The CPT^{online} is obtained by averaging over ten independent simulation runs, and the simulations are run using MATLAB version R2010a on a computer with Core i5 (2.6 GHz), 8 GB RAM, and Windows 7 OS.

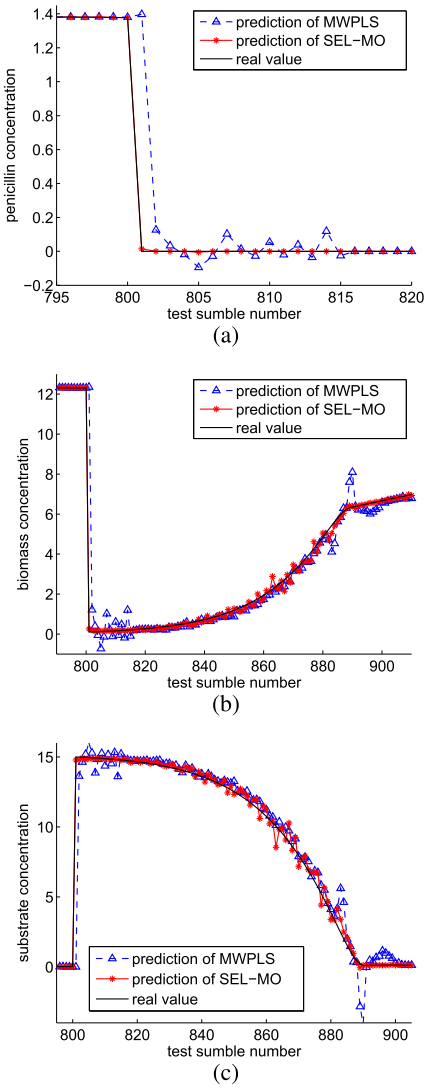


FIGURE 21. Comparisons between the predictions of the MWPLS and the SEL-MO with the noise level 30% for: (a) penicillin concentration, (b) biomass concentration, and (c) substrate concentration.

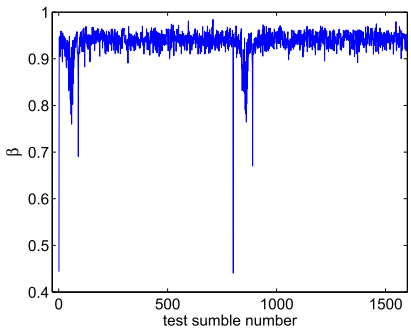


FIGURE 22. Time plot of β given the noise level 15% for the penicillin fermentation process.

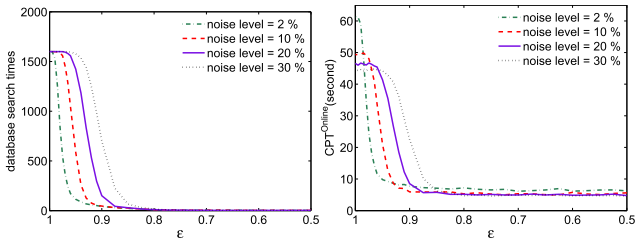


FIGURE 23. Influence of the threshold value ε on: (a) database search times, and (b) CPT^{online}, for the penicillin fermentation process.

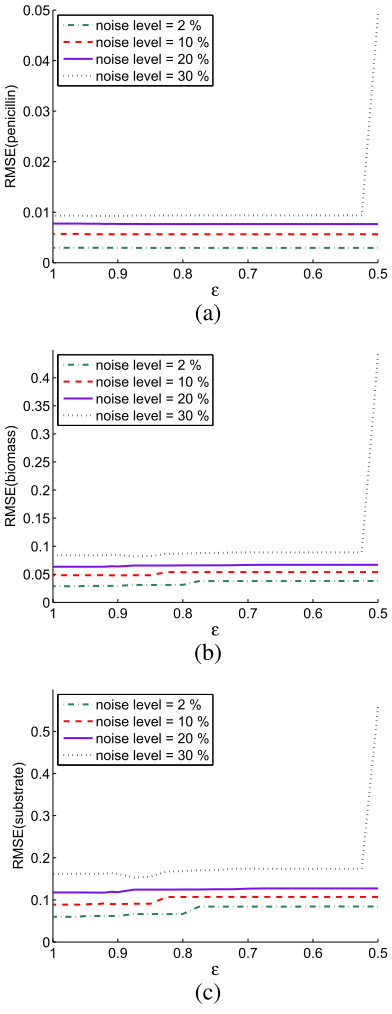


FIGURE 24. Influence of the threshold value ε on: (a) RMSE for the penicillin concentration, (b) RMSE for the biomass concentration, and (c) RMSE for the substrate concentration.

For various noise levels, the search times and CPT^{online} as the functions of ε are depicted in Fig. 23, while the prediction RMSEs as the functions of ε are shown in Fig. 24. It can be seen from Fig. 23 that the online computational efficiency improves significantly when ε is reduced from the peak value 1 to 0.8, and reaches the minimum online complexity when $\varepsilon < 0.8$. By contrast, the prediction accuracies for the three primary variables deteriorate very slightly as ε reduces. The exception is for the noise level 30%, where the predicted RMSEs suddenly increase sharply when ε is lower than approximately 0.52. This is because under this situation, the information related to the query sample is neglected even when abrupt changes occur. Table 3 compares the achievable performance of the SEL-MO given $\varepsilon = 1$ and 0.9 for the noise level 10%. It can be clearly seen that by applying the insensitivity strategy with an appropriately chosen value of $\varepsilon < 1$, the online computational complexity can be reduced dramatically, while the prediction accuracies are hardly affected.

TABLE 3. Performance comparison for the fermentation process by the SEL-MO given $\varepsilon = 1$ and 0.9 with the noise level 10%.

	$\varepsilon = 1$	$\varepsilon = 0.9$
RMSE of penicillin	0.0057	0.0056
RMSE of biomass	0.0483	0.0480
RMSE of substrate	0.0890	0.0901
database search times	1600	46
CPT^{online} (s)	49.39	5.93

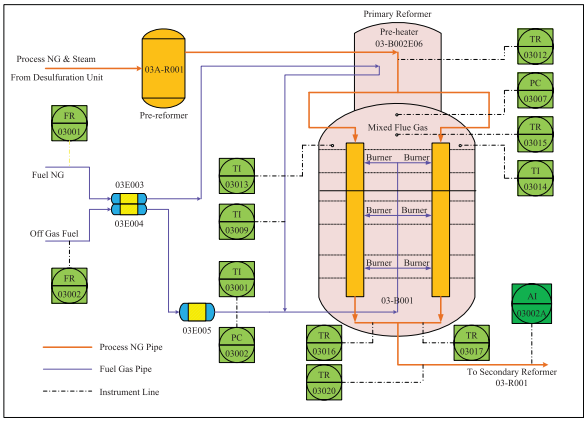


FIGURE 25. Flow chart of the primary reformer.

TABLE 4. Descriptions of selected secondary variables for soft sensing concentrations of CH_4, CO, CO_2 and H_2.

Tags	Descriptions
FR03001.PV	Flow of fuel NG into 03B001
FR03002.PV	Flow of fuel off gas into 03B001
PC03002.PV	Pressure of fuel off gas at 03E005's exit
PC03007.PV	Pressure of furnace flue gas at 03B001's exit
TI03001.PV	Temperature of fuel off gas at 03E005's exit
TI03009.PV	Temperature of fuel NG at 03B002E06's exit
TR03012.PV	Temperature of process gas at 03B001's entrance
TI03013.PV	Temperature of furnace flue gas at 03B001's top left
TI03014.PV	Temperature of furnace flue gas at 03B001's top right
TR03015.PV	Temperature of mixed furnace flue gas at 03B001's top
TR03016.PV	Temperature of transformed gas at 03B001's left exit
TR03017.PV	Temperature of transformed gas at 03B001's right exit
TR03020.PV	Temperature of transformed gas at 03B001's exit

B. PRIMARY REFORMER

Fig. 25 depicts the primary reformer, a unit of the ammonia (NH_3) synthesis process (ASP) [43]. Processed gases are fed

into the primary reformer for producing hydrogen (H_2). The chemical reactions carried out in the primary reformer using nickel catalyst are as follows:

$$\begin{aligned} C_nH_{2n+2} + nH_2O &\xleftrightarrow{\Delta} nCO + (2n+1)H_2, \\ CH_4 + H_2O &\xleftrightarrow{\Delta} CO + 3H_2, \\ CO + H_2O &\xleftrightarrow{\Delta} CO_2 + H_2. \end{aligned} \tag{37}$$

The hydrogen made from the primary reformer is the key source material for the ASP to yield NH_3. Reaction temperature is crucial for reactions in (37). In the ASP, the reaction temperature is controlled by adjusting the burning condition of fuel gases. According to (37), the outlet gases of the primary reformer consist of H_2, methane (CH_4), carbon monoxide (CO) and carbon dioxide (CO_2), which are the source gases for the down streaming units, including the high-low temperature transformer, CO_2 absorption column and ammonia synthesis unit, etc. These gases have significant impacts on the down streaming operations, such as energy consumption, production yield and process safety. Thus, the concentrations of these outlet gases need to be strictly monitored. Conventionally, they are measured by a mass spectrometer AI03002A, marked with deep green rectangle in Fig. 25, which is not only expensive but also often malfunctions. Therefore, a soft sensor is desired for online predicting the concentrations of these gases.

Secondary variables of the soft sensor for predicting the concentrations of CH_4, CO, CO_2 and H_2 were selected using the expert knowledge by field engineers, which are listed in Table 4. Due to the complex burning conditions and variations in source natural gases, the primary reformer is nonlinear and time-varying. Dataset for developing soft sensors for the target gases was collected from the database of the distributed control systems for a real-life industrial ASP. The sampling rate for the concentrations of the target gases is 10 minutes, and 6400 samples have been collected. In our investigation, the training set contains 1600 samples that are evenly selected from the first half of the dataset, and the second half of the dataset is partitioned evenly into the validation dataset and the testing dataset. Similar to the first case study, in addition to the proposed SEL-MO, the LSSVR, LW-PLS, JIT-PLS, MW-PLS and four SEL-SOs are also employed to develop soft sensors for this primary reformer, and all the algorithmic parameters of the six soft sensing approaches were optimized by the PSO using the validation dataset. Note that in this case, the lower limits of the window size or neighborhood size for the LW-PLS, MW-PLS, SEL-SO and SEL-SOs were set to 20, and the four primary variables share the same importance. After optimization by the PSO, the algorithmic parameters of the SEL-MO were set as: $W = 22$, $A = 2$, $K = 14$, $\varphi = 0.5375$, and $\delta = 0.3135$.

The accuracies of various soft sensors for all the four primary variables are quantitatively presented in Table 5, in terms of generalization RMSE. As expected, the proposed SEL-MO based soft sensor produces the best results. Moreover, the computational complexity of the proposed SEL-MO

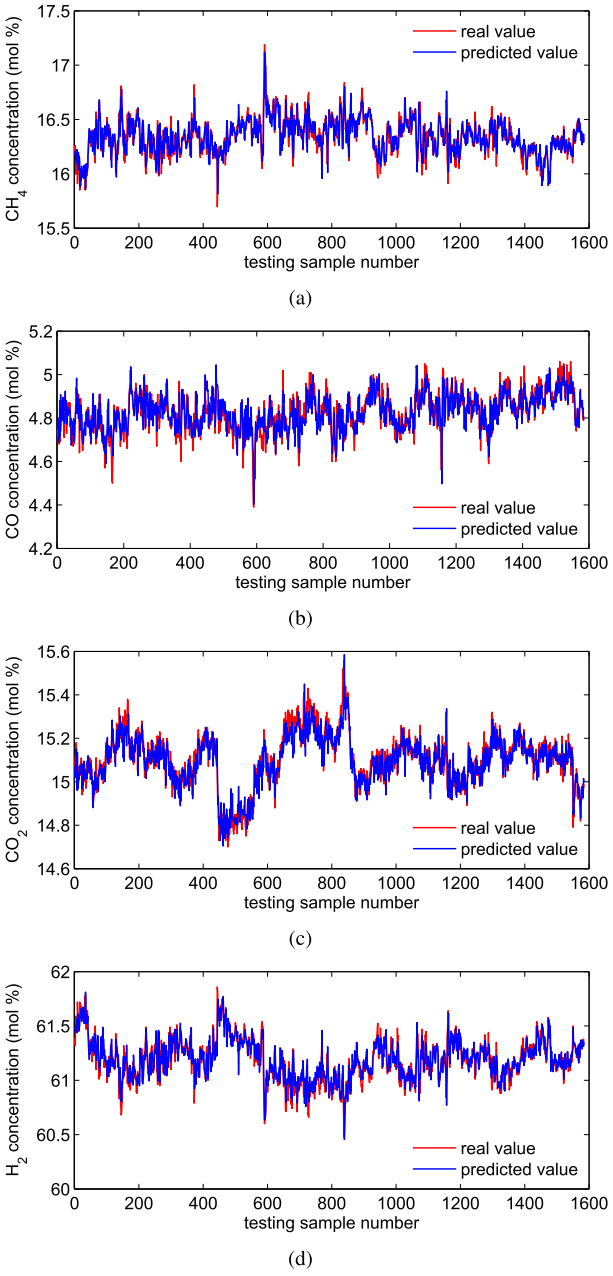


FIGURE 26. Predictions of the SEL-MO for: (a) CH_4 concentration, (b) CO concentration, (c) CO_2 concentration, and (d) H_2 concentration.

for offline model construction and online prediction is only 25% of the complexity imposed by our previous SEL-SO approach. Predictions of the concentrations of CH_4, CO, CO_2 and H_2 achieved by the SEL-MO based soft sensor on the testing dataset are visualised in Fig. 26. As can be seen, the predicted values can well track the true values, indicating that the SEL-MO is able to provide good estimations for the four gas concentrations. Note that the SEL-SO soft sensor obtains slightly better performance for the CO, and the reasons have been analyzed in the previous subsection.

The prediction performance of the SEL-MO listed in Table 5 was obtained with $\varepsilon = 1$. We also apply the

TABLE 5. Predicted RMSEs of various soft sensors for the primary reformer.

primary variable	LSSVR	LW-PLS	MWPLS	JIT-PLS	SEL-SOs	SEL-MO
CH_4	0.1688	0.1250	0.1266	0.1373	0.0870	**0.0864**
CO	0.1136	0.0741	0.0726	0.0806	**0.0586**	0.0590
CO_2	0.1735	0.0830	0.0646	0.0917	0.0589	**0.0573**
H_2	0.2094	0.1425	0.1250	0.1572	0.0912	**0.0907**

TABLE 6. Performance comparison for the primary reformer by the SEL-MO given various values of ε.

	$\varepsilon = 1$	$\varepsilon = 0.95$	$\varepsilon = 0.9$	$\varepsilon = 0.7$
RMSE of CH_4	0.0864	0.0867	0.0879	0.0909
RMSE of CO	0.0590	0.0593	0.0954	0.0602
RMSE of CO_2	0.0573	0.0573	0.0572	0.0571
RMSE of H_2	0.0907	0.0909	0.0912	0.0937
database search times	1584	666	204	0
CPT^{online} (s)	23.9	16.4	11.3	8.5

insensitivity strategy to the SEL-MO, and the performance of the SEL-MO with various values of ε are summarized in Table 6. As can be seen, as ε is reduced from 1 to 0.7, the data base search times and online computational burden decrease considerably but the estimation accuracy hardly changes. Therefore, by appropriately choosing the value of ε, a significantly decreasing in the online computational burden can be achieved with only a slight loss in prediction accuracy. For example, by using $\varepsilon = 0.95$, the online database search times and CPT^{online} are reduced by 58% and 31%, respectively, while the prediction accuracy deteriorations for the CH_4, CO, CO_2 and H_2 are only 0.35%, 0.51%, 0.0% and 0.22%, respectively, in comparison to $\varepsilon = 1$. Note that the improvement of online computational efficiency provided by our insensitivity strategy will be more prominent when the historical database is large.

V. CONCLUSIONS

In this paper, an adaptive soft sensor modeling approach, referred to as the SEL-MO, has been developed for nonlinear and time-varying industrial processes. Our novel contribution has been twofold. Firstly, a generic localization scheme and a selective ensemble learning framework have been developed for industrial processes with multiple primary variables. Secondly, a new insensitivity strategy for the SEL-MO based soft sensing method has been proposed, which is capable of significantly reducing the online computational load while maintaining high prediction accuracy performance. Two case studies have been conducted, involving a simulated industrial process and a real-life industrial process. Our results have demonstrated that the proposed SEL-MO soft sensor outperforms the existing state-of-the-art adaptive soft sensors for nonlinear and time-varying industrial processes. In particular, it has been shown that our SEL-MO is very robust to the measurement noise and can deal with effectively the process nonlinearity as well as both gradual and abrupt changes in the process characteristics. The effectiveness of the proposed novel insensitivity strategy has also been verified.

REFERENCES

[1] W. Shao, X. Tian, and P. Wang, "Soft sensor development for nonlinear and time-varying processes based on supervised ensemble learning with improved process state partition," *Asia–Pacific J. Chem. Eng.*, vol. 10, no. 2, pp. 282–296, 2015.

[2] J. Yu, "Multiway Gaussian mixture model based adaptive kernel partial least squares regression method for soft sensor estimation and reliable quality prediction of nonlinear multiphase batch processes," *Ind. Eng. Chem. Res.*, vol. 51, no. 40, pp. 13227–13237, 2012.

[3] H. Jin, X. Chen, L. Wang, K. Yanga, and L. Wua, "Dual learning-based online ensemble regression approach for adaptive soft sensor modeling of nonlinear time-varying processes," *Chemometrics Intell. Lab. Syst.*, vol. 151, pp. 228–244, Feb. 2016.

[4] P. Kadlec, B. Gabrys, and S. Strandt, "Data-driven soft sensors in the process industry," *Comput. Chem. Eng.*, vol. 33, no. 4, pp. 795–814, Apr. 2009.

[5] Z. Ge, Z. Song, and P. Wang, "Probabilistic combination of local independent component regression model for multimode quality prediction in chemical processes," *Chem. Eng. Res. Des.*, vol. 92, no. 3, pp. 509–521, 2014.

[6] P. Cao and X. Luo, "Modeling of soft sensor for chemical process," *J. Chem. Ind. Eng. Soc. China*, vol. 64, no. 3, pp. 788–800, 2013.

[7] M. Kano and M. Ogawa, "The state of the art in chemical process control in Japan: Good practice and questionnaire survey," *J. Process Control*, vol. 20, no. 9, pp. 969–982, Oct. 2010.

[8] S. Kim, M. Kano, S. Hasebe, T. Seki, and A. Takinami, "Long-term industrial applications of inferential control based on just-in-time soft-sensors: Economical impact and challenges," *Ind. Eng. Chem. Res.*, vol. 52, no. 35, pp. 12346–12356, 2013.

[9] S. J. Qin, "Recursive PLS algorithms for adaptive data modeling," *Comput. Chem. Eng.*, vol. 22, nos. 4–5, pp. 503–514, 1998.

[10] S. Mu, Y. Zeng, R. Liu, P. Wu, H. Su, and J. Chu, "Online dual updating with recursive PLS model and its application in predicting crystal size of purified terephthalic acid (PTA) process," *J. Process Control*, vol. 16, no. 6, pp. 557–566, 2006.

[11] J. Tang, W. Yu, T. Chai, and L. Zhao, "On-line principal component analysis with application to process modeling," *Neurocomputing*, vol. 82, pp. 167–178, Apr. 2012.

[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," *Ind. Eng. Chem. Res.*, vol. 48, no. 12, pp. 5731–5741, 2009.

[13] J. Liu, D.-S. Chen, and J.-F. Shen, "Development of self-validating soft sensors using fast moving window partial least squares," *Ind. Eng. Chem. Res.*, vol. 49, no. 22, pp. 11530–11546, 2010.

[14] W. Ni, S. K. Tan, W. J. Ng, and S. D. Brown, "Moving-window GPR for nonlinear dynamic system modeling with dual updating and dual preprocessing," *Ind. Eng. Chem. Res.*, vol. 51, no. 18, pp. 6416–6428, 2012.

[15] K. Fujiwara, M. Kano, S. Hasebe, and A. Takinami, "Soft-sensor development using correlation-based just-in-time modeling," *AIChE J.*, vol. 55, no. 7, pp. 1754–1765, 2009.

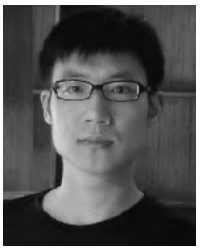
[16] S. Kim, R. Okajima, M. Kano, and S. Hasebe, "Development of soft-sensor using locally weighted PLS with adaptive similarity measure," *Chemometrics Intell. Lab. Syst.*, vol. 124, pp. 43–49, May 2013.

[17] L. Xie, J. Zeng, and C. Gao, "Novel just-in-time learning-based soft sensor utilizing non-Gaussian information," *IEEE Trans. Control Syst. Technol.*, vol. 22, no. 1, pp. 360–368, Jan. 2014.

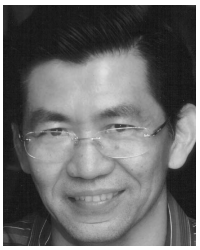
[18] P. Kadlec, "On robust and adaptive soft sensors," Ph.D. dissertation, Fac. Sci. Technol., Bournemouth Univ., Poole, U.K., 2009.

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

[20] S. Khatibisepehr, B. Huang, F. Xu, and A. Espejo, "A Bayesian approach to design of adaptive multi-model inferential sensors with application in oil sand industry," *J. Process Control*, vol. 22, no. 10, pp. 1913–1929, 2012.
[21] J. Tang, T. Chai, W. Yu, and L. Zhao, "Modeling load parameters of ball mill in grinding process based on selective ensemble multisensor information," *IEEE Trans. Autom. Sci. Eng.*, vol. 10, no. 3, pp. 726–740, Jul. 2013.
[22] H. Kaneko and K. Funatsu, "Adaptive soft sensor based on online support vector regression and Bayesian ensemble learning for various states in chemical plants," *Chemometr. Intell. Lab. Syst.*, vol. 137, pp. 57–66, Oct. 2014.
[23] R. Grbić, D. Slišković, and P. Kadlec, "Adaptive soft sensor for online prediction and process monitoring based on a mixture of Gaussian process models," *Comput. Chem. Eng.*, vol. 58, no. 22, pp. 84–97, 2013.
[24] H. Jin, X. Chen, L. Wang, K. Yang, and L. Wu, "Adaptive soft sensor development based on online ensemble Gaussian process regression for nonlinear time-varying batch processes," *Ind. Eng. Chem. Res.*, vol. 54, no. 30, pp. 7320–7345, 2015.
[25] H. Kaneko and K. Funatsu, "Smoothing-combined soft sensors for noise reduction and improvement of predictive ability," *Ind. Eng. Chem. Res.*, vol. 54, no. 50, pp. 12630–12638, 2015.
[26] W. Shao, X. Tian, P. Wang, X. Deng, and S. Chen, "Online soft sensor design using local partial least squares models with adaptive process state partition," *Chemometrics Intell. Lab. Syst.*, vol. 144, pp. 108–121, May 2015.
[27] W. Shao, and X. Tian, "Adaptive soft sensor for quality prediction of chemical processes based on selective ensemble of local partial least squares models," *Chem. Eng. Res. Des.*, vol. 95, pp. 113–132, 2015.
[28] Y. Liu and Z. Gao, "Real-time property prediction for an industrial rubber-mixing process with probabilistic ensemble Gaussian process regression models," *J. Appl. Polym. Sci.*, vol. 132, no. 6, pp. 41432–41440, 2015.
[29] H. Kaneko and K. Funatsu, "Ensemble locally weighted partial least squares as a just-in-time modeling method," *AIChE J.*, vol. 62, no. 3, pp. 717–725, 2016.
[30] L.-J. Zhao, T.-Y. Chai, and D.-C. Yuan, "Selective ensemble extreme learning machine modeling of effluent quality in wastewater treatment plants," *Int. J. Autom. Comput.*, vol. 9, no. 6, pp. 627–633, 2012.
[31] Z.-H. Zhou, J. Wu, and W. Tang, "Ensembling neural networks: Many could be better than all," *Artif. Intell.*, vol. 137, no. 1, pp. 239–263, 2002.
[32] R. Lindgren, P. Geladi, and S. Wold, "The kernel algorithm for PLS," *J. Chemometrics*, vol. 7, no. 1, pp. 45–59, 1993.
[33] M. Kano and K. Fujiwara, "Virtual sensing technology in process industries: Trends and challenges revealed by recent industrial applications," *J. Chem. Eng. Jpn.*, vol. 46, no. 1, pp. 1–17, 2013.
[34] Z. Ge and Z. Song, "A comparative study of just-in-time-learning based methods for online soft sensor modeling," *Chemometr. Intell. Lab. Syst.*, vol. 104, no. 2, pp. 306–317, 2010.
[35] W. Shao, P. Wang, and X. Tian, "Local partial least squares based online soft sensing method for multi-output processes with adaptive process states division," *Chin. J. Chem. Eng.*, vol. 22, no. 7, pp. 828–836, Jul. 2014.
[36] J. Kennedy and R. C. Eberhart, *Swarm Intelligence*. San Mateo, CA, USA: Morgan Kaufmann, 2001.
[37] A. Ratnaweera, S. K. Halgamuge, and H. C. Watson, "Self-organizing hierarchical particle swarm optimizer with time-varying acceleration coefficients," *IEEE Trans. Evol. Comput.*, vol. 8, no. 3, pp. 240–255, Jun. 2004.
[38] S. Chen, X. Hong, B. L. Luk, and C. J. Harris, "Non-linear system identification using particle swarm optimisation tuned radial basis function models," *Int. J. Bio-Inspired Comput.*, vol. 1, no. 4, pp. 246–258, 2009.
[39] J. A. K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, and J. Vandewalle, *Least Squares Support Vector Machines*. Singapore: World Scientific, 2002.
[40] *A Web Based Program for Dynamic Simulation of Fed-Batch Penicillin Production*. Accessed: Jun. 2017. [Online]. Available: http://simulator.iit.edu/web/pensim/simul.html
[41] X. Yuan, Z. Ge, and Z. Song, "Spatio-temporal adaptive soft sensor for nonlinear time-varying and variable drifting processes based on moving window LWPLS and time difference model," *Asia–Pacific J. Chem. Eng.*, vol. 11, no. 2, pp. 209–219, 2015.
[42] H. Kaneko, M. Arakawa, and K. Funatsu, "Novel soft sensor method for detecting completion of transition in industrial polymer processes," *Comput. Chem. Eng.*, vol. 35, no. 6, pp. 1135–1142, 2011.
[43] L. Yao and Z. Ge, "Moving window adaptive soft sensor for state shifting process based on weighted supervised latent factor analysis," *Control Eng. Pract.*, vol. 61, pp. 72–80, Apr. 2017.

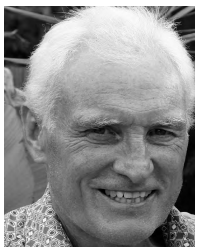


WEIMING SHAO received the B.Eng. and Ph.D. degrees from the College of Information and Control Engineering, China University of Petroleum, Qingdao, China, in 2009 and 2016, respectively. He was a Visiting Research Associate with the Department of Electrical Engineering, Petroleum Institute, Abu Dhabi, United Arab Emirates, from 2014 to 2015. He is currently a Post-Doctoral Research Fellow with the State Key Laboratory of Industrial Control Technology, College of Control Science and Engineering, Zhejiang University, Hangzhou, China. His research interests include machine learning and statistical learning methods and their applications to semisupervised, robust, and adaptive soft sensor development.



SHENG CHEN (M'90–SM'97–F'08) received the B.Eng. degree in control engineering from the East China Petroleum Institute, Dongying, China, in 1982, the Ph.D. degree in control engineering from the City University of London, London, in 1986, and the D.Sc. degree from the University of Southampton, Southampton, U.K., in 2005. From 1986 to 1999, he held research and academic appointments with The University of Sheffield, Edinburgh, U.K., and The University of Sheffield, Portsmouth, U.K. Since 1999, he has been with the School of Electronics and Computer Science, University of Southampton, U.K., where he is currently a Professor of intelligent systems and signal processing.

He is also a Distinguished Adjunct Professor with King Abdulaziz University, Jeddah, Saudi Arabia. He is an ISI highly cited researcher in engineering in 2004. He has published over 600 research papers. He is a fellow of the Royal Academy of Engineering, U.K., and IET. His research interests include adaptive signal processing, wireless communications, modeling and identification of nonlinear systems, neural network and machine learning, intelligent control system design, evolutionary computation methods, and optimization.



CHRIS J. HARRIS received the B.Sc. degree from the University of Leicester, the M.A. degree from the University of Oxford, U.K., the Ph.D. degree from the University of Southampton, U.K, in 1972, and the D.Sc. degree from the University of Southampton in 2001. He is currently a Emeritus Research Professor with the University of Southampton, having previously held senior academic appointments at Imperial College, Oxford, and Manchester Universities. He is also a Deputy Chief Scientist with the U.K. Government. He was elected to the Royal Academy of Engineering, U.K., in 1996. He has co-authored over 450 scientific research papers during the 45-year research career. He received the IEE Senior Achievement Medal for data fusion research and the IEE Faraday Medal for distinguished international research in machine learning.