Quality Variable Prediction for Dynamic Process Based on Adaptive Principal Component Regression with Selective Integration of Multiple Local Models

Ying Tian[1*] and Yuting Zhu[2]
[1] School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology
Shanghai 200093, China
[e-mail: tianying@usst.edu.cn]
[2] School of Mechanical Engineering, University of Shanghai for Science and Technology
Shanghai 200093, China
[e-mail: 903301874@qq.com]
[*]Corresponding author: Ying Tian

Abstract

The measurement of the key product quality index plays an important role in improving the production efficiency and ensuring the safety of the enterprise. Since the actual working conditions and parameters will inevitably change to some extent with time, such as drift of working point, wear of equipment and temperature change, etc., these will lead to the degradation of the quality variable prediction model. To deal with this problem, the selective integrated moving windows based principal component regression (SIMV-PCR) is proposed in this study. In the algorithm of traditional moving window, only the latest local process information is used, and the global process information will not be enough. In order to make full use of the process information contained in the past windows, a set of local models with differences are selected through hypothesis testing theory. The significance levels of both T - test and χ^2 - test are used to judge whether there is identity between two local models. Then the models are integrated by Bayesian quality estimation to improve the accuracy of quality variable prediction. The effectiveness of the proposed adaptive soft measurement method is verified by a numerical example and a practical industrial process.

---

1. Introduction

With the rapid development of modern industry, factories have higher requirements on the production quality and efficiency of products, which makes the accurate measurement of the relevant important parameters in the process become particularly critical [1]. Soft measurement technique is a method to predict and estimate quality variables by establishing mathematical models between key quality variables and easily measurable process variables. The traditional measurement method is mainly based on mechanism modeling, which requires a clear understanding of the mathematical relationship between process variables and quality variables. In fact, most industrial processes have complex platform structures and technological processes, so it is often difficult to obtain the process mathematical models accurately [2]. Data-driven modeling methods do not need to accurately understand and describe the industrial process, but directly use the internal information between the process data to establish the mathematical model between the leading variable and the auxiliary variable, which is also known as the black box modeling method [3]. With the installation of a large number of sensors on the process platform and the application of intelligent instruments, distributed control system (DCS) and computer storage technology, massive process data can be collected and stored, which lays a foundation for data-driven modeling methods.

Nowadays, the widely used linear regression modeling methods mainly include principal components regression (PCR) [4] and partial least squares (PLS) [5]. PCR is a modeling method that uses principal component analysis (PCA) [6] to extract the characteristics of the input variable space and then utilizes the least square regression (LSR) to establish the regression model between the quality variable and the principal component characteristics. PLS is also a feature dimensionality reduction regression method, which simultaneously extracts the principal component features from the input and output space and guarantees the maximization of the covariance of the feature vectors of the input and output space [7]. The advantages of PCR and PLS are simple structure and easy implementation. In addition, there are a lot of commonly used nonlinear modeling methods, such as artificial neural network (ANN) [8] and Gaussian process regression (GPR) [9]. Compared with linear modeling, the nonlinear has higher complexity but stronger generalization ability. ANN does not need the prior knowledge of the research object, and it has great advantages to deal with complex and changeable industrial processes. GPR is a new machine learning method based on Bayesian theory and statistical learning theory, which is suitable for dealing with complex regression problems such as high dimension, small sample and nonlinearity and its output has probabilistic significance, however, it is necessary to consider the complexity of calculation, noise following Gaussian distribution and other problems [10].

In the recent literature, Bayesian regression [11] and deep learning have gradually become the focus of research. The quality variable prediction model based on bayesian network (BN) [12, 13] can solve the uncertainty relationship in the process and effectively solve the problem of missing data. However, in the process of model construction, the training process of bayesian network parameters is quite complex, even if the gaussian mixture model is used to approximate the conditional probability in BN, it is still in a state of huge computation [14]. Constantly updating the model requires constant recalculation of network parameters, which is obviously uneconomical and impractical. For deep learning [15], a large number of data samples are needed to train the model in the construction of regression model, which has a high requirement for the number of data samples. Moreover, when the data volume reaches a certain degree, whether a series of drift has occurred in the industrial process will affect the prediction accuracy [16, 17].

In the process of constructing the model of quality variable prediction, PCR is one of the most popular methods, because of its simple structure and fast response time. For factory production, high efficiency and low cost are important factors for choosing principal component regression as soft measurement method [18]. With the increasing complexity of industrial processes, process variables inevitably show a high-dimensional trend. However, there is correlation between these variables, so dimensionality reduction of data to find the main causal relationship between variables is essential [19]. By using linear transformation, PCA replace multiple variables in the data with a set of independent comprehensive indicators. In this way, important information will not be lost, and the collinearity problem between variables is avoided, which is convenient for further analysis. At the same time, the computational complexity of soft measurement process can be reduced to some extent.

One of the most important problems need to be overcome for quality variable prediction is the model performance degradation caused by the drift of process characteristics [20, 21]. The most obvious result of model degradation is a decline in prediction accuracy, mainly due to factors such as catalyst deactivation, mechanical aging, changes in operating environment, and even climate change [22]. Therefore, in order to improve the prediction accuracy, an adaptive soft measurement method is proposed to update the model with new measurement data in industrial applications [23, 24]. Up to now, most adaptive soft measurement processes have been built on just-in-time learning (JIT) [25-27], time difference (TD) [28] or moving window (MW) [29, 30]. JIT model is widely used in soft sensor with selected sample data set. The local prediction model is established by selecting training samples similar to the query samples in the historical sample set. As the computer's data storage capacity grows, the JIT needs to select the data set containing the entire process information from the distributed control system (DCS) database. In general, the size of the selected dataset is very large, and searching for similar samples from it is always time-consuming. Meanwhile, the JIT model ignores the correlation between process variables. And the prediction performance depends on the similarity measurement of relevant samples to a great extent. Over the past few decades, different similarity measures have been developed for sample selection. However, each method focuses on only one aspect of sample similarity and has its own limitations. Moreover, the JITL model based on correlation may sometimes be unable to select the appropriate model to analyze the actual process data. The TD method is able to generate the output without reconstructing the model and eliminate the variation of the value due to the deviation caused by the drift of some process variables. However, traditional TD models cannot take into account nonlinear processes, except in combination with physical models or some nonlinear modeling techniques, when the relationship between process variables and quality variables changes, TD methods are no longer applicable. MW updates the window dataset by merging the latest data samples while discarding the oldest. Each time the window is updated, the latest information of the industrial process is obtained, and the new model can be built to describe the current state more effectively, even if the working conditions begin to change.

As mentioned above, the final choice is to propose improvements based on the strategy of MW. In the traditional moving window, only the window data closest to the query sample is selected, and a local model is established to predict the quality variable. Whenever the window moves, create a new local model and discard the old model. A single local model contains only part of the process information, which can lead to the deviation of quality variable prediction. Therefore, in order to cover as much process information as possible, the concept of selective integration local models is proposed [31, 32]. There are two advantages to an adaptive strategy based on MW. On the one hand, considering that the working state will drift with time, the latest window can grab the information of the current process state in

industrial production. On the other hand, MW can be used to divide the data according to the characteristics of the sequential data set, when the state is stable in a certain period. Then, multiple local state data areas can be divided and corresponding local models can be established. This can lay the foundation for subsequent model integration phases. Divide-and-conquer strategy adopted in the integrated learning method can solve both nonlinear and time-varying process problems [33]. The integration process is to obtain more process information to improve the generalization ability of the model [34]. Then, the selection of this set of local models need to be considered. If only a group of adjacent local models are integrated, there are multiple models containing similar local process states, which is not conducive to the description of the global process and will result in the poor predictive ability of the integrated model. Therefore, it is necessary to use the hypothesis testing principle to screen out a set of local models with differences, which can make the integration model to have better generalization capability.

To sum up, in order to solve the problem of dynamic in industrial process, a method for selective integrated moving windows based on PCR (SIMV-PCR) is proposed to improve the accuracy of quality variable prediction. At first, by using the adaptive strategy of MW, the global process is divided into several local modules. Then, for each local module, PCR is used to establish the multivariate linear regression relationship between process variables and quality variables based on rapidity and economic factors. By using the hypothesis testing principle, the identity of two local modules is compared and a group of different local modules is selected, which means that the integrated model can contain more process information. After the local modules are selected, their corresponding models are integrated using Bayesian estimation rules. Finally, the integrated model is used to predict the query samples.

The rest of this article is organized as follows. The second part briefly introduces the related work, including the principle of hypothesis testing and the principle of multiple linear regression based on PCA. In the third part, the soft sensor development based on SIMV-PCR is described and analyzed in detail, including selective update strategy of local model and integrated learning strategy. In the fourth part, the numerical case and industrial process case are used to study and the performance comparison between SIMV-PCR and other adaptive soft measurement methods is reported. Finally, our work ends in section 5.

2. Preliminaries

This section will briefly describe the principle of hypothesis testing and briefly review multiple linear regression prediction methods based on PCA.

2.1 Hypothesis Testing

Hypothesis testing is a method in mathematical statistics to infer the population from samples based on certain assumptions, the basic idea of which is the thought of small probability proof by contradiction. The steps are as follows:

(a) Make an assumption about the population under study according to the needs of the problem, and call it H_0;

(b) Select the appropriate statistic so that the distribution is known when H_0 is assumed to be true;

(c) From the measured samples, the values of the statistics are calculated and tested according to the significance level given in advance to make the judgment of rejecting or accepting the hypothesis H_0.

Hypothesis testing, also known as significance testing, rejects H_0 when the actual data deviates significantly from the theoretical hypothesis. Deviation to a significant degree is usually specified by a very small positive number α such as 0.05, 0.01 so that when H_0 is correct, its probability of rejection is no more than α, which is called the significance level.

2.2 Multiple Linear Regression

PCR is a modeling method that uses PCA to extract the characteristics of the input variable space and then utilizes multiple linear regression to establish the relationship between the characteristic matrix and the quality variable. Through linear transformation, principal component method combines the original multiple indicators into a few independent indicators that can fully reflect the overall information, so as to avoid collinearity among variables without losing important information and facilitate further analysis. Each principal component extracted from principal component analysis is a linear combination of the original multiple indexes.

First of all, assuming the number of data samples collected is n, the number of process variables is s, then, the process variables can be expressed as $X = [x_1, x_2, ..., x_s]$.

By establishing the initial factor load matrix, the principal components can be expressed as

$$\begin{cases} z_1 = a_{11}x_1 + a_{12}x_2 + ... + a_{1s}x_s \\ z_2 = a_{21}x_1 + a_{22}x_2 + ... + a_{2s}x_s \\ \quad\quad\vdots \\ z_s = a_{s1}x_1 + a_{s2}x_2 + ... + a_{ss}x_s \end{cases} \tag{1}$$

in the formula:

z_i —— Intermediate variable obtained by process variables through a specific linear combination relationship;

a_{ij} —— The coefficient of x_j for z_i.

By standardizing the sample data, and the correlation coefficient R is obtained.

$$R_{ij} = \frac{\sum_{o=1}^{n}(x_{oj} - x_i)(x_{oj} - x_j)}{\sqrt{\sum_{o=1}^{n}(x_{oj} - x_i)^2(x_{oj} - x_j)^2}} (1 \le i \le n; 1 \le j \le s) \tag{2}$$

Then, according to the covariance matrix R, the eigenvalue, principal component contribution rate and cumulative variance contribution rate can be calculated, and the number of principal components can be determined. Because R is a positive definite matrix, the eigenvalues are positive, namely, $\lambda_1 > \lambda_2 > ... > \lambda_n > 0$, and the corresponding normalized eigenvectors are $v_1, v_2, ..., v_3$. The eigenvalue is the variance of each principal component, and its value reflects the influence of each principal component. The contribution rate of the principal component can be calculated as follows:

$$\psi_i = \lambda_i / \sum_{i=1}^{n} \lambda_i \tag{3}$$

in the formula:

ψ_i — the contribution rate of the principal component z_i.

Based on multiple linear regression theory, the following model can be established:

$$Y = b_0 + b_1 x_1 + b_2 x_2 + ... + b_s x_s + \mu \tag{4}$$

The quality variable Y was regressed with respect to d $(0<d<p)$ principal components:

$$\hat{Y} = \hat{\alpha}_1 z_1 + \hat{\alpha}_2 z_2 + ... + \hat{\alpha}_d z_d \tag{5}$$

At this point, put (1) into (5), the relationship between process variables and quality variables is expressed as:

$$\hat{Y} = \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + ... + \hat{\beta}_s x_s \tag{6}$$

The relationship between the coefficient $\hat{\beta}_i$ and the parameter b_i in the original model:

$$b_i = \frac{S_Y}{S_i} \hat{\beta}_i \qquad (i = 1, 2, ..., s) \tag{7}$$

$$b_0 = \bar{Y} - \sum_{i=1}^{s} b_i \bar{x}_i \tag{8}$$

in the formula:

S_Y —Standard deviation of Y;

S_i —Standard deviation of $\hat{\beta}_i$;

$\bar{Y}$ —Mean value of Y;

$\bar{X}_i$ —Mean value of x_i .

Therefore, the parameters in the original regression model can be calculated. Then, put (7) and (8) into (4), the final regression model can be obtained.

3. Adaptive Principal Component Regression with Selective Integration of Multiple Local Models

In this study, in order to solve the problem of dynamic in industrial process, SIMV-PCR is proposed. The dynamic problem of industrial process brings about a series of problems such as model degradation. The traditional MW model can capture the latest local state of the process. However, the industrial process information obtained in this way has limitations and the prediction of quality variables is not comprehensive enough. Therefore, the selective integration of local models can effectively utilize the past window information. In this way, we can not only effectively use the latest process state information to avoid model degradation, but also prevent a single local model from causing prediction bias. In this part, SIMV-PCR is introduced in detail from the two aspects of local model set update and integration.

3.1 Adaptive Strategy

In terms of model updates, appropriate window length is selected to divide the whole process state into local model regions, and a concise local PCR model set is obtained firstly. As the

factory continues to run, providing continuous time samples directly causes the information contained in the current process state between adjacent windows to be too similar, in which case the corresponding sub-model can be considered identical, resulting in computational redundancy. If the most similar model is built repeatedly and the original model with different properties is abandoned, the prediction accuracy of the quality variable will be unsatisfactory.
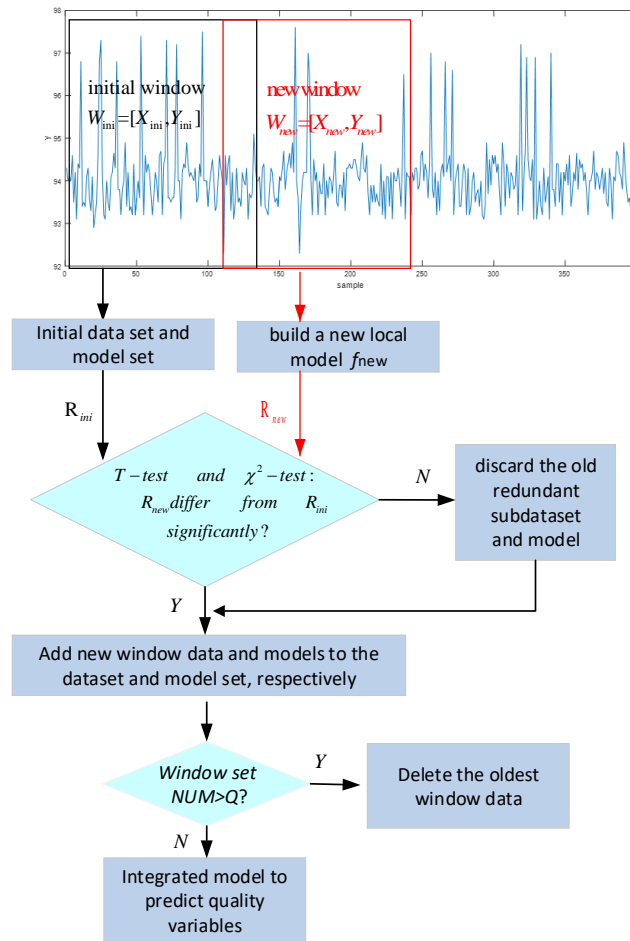

Fig. 1. Flowchart of the adaptive approach based on statistical hypothesis testing and moving window

With statistical hypothesis testing and moving window as the main methods, this adaptive strategy considers the variance of the first and second order information of the prediction residuals, using t-test and χ^2-test respectively, which is shown in **Fig. 1**. In this study, we consider only the single output process, that is, $r = 1$, where r is the number of quality variables. At the beginning, the initial window dataset $W_{ini} = \{X_{ini}, Y_{ini}\}$ was used, and the length of the window dataset was set as w, indicating that there were w continuous time samples, on which the regression model f_{ini} was constructed, where $X_{ini} \in R^{w \times s}$ and $Y_{ini} \in R^{w \times 1}$ represent the input matrix and output matrix respectively. Assuming we have identified and stored $K(K \geq 1)$ regression models $\{f_1, f_2, ..., f_K\}$, this means that there are $K(K \geq 1)$ data regions $\{W_1, W_2, ..., W_K\}$,

each containing W continuous time samples. On the premise of keeping the generality, the local model was identified in the previous local model region extraction and obtained $W_{ini} = W_K$ and $f_{ini} = f_K$. Then, W_{ini} is moved a sample step forward to get a new moving window dataset $W_{new} = \{X_{new}, Y_{new}\}$.

The predicted residuals R_{ini} and R_{new} calculated based on f_{ini} and f_{new} respectively can be expressed as

$$R_{ini} = f_{ini}(X_{ini}) - Y_{ini} \tag{9}$$

$$R_{new} = f_{new}(X_{ini}) - Y_{ini} \tag{10}$$

If the difference between R_{new} and R_{ini} is not significant, the performance of f_{new} on W_{new} is considered to be the same as that of f_{ini} on W_{ini}. That is, the samples in W_{ini} and W_{new} come from the same local process state. Therefore, the old model can be replaced by the establishment of a new local model based on the new window dataset. Once the R_{new} deviates significantly from the R_{ini}, a new local process state distinct from the representation of W_{ini} is identified. Then, add it to the current window dataset and model set. And the window will be continuously shifted forward to calculate the new R_{new} to update the stored dataset and model set in this way. The key question that needs to be addressed is how to determine whether there is a difference in evidence between R_{new} and R_{ini}. According to the idea of hypothesis testing, we transformed this problem into testing whether the mean and variance of two residuals were significantly different. In this paper, two statistical testing methods, t-test and χ^2-test, were used simultaneously.

Assuming that both R_{ini} and R_{new} are normally distributed, we first construct T statistics and χ^2 statistics as follows:

$$T = \sqrt{w}(\overline{R}_{new} - \overline{R}_{ini}) / \sigma_{new} \tag{11}$$

$$\chi^2 = (w-1)\sigma_{new}^2 / \sigma_{ini}^2 \tag{12}$$

In the formula:
$\overline{R}_{new}$ — the mean of R_{new};
σ_{new} — the standard deviation of R_{new};
$\overline{R}_{ini}$ — the mean of R_{ini};
σ_{ini} — the standard deviation of R_{ini};
w — the number of window samples;
According to the hypothesis statistical theory, the model redundancy can be checked by the following equations：

$$H_1 : \overline{R}_{new} = \overline{R}_{ini} \tag{13}$$

$$H_2 : \sigma_{new} = \sigma_{ini} \tag{14}$$

If (13) and （14） are both satisfied, then, the distribution of R_{new} and is R_{ini} considered to be the same. How to verify that the hypothesis is satisfied, generally, there are two conditions that need to be fulfilled:

$$|T| < \theta_t \quad and \quad \chi^2 < \theta_\chi \tag{15}$$

in the formula:

θ_t —— The threshold value of the T - statistic for the given significance level α_t ;

θ_χ —— The threshold value of the χ^2 - statistic for the given significance level α_χ ;
That means the probability satisfies

$$P\{|T| < \theta_t\} = 1 - \alpha_t \tag{16}$$

$$P\{\chi^2 < \theta_\chi\} = 1 - \alpha_\chi \tag{17}$$

Therefore, only when both T and χ^2 are in their in the receiving domain, respectively, R_{new} and R_{ini} are considered consistent enough. Otherwise, W_{new} is supposed to differ from W_{ini}, namely, there is a new local process state. Then, a new model expressed by f_{new} can be constructed based on the corresponding local data set $W_{new} = \{X_{new}, Y_{new}\}$.

As mentioned above, the identity of two local models is verified by hypothesis testing. In the process of online prediction, an integrated approach is adopted in order to include more process information, that is, a set of local models with difference needs to be saved here. In this way, the creation of a sub-model based on the new local data obtained by moving a sample step, requires an identity comparison with each sub-model in the stored model set. The comparison process is repeated K times for each new window. If a local model in the model set is identical with the new one, the new one will replace the original one. If it does not exist, the new model is added directly to the model set. A potential problem with this approach, however, is that the number of local models is increasing as the factory continues to run and provide continuous data every day. If the number of models is too large, the computational complexity will increase. Therefore, there is a need to set a maximum value for the number of model sets. Here, Q represents the maximum value of the model set. $\{W_{new}, f_{new}\}$ represents a truly new local process state only if it is different from all the models in the model set, namely, $\{W_k, f_k\}$ for $1 \le k \le K$. In this case, $\{W_{new}, f_{new}\}$ is added to the local model set as the latest process state, meanwhile, the oldest is discarded. However, if there is a local process state $\{W_k, f_k\}$ similar to $\{W_{new}, f_{new}\}$, where $k \in \{1, 2, ..., K\}$, then $\{W_k, f_k\}$ can be replaced by $\{W_{new}, f_{new}\}$. In this way, the updated model set can be obtained.

3.2 Selective Integration Based on PCR

In terms of integration of local models, the sub-models prepared after the model update are selectively combined to estimate the output of each query sample. In this part, the local model is multiple linear regression model based on PCA. The concept of traditional integration learning means that during the prediction process of the query samples, the output results of

all local models selected in the adaptive update process need to be combined, which is

$$f(x_{new}) = \sum_{k=1}^{K} \eta_k f_k(x_{new}) \tag{18}$$

in the formula:

$f(x_n)$ — The final prediction value;

$f_k(x_{new})$ — The output value of the kth local model.

Given x_{new}, η_k represents the combination weight assigned to the *kth* model f_k, which satisfies:

$$\eta_k \geq 0 \quad and \quad \sum_{k=1}^{K} \eta_k = 1 \tag{19}$$

However, it is a difficult task to find an optimal weight for each local model. In this part, η_k can be calculated by Bayes' rule in the following formula:

$$\eta_k = P(f_k \mid x_{new}) = \frac{p(x_{new} \mid f_k)p(f_k)}{\sum_{k=1}^{K} p(x_{new} \mid f_k)p(f_k)} \tag{20}$$

In this paper, it is assumed that the prior probability of each model is the same, given as follows:

$$p(f_k) = K^{-1} \tag{21}$$

And the $P(f_k \mid x_{new})$ is used to describe the predictive power of the kth window model in the model set. Therefore, $P(f_k \mid x_{new})$ can be standardized, and given as follows:

$$p(x_{new} \mid f_k) = \varphi_k \sqcap (\sum_{k=1}^{K} \varphi_k)^{-1} \tag{22}$$

$$\varphi_k = rmse_k^{-2} \tag{23}$$

in the formula:

$rmse_k$ — The root mean square error of the *kth* window model. And $rmse_k$ can be used to estimate the predictive power of the model.

In fact, selective integration learning forces the unselected model to be zero to filter out the negative effects of the prediction of quality variables. In this way, the accuracy of quality variable prediction can be improved. SIMV-PCR captures the local latest state of a process by using the moving window strategy. And it makes full use of the previous window information, select a group of local states with differences to establish multiple local models. Finally, these local models are integrated to predict the quality variables through Bayesian estimation rules. Thus, it can be seen that for the time-varying problem of industrial processes, SIMV-PCR can not only deal with the model degradation problem by moving the window, but also reduce the prediction error by using the integration method. In addition, in order to quantitatively evaluate the predictive performance, the following aspects are analyzed:

RMSE— the root mean square error of query sample;
MAE— the maximum relative error of query sample;
MRE— the mean relative error of query sample.

4. Experimental Analysis

In order to effectively verify the effectiveness of SIMV-PCR, this section will carry out experiments from two aspects: numerical case and actual industrial data measurement.

4.1 Numerical Example

In this section, a numerical example is used to verify the feasibility and effectiveness of the proposed SIMV-PCR method. The numerical example data is generated by the following formula:

$$y = 5x_1 * 10^{-2} + \sin(x_2) / x_3 + ax_4 + \cos(x_5) \tag{24}$$

in the formula:

y — The output value;

x_1 — Input variable 1, value range from -10 to 10;

x_2 — Input variable 2, value range from -10 to 10;

x_3 — Input variable 3, value range from -10 to 10;

x_4 — Input variable 4, random numbers that belong to the standard normal distribution;

x_5 — Input variable 5, the value is equal to x_2 plus x_3;

a — The coefficient of x_4, by changing its size to produce deviation.

Here, the range of input variable 1, 2, and 3 is set from -10 to 10 for no special reason，which can be changed. In the numerical example, the data drift is generated by changing input variable 4 to simulate the data drift caused by time change in the industrial process. In order to achieve it, the coefficient a is going to change over time.

In the numerical case, the size of the window is set to 20 and the maximum value of the model set is set to 3. The number of principal components is also set to 3. The value of a is changed every 50 samples. A total of 120 samples were selected during the experiment. The first 20 samples were used to establish the initial window model, and the last 100 samples were used as query samples.

**Table 1.** Prediction performance based on various soft sensors in numerical example

Soft Sensor	MAE	RMSE	MRE(%)
PCR	1.2003	0.6364	53.62
MV-PCR	1.0378	0.3257	20.71
GP	1.7697	0.9923	86.67
SIMV-PCR	0.3910	0.2455	16.32

As is shown in **Table 1**, the prediction results of PCR, grey prediction (GP), moving window-based principal component regression (MV-PCR) and SIMV-PCR are compared comprehensively. From the comparison of the data, it is not difficult to see that SIMV-PCR is the best of the four methods.

(a)   Predicted y-value by PCR

(b) Predicted y-value by MV-PCR

(c) Predicted y-value by SIMV-PCR
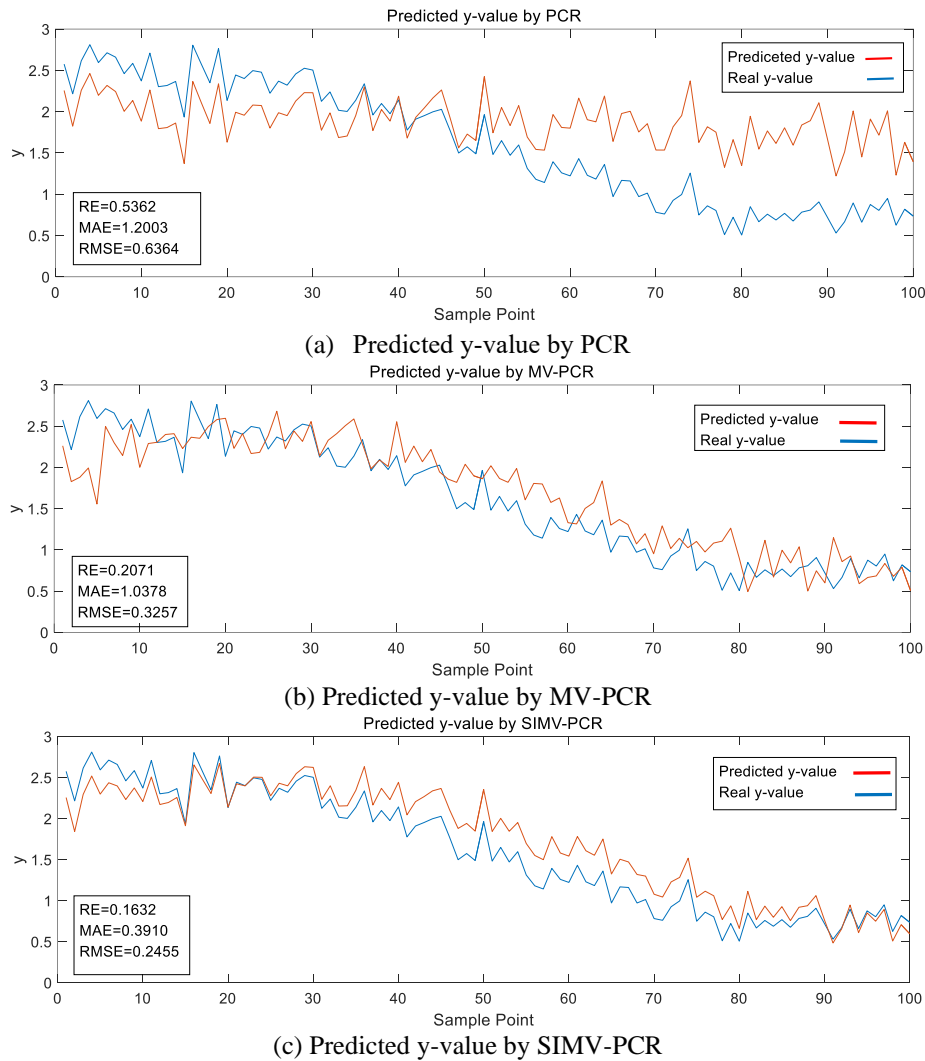
Fig. 2.   Comparison of prediction results of numerical cases

  For a further visual comparison, the prediction curves of several methods are shown in **Fig. 2**. With the drift of process state, it is obvious that the model based on PCR degrades seriously shown in **Fig. 2(a)**. And by MV-PCR, the model degradation problem has been improved. As shown in **Fig. 2(b)**, the output prediction curve of MV-PCR can follow the real value well. However, RE, MAE, RMSE can be improved further. It can be clearly seen that the prediction of output value using SIMV-PCR has been greatly improved in **Fig. 2(c)**. With regard to RE, MV-PCR was reduced by 61.4% compared to PCR, yet SIMV-PCR was reduced by 69.6%. The MAE of MV-PCR was down to 1.0378, yet that of SIMV-PCR was 0.3910. In other words, the MAE of SIMV-PCR was down 62.3% compared to MV-PCR. About the RMSE, SIMV-PCR is 24.6% less than MV-PCR. It can be seen that SIMV-PCR has strong generalization ability.
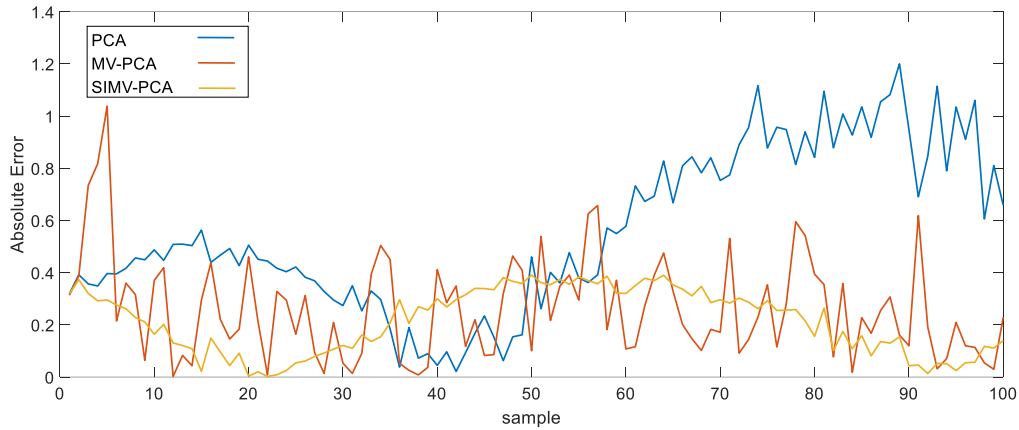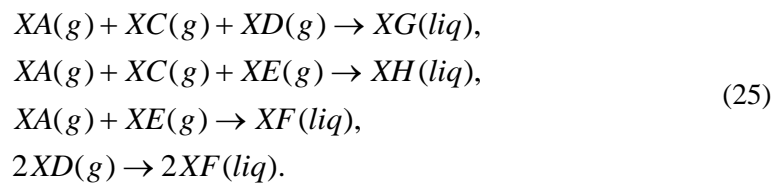
Fig. 3. Comparison of absolute errors of different methods in numerical examples

To further give sufficient inference, the absolute error curves of PCA, MV-PCA and SIMV-PCA are compared in **Fig. 3**. It can be seen that the absolute error curve of SIMV-PCA is the most stable and the fluctuation is the least. PCA and MV-PCA have relatively weak generalization ability in the face of state offset.

4.2 Tennessee Simulation Model

Downs et al. established the Tennessee Eastman (TE) simulation system based on the actual chemical process, and the detailed process structure is shown in **Fig. 4**. In the research field of process control, TE process is often used in the research of quality variable prediction, process monitoring and fault detection because it can better reflect many typical characteristics of the actual industrial production process.

TE process is mainly composed of five modules: reactor, condenser, compressor, gas-liquid separator and stripper. The data set consists of measurement variables and operation variables. The whole reaction process involves eight substances, including XA, XB, XC, XD, XE, XF, XG and XH. Among them, XG and XH are the final products. The reactions involved are as follows:

$$
\begin{aligned}
XA(g) + XC(g) + XD(g) &\rightarrow XG(liq), \\
XA(g) + XC(g) + XE(g) &\rightarrow XH(liq), \\
XA(g) + XE(g) &\rightarrow XF(liq), \\
2XD(g) &\rightarrow 2XF(liq).
\end{aligned}
\tag{25}
$$

The whole TE process involves 11 operating variables and 41 measurement variables, among which the measurement variables can be further subdivided into 22 continuous process variables and 19 discontinuous component variables. In the process, real-time monitoring of XG and XH of the final product lays a foundation for effectively improving the output product quality control system. This paper mainly studies the nonlinear system with multiple inputs and single outputs. All the 19 discontinuous component variables are obtained by the component analyzer, which is relatively complex. Therefore, a 33-dimensional variable is selected as the input, that is, it is composed of 11 operating variables and 22 continuous process variables, and the final XG component content is selected as the output variable, so as to establish the prediction model.
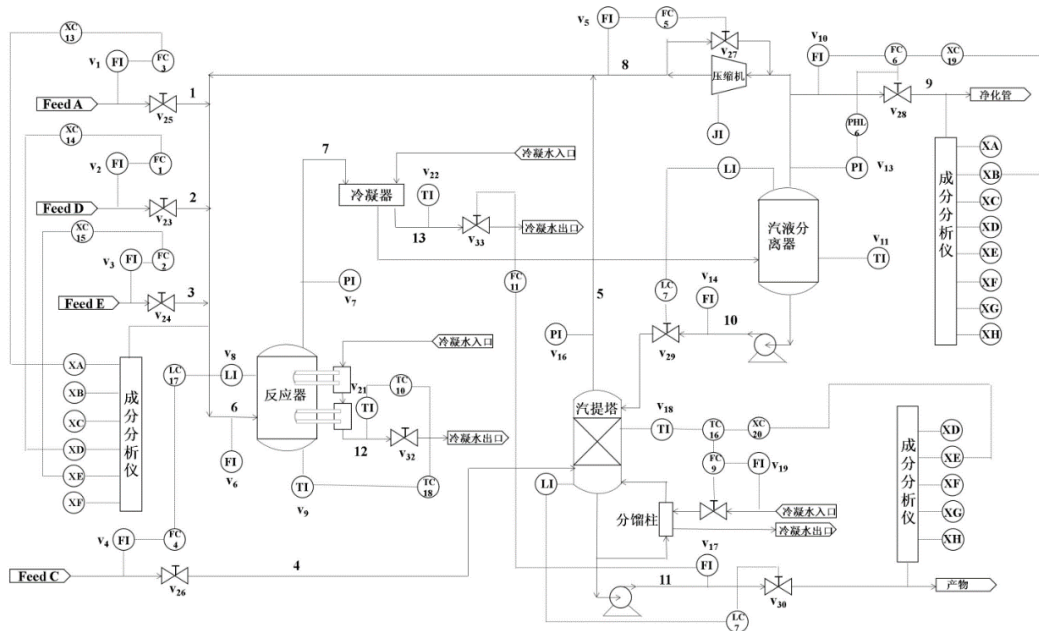
Fig. 4. TE process flow chart

Table 2. Prediction performance based on various soft sensors in Tennessee example

| Soft Sensor	PCR	MV-PCR	SIMV-PCR
RMSE	0.5814	0.5893	0.5738
MAE	1.6116	1.9987	1.5966
MRE(%)	0.89	0.87	0.87
R^2	-0.3213	-0.3575	-0.2872

In this section, the Tennessee simulation data are used to verify the above method. In the verification process, PCR and MV-PCR were mainly compared, and it was found that the prediction accuracy of the three was not significantly different. The prediction results of PCR, MV-PCR and SIMV-PCR under one of the same conditions are shown in **Fig. 5**, where the number of principal components is set as 17. And under this condition, the evaluation indexes of the three prediction models are also given in **Table 2**. R^2 stands for the determination coefficient, which can reflect the quality of the fitting. The R^2 of all three methods is negative, which indicates that the Tennessee simulation data is not suitable for using SIMV-PCR for regression.
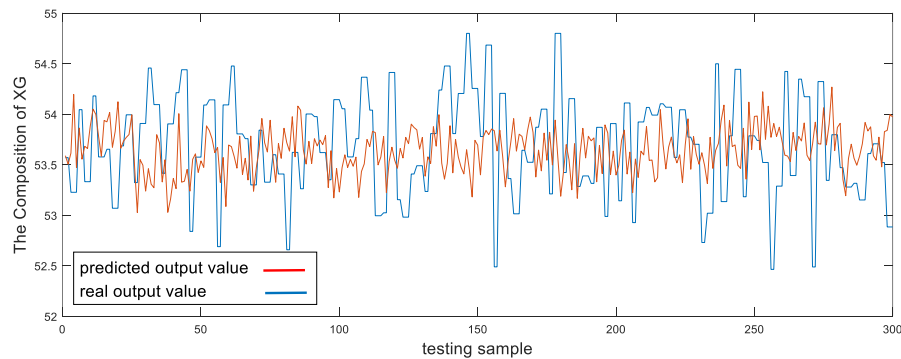
(a) Predicted the composition of XG by PCR
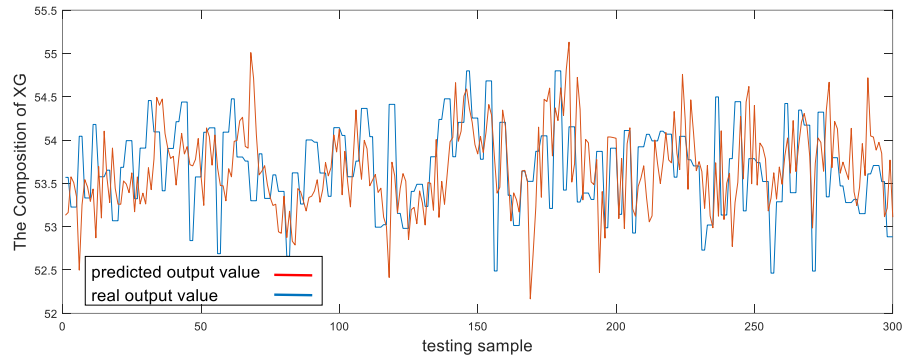
(b) Predicted the composition of XG by MV-PCR
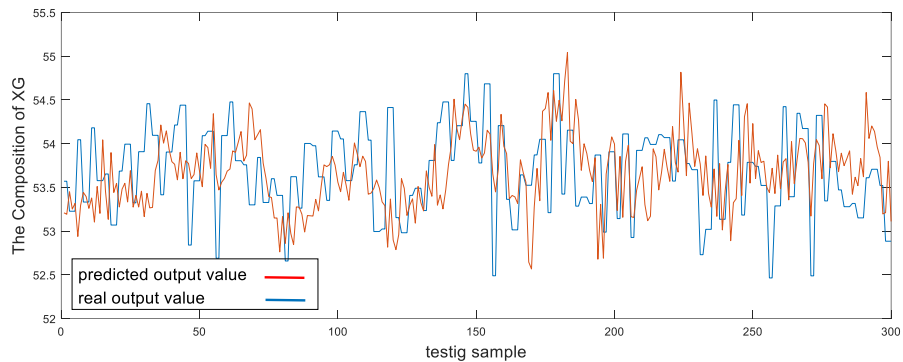
(c) Predicted the composition of XG by SIMV-PCR
Fig. 5. Comparison of prediction results of TE data

4.3 Octane Value Prediction

In order to verify the validity of the method by measuring the industrial data, the octane number of gasoline was used in this selection. Because gasoline is an extremely flammable and explosive liquid, it is important to control the intensity of combustion in order to allow it to burn smoothly in the engine. This shows the anti-knock index of gasoline, also known as the anti-knock index, and the anti-knock performance is proportional to the amount of isooctane. For ease of use, the amount of octane in gasoline is used to represent the antiknock index and also the type of gasoline. As the most important quality index of gasoline, octane value need to be measured accurately. However, the traditional laboratory testing methods whose problems contains large sample consumption, long test cycle and high cost, cannot be suited to production control, especially the online test. The near infrared (NIR) spectroscopic analysis method, as a quick analysis method, has been widely used in agriculture, pharmaceutical, biological, chemical, petroleum products and other fields. Its advantages are nondestructive testing, low cost, no pollution, online analysis, more suitable for production and control needs.

In this study, a soft octane value measurement model is established based on the near infrared spectrum data of gasoline. In addition, for the training data set and the test data set, the corresponding octane value of near-infrared samples was obtained through laboratory tests.

4.3.1 Parameter Selection

During the SIMV-PCR experiment, a total of 460 test samples obtained in continuous time

were selected. During the training process, the first 160 samples were selected as training samples in the offline operation stage (train –X and train –Y) to train the original stored data set and model set，where statistical hypothesis testing and moving window theory were used here. Then, the remaining 300 samples are used as query samples. The window size of the method is set to 100, and the training dataset is updated as the window slides along the entire dataset.

Table 3. Root mean square error and maximum absolute error results under various parameters

Number of Compoments	Number of Models	20	25	30	35	40	45	50
21	RMSE	0.3510	0.3483	0.3450	0.3417	0.3405	0.3391	0.3389
	R^2	0.7812	0.7846	0.7886	0.7926	0.7940	0.7958	0.7960
	MAE	1.7450	1.7016	1.6369	1.5868	1.6736	1.6928	1.7448
22	RMSE	0.3543	0.3510	0.3471	0.3449	0.3432	0.3428	0.3436
	R^2	0.7770	0.7811	0.7859	0.7887	0.7907	0.7913	0.7903
	MAE	1.8504	1.8412	1.7950	1.8072	1.8509	1.9153	2.0025
23	RMSE	0.3443	0.3418	0.3397	0.3394	0.3375	0.3371	0.3381
	R^2	0.7894	0.7925	0.7950	0.7954	0.7977	0.7981	0.7969
	MAE	1.5769	1.5942	1.5741	1.6528	1.6752	1.7543	1.8580
24	RMSE	0.3421	0.3379	0.3331	0.3332	0.3351	0.3334	0.3350
	R^2	0.7922	0.7972	0.8029	0.8028	0.8006	0.8026	0.8006
	MAE	1.6038	1.6123	1.5717	1.5543	1.5390	1.5244	1.6703
25	RMSE	0.3346	0.3308	0.3267	0.3243	0.3258	0.3262	0.3265
	R^2	0.8011	0.8057	0.8104	0.8132	0.8115	0.8110	0.8107
	MAE	1.6099	1.6166	1.5713	1.5431	1.5237	1.5094	1.4998
26	RMSE	0.3432	0.3338	0.3253	0.3187	0.3150	0.3131	0.3112
	R^2	0.7908	0.8021	0.8120	0.8198	0.8238	0.8259	0.8280
	MAE	1.7152	1.5827	1.5417	1.5152	1.4957	1.4783	1.4631
27	RMSE	0.3481	0.3378	0.3269	0.3175	0.3115	0.3091	0.3073
	R^2	0.7848	0.7973	0.8102	0.8209	0.8276	0.8302	0.8323
	MAE	2.1940	1.8534	1.5130	1.4640	1.4287	1.3933	1.3632
28	RMSE	0.3615	0.3512	0.3418	0.3338	0.3252	0.3208	0.3170
	R^2	0.7679	0.7809	0.7925	0.8021	0.8121	0.8172	0.8215
	MAE	2.5219	2.2121	1.8836	1.5016	1.3679	1.3144	1.3758
29	RMSE	0.3812	0.3661	0.3567	0.3468	0.3351	0.3299	0.3267
	R^2	0.7419	0.7619	0.7740	0.7863	0.8005	0.8067	0.8104
	MAE	3.0528	2.6543	2.3827	2.0152	2.0860	1.4449	1.4707
30	RMSE	0.3840	0.3741	0.3658	0.3554	0.3440	0.3382	0.3390

R^2	0.7381	0.7514	0.7623	0.7757	0.7898	0.7968	0.7959
MAE	3.3746	3.1749	2.9410	2.5849	2.0627	1.5753	1.8604

During the training, in order to make full use of the past windows, selective integrated moving window PCR（SIMW-PCR） was developed. We set the maximum number of reserved windows to be no more than Q, and the Q here needs to undergo multiple experiments. Proper window size can better monitor the drift of process state and achieve better prediction effect. If the window size is too large, it will lead to calculation redundancy and increase the amount of calculation. If the window size is too small, it will also lead to loss of important sample information. At the same time, in order to make the prediction more accurate, there are certain requirements for the selection of principal component number d. Trough **Table 3**, the setting of parameters Q and d of the experimental process can be optimized. After the training, 300 test sets were used to test the model, and MAE, R^2 and RMSE were used to measure the performance of the soft measurement model. MAE and RMSE take into account the difference between the predicted value and the true value, and by using R^2 analysis, the difference between the true value of the problem itself can be further considered.

In the experiment, only a single output path was considered, and the number of process variables in the near infrared data of gasoline octane number was 201. It was found that it was a watershed for the selection of the number of principal components, when the cumulative contribution rate of the principal component reached 99.9%. When the contribution rate was lower than that, the prediction accuracy significantly decreased, and the number of principal elements was 20. Then, taking five models as a unit, in **Table 1** the experimental prediction results are only partially given, where the number of principal elements is from 21 to 30, meanwhile, the number of model sets from 20 to 50. The comparison shows that the result is better when the number of principal elements is 27.

However, as can be seen from the **Table 3**, when the number of principal components is fixed, the prediction effect presents a curve that rises first and then falls as the number of models in the model set increases. Therefore, when the number of pivot elements is 27, the number of models is set to 50, which is not optimal. In order to get the best results, we continue the experiment and finally set the maximum number of models in the model set to 60. As shown in **Fig. 6**, when the number of principal components is 27, the trend of RMSE and MAE is given with the change of the maximum value of the local model. With the increase of Q, RMSE and MAE decreased rapidly at the beginning and became stable at $Q =40$. As shown in **Fig. 6**, $Q=60$ is finally selected, meanwhile, R^2 is as high as 0.8350. This not only makes the generalization ability of the model better, but also avoids complex computation problems.
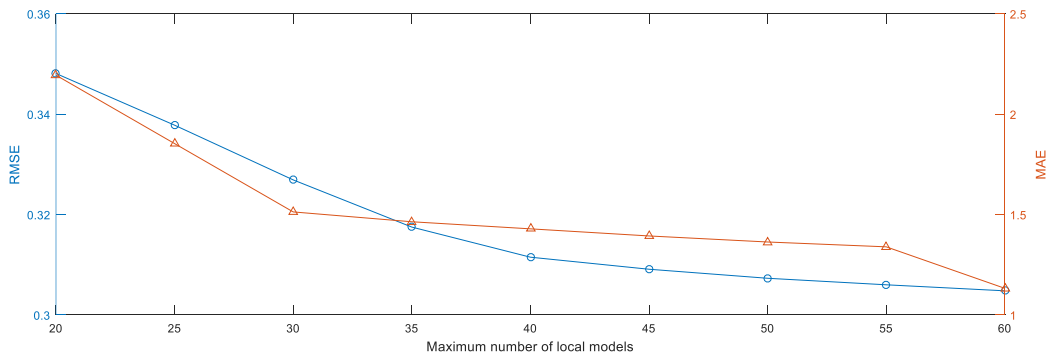
Fig. 6. Impact of Q with $d=27$

4.3.2 Analysis of Predictive Indexes

Here, to prove the superiority of SIMV-PCR in solving the time-varying problem of industrial process, PCR, GP and MV-PCR are also compared in this case. In addition, kernel principal component analysis (KPCA), PCA-GP are also added. For the selection of the number of principal components, all of PCR, MV-PCR, PCA-GP and KPCA are determined to be 27. The prediction results of the SIMW-PCR and the other four methods are shown in **Table 4**. In order to effectively illustrate the higher prediction accuracy of SIMV-PCR, four indexes including MAE, R^2, RMSE and MRE are comprehensively compared here. As is shown in the **Table 4**, the MAE of PCR is up to 4.8926. For the maximum absolute error, the MV-PCR model based on the latest window data is 23.9% lower than PCR. However, the predicted value obtained by using SIMV-PCR, MAE was reduced by a full 76.9 percent. Then, for RMSE, MV-PCR decreases by only 12% compared with PCR, while SIMV-PCR decreases by 34.8%. In addition, regarding the improvement of mean relative error, MV-PCR has no optimization compared with PCR, but MRE is reduced by SIMV-PCR. Too many process variables lead to too large deviation of the predicted results of GP. After principal component extraction, the generalization performance was greatly improved. Then, MW is added into PCA-GP to further improve the predictive power. However, it was still not as good as SIMV-PCR. Similarly, SIMV-PCR also has great advantages for R^2. From what has been mentioned above, it is obvious that SIMV-PCR is better than other soft measurement methods in predicting quality variables.

Table 4. Prediction performance based on various soft sensors

| Soft Sensor	MAE	RMSE	MRE(%)	R^2
PCR	4.8926	0.4678	0.27	0.6113
MV-PCR	3.7217	0.4113	0.27	0.6996
GP	32.2743	2.0674	0.46	-0.5921
PCA-GP	2.4989	0.3651	0.25	0.8012
MV-PCA-GP	2.3174	0.3543	0.25	0.8033
KPCA	2.9830	0.3261	0.24	0.8336
SIMV-PCR	1.1324	0.3048	0.23	0.8350

4.3.3 Comparison of Prediction Curves

The predicted indicators can only show the average process of the whole, but cannot show the forecast situation of each sample. Therefore, this part will selectively observe the prediction curve of several soft measurement methods.

As shown in **Fig. 7**, the blue curve represents the prediction results of the PCR model. As time goes by, the error of the model established with the initial data starts to increase due to data drift, machine aging, etc. The red curve represents the prediction results of the MV-PCR model. As the latest window data is used to build the model, MV-PCR is much better than PCR. However, depending on a single window, the information of the industrial process is not contained enough, which is why improvements in prediction accuracy are limited. As for sample 180, PCR and MV-PCR both produced large errors, PCR derived from model degradation, while MV-PCR errors were caused by insufficient global characteristics contained in local information. SMV-PCR results well corrected this error. Therefore, through the integration window approach, more effective information in the industrial process can be retained. Obviously, it can be seen that the yellow prediction curve representing SIMV-PCR fluctuates smoothly and does not show excessive peaks.
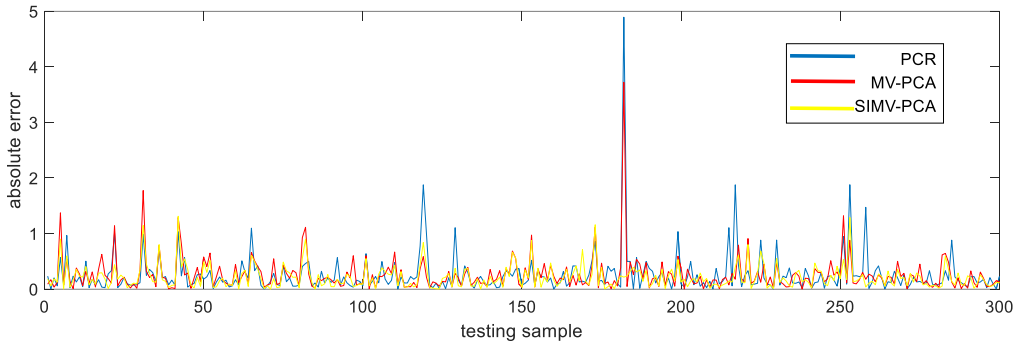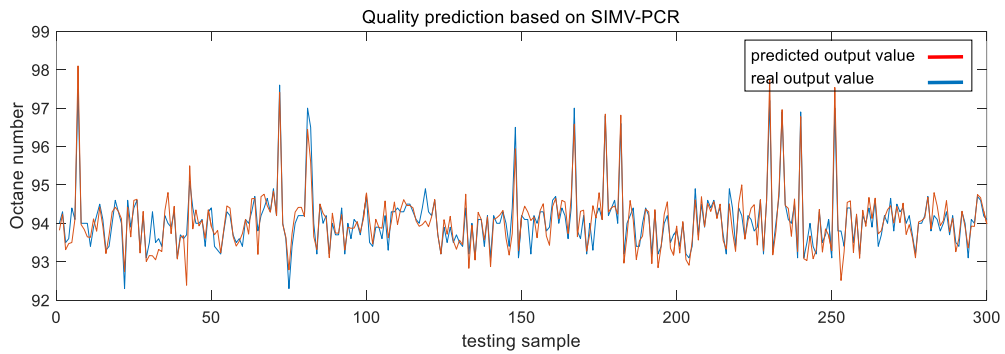
Fig. 7. The testing error curves of PCR, MV-PCA and SIMV-PCA
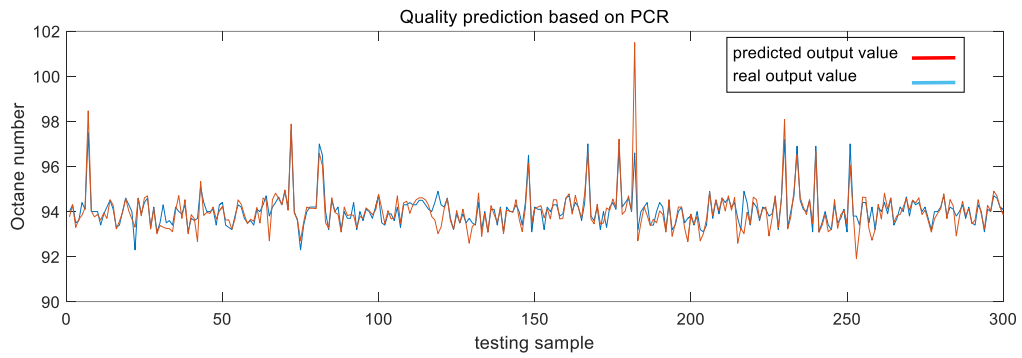
Fig. 8. Quality prediction based on PCR
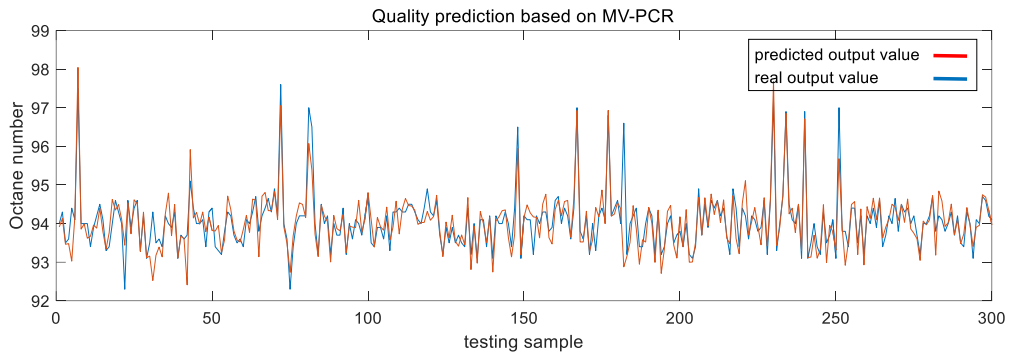
Fig. 9. Quality prediction based on MV-PCR

Fig. 10. Quality prediction based on SIMV-PCR

Fig. 8, **Fig. 9** and **Fig. 10** respectively show the changes of PCR, MV-PCR and SIMV-PCR on the tracking process state. With time migration, the prediction model established based on PCA is not good at dealing with the state migration and tracking the state migration with a large deviation. Then, MV-PCR can obtain better prediction results than PCR by combining the latest state information to track the process state. For SIMV-PCR, not only are the latest data samples included for model updates, but previous information Windows are also taken into account. Finally, SIMV-PCR can be used to obtain the process state tracking curve with a relatively small deviation.

As mentioned above, the adaptive soft measurement model based on SIMV-PCA can achieve satisfactory predictive performance in the industrial process.

5. Conclusion

Aiming at the nonlinear and time-varying problems of industrial processes, an adaptive soft measurement method has been proposed, referred to as the SIMW-PCR, which combines local learning strategy and selective integrated learning strategy at the same time. First of all, an adaptive localization method based on combining PCR and moving window is proposed to construct local model set. The move window is used to partition local areas. Because of the redundancy of retaining window information, the soft measurement accuracy is improved, but the computational complexity is large. Then, based on the theory of hypothesis testing, a selection method is proposed to deal with the problem with increasing the number of models, which can determine whether the window dataset should be retained or not. After that, most information models with some difference between each other can be integrated into an effective prediction model through Bayesian's quality evaluation rules. A real industrial case has demonstrated the priorities of the proposed adaptive quality prediction method through comparing with several existed adaptive soft sensors. However, the selection of some parameters, such as the default value of the number of windows, needs further study. In addition, the industrial data types applicable to this method also need further analysis.

Acknowledgement

References

[1]   S. Khatibisepehr, B. Huang, and S. Khare, "Design of inferential sensors in the process industry: A review of Bayesian methods," *Journal of Process Control*, vol. 23, no. 10, pp. 1575-1596, Nov. 2013. Article (CrossRef Link)

[2]   Z. Ge, "Review on data-driven modeling and monitoring for plant-wide industrial processes," *Chemometrics and Intelligent Laboratory Systems*, vol. 171, no. 15, pp. 16-25, Dec. 2017. Article (CrossRef Link)

[3]   Z. Ge, Z. Song, S. X. Ding, and B. Huang, "Data Mining and Analytics in the Process Industry: The Role of Machine Learning," *IEEE Access*, vol. 5, pp. 20590-20616, Sep. 2017. Article (CrossRef Link)

[4]   Z. Ge, F. Gao, and Z. Song, "Mixture probabilistic PCR model for soft sensing of multimode processes," *Chemometrics & Intelligent Laboratory Systems*, vol. 105, no. 1, pp. 91-105, Jan. 2011. Article (CrossRef Link)

[5]   Y. P. Du, Y. Z. Liang, J. H. Jiang, R. J. Berry, and Y. Ozaki, "Spectral regions selection to improve prediction ability of PLS models by changeable size moving window partial least squares and searching combination moving window partial least squares," *Analytica Chimica Acta*, vol. 501, no. 2, pp. 183-191, Jan. 2004. Article (CrossRef Link)

[6]   E. Barshan, A. Ghodsi, Z. Azimifar, and Z. Jahromi, "Supervised principal component analysis: Visualization, classification and regression on subspaces and submanifolds," *Pattern Recognition*, vol. 44, no. 7, pp. 1357-1371, July 2011. Article (CrossRef Link)

[7]   J. Liu, D. S. Chen, and J. F. Shen, "Development of Self-Validating Soft Sensors Using Fast Moving Window Partial Least Squares," *Industrial & Engineering Chemistry Research*, vol. 49, no. 22, pp. 11530-11546, Oct. 2010. Article (CrossRef Link)

[8]   A. Rani, V. Singh, and J. Gupta, "Development of soft sensor for neural network based control of distillation column," *ISA Transactions*, vol. 52, no. 3, pp. 438-449, May 2013. Article (CrossRef Link)

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

[10]   X. Yuan, Z. Ge, and Z. Song, "Soft sensor model development in multiphase/multimode processes based on Gaussian mixture regression," *Chemometrics and Intelligent Laboratory Systems*, vol. 138, pp. 97-109, Nov. 2014. Article (CrossRef Link)

[11]   J. Zheng, J. Zhu, G. Chen, Z. Song, and Z. Ge, "Dynamic Bayesian network for robust latent variable modeling and fault classification," *Engineering Applications of Artificial Intelligence*, vol. 89, p.103475, Mar. 2020. Article (CrossRef Link)

[12]   W. Shao, Z. Ge, L. Yao, and Z. Song, "Bayesian Nonlinear Gaussian Mixture Regression and its Application to Virtual Sensing for Multimode Industrial Processes," *IEEE Transactions on Automation Science and Engineering*, vol. 17, no. 2, pp. 871-885, Apr. 2020. Article (CrossRef Link)

[13]   Z. Yang, L. Yao, and Z. Ge, "Streaming parallel variational Bayesian supervised factor analysis for adaptive soft sensor modeling with big process data," *Journal of Process Control*, vol. 85, pp. 52-64, Jan. 2020. Article (CrossRef Link)

[14]   W. Shao, Z. Ge, and Z. Song, "Bayesian Just-in-Time Learning and Its Application to Industrial Soft Sensing," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 4, pp. 2787-2798, Apr. 2020. Article (CrossRef Link)

[15]   L. Yao and Z. Ge, "Distributed parallel deep learning of Hierarchical Extreme Learning Machine for multimode quality prediction with big process data," *Engineering Applications of Artificial Intelligence*, vol. 81, pp. 450-465, May 2019. Article (CrossRef Link)

[16]   Q. Sun and Z. Ge, "Deep Learning for Industrial KPI Prediction: When Ensemble Learning Meets Semi-Supervised Data," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 1, pp. 260-269, Jan. 2021. Article (CrossRef Link)

[17]   Z. Chen, J. Hu, G. Min, A. Y. Zomaya, and T. El-Ghazawi, "Towards Accurate Prediction for High-Dimensional and Highly-Variable Cloud Workloads with Deep Learning," *IEEE Transactions on Parallel and Distributed Systems*, vol. 31, no. 4, pp. 923-934, Apr. 2019. Article (CrossRef Link)

[18]   M. Kano and M. Ogawa, "The state of the art in chemical process control in Japan: Good practice and questionnaire survey," *Journal of Process Control*, vol. 20, no. 9, pp. 969-982, Oct. 2010. Article (CrossRef Link)

[19]   J. Zhu, Z. Ge, and Z. Song, "Distributed Parallel PCA for Modeling and Monitoring of Large-scale Plant-wide Processes with Big Data," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 4, pp. 1877-1885, Jan. 2017. Article (CrossRef Link)

[20]   H. Kaneko and K. Funatsu, "Classification of the Degradation of Soft Sensor Models and Discussion on Adaptive Models," *AIChE Journal*, vol. 59, no. 7, pp. 2339-2347, Jan. 2013. Article (CrossRef Link)

[21] J. Huaiping, X. Chen, J. Yang, H. Zhang, L. Wang, and L. Wu, "Multi-model adaptive soft sensor modeling method using local learning and online support vector regression for nonlinear time-variant batch processes," *Chemical Engineering Science*, vol. 131, pp. 282-303, July 2015. [Article (CrossRef Link)](#)

[22] Z. Ge and Z. Song, "A comparative study of just-in-time-learning based methods for online soft sensor modeling," *Chemometrics and Intelligent Laboratory Systems*, vol. 104, no. 2, pp. 306-317, Dec. 2010. [Article (CrossRef Link)](#)

[23] Z. Liu, Z. Ge, G. Chen, and Z. Song, "Adaptive soft sensors for quality prediction under the framework of Bayesian network," *Control Engineering Practice*, vol. 72, pp. 19-28, Mar. 2018. [Article (CrossRef Link)](#)

[24] K. Li, Y. Han, and H. Huang, "Soft sensor method for moisture content of well oil based on automatic spectral clustering and multiple extreme learning," *CIESC Journal,* vol. 67, no. 7, pp. 2925-2933, May 2016. [Article (CrossRef Link)](#)

[25] X. Yuan, Z. Ge, B. Huang, Z. Song, and Y. Wang, "Semisupervised JITL Framework for Nonlinear Industrial Soft Sensing Based on Locally Semisupervised Weighted PCR," *IEEE Transactions on Industrial Informatics*, vol. 13, no 2 , pp. 532-541, Apr. 2017. [Article (CrossRef Link)](#)

[26] B. Pan, H. Jin, L. Wang, B. Qian, X. Chen, S. Huang, and J. Li, "Just-in-time learning based soft sensor with variable selection and weighting optimized by evolutionary optimization for quality prediction of nonlinear processes," *Chemical Engineering Research and Design*, vol. 144, pp. 285-299, Apr. 2019. [Article (CrossRef Link)](#)

[27] X. Yuan, J. Zhou, Y. Wang, and C. Yang, "Multi-similarity measurement driven ensemble just-in-time learning for soft sensing of industrial processes," *Journal of Chemometrics*, vol. 32. no. 10, pp. 1-14, 2018. [Article (CrossRef Link)](#)

[28] H. Kaneko and K. Funatsu, "Maintenance-free soft sensor models with time difference of process variables" *Chemometrics & Intelligent Laboratory Systems*, vol. 107, no. 2, pp. 312-317, July 2011. [Article (CrossRef Link)](#)

[29] L. Yao and Z. Ge, "Moving window adaptive soft sensor for state shifting process based on weighted supervised latent factor analysis," *Control Engineering Practice*, vol. 61, pp. 72-80, Apr. 2017. [Article (CrossRef Link)](#)

[30] H. Kaneko and K. Funatsu, "Moving Window and Just-in-Time Soft Sensor Model Based on Time Differences Considering a Small Number of Measurements," *Industrial & Engineering Chemistry Research*, vol. 54, no. 2, pp. 700-704, Jan. 2015. [Article (CrossRef Link)](#)

[31] Y. Liu, C. Li, and Z. Gao, "A novel unified correlation model using ensemble support vector regression for prediction of flooding velocity in randomly packed towers," *Journal of Industrial and Engineering Chemistry*, vol. 20, no. 3, pp. 1109-1118, May 2014. [Article (CrossRef Link)](#)

[32] H. Kaneko and K. Funatsu, "Ensemble locally-weighted partial least squares as a just-in-time modeling method," *AIChE Journal*, vol. 62, no. 3, pp. 717-725, Oct. 2015. [Article (CrossRef Link)](#)

[33] S. Weiming and X. Tian, "Adaptive soft sensor for quality prediction of chemical processes based on selective ensemble of local partial least squares models," *Chemical Engineering Research and Design*, vol. 95, pp. 113-132, Mar. 2015. [Article (CrossRef Link)](#)

[34] H. Jin, X. Chen, L. Wang, K. Yang, and L. Lu, "Adaptive soft sensor development based on online ensemble Gaussian process regression for nonlinear time-varying batch processes," *Industrial & Engineering Chemistry Research*, vol. 54, no. 30, pp. 7320-7345, July 2015. [Article (CrossRef Link)](#)

Ying Tian received the B.S. degree in electrical engineering and automation from East China University of Science and Technology, Shanghai, China, in 2010.And receive Ph.D. degree in control theory and control engineering from East China University of Science and Technology, Shanghai, China, in 2015.She is currently an associate professor in the School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology. Her research interests include process monitoring, system modeling, advanced control and process optimization.

Yuting Zhu received the B.S. degree in electrical engineering and automation from Tianjin University of Technology, Tianjin, China, in 2018. She is now pursuing the M.S. degree in the Electrical Engineering Department, University of Shanghai for Science and Technology (USST). Her current research interests include machine learning and deep learning.