Early Identification of Process Deviation Based on the Spatial Correlation of Measurements

Fangyuan Ma[a], Xianyao Han [a], Dexi Lin[b], Jingtao Zhong[b], Jingde Wang[c], Wei Sun[a*]

[a] Beijing Key Lab of Membrane Science and Technology, College of Chemical Engineering, Beijing University of Chemical Technology, 100029 Beijing, China
[b] Sinochem Quanzhou Petrochemical Co., Ltd, 362103, Quanzhou, China
[c] Center for Process Simulation and Optimization, College of Chemical Engineering, Beijing University of Chemical Technology, 100029 Beijing, China
sunwei@mail.buct.edu.cn

Almost all equipment in chemical process are three-dimensional, where mass and heat transfer are conducted, sometimes reaction as well. The corresponding progress can be reflected by the measurements at different locations in a unit, such as composition, temperature and pressure. In a vertical fixed bed reactor, the inlet flow goes from top to bottom and reacts in the reactor. To ignore the difference at the horizontal direction at the same bed height, the measurements at vertical direction will be the only index of reaction progress. If these measurements are available, the reaction can be well monitored.

It is critical to make sure the reaction operate according to the design, i.e., the normal operating condition. The composition of key component is the direct measurement of it, but it is hard to obtain in real time, instead, the acquisition of pressure and temperature measurements is much easier. Based on the understanding from chemical engineering point of view, they cannot tell the reaction progress directly, but they are the consequence of it. As the reaction gets further along the vertical direction of the unit, the temperature also changes accordingly. The spatial correlation among temperature measurements can be extracted when the process is under normal operational condition. Once this spatial correlation changes, it indicates that the process has deviated from the normal operating state. Therefore, early identification of process deviation can be achieved by monitoring the easily-obtained variable spatial correlation, such as temperature, if a proper data analysis method is employed.

In this work, a pre-reforming reactor of hydrogen production units is investigated. The spatial correlation among temperature data is extracted by using a multivariate statistical method. The results show that the process deviation can be detected 10 hours ahead of human operator.

1. Introduction

With the nature of multi-step processing and dynamics, the petrochemical production process is becoming more and more complex. Once a certain part of the system malfunctions, serious impacts will be resulted to both its upstream and downstream, and then become hard to isolate later. Therefore, it is significant to detect process deviation accurately and timely. Benefited from the wide use of distributed control system (DCS) in industrial process, large amounts of data have been collected, which have greatly accelerated the development of data-based process monitoring methods. The outstanding capability of multivariate statistical analysis (Kresta J V, MacGregor J F, Marlin T E, 1991) to detect anomalies has been recognized in many domains, which includes industrial monitoring (Kourti T, Lee J, Macgregor J F, 1996). Particularly, the multivariate statistical process control (MSPC) is one of the most popular data-based methods for process monitoring and is widely used in various industrial areas. According to literature, multivariate statistical methods mainly include Partial Least Squares (PLS), Principal Component Analysis (PCA), Canonical variate analysis (CVA), Independent Component Analysis (ICA), Artificial Neural Network (ANN), and so on (Cinar A, Palazoglu A, Kayihan F, 2007). Among them, PLS and PCA are common used methods in characterizing the

normal condition from data. For continuous steady operation，monitoring models are established by extracting cross-correlations among data (Qin S J, 2012). Furthermore, with the consideration of time dependence, dynamic PLS and dynamic PCA are proposed to extract the structure of autocorrelation for time variant operation (Dong Y, Qin S. J, 2017). However, in the chemical production process, some of the transfer and reaction conducted in the three-dimensional equipment with certain spatial distribution. In previous study, all variables are considered either by cross correlation or by autocorrelation, while the spatial correlation was neglected. With no doubt, spatial correlation is also a unique feature of a process object. If spatial correlation of measurements is properly considered, better process monitoring could be realized.

In this work, a pre-reforming reactor of hydrogen production units is analysed. On the basis of considering spatial correlation among temperature measurements, the multivariate statistical method based on PLS is used to realize early identification of process deviation.

2. Methodology

2.1 Partial Least Squares (PLS) method

Multivariate statistical analysis is a common method for variable spatial correlation analysis. In multiple linear regression method, PLS is a standard regression method, which is initially proposed by the Swedish statistician Herman O. A. Wold. Suppose a set of predictor variables $X \in R^{n \times m}$ and a set of predicted variables $Y \in R^{n \times p}$ are standardized matrices. PLS is a technique based on latent variables (LV) which develops a biased regression model between X and Y. PLS selects latent variables so that variation in X which is most predictive of Y is extracted (Tian S, Han X, Sun W, et al., 2017). The matrix X and Y can be decomposed respectively as follows:

$$X = TP^T + E = \sum_{j=1}^{\alpha} t_j p_j^T + E \tag{1}$$

$$Y = UQ^T + F = \sum_{j=1}^{\alpha} u_j q_j^T + F \tag{2}$$

where T is a score matrix, P is a loading matrix, α is the number of latent variables, E is residual matrix, t_j is score vector and P_j is loading vector. In equation (2), U is a score matrix, Q is a loading matrix, F is a residual matrix, u_j is a score vector and q_j is a loading vector. The inner relationship between t_j and u_j can be obtained through a univariate regression as follows:

$$u_j = b_j t_j \tag{3}$$

where b_j is the regression coefficient. To maximize the covariance between X and Y, the optimal number of latent variables can be extracted by testing the models.

2.2 Assessing the Accuracy of a Model

The quality of a linear regression fit is typically assessed using two related quantities: the residual standard error and R^2 statistic (James G, Witten D, Hastie T, et al., 2013).

2.2.1 Residual Standard Error (RSE)

The RSE is considered a measure of the lack of fit of the established model to data, which is computed using the formula as follow:

$$RSE = \sqrt{\frac{1}{n-2} RSS} = \sqrt{\frac{1}{n-2} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{4}$$

where $\hat{y}_i$ indicates the prediction value of model. In equation (4), the residual sum of squares (RSS) is given by the formula

$$RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{5}$$

The smaller value of RSE means the better fitting result of model.

2.2.2 R^2 Statistic

The R^2 statistic provides the proportion of variance explained measure of fit. It can be calculated as following:

$$R^2 = \frac{TSS - RSS}{TSS} = 1 - \frac{RSS}{TSS} \tag{6}$$

$$TSS = \sum (y_i - \overline{y})^2 \qquad (7)$$

where total sum of squares (*TSS*) is the total sum of squares, and *RSS* is defined in equation (5).
An R^2 statistic that is close to 1 indicates that a large proportion of the variability in the response has been explained by the regression, which means the model is good.

2.3 Control limit determination

 In this paper, the control limit (*CL*) of the model is determined based on validation data by the maximum absolute difference between the predicted value and the actual value. It can be calculated as following:

$$CL = \max(|y_i - \hat{y}_i|) \qquad (8)$$

2. Process Deviation Monitoring Model Construction Based on the Spatial Correlation of Measurements

3.1 Data description

In the current study, a pre-reforming reactor of hydrogen production is analysed. The pre-reforming reactor is a fixed-bed adiabatic reactor. In an adiabatic reactor bed, the macromolecular hydrocarbons are pre-reformed to form methane-rich gas, thereby reducing the reaction strength and heat load of the reformer. The gas fluid flows through the bed from top to bottom in the axial direction and a chemical reaction takes place under the action of the solid catalyst.
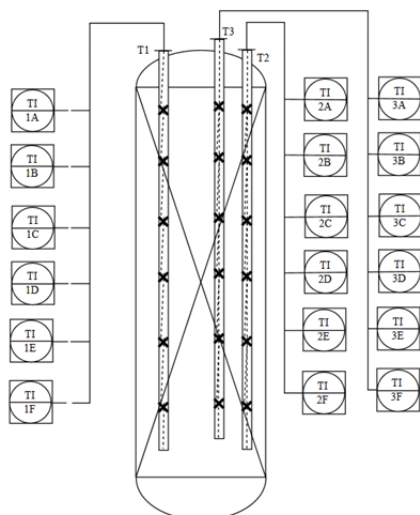

Figure 1: Thermocouple point distribution of pre-reforming reactor

Normally, the temperature data is easier to obtain than the component data. Thus, in the actual industry, the engineers mainly use the temperature data to identify reaction progress. The thermocouple point distribution of this pre-reforming reactor is shown in Figure 1. 18 thermocouple temperature measuring instruments are distributed over 6 different bed heights, and there are 3 temperature measuring points in each bed height.
In this work,18 temperature variables from bed section are considered in monitoring. The process data are collected from October 1, 2016 to December 31, 2016, which are sampled every minute for investigation.

3.2 Modelling

Temperature is an important indicator for evaluating the progress of the reaction. If the process is under a normal steady state, the spatial correlation of temperature can be approximated by a linear relationship. As mentioned above, the spatial correlation among temperature measurements is extracted by PLS algorithm. By Considering the spatial characteristics of the flow from top to bottom, the 12 temperature variables of the beds A, B, C, and D are considered as predictor variables X and the 6 temperature variables of the beds E, and F are regarded as predicted variables Y. For 6 predicted variables Y, PLS regression models are established respectively.
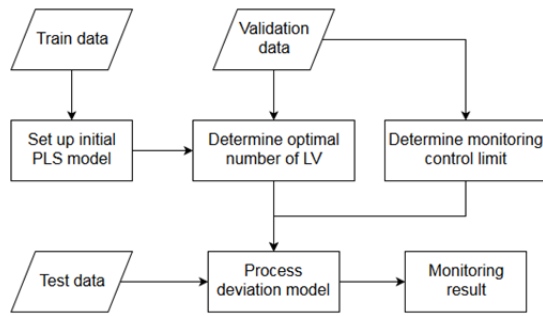
Figure 2: Process deviation monitoring procedure

As shown in the Figure 2, the PLS model is first built using the training data. Then the number of latent variables is obtained based on validation data by calculating the residual standard error and the R^2 statistic of the model. Meanwhile, the control limit value is determined by validation data. By combining PLS model and control limit, process deviation monitoring model based on the spatial correlation of measurements is established. In this work, 3500 samples are selected as training data, 1000 sample are selected as validation data, and 2300 sample are selected as test data.
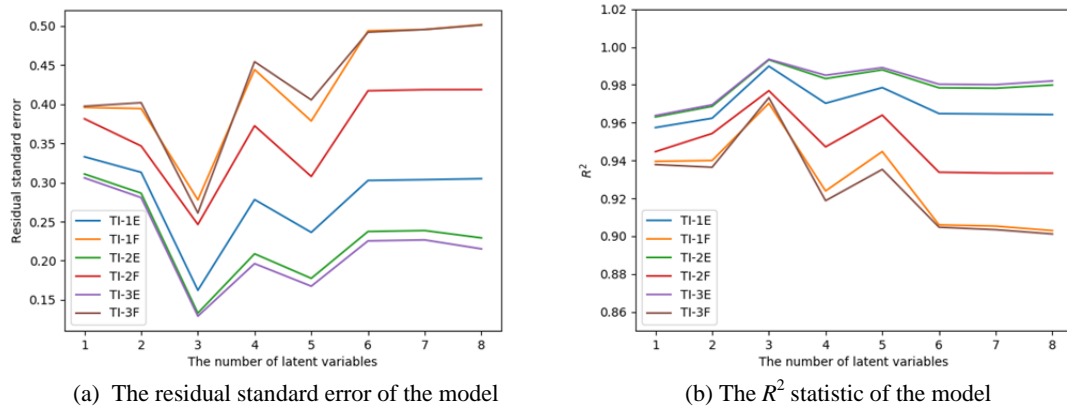

(a) The residual standard error of the model   (b) The R^2 statistic of the model

Figure 3: Selection for the number of latent variables

It can be seen that the residual standard error of the model is the smallest and the R^2 statistic is closest to 1 when the number of latent variables is 3. Hence, the optimal number of latent variables for each PLS model all is 3. The control limit is calculated by the method mentioned in 2.3.

3. Results and Discussion

In this work, a multivariate regression model is established to extract the spatial correlation among temperature data. By comparing the deviation between the predicted value and the actual value of the model, the spatial correlation deviation variation of the variable is monitored. Once the deviation exceeds the control limit, the process is identified as deviating from normal condition.

Table 1: The RE for the six PLS models and the CL for six process deviation models

	TI-1E	TI-2E	TI-3E	TI-1F	TI-2F	TI-3F
RE	0.0021	0.0017	0.0016	0.0034	0.0033	0.0031
CL	1.0227	0.8430	0.8045	1.6966	1.6350	1.5393

Based on the method above, six process deviation models based on measurement spatial correlations are established, which are TI-1E model, TI-2E model, TI-3E model, TI-1F model, TI-2F model and TI-3F model respectively. Based on validation data, the relative errors (*RE*) for the six PLS models and the calculation results of monitoring control limits (*CL*) for six process deviation models are shown in Table 1. The relative errors of the model are all less than 0.35%. In other words, models all have a good regression performance.
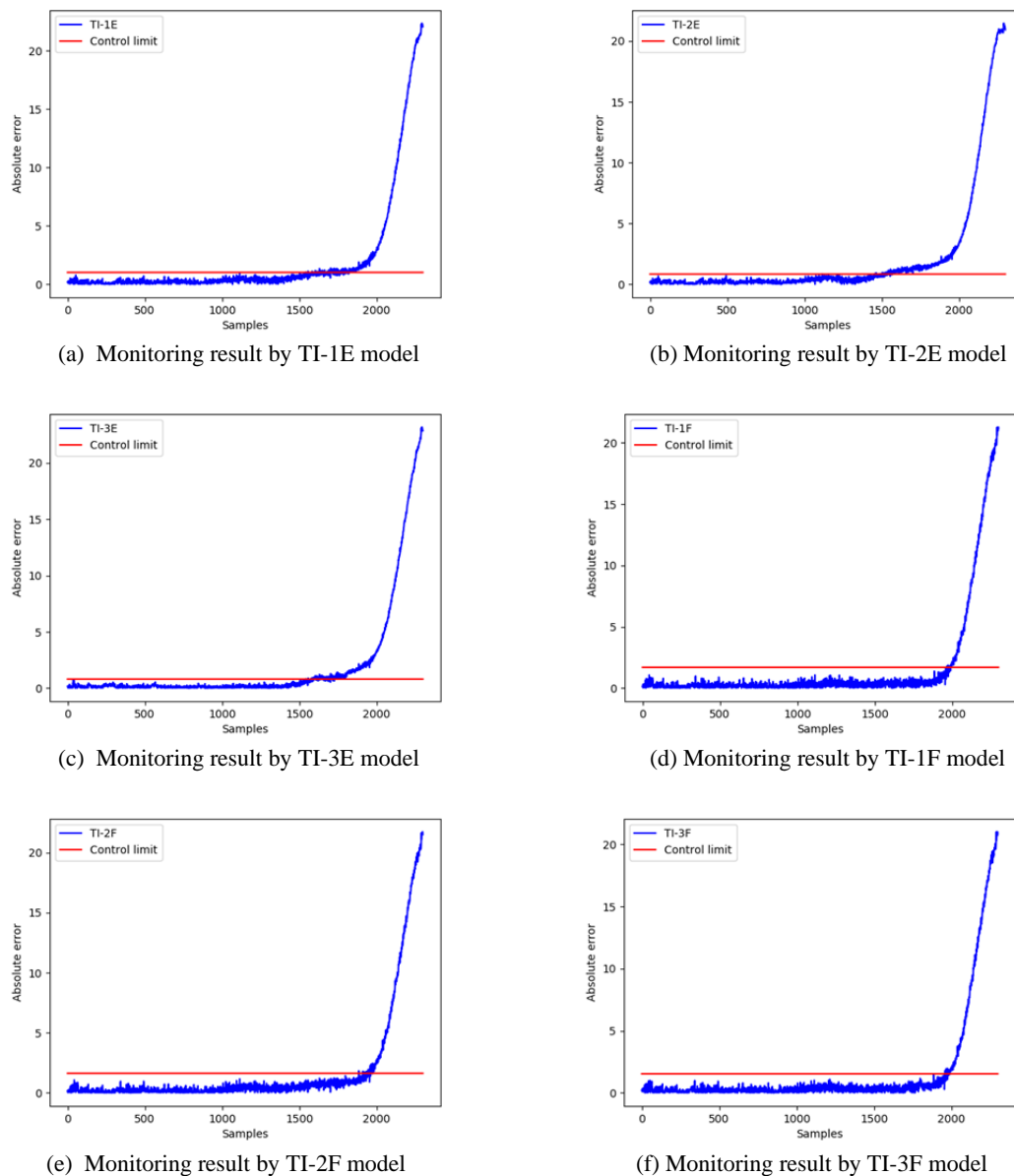
(a) Monitoring result by TI-1E model

(b) Monitoring result by TI-2E model

(c) Monitoring result by TI-3E model

(d) Monitoring result by TI-1F model

(e) Monitoring result by TI-2F model

(f) Monitoring result by TI-3F model

Figure 4: Monitoring results by six process deviation models

According to established process deviation monitoring models, the monitoring results of the test data are shown in Figure 4. When the process is under normal working condition, the absolute error between the predicted value of the PLS model and the actual measured value is smaller than the control limit. Once the absolute error exceeds the control limit, it shows that the process has deviated from the normal operating state. The early warning time results of this process deviation are shown in Table 2. The earliest process deviation is detected in TI-2E model and the time is 2016/11/9 18:33. The latest process deviation is detected in TI-1F model and the time is 2016/11/10 01:32.

Table 2: The early warning time results of process deviation

	TI-1E	TI-2E	TI-3E	TI-1F	TI-2F	TI-3F
Time	22:41 2016/11/9	18:33 2016/11/9	21:12 2016/11/9	01:32 2016/11/10	01:05 2016/11/10	01:30 2016/11/10

By checking the history record of the pre-reforming reactor operation, it can be found that engineers discovered a process deviation of the reactor at 4:20 on November 10, 2016 due to an abnormal drop in the temperature of reactor bed C and D. This also explains why TI-2E mode is the first to detect the process deviation. After discovering that the reactor was under an abnormal working condition, engineers analyzed the components of the reaction gas raw materials and found that the process deviation cause was the high sulfur content in the diluted steam, which resulted catalyst poisoning.

It can be seen that the earliest alarm time of the model is nearly 10 hours earlier than the time when the engineer found the reactor abnormality. The latest alarm time is nearly 3 hours earlier. In other words, this process deviation monitoring method based on spatial correlation of measurement has good monitoring performances on early identification of process deviation.

5. Conclusion

In this work, the spatial correlation of measurements in chemical production process is considered. The pre-reforming reactor of hydrogen production unit with catalyst poisoning is investigated. The spatial correlation among variables can be extracted by multivariate regression algorithm. Based on proposed method, early identification of process deviation can be realized. The results show that the model can be found 10 hours earlier than engineers which could provide a reference for engineers.

Acknowledgements

References

Cinar A., Palazoglu A., Kayihan F., 2007, Chemical process performance evaluation. Boca Raton: CRC press.

Dong Y., Qin S. J., 2017, A novel dynamic PCA algorithm for dynamic data modelling and process monitoring. Journal of Process Control, S095915241730094X.

James G., Witten D., Hastie T., et al., 2013, An Introduction to Statistical Learning: with Applications in R// An Introduction to Statistical Learning.

Kresta J. V., MacGregor J. F., Marlin T. E., 1991, Multivariate statistical monitoring of process operating performance. The Canadian Journal of Chemical Engineering, 69(1): 35-47.

Kourti T., Lee J., Macgregor J. F., 1996, Experiences with industrial applications of projection methods for multivariate statistical process control. Computers & chemical engineering, 20: S745-S750.

Qin S. J., 2012, Survey on data-driven industrial process monitoring and diagnosis. Annual Reviews in Control, 36(2):220-234.

Tian S., Han X., Sun W., et al., 2017, Coking analysis based on the prediction of coil surface temperature in radiation section of ethylene cracking furnace. Chemical Engineering Transactions, 61: 505-510.