Preprints, 10th IFAC International Symposium on
Advanced Control of Chemical Processes
Shenyang, Liaoning, China, July 25-27, 2018

LWS based PCA subspace ensemble model for soft sensor development

Xudong Shi [†], **Weili Xiong** [†,‡*]

[†] *School of Internet of Things Engineering, Jiangnan University, Wuxi, Jiangsu Province, China*
[‡] *Key Laboratory of Advanced Process Control for Light Industry of Ministry of Education, Jiangnan University, Wuxi, Jiangsu Province, China (e-mail: *greenpre@163.com)*

Abstract: Most regression approaches, such as principal component analysis (PCA), are based on an assumption that the process data follow a Gaussian distribution. However, the process data usually dissatisfy that assumption. Thus, the locally weighted standardization (LWS) method is employed for transforming data into an approximate Gaussian distribution. Furthermore, the LWS based subspace PCA ensemble modeling method is developed. The subspace PCA can select important variables in each subspace for ensemble modeling. As a result, the proposed method gives a weaker assumption constrain and a better regression performance. The effectiveness of this approach is testified by two study cases.
Keywords: Soft sensor; Non-Gaussian data; Principal component Analysis; Subspace ensemble learning;

1 INTRODUCTION

In modern process industries, how to measure and control key product variables plays an important role in producing high-quality products. To obtain reliable and real-time prediction of key quality variables, soft sensors are widely adopted. The data-driven regression methods are commonly used for soft sensing, which aims to find a functional relationship between the input and output variables. The category of popular data-driven soft sensors includes: principal component analysis (PCA) (Ge 2014), partial least square (PLS) (Zhao and Wang et al. 2010), support vector machine (SVM) (Xu and Du et al. 2007), and Gaussian process regression (GPR) (Xiong and Zhang et al. 2016; Ranjan and Huang 2016).

However, most of the conventional soft sensor models have a fundamental assumption that the collected data obey a Gaussian distribution. If these models are applied to the process data directly, the predicted performance of soft sensors will be poor, because the distribution of the process data may be not necessarily Gaussian. Dealing with non-Gaussian data modeling is a common issue in soft sensing, and its solution can be categorized into two types: the multiple local modeling (Zhao and Zhang et al 2006) and the global modeling (Hwang and Han 1999; Lane and Martin et al 2001). The methodology of multiple local modeling is to partition the whole dataset into several counterparts, and each counterpart obeys a single Gaussian distribution, such as Gaussian mixture model (Yu 2012). On the other hand, constructing a global model is an audacious attempt, which alters the data into a Gaussian distribution approximately. Generally, the global method transforms the non-Gaussian distributional data into a Gaussian distribution by data preprocessing approaches. The data preprocessing has been applied into process monitoring, since the assumption of Gaussian distribution is very common in process control (Ma and Hu et al. 2013; Wang and Liu et al. 2015; Zhao and Song et al. 2016). This paper focuses on the study of the non-Gaussian data transformation and the K-nearest neighborhood methods is employed to implement this transformation. Unlike the traditional Z-Score standardization method in which only a global set of mean and standard deviation is utilized for each data sample in the data set, we aim to standardize each sample through its K-nearest neighborhood. Therefore, the mean and standard deviation is unique to each data sample, and this method is analogous to the locally weighted modeling in just in time learning. However, the locally weighted approach is an online modeling method, in which the standardization is focused on testing data, not suitable for application upon non-Gaussian data modeling directly. Thus, in this work, we employ the locally weighted to standardize both the training and testing data, and call this method locally weighted standardization (LWS). Using K-nearest neighborhood for standardization is an intuitional practice, because the similar samples are doomed to share the same process information. Initially this standardization is simply replacing K-nearest neighborhood with the whole dataset for calculating mean the standard deviation, and this approach is called local neighborhood standardization (LNS). The difference between LWS and LNS is the usage of weights, by which the locally weighted method can be applied to LNS. As mentioned, the LNS focused on the common information of neighborhood samples, in order to make full use of local information the weights is necessary. In LWS, the mean the standard deviation is a locally weighted result of the dataset, and the more similar sample has a more important effect on the result. In that case, the multimodal distribution would be eliminated, because the peak value is subtracted by a large mean value owing to the smoothness of the data. As a result, the data preprocessed by LWS is probable to follow a distribution without multimodal feature, which is a good approximate form for a single Gaussian distribution.

For the sake of better regression performance, the principal component subspace ensemble method (PCSE) is employed. Ensemble learning is to construct several sub-models by partitioning the original training dataset into sub datasets and combing the corresponding sub-results for the final prediction of quality variable, which has been proven to be

theoretically and practically effective to improve the regression performance compared to the single-model (Ge 2014; Zhang and Fu et al. 2012; Wang and Liu et al. 2015). Subspace ensemble is strategy to partition dataset in the dimension of input variables, in which the subspaces is consisted of randomly chosen variables. Therefore the performance of the ensemble learning based soft sensor may be poor, because the two significant issues, namely the diversity and accuracy of sub-models cannot be guaranteed. Principal component analysis (PCA) is a conventional data analysis method, in which the loading vectors denote the largest covariance directions, and the loading vectors are orthogonal with each other. Owing to the orthogonality of PCA, the principal component subspaces spanned by the loading vectors are capable to satisfy the diversity requirement. Furthermore, the variables in each subspace are selected by the importance, which is evaluated by the element values of the corresponding loading vector. Since we only select the important variables and discard the irrelevant ones, the accuracy of the sub-models can be also improved. Hence, the principal component subspaces ensemble (PCSE) method takes good care of both the diversity and accuracy of sub-models.

Motivated by employing ensemble learning method to improve the prediction performance of non-Gaussian process modeling, this paper proposes a novel soft sensing method. In the proposed method, the non-Gaussian data is transformed into an approximate Gaussian distribution by LWS. Furthermore, a PCSE based Gaussian process regression model is developed for more accurate prediction of the quality variables. It should be noted that both the PCA and GPR model have the assumption that the collected data obey a single Gaussian distribution, and the interesting point is that the LWS is opportunely to transform the data into a single Gaussian distribution. Therefore, the PCA and GPR model are fortunate to be satisfied the fundamental Gaussian assumption.

The remainder of this paper is organized as follows. In section 2, the LWS method is introduced and a numerical example is provided. We revisit the GPR model briefly in section 3. The detailed information of the PCSE strategy is given in section 4, followed two case studies in section 5. Finally, we make conclusions in section 6.

2 NON-GAUSSIAN CHARACTERS AND DATA STANDARDIZATION STRATEGY

In this section, the locally weighted standardization (LWS) strategy is employed to overcome the drawbacks of the conventional Z-Score approach in non-Gaussian process. Different from the Z-Score approach, the proposed LWS strategy standardizes each sample by utilizing the weighted mean and standard deviation of its neighborhood interval rather than the average statistics of the whole dataset. Due to the smoothing effect of the weights, the preprocessed data can be altered into an Gaussian distribution approximately.

The detailed steps are given as follows:

For the given training dataset $\mathbf{X}(N\times J)$, in which the rows are samples and the columns are variables. Before we introduce the proposed LWS approach, several definitions are required to be given:

(1): The j-th nearest neighborhood of x_i is defined as $\mathrm{n}_j(\mathrm{x}_i)$,

(2): The neighborhood intervals are defined as: $\mathrm{I}_k(\mathrm{x}_i)=\{\mathrm{x}_i,\mathrm{n}_1(\mathrm{x}_i),\mathrm{n}_2(\mathrm{x}_i),...,\mathrm{n}_k(\mathrm{x}_i)\}$ and $k=1,2,...,N-1$.

For a certain sample $\mathrm{x}_i(i=1,2,...,N)$, the LWS formula is given as follow:

$$\breve{\mathbf{x}}_i=\frac{\mathbf{x}_i-\mathrm{m}_{\mathbf{x}_i}}{\mathrm{s}_{\mathbf{x}_i}}, \tag{1}$$

where $\breve{\mathbf{x}}_i$ is the LWS standardized data, $\mathrm{m}_{\mathbf{x}_i}=\mathbf{w}_{i,j}^{\mathrm{T}}\mathbf{m}_j$ and $\mathrm{s}_{\mathbf{x}_i}=\mathbf{w}_{i,j}^{\mathrm{T}}\mathbf{s}_j$,while $\mathbf{m}_j$ and $\mathbf{s}_j$ is the mean and standard deviation of the j-th interval I_j; $\mathbf{w}_{i,j}=[w_{i,1},w_{i,2},...,w_{i,N-1}]$ is the weights of all the neighborhoods for x_i whose elements can be calculated as follow:

$$w_{ij}=\left[\frac{1}{\mathrm{d}\left(\mathbf{x}_i,\mathrm{n}_j(\mathbf{x}_i)\right)}\right]\Big/\left[\sum_{m=1}^{N-1}\frac{1}{\mathrm{d}\left(\mathbf{x}_i,\mathrm{n}_m(\mathbf{x}_i)\right)}\right], \tag{2}$$

where $\mathrm{d}(\mathbf{x}_i,\mathbf{x}_j)$ is Euclidean the distance between samples $\mathbf{x}_i$ and $\mathbf{x}_j$.

Form Eq. (1), it can be found that the commonly used Z-Score standardization is a special case of the LWS method. Thus, the LWS is a modified and general form of the conventional Z-Score strategy. For a good visual presentation, a numerical example is represented to show the superiority of the LWS to the Z-Score standardization. A total of 400 samples are generated by Eq. (3), which apparently does not obey a Gaussian distribution.

$$x:\begin{cases}\#001-200\sim N(8,0.5)\\ \#201-400\sim N(15,0.6)\end{cases}. \tag{3}$$

These samples standardized by the two different strategies are plotted in Fig. 1. For more intuitional comprehension, the kernel density estimation method is employed, and the estimated results are shown in Fig. 2. From Fig. 1, we can see that the data processed by Z-Score and LWS are obviously different, therefore their probability density estimations are simultaneously different. As shown in Fig. 2, the Z-Score treated data cannot be depicted by a single Gaussian distribution, but they can be depicted by two Gaussian distributions perfectly. In addition, a single Gaussian distribution can fit the data processed by the LWS method approximately. Now, we can conclude that the Z-Score method cannot eliminate the non-Gaussian characteristic of the original data; meanwhile the LWS method alters the data into an approximate Gaussian distribution.

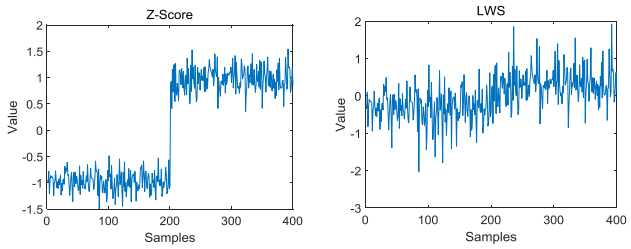


Fig. 1 Plot of the standardized data samples

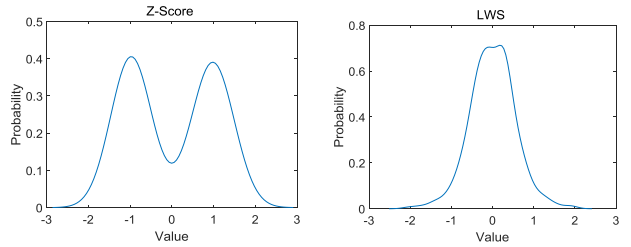


Fig. 2 kernel density estimation of the standardized data

3. GAUSSIAN PROCESS REGRESSION

Suppose the training data can be represented as $\mathbf{X}(N\times J)$ and $\mathbf{y}(N\times 1)$, where N is the number of observations; J is the number of the input variables. The Gaussian process regression is assumed as the regression function with zero-mean Gaussian prior distribution:

$$y=\{f(\mathbf{x}_1),f(\mathbf{x}_2),...,f(\mathbf{x}_n)\}\sim \mathrm{GP}(\mathbf{0},\mathbf{K}), \tag{4}$$

where $\mathbf{x}_i\,(i=1,2,...,N)$ is the i-th observation; $\mathbf{K}$ is the covariance matrix with its ij-th element $k_{ij}=\sigma_f^2\exp^{\left\{-\frac{1}{2l^2}(\mathbf{x}_i-\mathbf{x}_j)^{\mathrm{T}}(\mathbf{x}_i-\mathbf{x}_j)\right\}}+\delta_{ij}\sigma_{\mathrm{n}}^2$. In which $\delta_{ij}=1$ only when $i=j$, otherwise, $\delta_{ij}=0$; l is length-scale, σ_f^2 and σ_{n}^2 are signal and noise variance respectively. The hyper-parameters $\Theta=[\sigma_n^2,\sigma_f^2,l]$ can be optimized by maximizing the following likelihood estimation:

$$\begin{aligned}\Theta^{*}&=\arg\max_{\Theta}\ln p(\mathbf{y}\mid\mathbf{X},\Theta)\\&=\arg\max_{\Theta}\ln\int_f p(\mathbf{y}\mid f,\sigma_n^2)p(f\mid\mathbf{X},\sigma_f^2,l).\end{aligned} \tag{5}$$

When a quarry sample $\mathbf{x}_{\mathrm{new}}$ comes, the posterior distribution of the output $(y_{\mathrm{new}}\mid\mathbf{X},y,\mathbf{x}_{\mathrm{new}})\sim \mathrm{N}(\mu,\Sigma)$ can be determined, and take the expectation μ as the prediction $\hat{y}_{\mathrm{new}}$ of the GPR model, which can be calculated as follows:

$$\mu=\mathbf{K}(\mathbf{x}_{\mathrm{new}},\mathbf{X})[\mathrm{K}(\mathbf{X},\mathbf{X})+\sigma_n^2\,\mathrm{I}]^{-1}y. \tag{6}$$

4 PRINCIPAL COMPONENT SUBSPACE ENSEMBLE STRATEGY FOR SOFT SENSING

In this section, the principal component subspace ensemble (PCSE) learning approach is employed for a more accurate predicted result. It should be noted that the more diverse and more accurate sub-models are, the more reliable predicted result will be obtained. In order to improve the diversity of the sub-models, PCA is used to partition the original dataset into several sub datasets whose corresponding sub-space are orthogonal with each other. Furthermore, for accuracy enhancement the important variables are selected for modeling, while the unrelated variables are discarded . Therefore, the PCSE is a well-designed method which satisfies the diversity and accuracy requirement of the ensemble learning. The specific procedures of the PCSE are provided as follows:

Denote the input variables as $\mathbf{X}(N\times J)$, where N is the number of data samples, J is the number of variables. The corresponding quality variable is given as $\mathbf{y}(N\times 1)$. The conventional form of PCA can be expressed as follows:

$$\mathbf{X}=\mathbf{T}\mathbf{P}^{\mathrm{T}}+\mathbf{E}, \tag{7}$$

$$\mathbf{T}=\mathbf{X}\mathbf{P}, \tag{8}$$

where $\mathbf{P}(J\times A)$ is the loading matrix, $\mathbf{T}(N\times A)$ is the principal component score matrix, and A is the selected number of latent variables.

Based on the A orthogonal principal component directions, the corresponding subspace models can be determined. However, important variables need to be selected carefully for each subspace model, otherwise all the sub-spaces are homogeneous. The important variables should be selected, while the useless variables will be discarded. To evaluate the importance of each input variable in a certain subspace, a variable contribution index (VCI) is defined as:

$$\mathrm{VCI}(i,j)=\frac{|p_{ij}|}{|p_{1j}|+...+|p_{ij}|+...+|p_{mj}|}, \tag{9}$$

Where $i=1,2,...,J$, , p_{ij} is the i-th element in the j-th loading vector. Thus, in a certain sub-space, the importance of each variable can be determined, furthermore input variables can be selected regularly. Then, the dataset is divided into A sub-datasets, according the sub regression models is built:

$$\mathbf{X}\xrightarrow[\text{PC directions}]{\text{division}}\begin{cases}\#1\ \text{Subspace}(\mathbf{X}_1)\rightarrow \text{Mdl1}\\ \#2\ \text{Subspace}(\mathbf{X}_2)\rightarrow \text{Mdl2}\\ \vdots\\ \#A\ \text{Subspace}(\mathbf{X}_A)\rightarrow \text{Mdl}A\end{cases},$$

where $\mathbf{X}_i(i=1,2,...,A)$ is i -th subspace developed through the corresponding principal component direction, and the variables are selected based on the VCI. Finally, the final prediction $\hat{\mathbf{y}}_{\mathrm{final}}$ is gained by the mean value of the sub-prediction results $\hat{\mathbf{y}}_i$ of each sub-models, in other words $\hat{\mathbf{y}}_{\mathrm{final}}=\frac{1}{A}\sum_{i=1}^{A}\hat{\mathbf{y}}_i$.

The steps of the LWS based PCSE soft senor modeling is summarized as follows:

(1) Collect the training data $\mathbf{X}$ and $\mathbf{y}$,

(2) For each training sample x_i, standardize it by LWS,

(3) Employ PCSE to the standardized dataset and construct sub- models,

(4) When quarry sample x_{new} comes, using LWS to standardize it, and predict the corresponding quality variable by each sub-model,

(5) Finally fuse the predictions to obtain the final prediction.

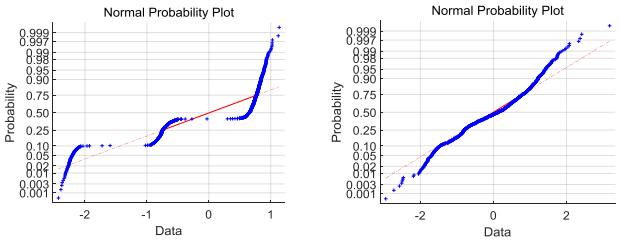
5 CASE STUDIES

In this part, two case studies on sulfur recover unit and a blast furnace ironmaking process will be carried out to verify the validity and effectiveness of LWS-PCSE method.

5.1 Application to the sulfur recover unit

The SRU process aims to recycle the sulfur and remove the pollution that cause harm to the environment and human health (Fortuna and Graziani et al 2007). The SRU plant mainly deals with two types of acidic gases: one is monoethanolamine (MEA) which is abundant with H_2S; the other is sour water stripper (SWS) which is abundant with H_2S and NH_3. The main combustion chamber is used for treating MEA gas and can be fully combusted when sufficient air (AIR-MEA) is used; another combustion chamber is used for treating SWS gas, and the air flow entering it can be written as AIR-SWS. The detailed description of the SRU can be found in (Fortuna and Graziani et al 2007). During the SRU process, the H_2S is transformed into product sulfur with the SO_2 generated. The concentration of H_2S and SO_2 should be monitored, since they are harmful to the environment and human body. In this case, we develop 5 soft sensors to estimate the concentration of SO_2, where the 5 input variables are: MAE gas flow, first air flow, second air flow, gas flow in SWS zone, air flow in SWS zone. A total of 1500 samples collected form the SRU dataset are divided into two parts: the training dataset consisting of 1000 data samples, and the testing dataset which contains the remaining 500 data samples.

First, let us test the effectiveness of the LWS method for altering data into a Gaussian distribution. Normal probability plot is employed to implement that test, in this plot if the blue is more close to the red line indicates that the probability density of the variable is closer to the normal distribution. Fig. 3 is the testing result of the No. 4 variable. It can be seen in Fig. 3 the Z-Score treated data has a strong non-Gaussian characteristic since the blue line is not close to the red line. Meanwhile, the blue line in Fig. 3 (b) clusters much closer to the red line compared to Fig. 3 (a). Therefore, the LWS method is able to reduce the non-Gaussian characteristic from the original data. Although the data treated by LWS method do not obey a Gaussian distribution strictly, the Gaussian distribution is well approximate form for the real data probability density. Furthermore, the LWS treated data is more suitable for the PCA and GPR method than the conventional Z-Score method, due to the Gaussian assumption of the two methods.



(a) Z-Score (b) LWS

Fig. 3. Normal probability plot of the No.4 variable treated by Z-Score and LWS respectively

Then the processed data are used for soft sensor construction with the basic GPR model, and the two methods are called ZSS-GPR and LWS-GPR, respectively. In addition another approach that can deal with the non-Gaussian characteristic, namely, Gaussian mixture model (GMM) is used for comparison. Then, the PCSE strategy is applied to the modeling, and 3 principal components are selected. Hence, a total of 3 subspace models are built up through the chosen principal component directions, and the variables in each subspace is selected by the VCI.

The predicted results using different methods are summarized in Table 1. We can find that LWS based method has a better performance than the Z-Score based method, because the prediction index of the former is superior to that of the latter. Also, it can be found from Table 1 that RMSE of the GMM-GPR model is smaller than the ZSS-GPR model, while it is bigger than the LWS-GPR model. Therefore, compared with the GMM method, LWS is a better method in addressing the data Non-Gaussian characters in this case. Although, the GMM method and the LWS method are both proposed to solve the data Non-Gaussian problem, the mechanisms of them are quite different. The GMM is a cluster algorithm, and each component of the GMM is assumed to obey a Gaussian distribution. The main drawback of the GMM method is that the number of the data in each Gaussian cluster is much smaller than the original dataset, especially when the number of the training samples are relatively small. The LWS is a data pretreatment method, this approach can transform non-Gaussian data into an approximately Gaussian distribution, thus the number of available data keeps the same. As shown in Table 1, the prediction performance obtains a further promotion with the PCSE method. Because it is difficult for a single model to capture the whole data features precisely. While the PCSE, as an ensemble learning method, can divide the whole data into several subspaces. In that case a single model is able to get a good regression performance. There are two points to explain the superiority of the prosed method.: first, the PCA is based on the assumption that the process data follows a Gaussian distribution; second, the accuracy of the subspace models based on the LWS method are higher than the models based on the Z-Score method.

Table 1. Predictive results of SRU

Model	RMSE	COR	MAE
ZSS-GPR	0.0337	0.7977	0.2462
GMM-GPR	0.0321	0.8175	0.2722
LWS-GPR	0.0303	0.8396	0.1479
ZSS-PCSE-GPR	0.0323	0.8150	0.2304
LWS-PCSE-GPR	0.0299	0.8472	0.1606

For clear comparison, the 251th-400th prediction results of the ZSS-GPR and the proposed LWS-PCSE-GPR method is shown in Fig. 4, while their predicted errors are shown in Fig. 5. From Fig. 4, we can see that the LWS-PCSE-GPR given red plots track more tightly with the black real samples, than the ZSS-GPR given blue plots. Correspondingly, the predicted errors of the LWS-PCSE-GPR are smaller than that of the ZSS-GPR as well. The comparison results show the effectiveness of the proposed methods.

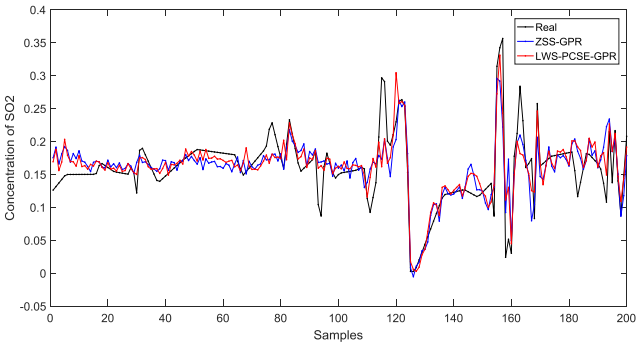


Fig. 5. Prediction plots of SRU

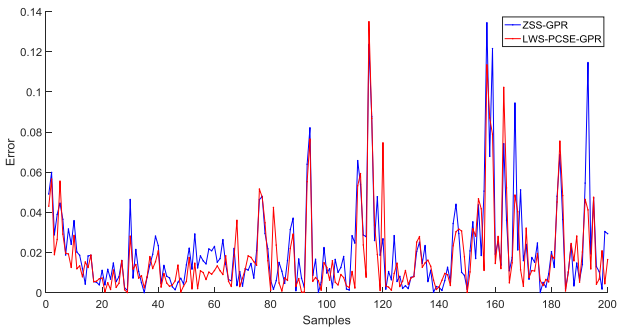


Fig. 6 Absolute value of prediction errors for SRU

5.2 Application to ironmaking process

In this subsection, the proposed LWS-PCSE strategy is also tested on a blast furnace ironmaking process. Due to the harsh environment in the blast furnace, it is difficult to measure the concentration of silicon directly (Zeng and Liu et al. 2008). Therefore, constructing a soft sensor to make a prediction is an efficient approach. In this paper, seven easy-to-measure variables are selected as: coal injection, air flow, air permeability, feed speed, differential pressure, and Oxygen enrichment. A total of 1000 samples are collected, whose change trend is plotted in Fig. 7.

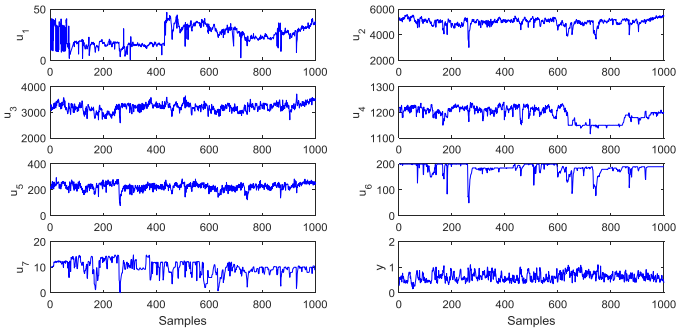
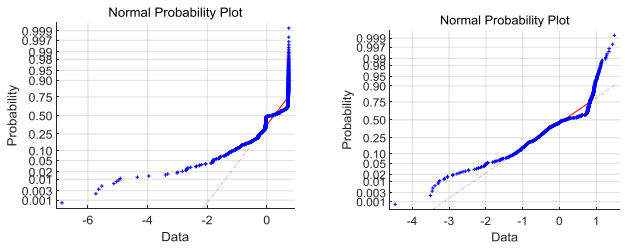


Fig. 7 Change trend plots of the ironmaking process



(a) Z-Score (b) LWS

Fig. 8. Normal probability plot of the No. 6 variable treated by Z-Score and LWS respectively

Table 2. Prediction results of ironmaking

Model	RMSE	COR	MAE
ZSS-GPR	0.1395	0.4453	0.3132
GMM-GPR	0.1379	0.4742	0.3478
LWS-GPR	0.1318	0.4692	0.3671
ZSS-PCSE-GPR	0.1330	0.4596	0.3388
LWS-PCSE-GPR	0.1262	0.4789	0.3745

These data are divided into two counterparts: the training dataset consisting of 700 data samples, and the testing dataset with the remaining 300 data samples. Again, the normality test is employed for comparison with the Z-Score and LWS method. The normality test results of the No. 6 input variable are shown in Fig. 8. It is obvious that the data treated by LWS method is closer to the normal distribution than the data treated by Z-Score method. The results illustrate the non-Gaussian characteristics of the process data is reduced by LWS. Then 5 soft models are then built up for quality prediction. Here, the number of the subspaces is set to be 6. The prediction results are listed in Table 2. For visual comparison, the first 50 real and predicted samples of ZSS-GPR and the proposed LWS-PCSE-GPR are plotted in Fig. 9 and their errors are provided in Fig. 10. It can be seen that LWS is superior to Z-Score, because the RMSE values of former is smaller than that of the latter. From Fig. 9 and 10 we can see that the LWS-PCSE-GPR performs better than the ZSS-GPR, the red curve tracks the real black curve better than the blue curve. In addition, the absolute value of the predicted error of the proposed method is smaller than the conventional ZSS-GPR method in general.

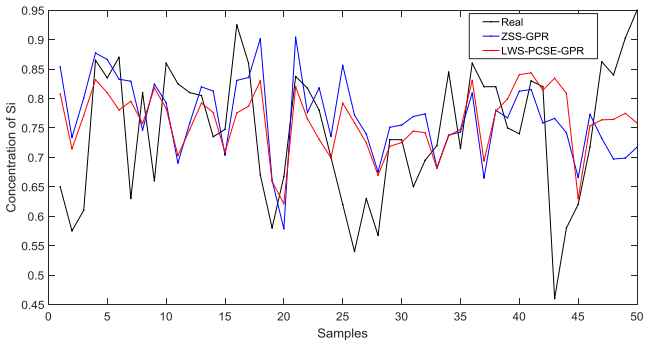


Fig.9 Prediction scatters of ironmaking

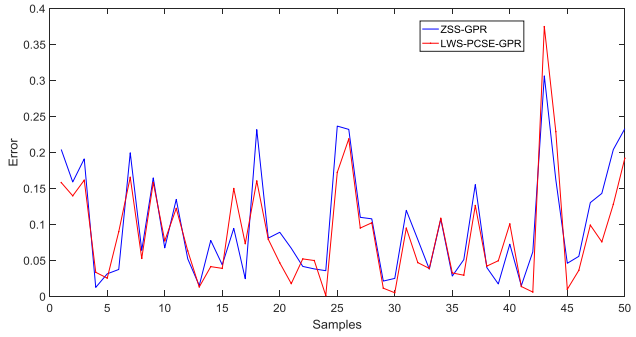


Fig. 10 Absolute value of prediction errors for ironmaking

6 CONCLUSIONS

In the present paper, a novel LWS approach is proposed to deal with the non-Gaussian problem through transforming the non-Gaussian data into an approximately Gaussian distribution. Furthermore, to improve the multivariate calibration performance, a PCSE strategy is developed. This method solves the problem of the inability of most soft sensor modeling methods to get satisfactory calibration performance when they are directly applied in non-Gaussian industrial processes. Finally, compared with the conventional soft sensor models, the LWS-PCSE method offers some kind of advantages and innovations.

ACKNOWLEDGEMENT

This work was supported by the National Natural Science Foundation of China (Nos: 61773182), the 111 Project (B12018). The authors greatly appreciate Dr. Jiusun Zeng for providing the data of the ironmaking process.

REFERENCE

Ge, Z. (2016) Supervised Latent Factor Analysis for Process Data Regression Modeling and Soft Sensor Application. IEEE Transactions on Control Systems Technology. 24, 1004-1011.

Ge, Z. (2014) Quality prediction and analysis for large-scale processes based on multi-level principal component modeling strategy. Control Engineering Practice. 31, 9-23.

Hwang, D. H., Han, C. H. (1999) Real-time monitoring for a process with multiple operating modes. Control Engineering Practice. 7, 891–902.

L. Fortuna, S. Graziani, et al. (2007). Soft Sensors for Monitoring and Control of Industrial Processes. Springer-Verlag, London.

Ma, H., Hu, Y., et al. (2013). Fault Detection and Identification Based on the Neighborhood Standardized Local Outlier Factor Method. Industrial & Engineering Chemistry Research. 52, 2389-2402.

R. Ranjan, B. Huang, et al. (2016). Robust Gaussian process modeling using EM algorithm. Journal of Process Control. 42, 125-136.

S. Lane, E. B Martin, et al. (2001) Performance monitoring of a multi-product semi-batch process. Journal of Process Control. 11, 1–11.

Wang, G., Liu, J., et al. (2015). A novel multi‐mode data processing method and its application in industrial process monitoring. Journal of Chemometrics. 29, 126-138.

Wang, L., Jin, H, et al (2016). Soft Sensor Development Based on the Hierarchical Ensemble of Gaussian Process Regression Models for Nonlinear and Non-Gaussian Chemical Processes. Industrial & Engineering Chemistry Research. 55 (28),7704-7719.

Xiong, W., Zhang, W., et al. (2016). JITL based MWGPR soft sensor for multi-mode process with dual-updating strategy. Computers & Chemical Engineering.90, 260-267.

Xu, Y., Du W., et al. (2007). Soft Sensor Modeling Based on KPCA and Least Square SVM. Journal of System Simulation. 19(17):3873-3855.

Yu, J. (2012) 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. 82, 22-30.

Zeng, J., Liu, X., et al. (2008). Wiener Model Identification of Blast Furnace Ironmaking Process. Isij International. 48, 1734-1738.

Zhang, W., Fu, Y., et al. (2012). Multi-model soft-sensor modeling based on improved clustering and weighted bagging. Ciesc Journal. 63, 2697-2702.

Zhao, C., Wang, F., et al. (2010). Quality prediction based on phase-specific average trajectory for batch processes. Aiche Journal. 54, 693-705.

Zhao, F., Song, B., et al. (2016). Multi-mode process monitoring based on a novel weighted local standardization strategy and support vector data description. Journal of Central South University. 23, 2896-2905.

Zhao, S., Zhang, J., et al. (2006). Performance monitoring of processes with multiple operating modes through multiple PLS models. Journal of Process Control. 16, 763–772.