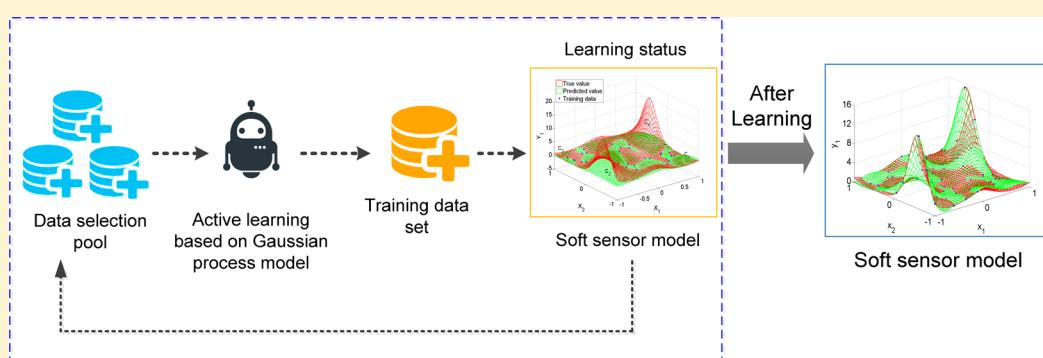


Active Selection of Informative Data for Sequential Quality Enhancement of Soft Sensor Models with Latent Variables

Yi Liu,[†] Qing-Yang Wu,[‡] and Junhui Chen^{*§,‡,ID}

[†]Institute of Process Equipment and Control Engineering, Zhejiang University of Technology, Hangzhou, 310014, People's Republic of China

[‡]Department of Chemical Engineering, Chung-Yuan Christian University, Chung-Li, Taoyuan, Taiwan, 32023, Republic of China



ABSTRACT: With training data of insufficient information, soft sensor models inevitably show some inaccurate predictions in their industrial applications. This work aims to develop an active learning method to sequentially select a data set with significant information to enhance latent variable model (LVM)-based soft sensors. Using the Gaussian process model to link the relationships between the score variables of LVM and the input process variables, the prediction variance can be formulated. And an uncertainty index of LVM is presented. It contains the variances of the predicted outputs and the changes of the predicted outputs per unit change in the designed inputs. Without any prior knowledge of the process, the index is sequentially used to adequately find out from which regions the new informative data should be adopted to enhance the model quality. Additionally, an evaluation criterion is proposed to monitor the active learning procedure. Consequently, the active learning procedures of exploration and exploitation analysis of the current model can effectively discover the meaningful data to be included into the soft sensor model. The proposed strategy can be applied to any types of LVMs. The effectiveness and the promising results are demonstrated through a numerical example and a real industrial plant in Taiwan with multiple outputs.

1. INTRODUCTION

One common objective in chemical industries is to produce high-quality products. However, it is often difficult to accurately measure key product qualities online. It is time-consuming and costly in the laboratory assaying/analyzing process. Different strategies for estimation of product qualities have been proposed in the literature to model real systems, depending on the level of a priori knowledge of the process. Models can be obtained either on the basis of mechanistic models or by data-driven based methods. Development of detailed mechanistic models often takes a long time. It is infeasible for agile responsive manufacturing because the products are typically short-lived and of small volume. With the rapid development of computer and communication technologies, a great amount of historical data is usually acquired for monitoring purposes. As a result, increasing data-driven soft sensor methods have been widely applied to infer/predict product qualities that are difficult to measure online.^{1–7} One main advantage of these data-driven soft sensor models is that they can generally be developed quickly without substantial understanding of the phenomenology involved. Common methods include latent variable models (LVMs) such

as principal component regression (PCR),⁸ partial least-squares (PLS),^{9–12} various neural networks (NN),^{13,14} fuzzy systems,¹⁵ support vector regression (SVR),^{16–19} and other kernel learning-based regression methods.^{20–23} Among them, LVM-based soft sensors have been widely applied in industrial processes because they can model highly correlated process input variables. Additionally, LVM-based soft sensors can be directly utilized for multi-input–multi-output (MIMO) processes.

In soft sensor applications, the modeling data are collected from the historical data in an existing process, and the selected models are built upon the collected data. Then, they are used for predicting the product quality of the process over time. The trial procedures that pick the training data up from the historical data would be performed repeatedly until clean and representative data are obtained. In fact, even if there are many soft sensor modeling methods available in the literature, high volume data

Received: November 29, 2016

Revised: March 2, 2017

Accepted: March 31, 2017

Published: March 31, 2017

with good quality are assumed to be available in industrial processes. This pragmatic and straightforward approach is used in many industrial applications although it is not ideal, and how to adaptively select the training data with significant information has been rarely investigated.

When big data are available in the database, it is usually inflexible and impractical to design soft sensor models using the whole data set. Conventional soft sensor algorithms often require a huge space to store the regressor matrix and a large computational load, particularly requiring matrix inversion. To overcome the aforementioned shortcomings, this paper aims to develop an active learning method to sequentially select a set of training data with the most significant information and to enhance existing LVM-based soft sensor models. Recently, by combining the process monitoring statistics, a PCR-based active learning strategy for the soft sensor development was proposed.⁸ However, the method cannot be directly applied to other LVM-based soft sensors. Additionally, most methods are not explicitly probabilistic and offer little guidance as to where estimates are accurate or where a model may perform poorly. Previously, prediction uncertainty has rarely been utilized as a useful tool in the soft sensor development. Therefore, probabilistic models are desirable as they provide uncertainty estimates as part of the inference/prediction process. As a recent probabilistic modeling method, the Gaussian process model (GPM) is capable of simultaneously providing the accurate predictive mean and variance.²⁴ There are increasing GPM-based soft sensor applications in chemical processes.^{25–30} In this study, GPM is adopted to quantify the prediction uncertainty of the LVMs-based soft sensors and then to develop an active learning method.

From the modeling point of view, the automatic exploration and exploitation analysis of the current soft sensor model can effectively discover informative data to be included in the model and enhance its quality. Recently, a GPM-based active learning method was proposed to label data samples.³⁰ The model is used for single-output processes only and the correlations among the input variables in the model construction are not considered. The computation of GPM for MIMO processes is still not efficient so far. Additionally, the correlated process input variables cannot be treated using GPM. To overcome these problems, an active learning approach based on LVM for MIMO processes is proposed. First, GPM is used to construct the relationships between the score variables of LVM and the process input variables. Under the given structure of LVM, the predictive means and the variances of the outputs can be derived. Second, the uncertainty index of LVM is presented. It contains the variances of the predicted outputs and the changes of the predicted outputs per unit change in the designable inputs. Without the prior knowledge of the process, the index is sequentially used to adequately find out from which regions the new data sample should be adopted to enhance the model quality. Third, an evaluation criterion of the active learning stage is proposed to judge whether the training procedure should be stopped. Moreover, with a general form, the proposed active learning approach can be applied to any types of LVMs-based soft sensors.

The remainder of this paper is organized as follows. Several LVMs are briefly reviewed, and the unification of these popular modeling methods is discussed in section 2. This unification would be applied to the proposed active learning strategy in the common framework for these modeling methods. In section 3, the GPM method is first introduced. Then, the

prediction uncertainty and its sequential enhancement of LVMs are derived and formulated, respectively. The evaluation criterion of the active learning stage is proposed. Also, detailed implementations of the sequential active learning approach are made in this section. In section 4, the effectiveness and promising results of the proposed method are demonstrated through its applications to a numerical example and a chemical plant in Taiwan. Finally, the conclusion is drawn in section 5.

2. SOFT SENSOR MODELS WITH LATENT VARIABLES

In chemical processes, PCR and PLS are two popular LVM methods for constructing soft sensors.^{8–12} They are basically multilinear algorithms which can handle the high dimensionality and collinearity by projecting the data information down onto a low dimensional space defined by a small number of orthogonal latent variables. The latent variables in the new dimension summarize the information contained in the original data set. According to a recent paper on the data-driven soft sensor methods,² PCR is the most widely used one (accounting for 23% of all the soft sensor models). PLS, as an extension of PCR, is a classical regression model based on the calculation of latent variables or factors. Unlike PCR, the PLS model finds in the input space the multidimensional directions that interpret the maximum variance of the multidimensional directions in the output space.^{1,2}

On the other hand, many industrial processes are nonlinear in nature. It is also desirable to develop effective methods which can solve the nonlinear problems. With the wide applications of support vector machines and kernel learning methods to pattern recognition and nonlinear regression (in the machine learning area), the kernel-based nonlinear forms of PCR and PLS models have been recently used in nonlinear chemical processes.^{20–22,31–33} The main purpose of kernel PCR (KPCR) is to map data points $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^N$ onto a highly dimensional feature space H , i.e., $\mathbf{x}_i \rightarrow \phi(\mathbf{x}_i)$, and then perform principal component analysis (PCA) in H .³¹ And the mapping $\phi(\mathbf{x}_i)$ does not need to be known because it is implicitly defined by the choice of the kernel function $K(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle$. Correspondingly, the original input data set \mathbf{X} can be represented by a new matrix denoted as Φ formed by the kernel functions. Then using KPCA as a data reprocessing step can create a linear model based on the extracted nonlinear principal components.^{31,32} Therefore, KPCR can be applied to modeling of nonlinear regression problems. Similarly, the kernel PLS (KPLS) model can be developed using PLS and the kernel learning methods.^{31–33}

So far, four LVMs, including PCR, PLS, KPCR, and KPLS, have been briefly introduced. The first two models are linear and the latter two are nonlinear kernel forms. Suppose the centralized and normalized data sets can be represented by $\{\mathbf{X}, \mathbf{Y}\}$, where $\{\mathbf{X} \in \mathbb{R}^{N \times D}\} = \{\mathbf{x}_i\}_{i=1}^N$ and $\{\mathbf{Y} \in \mathbb{R}^{N \times Q}\} = \{\mathbf{y}_i\}_{i=1}^N$ are the input and the output data sets with N samples, respectively. D and Q are the number of the measured input and output variables, respectively. First, define a unified matrix \mathbf{A} related to the input data set,

$$\mathbf{A} = \begin{cases} \mathbf{X} & \text{for PCR and PLS} \\ \Phi & \text{for KPCR and KPLS} \end{cases} \quad (1)$$

where \mathbf{X} is the input data set in the original space for linear LVMs of PCR and PLS; and Φ is the input data set in the feature space for nonlinear LVMs of KPCR and KPLS.^{31–33}

Then, these four LVMs can be formulated in a general form:

$$\begin{cases} \mathbf{A} = \mathbf{T}\mathbf{G} + \mathbf{E} \\ \mathbf{Y} = \mathbf{T}\mathbf{B} + \mathbf{F} \end{cases} \quad (2)$$

where $\mathbf{T} \in \mathbb{R}^{N \times R}$ denotes the score matrix, for example, the principal component matrix, and R is the selected number of principal components. The item $\mathbf{G} \in \mathbb{R}^{R \times D}$ is the loading matrix. And the item $\mathbf{B} \in \mathbb{R}^{R \times Q}$ can be considered as the regression matrix connected to the output data set \mathbf{Y} . $\mathbf{E} \in \mathbb{R}^{N \times D}$ and $\mathbf{F} \in \mathbb{R}^{N \times Q}$ are the residuals matrices with appropriate dimensions. The first formula in eq 2 describes the relationship between the input variables and the latent variables. And the second formula in eq 2 describes the relationship between the output variables and the latent variables. To describe eq 2 more clearly, the inner relationships of LVM-based soft sensor models for MIMO processes are shown in Figure 1. The meanings of the items ($\mathbf{A}, \mathbf{G}, \mathbf{T}, \mathbf{B}$) are related to eq 2.

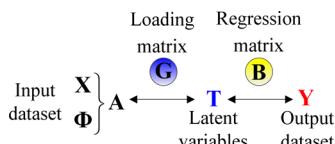


Figure 1. Inner relationships of LVM-based soft sensor models for MIMO processes.

With the above general form of LVMs, the relationship between several linear and nonlinear models can be simply revealed. Detailed implementations for training PCR and PLS models are not described here, and detailed algorithmic implementations of KPCR and KPLS models can be found in the references.^{31–33}

3. ACTIVE LEARNING FOR SEQUENTIAL TRAINING LVMs

How to select a set of informative training data to enhance the LVMs-based soft sensors is an important but difficult task. In this section, a sequentially active learning approach is developed. The learning procedures of exploration and exploitation analysis of the current model can effectively discover meaningful data to be included in the soft sensor model.

To clarify the main structure of the developed method, a modeling scheme is first shown in Figure 2. The GPM described in section 3.1 is adopted to design the active learning approach because the GPM can provide the probabilistic information for its own prediction. First, the current training data are used to construct an LVM. (Note that any type of LVM can be adopted.) Second, the prediction uncertainty of LVM is analyzed and formulated in section 3.2. Third, the uncertainty index of LVM is presented. It contains the variances of the predicted outputs and the changes of the predicted outputs per unit change in the designable inputs. In section 3.3 the index is sequentially used to find out from which regions the new informative data sample should be collected. Consequently, the quality of an LVM model can be sequentially enhanced. Finally, the evaluation criterion of the active learning strategy is proposed in section 3.4.

3.1. Gaussian Process Model (GPM). The GPM method for nonlinear process regression is used to learn a model f that approximates a training set $\{\mathbf{X}, \mathbf{y}\}$, where $\{\mathbf{X} \in \mathbb{R}^{N \times D}\} = \{\mathbf{x}_i\}_{i=1}^N$ and $\{\mathbf{y} \in \mathbb{R}^{N \times 1}\} = \{y_i\}_{i=1}^N$ are the input and the output data sets with N samples, respectively. Traditionally, the GPM modeling

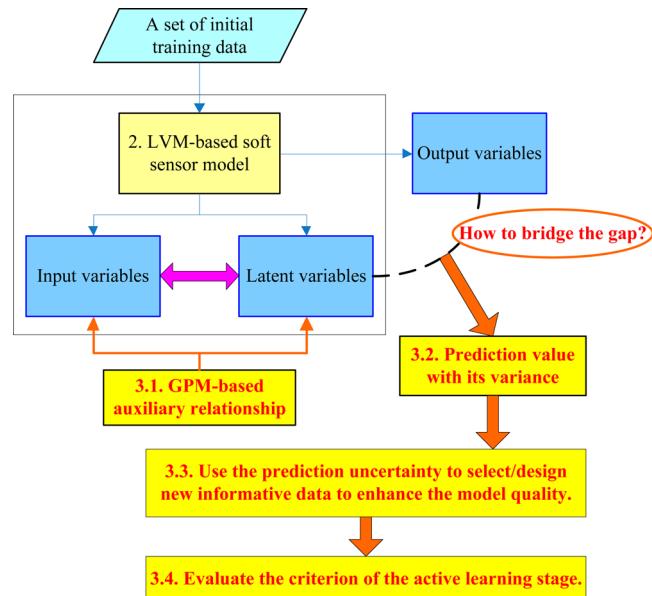


Figure 2. A brief scheme of the developed active learning method for sequential selection of new informative data to enhance the quality of LVM-based soft sensors (The numbers denote the sections; e.g., 3.1. denotes section 3.1.).

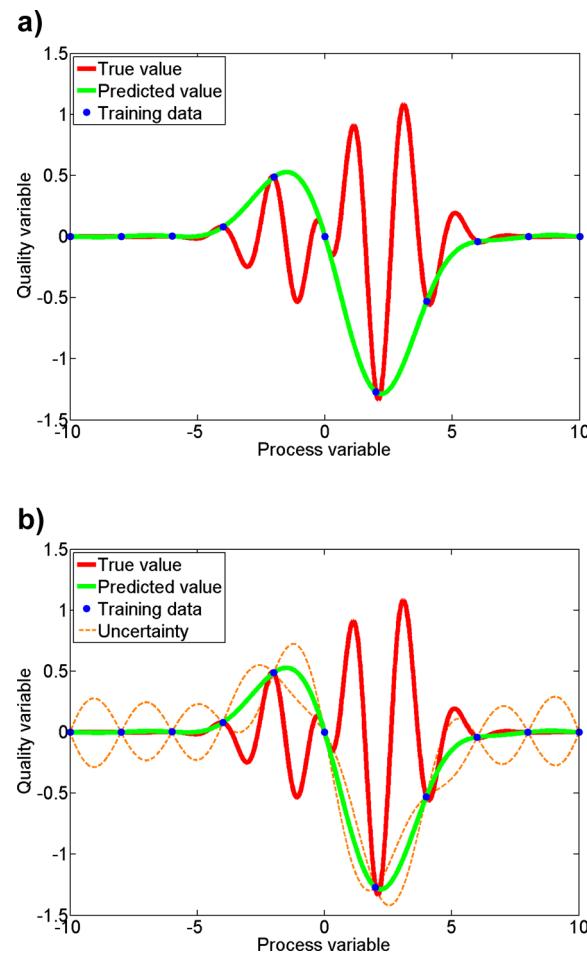


Figure 3. (a) Regression results of the traditional LVM for a simple nonlinear example. (b) Regression results of the proposed method for a simple nonlinear example. (The uncertainty information on the predicted output can also be obtained.)

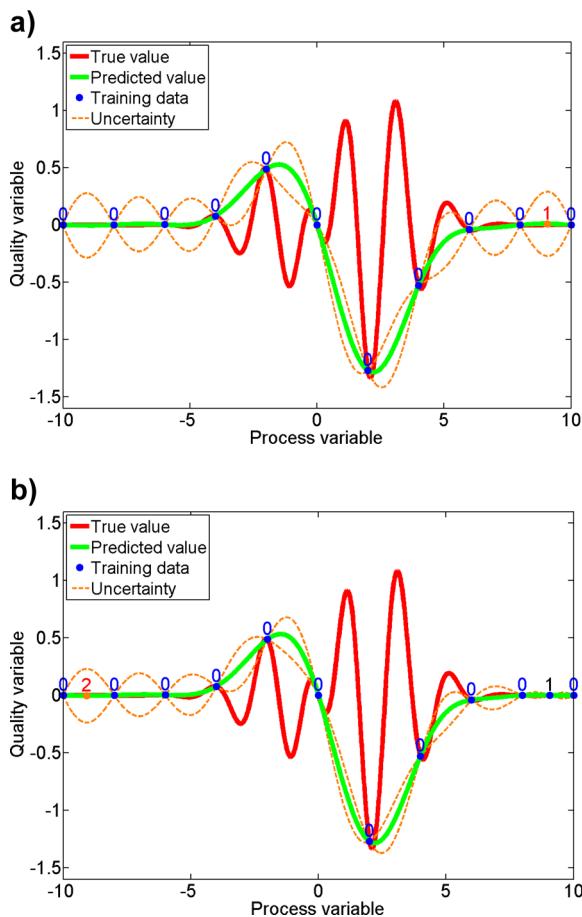


Figure 4. Implementation of the variance-based criterion in eq 12: Adding new data to the model for a simple nonlinear example. (a) Data marked with 0 are the original data in the model. The new data marked with 1 will be introduced into the model. (b) Data marked with 0 are the original data in the model. Data marked with 1 have been introduced into the model, and the data marked with 2 will next be introduced into the model.

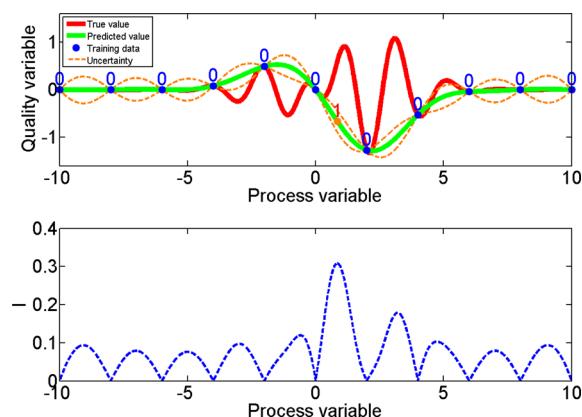


Figure 5. Implementation of the variance-and-nonlinearity-based criterion in eqs 13–15: Adding new data to the model for a simple nonlinear example (The data marked with 0 are the original data in the model. And the data marked with 1 will be introduced into the model first.).

method is applied to the multi-input–single-output process. GPM provides a prediction of the output variable for an input sample through Bayesian inference. For a single output variable of $\mathbf{y} = (y_1, \dots, y_N)^T$, GPM is the regression function with

Table 1. Main Implementation Steps of the Active Learning Approach to Sequentially Selecting New Informative Data into the LVM-Based Soft Sensor Model

Initialization: a set of original training data $\{\mathbf{X}, \mathbf{Y}\}$ collected from the historical data set

Step 1: Train an LVM-based soft sensor model.

Step 2: Learn several GPMs using \mathbf{X} and the r th latent variable \mathbf{t}_r , $r = 1, \dots, R$ (eqs 6–8).

Step 3: Derive the prediction value and the related variance (eqs 9 and 10).

Step 4: Adopt the criterion to evaluate the current model (eqs 13 and 15).

Step 5: Select the new informative data into the model while updating the training set.

Step 6: Obtain the sensitivity index and judge whether the active learning process should be continued. If $\xi_{(i)} < \rho$ (ρ can be chosen as a small positive value, e.g., 0.1), the active learning process can be stopped. Otherwise, go to Step 1.

a Gaussian prior distribution and zero mean, or in a discrete form²⁴

$$\mathbf{y} = (y_1, \dots, y_N)^T \sim G(0, \mathbf{C}) \quad (3)$$

where \mathbf{C} is the $N \times N$ covariance matrix with the ij th element defined by the covariance function, $C_{ij} = C(\mathbf{x}_i, \mathbf{x}_j)$.

Using a Bayesian approach, the hyperparameters of GPM can be estimated. Then the covariance function \mathbf{C} and GPM can be obtained. Detailed implementations on training the GPM were reported by Rasmussen and Williams.²⁴ Finally, for a test sample \mathbf{x}_b , the predicted output of y_t is also Gaussian with mean (\hat{y}_t) and variance ($\hat{\sigma}_{\hat{y}}^2$), calculated as follows:²⁴

$$\hat{y}_t = \mathbf{k}_t^T \mathbf{C}^{-1} \mathbf{y} \quad (4)$$

$$\hat{\sigma}_{\hat{y}}^2 = k_t - \mathbf{k}_t^T \mathbf{C}^{-1} \mathbf{k}_t \quad (5)$$

where $\mathbf{k}_t = [C(\mathbf{x}_b, \mathbf{x}_1), C(\mathbf{x}_b, \mathbf{x}_2), \dots, C(\mathbf{x}_b, \mathbf{x}_N)]^T$ is the covariance vector between the new input and the training samples, and $k_t = C(\mathbf{x}_b, \mathbf{x}_b)$ is the covariance of the new input. The vector $\mathbf{k}_t^T \mathbf{C}^{-1}$ denotes a smoothing term which weights the training outputs to make a prediction (eq 4) for the new input sample \mathbf{x}_t . In addition, eq 5 provides a confidence level on the model prediction, which is an appealing property of the GPM method.

3.2. Prediction Uncertainty of LVMs. This section mainly analyzes the relationship between the LVM-based soft sensor models and GPM. Correspondingly, the prediction uncertainty of LVM is derived. The prediction variance of an LVM-based soft sensor reveals which regions should be introduced by new informative data to enhance the quality of a model. The LVM-based soft sensor can be sequentially trained in an active manner.

As aforementioned, GPM provides the prediction of a single output variable. Here, a two-stage approach is proposed to describe the uncertainty of LVMs. Suppose there are altogether R latent variables \mathbf{t}_r , $r = 1, \dots, R$ ($\mathbf{T} = [\mathbf{t}_1, \dots, \mathbf{t}_R]$) when an LVM soft sensor is built for $\{\mathbf{X}, \mathbf{Y}\}$. In the first stage, the GPM method can be used to model a relationship between \mathbf{X} and one of the latent variables \mathbf{t}_r , $r = 1, \dots, R$. In the second stage, R latent variables can be related to the multiple outputs using eq 2. The prediction variance of the latent variables can be obtained using the GPM-based probabilistic inference. Therefore, the prediction uncertainty of multiple outputs \mathbf{Y} can be derived.

Like eq 3, the following GPM can be formulated for the set of input variables \mathbf{X} and the r th latent variable \mathbf{t}_r ,

$$\mathbf{t}_r = (t_{r,1}, \dots, t_{r,N})^T \sim G(0, \mathbf{C}_r) \quad (6)$$

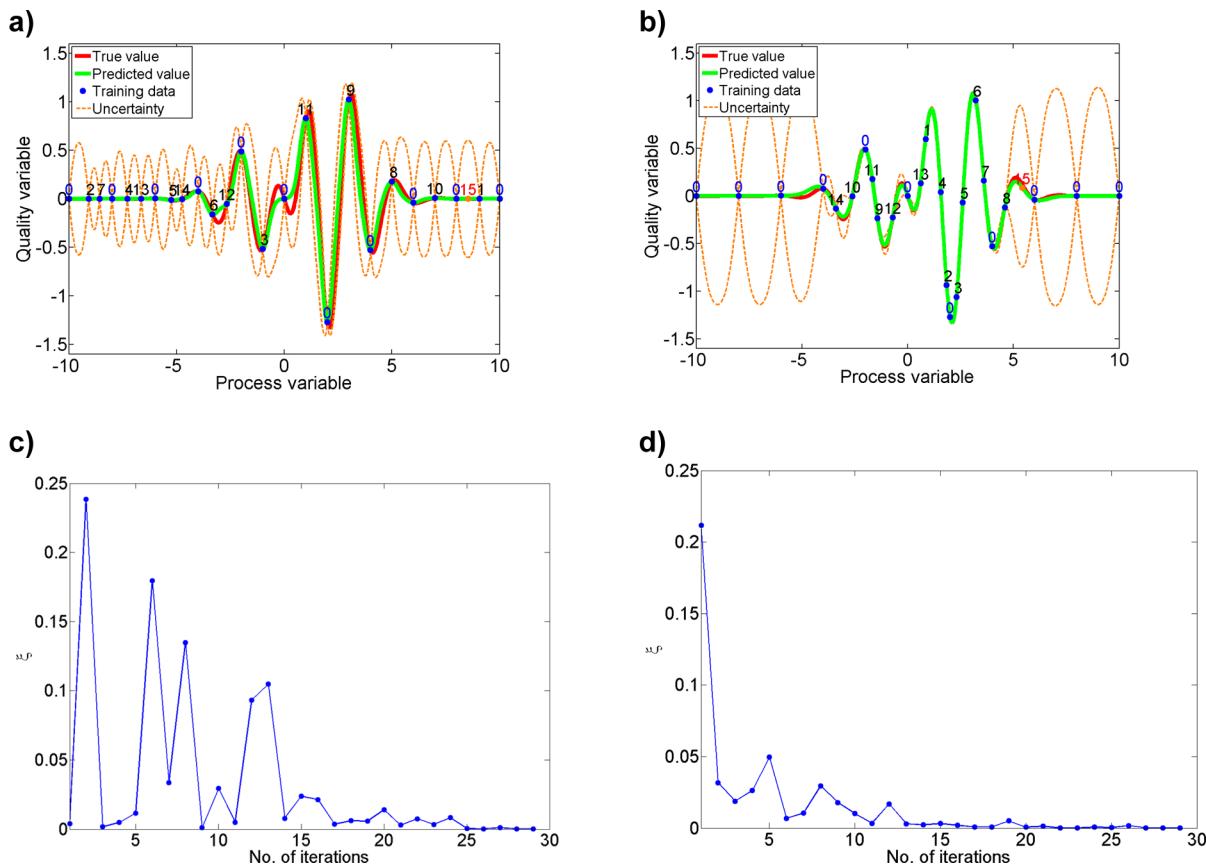


Figure 6. (a) Implementation of the variance-based criterion in eq 12: Adding new data to the model for a simple nonlinear example (15 iterations). (b) Implementation of the variance-and-nonlinearity-based criterion in eqs 13–15: Adding new data to the model for a simple nonlinear example (15 iterations). (c) Sensitivity index in the implementation of the variance-based criterion in eq 12: Adding new data to the model for a simple nonlinear example. (d) Sensitivity index in the implementation of the variance-and-nonlinearity-based criterion in eqs 13–15: Adding new data to the model for a simple nonlinear example.

where \mathbf{C}_r is the $N \times N$ covariance matrix with the ij th element defined by the covariance function, $C_{r,ij} = C_r(\mathbf{x}_i, \mathbf{x}_j)$.

For a test sample \mathbf{x}_t , the predicted mean ($\hat{t}_{r,t}$) and the variance ($\sigma_{\hat{t}_{r,t}}^2$) of the latent variable $t_{r,t}$ can be calculated, respectively:

$$\hat{t}_{r,t} = \mathbf{k}_{r,t}^T \mathbf{C}_r^{-1} \mathbf{t}_r \quad (7)$$

$$\sigma_{\hat{t}_{r,t}}^2 = k_{r,t} - \mathbf{k}_{r,t}^T \mathbf{C}_r^{-1} \mathbf{k}_{r,t} \quad (8)$$

where $\mathbf{k}_{r,t} = [C_r(\mathbf{x}_t, \mathbf{x}_1), C_r(\mathbf{x}_t, \mathbf{x}_2), \dots, C_r(\mathbf{x}_t, \mathbf{x}_N)]^T$ is the covariance vector between the new input and the training samples, and $k_{r,t} = C_r(\mathbf{x}_t, \mathbf{x}_t)$ is the covariance of the new input.

Correspondingly, for the multiple output $\mathbf{y}_t = [y_{t,1}, \dots, y_{t,q}, \dots, y_{t,Q}]^T$ of this sample \mathbf{x}_t , according to eqs 2, 7, and 8, the following formulas can be obtained.

$$\hat{y}_{t,q} = \sum_{r=1}^R \hat{t}_{r,t} b_{r,q} \quad (9)$$

$$\sigma_{\hat{y}_{t,q}}^2 = \sum_{r=1}^R \sigma_{\hat{t}_{r,t}}^2 b_{r,q}^2 \quad (10)$$

where $\hat{y}_{t,q}$ and $\sigma_{\hat{y}_{t,q}}^2$ are the predicted mean and the variance for the output, respectively; and $b_{r,q}$ is the related element of the regression matrix \mathbf{B} in eq 2. It should be noticed that eq 10 holds true because R latent variables \mathbf{t}_r , $r = 1, \dots, R$ in LVM-based soft sensor models are orthogonal.

To illustrate the difference between the traditional LVMs-based regression method and the proposed one, a simple single-input-single-output example shown in eq 11 is first adopted. Multi-input–multioutput examples will be investigated in detail in section 4.

$$f(x) = -0.85 \cos(3x) \cdot x \cdot \exp[-(0.8x - 0.4)^2] \quad (11)$$

Without any prior knowledge of the process, the training samples are even samplings of the input variable. Initially, as shown in Figure 3a, suppose there are only a few training samples evenly located in the range of $[-10, 10]$. Using a KPLS model for this nonlinear case, the regression results are shown in Figure 3b. The fitting results of available training data are very good. For those linear regions ($[-10, -5]$ and $[6, 10]$), the prediction results are good. However, for those nonlinear regions (about $[-4, 5]$), it is clearly shown that the prediction results are inaccurate.

Using the above analysis, for LVMs, the uncertainty information on the predicted output can also be obtained based on eq 10. The upper dashed line and lower dashed line show $\hat{y}_t + 3\sigma_{\hat{y}_t}$ and $\hat{y}_t - 3\sigma_{\hat{y}_t}$, respectively. (\hat{y}_t is the prediction value of $f(x_t)$ shown in eq 11.) It indicates that the data of the regions far away from the training data have a relatively large prediction uncertainty. The prediction uncertainty is important for soft sensors mainly because the prediction can show the users (e.g., operators/engineers) the reliability. In the following sections, how to use the prediction uncertainty to improve the quality of the traditional LVMs in an efficient way will be described.

3.3. Sequential Enhancement of LVMs. In this section, with the active learning method, the prediction uncertainty is used to enhance the quality of the LVMs-based soft sensors. As analyzed in [section 3.2](#), the regions with a relatively large prediction uncertainty should be enhanced to be more informative. The new data in these regions should be introduced into the current soft sensor model. Intuitively, a simple variance-based criterion of adding new data into the model can be described as

$$\mathbf{x}_t = \arg \max_{\mathbf{x}_t \in \mathbf{X}^{\text{des}}} \sigma_{\hat{y}_{t,q}}^2 \quad (12)$$

where \mathbf{X}^{des} is the designable input data set for searching candidate data. And the searchable data can be simply sampled using a grid-search method.

With the variance-based criterion in [eq 12](#), adding new data into the model for the above nonlinear example can be shown in [Figure 4](#) panels a and b, respectively. The data marked with 0 are the original data in the model. In [Figure 4a](#), the data sample marked with 1 (located nearby $x = 9$) has been introduced into the model because it has the largest prediction variance. Sequentially, in [Figure 4b](#), after adding the data marked with 1 into the model, the uncertainty of this region can become smaller. In a sequential step, the data marked with 2 (located nearby $x = -9$) will be introduced into the model because it has the largest prediction variance.

It should be noticed that the new data marked with 1 introduced into the model can reduce the prediction variance of the related region. However, the quality of the model has not been improved obviously mainly because the variance-based criterion only considers the prediction variance. Actually, as shown in [Figure 4a,b](#) that the input regions of $[0, 5]$ should be improved first as the strong nonlinearity should be learned by the model. For real processes, how to describe the process characteristics, for example, the nonlinearity, and then introduce the significant information into the criterion to enhance the model quality is important.

To overcome the drawback of the variance-based criterion in [eq 12](#), a **variance-and-nonlinearity-based criterion** can be expressed by

$$\mathbf{x}_t = \arg \max_{\mathbf{x}_t \in \mathbf{X}^{\text{des}}} I_{\hat{y}_{t,q}} \quad (13)$$

$$I_{\hat{y}_{t,q}} = \sigma_{\hat{y}_{t,q}}^2 + \lambda \cdot \sigma_{\hat{y}_{t,q}}^2 \cdot \Delta m_{\hat{y}_{t,q}} \quad (14)$$

$$\Delta m_{\hat{y}_{t,q}} = \left(\frac{\partial \hat{y}_{t,q}}{\partial x_1} \right)^2 + \dots + \left(\frac{\partial \hat{y}_{t,q}}{\partial x_D} \right)^2 \quad (15)$$

where $I_{\hat{y}_{t,q}}$ is the new index, including the information on both the prediction variance and the process nonlinearity. Here, with little prior knowledge, the process information mainly relates to the effect of the input variables on the prediction value; that is, $\Delta m_{\hat{y}_{t,q}}$. The prediction items of LVM-based soft sensor models are related to the input variables. Consequently, the information on $\Delta m_{\hat{y}_{t,q}}$ can be obtained using [eq 15](#). The detailed formulas of $\Delta m_{\hat{y}_{t,q}}$ for different LVM-based models are not provided here because different LVM-based model structures have different prediction items. [Equation 14](#) consists of two parts, the prediction variance and the process nonlinearity. The weighting parameter is $\lambda \geq 0$. If $\lambda = 0$, only the variance-based criterion is considered. For highly nonlinear processes, a relatively large value of λ should be chosen. To effectively balance these two parts, in this study, λ is adjusted based on the estimated

$\sigma_{\hat{y}_{t,q}}^2$; this would not cause either the prediction variance or only the process nonlinearity to be considered.

This variance-and-nonlinearity-based criterion indicates that, on the one hand, the input regions with a relatively large prediction uncertainty ($\sigma_{\hat{y}_{t,q}}^2$) should be enhanced to be more informative. On the other hand, those regions with more complex relationship between the input variables and the prediction output ($\Delta m_{\hat{y}_{t,q}}$) should also be considered.

With the variance-and-nonlinearity-based criterion in [eqs 13–15](#), adding new data into the model for the simple nonlinear example is shown in [Figure 5](#). The data marked with 0 are the original data in the model. And the data point marked with 1 has the largest value of index $I_{\hat{y}_{t,q}}$ (shown at the bottom of [Figure 5](#))

and it will be introduced into the model. Unlike [Figure 4a](#), the first data sample introduced into the model locates near $x = 1$. The relationship between the input and the output in this region is more complex than the region near $x = 9$ (shown in [Figure 4a](#)). Considering the effect of the input variables on the prediction value, the index becomes more informative for mining the new data. Consequently, compared with the variance-based criterion in [eq 12](#), the variance-and-nonlinearity-based criterion in [eqs 13–15](#) can enhance the quality of the model in a more efficient manner.

3.4. Model Evaluation Criterion. In this section, an evaluation criterion of the active learning stage is proposed to judge whether any new data should be introduced into the model. A sensitivity index is designed as follows:

$$\xi_{(i)} = \sqrt{\frac{1}{Q} \int_{\mathbf{x}_t \in \mathbf{X}^{\text{des}}} \int_{\mathbf{x}_t \in \mathbf{X}^{\text{des}}} (\hat{y}_{(i-1)} - \hat{y}_{(i)})^T (\hat{y}_{(i-1)} - \hat{y}_{(i)}) d\mathbf{x}} \quad (16)$$

where $\xi_{(i)}$ is the sensitivity index of the i th iteration of the active learning procedure for the process output. The items of $\hat{y}_{(i-1)}$ and $\hat{y}_{(i)}$ are the average prediction values of all the output variables in the $(i-1)$ th and the i th iterations, respectively. It means that, for the designable input data set \mathbf{X}^{des} , if the predictions of the i th iteration and the $(i-1)$ th iteration are almost the same after several iterations, the active learning procedure can be finished. This is mainly because most of the information in the designable input data set has been introduced into the model. Otherwise, if the sensitivity index $\xi_{(i)}$ still changes to some extent, the active learning procedure should be continued to further explore the input data set. A user-defined parameter ρ is chosen to judge if the sensitivity index $\xi_{(i)} < \rho$. A small positive value can be chosen as the value of ρ .

In summary, the implementation steps of the active learning method, that sequentially select the new informative data into the model, are listed in [Table 1](#). To show the effectiveness of the active learning method, the iterations of the implementation steps using the two criteria listed in [section 3.3](#) are shown in [Figure 6](#) panels a–d. In [Figure 6](#) panels a and b, the data marked with the numbers denote the iteration numbers. With 15 iterations, the active learning method with the variance-and-nonlinearity-based criterion in [eqs 13–15](#) can enhance the model better than using the variance-based criterion in [eq 12](#). Additionally, in [Figure 6](#) panels c and d, the iterations of the sensitivity index are compared. The panels show that the variance-and-nonlinearity-based criterion in [eqs 13–15](#) make the active learning process more efficient. This is mainly because more informative data are explored and then quickly introduced into the soft sensor model.

Table 2. Main Information, Including the Mean and the Variance of the Designable Input Variables of x_1 and x_2 (the MIMO Case in Section 4.1)

operating conditions	x_1	x_2	no. of original training data
C_1	$N(0.5, 0.2)$	$N(-0.5, 0.2)$	10
C_2	$N(-0.5, 0.2)$	$N(-0.5, 0.2)$	5
C_3	$N(-0.5, 0.2)$	$N(0.5, 0.2)$	10
C_4	$N(0.5, 0.2)$	$N(0.5, 0.2)$	5

Remark 1: The proposed method aims to actively select the most informative samples for model training. The samples introduced into the model are only part of the historical set. This method resembles sparse learning to some extent. The proposed method is different from the sparse learning strategy for model training as the latter is used to obtain a sparse structure of the model. For example, for parsimonious modeling from the data, the orthogonal least-squares algorithm has been used to train the radial basis function networks and several

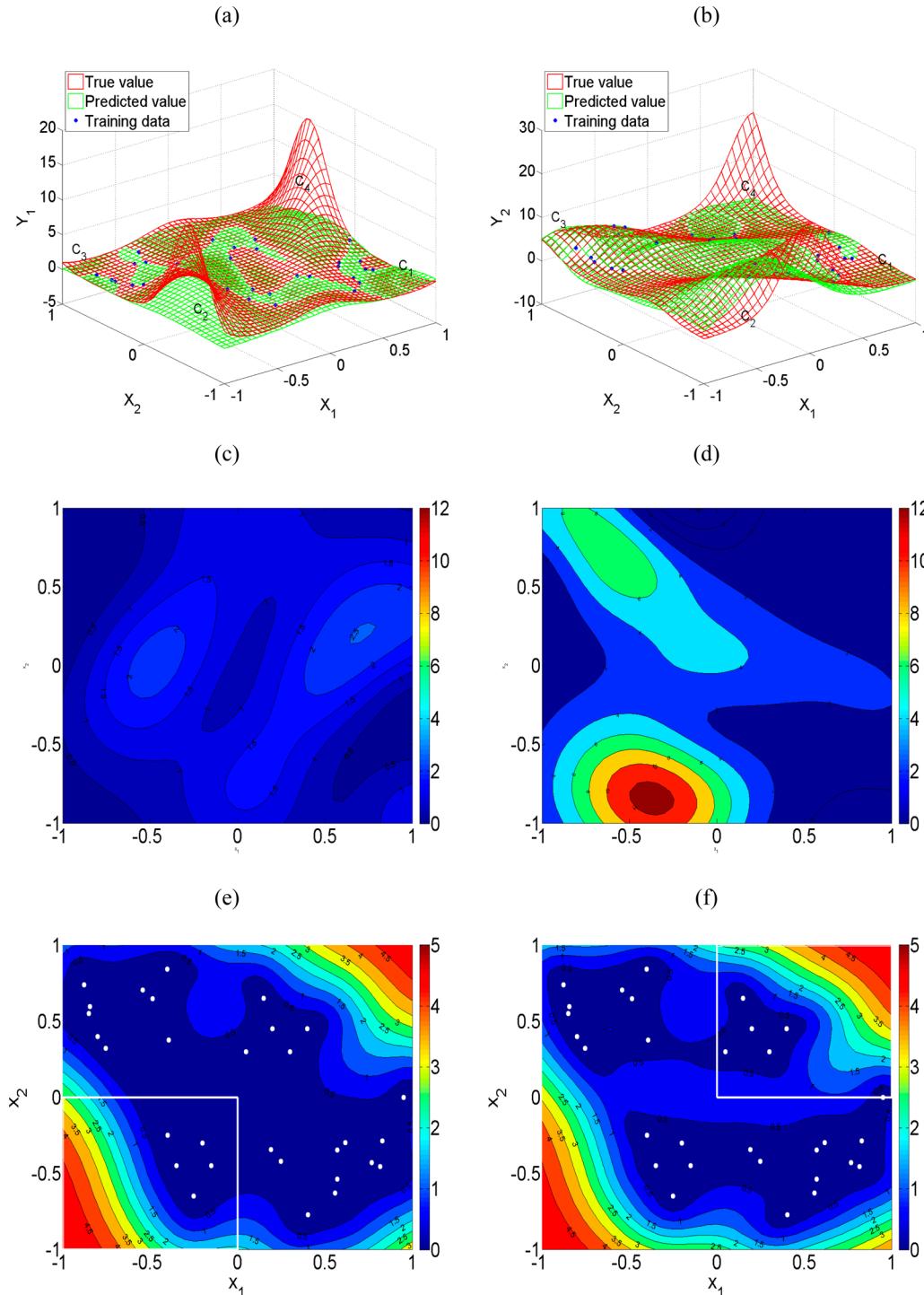


Figure 7. Initial training results for the MIMO numerical example: prediction results of (a) y_1 and (b) y_2 ; process nonlinearity information described in eq 15 for (c) y_1 and (d) y_2 ; prediction variance information on (e) y_1 and (f) y_2 .

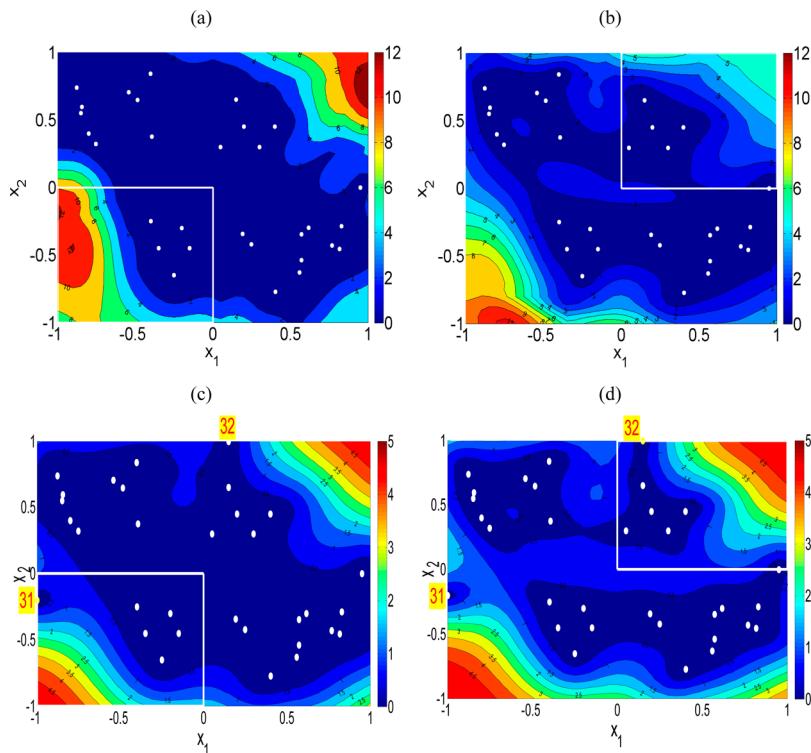


Figure 8. Contour lines using the improvement indices of eq 20 for (a) y_1 and (b) y_2 . At the first iteration, the new informative data denoted as 31 and 32 included in the model are shown in panels c and d, respectively.

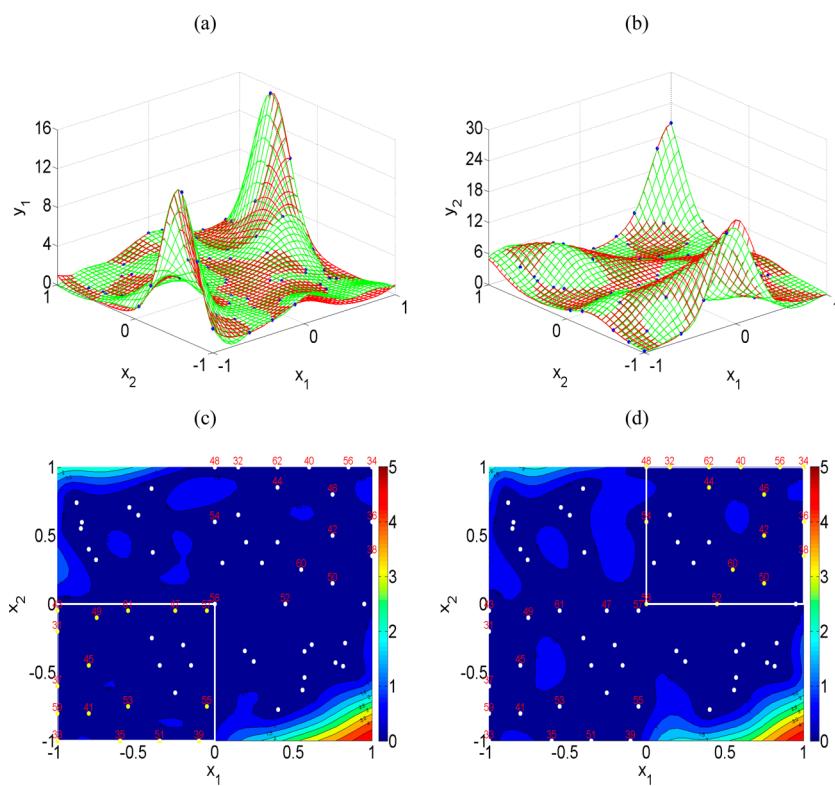


Figure 9. Training results after 16 iterations in the MIMO numerical example: prediction results of (a) y_1 and (b) y_2 ; introducing data and contour lines using the improvement indices of eq 23 for (c) y_1 and (d) y_2 .

regression models.^{34,35} The sparse learning strategy focuses on obtaining a sparse structure with a fixed number of data. The aim of our work is to actively select informative data to sequentially enhance the quality of LVM-based soft sensor

models. The active learning scheme concentrates on selecting informative data from the historical data set. The number of training data is unfixed. This means that there are rich data to be selected.

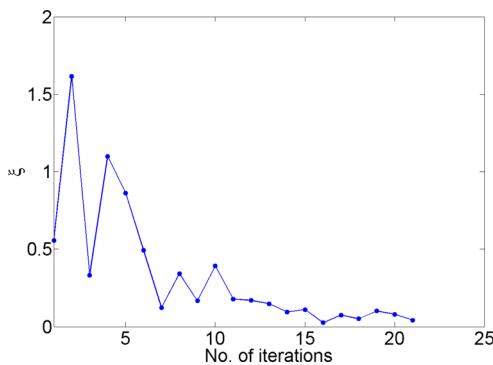


Figure 10. Evolution of the sensitivity index in the MIMO numerical example (a total of 21 iterations).

Table 3. Values of RMSE and RRMSE Using Different LVM-Based Soft Sensor Models (the MIMO Case in Section 4.1)

	PLS ¹	QPLS ¹⁵	NNPLS ¹⁵	Lin ¹⁵	this work
RMSE	0.0393	0.0374	0.0213	0.0208	0.0159
RRMSE (%)	25.51	24.30	13.85	13.53	11.07
no. of training data	400	400	400	400	72

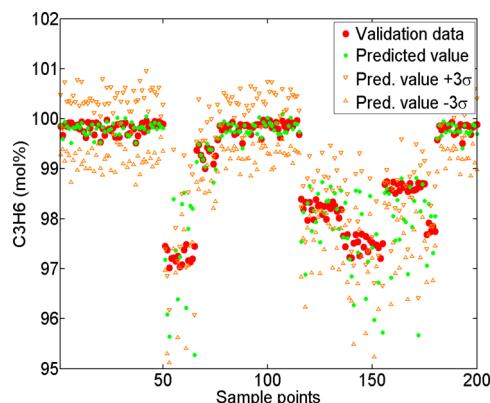


Figure 11. Prediction results of C_3H_6 on the testing data using the traditional training LVM-based soft sensor model for an industrial FCCU process.

4. ILLUSTRATIVE EXAMPLES

In this section, from different aspects and for comparison purposes, two MIMO examples are investigated. The first simulated case is used to illustrate its active learning steps for multiple outputs. If the established soft sensor with initial samples is not accurate, it should be improved by adding new representative data. The proposed method can be used to select those informative data to sequentially improve the LVM-based model. Other LVM soft sensors with the traditional training method are also compared. The second MIMO case further evaluates the characteristics using an industrial data set. It shows how to establish a good LVM soft sensor with those informative data hidden in the historical set.

4.1. A Numerical MIMO Example. First, a numerical MIMO example is investigated to show how the proposed active learning approach enhances the LVM-based soft sensor model in a sequential and efficient way. In this example, four input variables (simply noted as x_1 , x_2 , x_3 , and x_4) and two output variables (simply noted as y_1 and y_2)¹⁵ are used. Suppose two input variables are designable, that is, x_1 and x_2 . The relationships of this numerical example can be shown in eq 17.

$$\begin{aligned} x_3 &= x_1 + x_2 \\ x_4 &= x_1 - x_2 \\ y_1 &= \exp[2x_1 \sin(\pi x_4) + \sin(x_2 x_3)] \\ y_2 &= \exp[2x_2 \cos(\pi x_3) + \cos(x_1 x_4)] \end{aligned} \quad (17)$$

In this example, assume that some operating conditions are relatively important because they are frequently run; there are four operating conditions noted as C_1 , C_2 , C_3 , and C_4 . And the data for constructing a soft sensor model can be generated from these operating areas using the designable inputs of x_1 and x_2 .

The means and the variances of the designable input variables of x_1 and x_2 are shown in Table 2. Suppose in the initial stage, there is a total of 30 data points for training an LVM-based soft sensor model. Here the KPLS model is adopted because of its nonlinear modeling ability. The original training results and basic information are shown in Figure 7 with six subplots. The prediction results for y_1 and y_2 are shown in panels a and b, respectively. It is found that the prediction results for some areas in the operating conditions of C_2 and C_4 are not accurate. One main reason is that the training samples are relatively fewer in the areas of C_2 and C_4 (Table 2). Also, the process nonlinearity information described in eq 15 for y_1 and y_2 is shown in Figure 7 panels c and d, respectively. The prediction variance information on y_1 and y_2 is shown in panels e and f, respectively.

The active learning approach improves the model by introducing new data. Correspondingly, the designable operating areas of Ω_{C_2} and Ω_{C_4} for searching new data can be chosen as follows:

$$\begin{aligned} \Omega_{C_2}: x_1 &\in [-1, 0], x_2 \in [-1, 0] \\ \Omega_{C_4}: x_1 &\in [0, 1], x_2 \in [0, 1] \end{aligned} \quad (18)$$

Like in eqs 13–15, the variance-and-nonlinearity-based indices and criteria for searching new informative data can be formulated in eqs 19 and 20, respectively, to enhance the model quality of y_1 and y_2 . Two new data can be introduced into the model at each iteration,

$$\mathbf{x}_t = \arg \max_{\mathbf{x}_t \in \Omega_{C_2}} I_{\hat{y}_{t,1}} \cup \arg \max_{\mathbf{x}_t \in \Omega_{C_4}} I_{\hat{y}_{t,2}} \quad (19)$$

where \mathbf{x}_t contains two selected data. It is the union of two informative solutions of y_1 ($I_{\hat{y}_{t,1}}$) and y_2 ($I_{\hat{y}_{t,2}}$). $I_{\hat{y}_{t,1}}$ and $I_{\hat{y}_{t,2}}$ are defined as

$$\begin{aligned} I_{\hat{y}_{t,1}} &= \sigma_{\hat{y}_{t,1}}^2 + \lambda \cdot \sigma_{\hat{y}_{t,1}}^2 \cdot \Delta m_{\hat{y}_{t,1}} \\ I_{\hat{y}_{t,2}} &= \sigma_{\hat{y}_{t,2}}^2 + \lambda \cdot \sigma_{\hat{y}_{t,2}}^2 \cdot \Delta m_{\hat{y}_{t,2}} \end{aligned} \quad (20)$$

Finally, when the active learning process is completed, two common indices of the root-mean-square error (RMSE) and relative RMSE (RRMSE) are adopted to assess the prediction performance, as shown in eqs 21 and 22.

$$\text{RMSE} = \sqrt{\frac{1}{T} \frac{1}{Q} \sum_{t=1}^T \sum_{q=1}^Q (\hat{y}_{t,q} - y_{t,q})^2} \quad (21)$$

$$\text{RRMSE\%} = \sqrt{\frac{1}{T} \frac{1}{Q} \sum_{t=1}^T \sum_{q=1}^Q \left(\frac{\hat{y}_{t,q} - y_{t,q}}{y_{t,q}} \right)^2} \times 100 \quad (22)$$

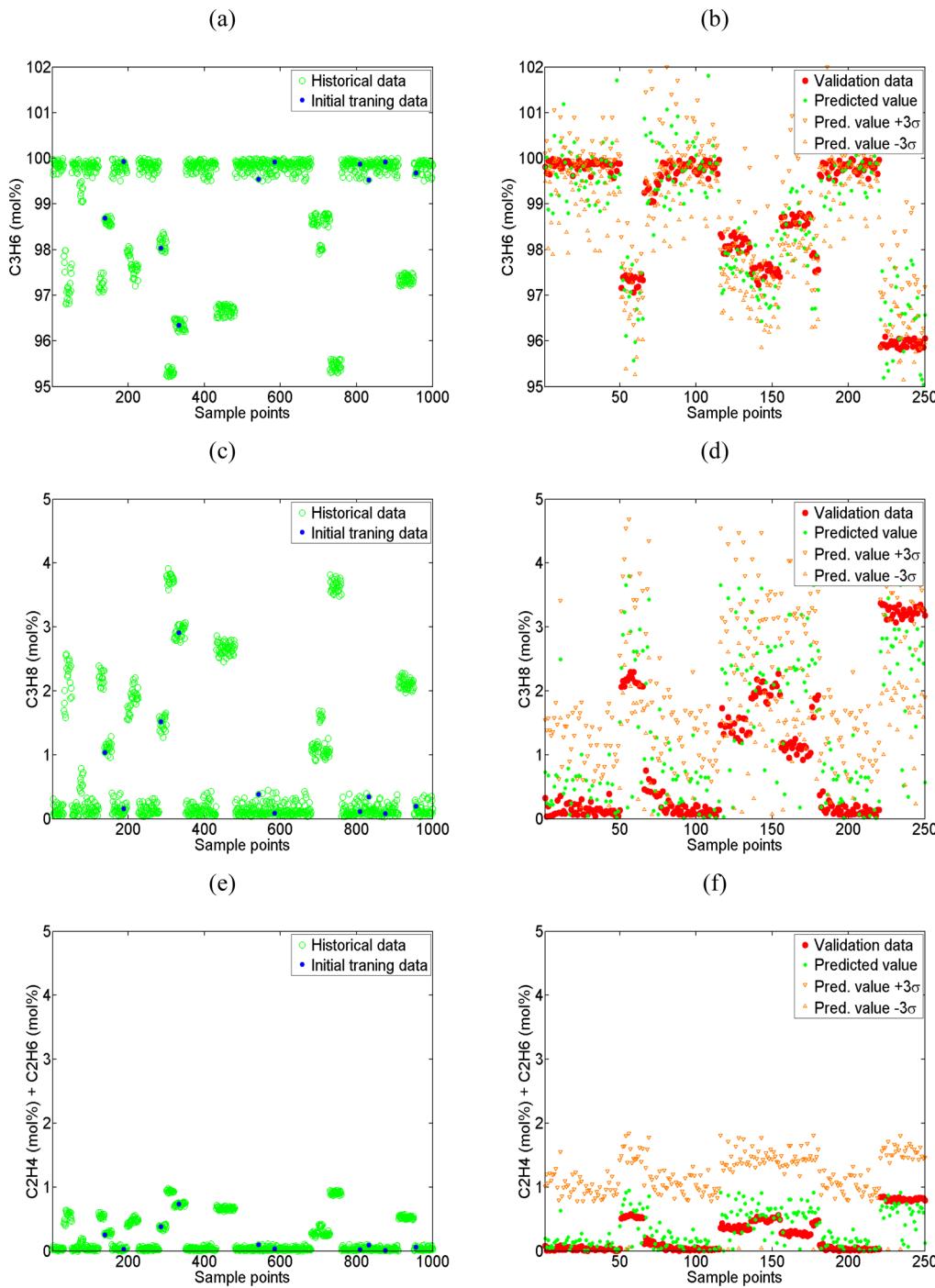


Figure 12. Original training and testing results of the proposed method for an industrial FCCU process: (a) the initial sampling data for C₃H₆ in the training set; (b) the prediction results of C₃H₆ on the testing data; (c) the initial sampling data for C₃H₈ in the training set; (d) the prediction results of C₃H₈ on the testing data; (e) the initial sampling data for C₂H₄ + C₂H₆ in the training set; (f) the prediction results of C₂H₄ + C₂H₆ on the testing data.

where T is the number of testing samples and Q is the number of outputs.

The indices of y_1 and y_2 in eq 20 are shown in panels a and b of Figure 8, respectively. Compared with the prediction variance information on y_1 and y_2 shown in panels e and f of Figure 7, the contour lines are different. This means that the informative data to be first introduced into the model are different. Note that there are initially 30 data points for the model to be trained. The corresponding results of the newly selected data are noted as 31 and 32 (Figure 8c,d), respectively. This is the first implementation of active data

selection to improve the model (introducing two new data in an iteration).

As aforementioned, the active learning process is implemented in a sequential manner. To exhibit the iterative steps more clearly, the active learning process after 16 iterations is shown in Figure 9 with four subplots. The prediction results for y_1 and y_2 are shown in panels a and b of Figure 9, respectively. And the indices of y_1 and y_2 in eq 20 are shown in panels c and d of Figure 9, respectively. It is found that the quality of the model has been improved greatly after 16 iterations. For this case, the active learning process can be stopped if the sensitivity index $\xi_{(i)} < 0.05$

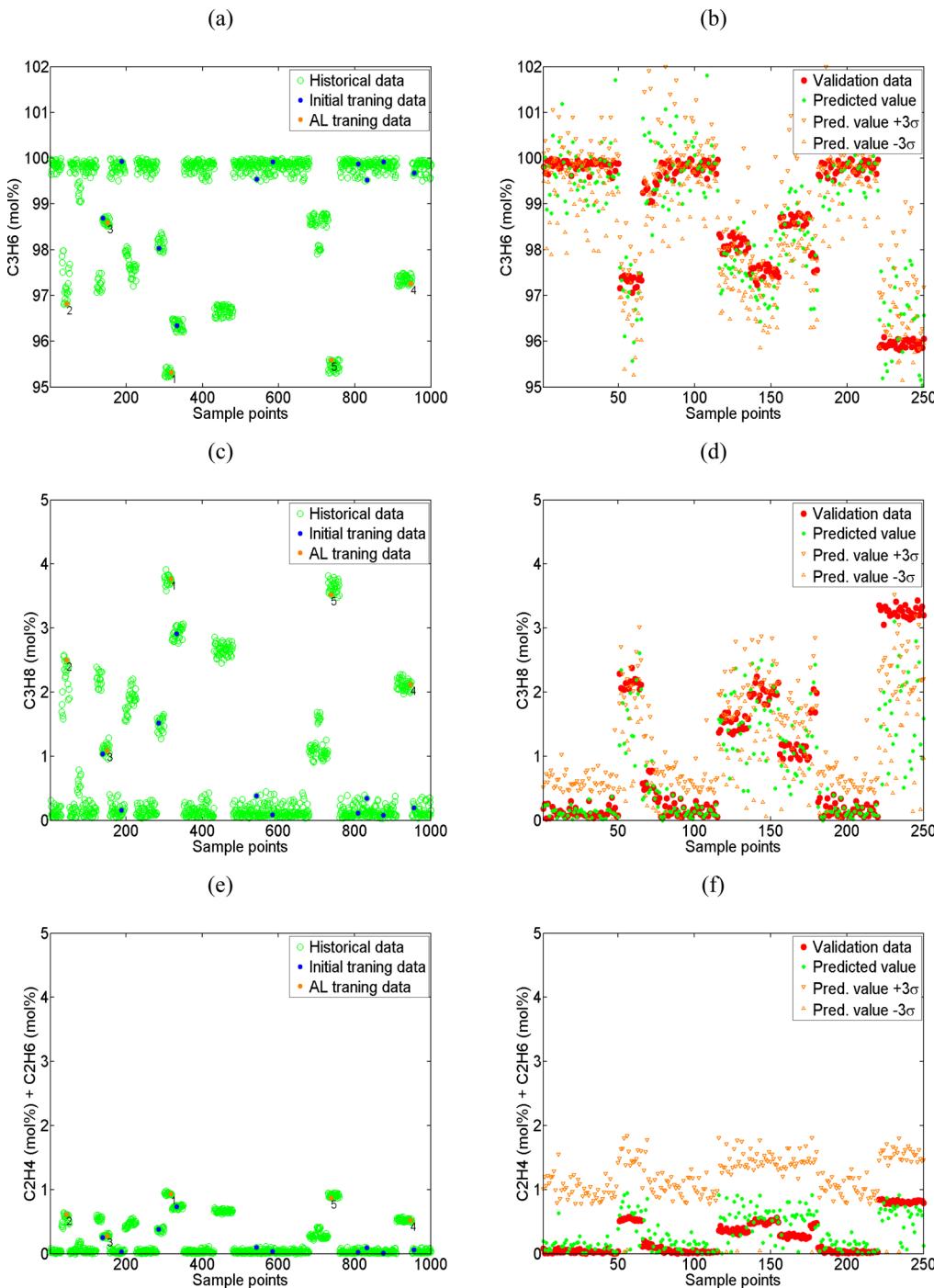


Figure 13. Fifth iteration of active and sequential selection of new informative data into the LVM-based soft sensor model for an industrial FCCU process: (a) the selected data for C_3H_6 in the training set; (b) the prediction results of C_3H_6 on the testing data; (c) the selected data for C_3H_8 in the training set; (d) the prediction results of C_3H_8 on the testing data; (e) the selected data for $C_2H_4 + C_2H_6$ in the training set; (f) the prediction results of $C_2H_4 + C_2H_6$ on the testing data.

is satisfied. Finally, the evolution of the sensitivity index with 21 iterations is shown in Figure 10. The figure shows that the trend is decreasing with iterations. Actually, after 10 iterations, the quality of the soft sensor model has been enhanced greatly.

Detailed comparisons between the proposed method and several existing PLS soft sensors¹⁵ are listed in Table 3. The comparison results show that the proposed method has the best prediction performance. The traditional PLS methods are not trained by an active learning approach. And all of them adopt

400 data while the proposed method only chooses 72 data, including the initial 30 and the latter 42 after the active data selection by 21 iterations. These data are delicately designed, so they are more informative when training a model. Therefore, this numerical MIMO example shows the superiority of the active learning method.

4.2. An Industrial Process. An industrial fluid catalytic cracking unit (FCCU) in Taiwan is investigated to further validate how the proposed active learning approach enhances the LVM-based soft sensor model in a sequential and efficient way.

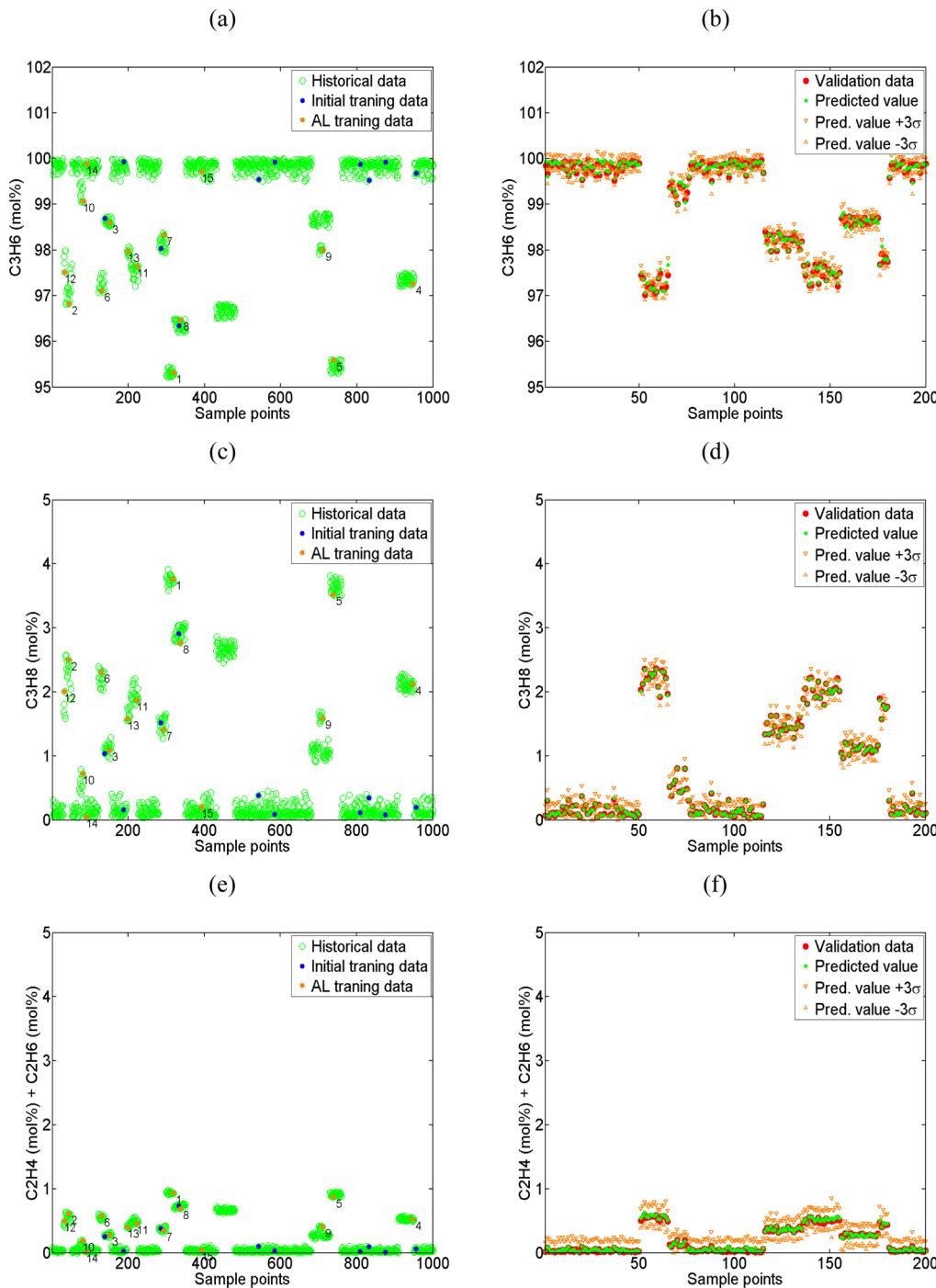


Figure 14. 15th iteration of active and sequential selection of new informative data into the LVM-based soft sensor model for an industrial FCCU process: (a) the selected data for C_3H_6 in the training set; (b) the prediction results of C_3H_6 on the testing data; (c) the selected data for C_3H_8 in the training set; (d) the prediction results of C_3H_8 on the testing data; (e) the selected data for $C_2H_4 + C_2H_6$ in the training set; (f) the prediction results of $C_2H_4 + C_2H_6$ on the testing data.

FCCU is the core unit of the oil secondary operation and an essential equipment of the refinery. A typical FCCU process is able to convert heavy gas oils into a range of lighter hydrocarbon products, that is, gasoline. It plays an important role in the overall economic performance of the refinery.^{35,36} The concentrations of C_3H_6 , C_3H_8 , C_2H_4 , and C_2H_6 of the depropanizer are the output products. However, they cannot be measured online, so soft sensors are necessary in practice. Among them, the concentration of C_3H_6 is considered as the most important one in the production process. According to the lab results, its

concentration is between 95 mol % and 100 mol %. The concentrations of C_3H_8 , C_2H_4 , and C_2H_6 are relatively low. To simplify the modeling flowchart, two outputs of C_2H_4 and C_2H_6 are combined as $C_2H_4 - C_2H_6$. Consequently, there are three main outputs, C_3H_6 , C_3H_8 , and $C_2H_4 - C_2H_6$, to be predicted by soft sensors. There are seven input measurable variables, including the feed temperature, the condenser temperature, the temperature of tray 1, the temperature of tray 21, the temperature of tray 40, the temperature of the reboiler, and the top pressure.

First, the traditional training method is investigated. Suppose 100 data are evenly sampled from the historical set. And the prediction performance is evaluated using a test data set of 200 samples. KPLS is used to construct a nonlinear soft sensor model. Two components for 95% of the cumulative percentage of variance have been selected. Take C_3H_6 as an example. The main prediction results for the testing data are shown in Figure 11. It is found that the prediction performance for most data around 99.5 mol % is accurate. However, the prediction performance for data in other areas is relatively inaccurate, and it consists of a larger variance. This is mainly because the training data are evenly sampled from the historical set using the traditional training method. It indicates that most of them are selected around 99.5 mol %.

Compared with the traditional training method, the KPLS soft sensor model with the active learning strategy is implemented as follows. Initially, only 10 data are evenly sampled to train the soft sensor models for C_3H_6 , C_3H_8 , and $C_2H_4-C_2H_6$ concentrations in Figure 12 panels a, c, and e, respectively. The prediction results of C_3H_6 , C_3H_8 , and $C_2H_4-C_2H_6$ concentrations are shown in Figure 12 panels b, d, and f, respectively. In any sample, the prediction value \hat{y}_t and its variance are both provided. The upper point and the lower point show $\hat{y}_t + 3\sigma_{\hat{y}_t}$ and $\hat{y}_t - 3\sigma_{\hat{y}_t}$, respectively. It is found that the prediction variances of most of the data are relatively large, mainly because initially there are not enough training data to capture the process characteristics.

The historical set can be a candidate data pool for selecting informative data to improve the model. As aforementioned, the active learning process is implemented in a sequential way. To exhibit the iteration steps more clearly, the fifth and the 15th iterations of adding new informative data into the model are shown in Figures 13 and 14, respectively. The sample marked with number means that it is selected as a meaningful one in the i th iteration process. It is found that the prediction uncertainty for all the three outputs becomes smaller. Take C_3H_6 for example. Compared with the original prediction results in Figure 12b, the predictions of the testing samples in Figure 14b become more accurate. The evolution of the sensitivity index is shown in Figure 15. The trend is decreasing with more iterations

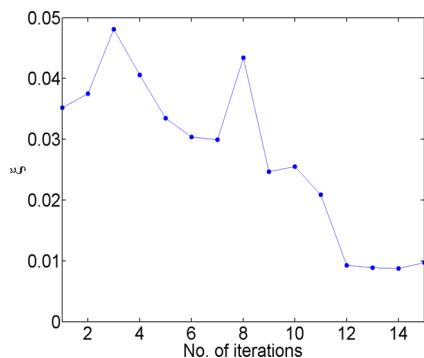


Figure 15. Evolution of the sensitivity index of an industrial FCCU process (a total of 15 iterations).

and it should be stopped when it is small enough, for example, $\xi_{(i)} < 0.01$ for this case. Finally, a total of 15 data points is selected to train the model. The proposed method requires significantly less data than the traditional training method.

When the active learning process is completed, two indices of RMSE and RRMSE are adopted to assess the prediction performance on the testing data. To make the results more

reasonable, the above training procedures are implemented for 30 times, each time with a different original set of training data. The average comparison results of the active learning method and the traditional one are listed in Table 4. It is found that the

Table 4. Average Comparisons between the Proposed Active Learning Method and the Traditional Training Method for 30 Times (the Industrial Case in Section 4.2)

	no. of training data	RMSE	RRMSE (%)	prediction variance
traditional training method	100	0.1236	32.15	0.2022
active learning	28	0.0252	26.02	0.0329

proposed one with much less training data has a better prediction performance and a smaller prediction variance. Therefore, the obtained results show that the proposed active learning method can be simply and efficiently implemented to sequential training of an LVM-based soft sensor model.

5. CONCLUSION

Many soft sensor modeling methods focus on how to develop suitable methods for different chemical processes and different process characteristics. Unlike the past research, this study addresses the topic of actively selecting informative data to sequentially enhance the quality of a soft sensor prediction model. Its distinguishing characteristics can be summarized in three main aspects:

- (1) The prediction uncertainty is explored and integrated into the soft sensor development procedure to enhance the quality of a prediction model.
- (2) An active learning method that sequentially selects a set of informative data into the model is developed. Without any prior knowledge of the process, it can still be implemented in an efficient way.
- (3) The idea can be simply extended to other LVM-based soft sensor models.

It should also be noted that, using the active learning method, much human effort can be saved on labeling/analyzing data in industrial processes. This is mainly because only those informative data that enhance the quality of a soft sensor model should be labeled. The superiority of the proposed active learning method has been validated through an MIMO numerical example and an industrial process. The proposed method considers LVM-based soft sensors only. To extend the active learning method to develop GPM-based soft sensors for MIMO processes with correlated input variables would be an interesting study for future research.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: jason@wavenet.cycu.edu.tw. Tel.: +886-3-2654107. Fax: +886-3-2654199.

ORCID

Junghui Chen: [0000-0002-9994-839X](http://orcid.org/0000-0002-9994-839X)

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors would like to gratefully acknowledge the Ministry of Science and Technology, R.O.C. (MOST 103-2221-E-033-068-MY3) and the National Natural Science Foundation of China (Grant No. 61640312).

ABBREVIATIONS

- GPM = Gaussian process model
 KPCR = kernel principal component regression
 KPLS = kernel partial least-squares
 LVMs = latent variable models
 MIMO = multi-input–multi-output
 NN = neural networks
 PCA = principal component analysis
 PCR = principal component regression
 PLS = partial least-squares
 RMSE = root-mean-square error
 RRMSE = relative root-mean-square error
 SVR = support vector regression

REFERENCES

- (1) Fortuna, L.; Graziani, S.; Rizzo, A.; Xibilia, M. G. *Soft Sensors for Monitoring and Control of Industrial Processes*; Springer-Verlag: New York, 2007.
- (2) Kadlec, P.; Gabrys, B.; Strandt, S. Data-driven soft sensors in the process industry. *Comput. Chem. Eng.* **2009**, *33*, 795–814.
- (3) Kadlec, P.; Grbic, R.; Gabrys, B. Review of adaptation mechanisms for data-driven soft sensors. *Comput. Chem. Eng.* **2011**, *35*, 1–24.
- (4) Kano, M.; Ogawa, M. The state of the art in chemical process control in Japan: Good practice and questionnaire survey. *J. Process Control* **2010**, *20*, 969–982.
- (5) Khatibisepehr, S.; Huang, B.; Khare, S. Design of inferential sensors in the process industry: A review of Bayesian methods. *J. Process Control* **2013**, *23*, 1575–1596.
- (6) Undey, C.; Cinar, A. Statistical monitoring of multistage, multiphase batch processes. *IEEE Control Syst. Mag.* **2002**, *22*, 40–52.
- (7) Yao, Y.; Gao, F. R. A survey on multistage/multiphase statistical modeling methods for batch processes. *Annu. Rev. Control* **2009**, *33*, 172–183.
- (8) Ge, Z. Q. Active learning strategy for smart soft sensor development under a small number of labeled data samples. *J. Process Control* **2014**, *24*, 1454–1461.
- (9) Qin, S. J. Recursive PLS algorithms for adaptive data modeling. *Comput. Chem. Eng.* **1998**, *22*, 503–514.
- (10) Zhao, C. H.; Wang, F. L.; Mao, Z. Z.; Lu, N. Y.; Jia, M. X. Quality prediction based on phase-specific average trajectory for batch processes. *AIChE J.* **2008**, *54*, 693–705.
- (11) Galicia, H. J.; He, Q. P.; Wang, J. Comparison of the performance of a reduced-order dynamic PLS soft sensor with different updating schemes for digester control. *Control Eng. Pract.* **2012**, *20*, 747–760.
- (12) Shao, W. M.; Tian, X. M.; Wang, P.; Deng, X. G.; Chen, S. Online soft sensor design using local partial least squares models with adaptive process state partition. *Chemom. Intell. Lab. Syst.* **2015**, *144*, 108–121.
- (13) Zhang, J.; Jin, Q.; Xu, Y. M. Inferential estimation of polymer melt index using sequentially trained bootstrap aggregated neural networks. *Chem. Eng. Technol.* **2006**, *29*, 442–448.
- (14) Mat Noor, R. A.; Ahmad, Z.; Mat Don, M.; Uzir, M. H. Modelling and control of different types of polymerization processes using neural networks technique: A review. *Can. J. Chem. Eng.* **2010**, *88*, 1065–1084.
- (15) Liu, J. L. On-line soft sensor for polyethylene process with multiple production grades. *Control Eng. Pract.* **2007**, *15*, 769–778.
- (16) Chitalekha, S. B.; Shah, S. L. Application of support vector regression for developing soft sensors for nonlinear processes. *Can. J. Chem. Eng.* **2010**, *88*, 696–709.
- (17) Kaneko, H.; Funatsu, K. Development of soft sensor models based on time difference of process variables with accounting for nonlinear relationship. *Ind. Eng. Chem. Res.* **2011**, *50*, 10643–10651.
- (18) Liu, Y.; Chen, J. Integrated soft sensor using just-in-time support vector regression and probabilistic analysis for quality prediction of multi-grade processes. *J. Process Control* **2013**, *23*, 793–804.
- (19) Liu, Y.; Gao, Z. L.; Chen, J. Development of soft-sensors for online quality prediction of sequential-reactor-multi-grade industrial processes. *Chem. Eng. Sci.* **2013**, *102*, 602–612.
- (20) Yuan, X. F.; Ge, Z. Q.; Song, Z. H. Locally weighted kernel principal component regression model for soft sensing of nonlinear time-variant processes. *Ind. Eng. Chem. Res.* **2014**, *53*, 13736–13749.
- (21) Zhang, Y. W.; Teng, Y. D.; Zhang, Y. Complex process quality prediction using modified kernel partial least squares. *Chem. Eng. Sci.* **2010**, *65*, 2153–2158.
- (22) Yu, J. 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.* **2012**, *51*, 13227–13237.
- (23) Liu, Y.; Gao, Z. L.; Li, P.; Wang, H. Q. Just-in-time kernel learning with adaptive parameter selection for soft sensor modeling of batch processes. *Ind. Eng. Chem. Res.* **2012**, *51*, 4313–4327.
- (24) Rasmussen, C. E.; Williams, C. K. I. *Gaussian Processes for Machine Learning*; MIT Press, 2006.
- (25) Chen, T.; Ren, J. H. Bagging for Gaussian process regression. *Neurocomput.* **2009**, *72*, 1605–1610.
- (26) Gregorcic, G.; Lightbody, G. Gaussian process approach for modeling of nonlinear systems. *Eng. Appl. Art. Intell.* **2009**, *22*, 522–533.
- (27) Yu, J.; Chen, K. L.; Rashid, M. M. A Bayesian model averaging based multi-kernel Gaussian process regression framework for nonlinear state estimation and quality prediction of multiphase batch processes with transient dynamics and uncertainty. *Chem. Eng. Sci.* **2013**, *93*, 96–109.
- (28) Liu, Y.; Gao, Z. L. Real-time property prediction for an industrial rubber-mixing process with probabilistic ensemble Gaussian process regression models. *J. Appl. Polym. Sci.* **2015**, *132*, 1905–1913.
- (29) Yang, K.; Jin, H. P.; Chen, X. G.; Dai, J. Y.; Wang, L.; Zhang, D. X. Soft sensor development for online quality prediction of industrial batch rubber mixing process using ensemble just-in-time Gaussian process regression models. *Chemom. Intell. Lab. Syst.* **2016**, *155*, 170–182.
- (30) Ge, Z. Active probabilistic sample selection for intelligent soft sensing of industrial processes. *Chemom. Intell. Lab. Syst.* **2016**, *151*, 181–189.
- (31) Schölkopf, B.; Smola, A. J.; Müller, K. J. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Comput.* **1998**, *10*, 1299–1399.
- (32) Schölkopf, B.; Smola, A. J. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press: Cambridge, MA, 2002.
- (33) Rosipal, R.; Trejo, L. J. Kernel partial least squares regression in reproducing kernel Hilbert space. *J. Mach. Learn. Res.* **2002**, *2*, 97–123.
- (34) Chen, S.; Cowan, C. F. N.; Grant, P. M. Orthogonal least squares learning algorithm for radial basis function networks. *IEEE Trans. Neural Netw.* **1991**, *2*, 302–309.
- (35) Wang, X. X.; Chen, S.; Lowe, D.; Harris, C. J. Sparse support vector regression based on orthogonal forward selection for the generalised kernel model. *Neurocomputing* **2006**, *70*, 462–474.
- (36) Alaradi, A. A.; Rohani, S. Identification and control of a riser-type FCC unit using neural networks. *Comput. Chem. Eng.* **2002**, *26*, 401–421.