Preprints of the
9th International Symposium on Advanced Control of Chemical Processes
The International Federation of Automatic Control
June 7-10, 2015, Whistler, British Columbia, Canada

PLS-based Similarity Analysis for Mode Identification in Multimode Manufacturing Processes

Ying Zheng*, S. Joe Qin, Fu-li Wang*****

** School of Automation, Huazhong University of Science and Technology, Wuhan, China*
*** Corresponding Author. The Chinese University of Hong Kong, Shenzhen, 2001 Longxiang Blvd., Longgang, Shenzhen, Guangdong, China. On leave from The University of Southern California, Los Angeles, CA 90089 USA (e-mail:sqin@usc.edu)*
**** College of Information Science and Engineering, Northeastern University, Shenyang, China*

Abstract: Many industrial manufacturing processes have multiple operation modes because of different strategy and varying feedstock. The traditional statistical process monitoring tools such as PCA and PLS cannot be applied since they assume that the process must have single mode operation region only. In this paper, all the factors that will affect the change of the mode are considered, a similarity factor including the similarity factor of PLS models and the mean shift of the external variables is introduced to measure the similarity of two sets of data. On basis of this similarity factor, a moving window is used and a mode identification approach for multimode process monitoring is proposed. The proposed approach is demonstrated on the benchmark Tennessee Eastman process.

Keywords: process monitoring, multiple operation mode, mode identification, partial least square(PLS), external analysis.

1. INTRODUCTION

Statistical process monitoring (SPM) is a powerful technology used in many industries to detect and identify changes and faults, which makes use of the input-output process data directly to monitor the process. Such methods as principal component analysis (PCA) and partial least square(PLS) are mostly frequently employed as key tools of SPM.

In general manufacturing industries, due to the alteration of the feedstock and production strategies, the manufacturing process will be operated in different modes. However, one of the drawbacks of PCA and PLS is that the monitored process must have one nominal operation region only, so that they behaves unsatisfactorily when applied to the process with multiple operation modes.

To handle the multimode problems, many methods have been proposed in recent years. Some literature considered global or mixture models for process monitoring. Yu and Qin proposed a multimode process monitoring approach based on finite Gaussian mixture models (FGMM) and Bayesian inference strategy. The Figueiredo–Jain (F–J) algorithm was adopted to optimize the number of Gaussian components and estimate their statistical distribution parameters. A Bayesian inference strategy was utilized to compute *a posteriori* probability of each monitored sample belonging to the multiple components and an integrated global probabilistic index was derived for fault detection of multimode processes. Besides, GMM was adopted and integrated with other techniques to deal with multimode process by Chen and Zhang, Feital et al. On the basis of the probabilistic PCA (PPCA) framework, Ge and Song introduced a Bayesian regularization method for performance improvement, thus a mixture Bayesian regularization method of PPCA was developed. They proposed a probabilistic strategy for combining results in different operation modes. Yu presented hidden Markov models (HMM)-based process monitoring models, which can quantify process states by combining local information (Mahalanobis distance) and global information (negative log likelihood probability) in HMM. Tong et al. offered a global modelling approach which was integrated with mode clustering, mode unfolding and an adaptive strategy for multimode processes.

A global model is simple but it loses information of each operation mode. Adaptive model based approach can represent most of the local mode information. Haghani et al. focused on the nonlinearity of processes and considered the multimode process as a piecewise linear system corresponding to each operating mode. Ma and Shi considered cross-mode correlations along with within-mode correlations in multimode processes and proposed a mixture factor analysis method to align all constructed models. Ng and Srinivasan developed an Adjoined PCA which allowed a smooth transition between two models. Choi et al. derived recursive formulas for updating a weighted mean and covariance matrix to recursively build a new model. Lee and Han used a signed digraph and statistical data analysis to generate the if–then rules from process knowledge.

Some scholars noticed there are common information among different modes, thus they developed methods which extract this information and designed monitoring scheme on basis of it. Zhang et al. extracted the common subspace of different modes by the subspace separation, and the kernel principal component models are built for the common and specific subspace respectively. Kano et al. proposed a method which is based on external analysis. They divided process variables into two parts: main variables and external variables which represented operation mode change. When the effect of external variables on the main variables is removed, the rest of the information from the main variables can be monitored. Ge et al. developed this method and proposed an online monitoring approach. They also constructed an additional regression model for soft sensing, which was robust to the change of the operation mode.

Since there are sufficient and efficient approaches for single mode process monitoring, the multimode process monitoring is easily implemented if the different modes are identified. PCA similar factor is adopted by Singhal and Seborg to identify a mode in a benchmark Tennessee Eastman challenge process. Tan et al. compared the loading matrices of the PCA model to identify the modes. Since PLS model employs the quality information, the similarity between different PLS models is also useful for quality-related multimode process monitoring. Zhao et al. presented the similarity metric for PLS models based on principal angles. Thus, multiple PLS models were developed for different operating modes and a principal angle was used to measure the similarities between any two PLS models. However, the similarity factor in this paper focused only on the PLS model structure. The mode changes due to setpoint changes and non-stationary disturbance changes may not affect the PLS model structure. Besides, the number of modes in Zhao et al. was assumed to be known as a priori.

To overcome these shortcomings, a new approach for mode identification is proposed in this paper. A moving window is adopted for the online process monitoring. Inspired by external analysis used in Keno et al and Ge et al (2008, 2014), a mean shift factor is designed to cover the mean change of the external variables by statistical methods. A similar factor which combines the PLS similar factor model in Zhao et al. and mean shift is proposed. A similarity factor of data points in the current window and the previous modes is derived to determine the current window's mode.

The rest of this paper is organized as follows. Section 2 introduces the similarity factor which combines the PLS similarity factor and the mean shift factor of external variables. Section 3 proposes the similarity-based online mode identification approach. The effectiveness of our approach is demonstrated in Section 4 on the benchmark Tennessee Eastman process. The last section gives conclusions.

2. SIMILARITY FACTOR

2.1 PLS similarity factor

Let columns of F and G be given in R^m with $p=\dim(\mathrm{F})\geq q=\dim(\mathrm{G})$. Let u and v be in the ranges of F and G, respectively. The q smallest principal angles between F and G are defined as

$$\theta_k=\arccos\left[\max_{u\in F,v\in G}\left(u^T v\right)\right]=\arccos\left[{u_k}^T v_k\right] \tag{1}$$

subject to $\|u\|_2=\|v\|_2=1, {u_i}^T u_j=0, {v_i}^T v_j=0$, $i\neq j$, where $(u_1,u_2,\ldots,u_q)$ and $(v_1,v_2,\ldots,v_q)$ are referred to as principal vectors. It is obvious with maximization that the principal angles satisfy:

$$0\leq\theta_1\leq\cdots\leq\theta_q\leq\frac{\pi}{2} \tag{2}$$

If $Q_F\in R^{m\times p}, Q_G\in R^{m\times q}$ are orthonormal bases for F and G respectively, then the overall similarity metric is 0,

$$\theta_0{}^2=\sum_{k=1}^{q}\cos^2\theta_k=\sum_{k=1}^{q}\lambda_k(Q_G^T Q_F Q_F^T Q_G)=trace(Q_G^T Q_F Q_F^T Q_G) \tag{3}$$

For normalized process input and output data matrices $X\in R^{n\times p}$ and $Y\in R^{n\times q}$, the PLS decomposition of X and Y results in

$$\begin{aligned} X&=\sum_{i=1}^{a}t_i p_i^T+E_a=TP^T+E_a\\ Y&=\sum_{i=1}^{a}u_i q_i^T+F_a=TBQ^T+F_a \end{aligned} \tag{4}$$

where $T=\left[t_1,\cdots,t_a\right]$ is the scores with $t_i=E_{i-1}w_i$, $P=\left[p_1,\cdots,p_a\right]$ is the loading vectors for X, $Q=\left[q_1,\cdots,q_a\right]$ is the loading vectors for Y, and a is the number of PLS factors which is usually determined by cross-validation. The score vectors T can be also be formulated as

$$T=XW\left(P^T W\right)^{-1} \tag{5}$$

where $W=\left[w_1,\cdots,w_a\right]$ is the weights in the PLS calculations and $WW^T=W^T W=\mathrm{I}$. Therefore, the weight can be chosen as the orthonormal basis.

For two different models, PLS1 and PLS2, with weights W_1 and W_2, and the numbers of factors a_1 and a_2, respectively, Zhao et al. introduced the similar factor between the two PLS models as

$$\theta_{1,2}=\sqrt{\frac{trace(W_1^T W_2 W_2^T W_1)}{\min(a_1,a_2)}}\in\left[0,1\right] \tag{6}$$

2.2 Mean shift factor of external variable

The PLS similarity factor (6) only focuses on the input-output relationship of PLS models. However, it does not consider the change of the operating setpoints or non-stationary disturbance changes which will also affect the process operation modes(McClure et al.).

Recently, a method called external analysis is adopted to monitor a multimode process. It divides the monitoring variables into two categories: external variables and main variables. The changes in external variables, such as throughput rate and set-points of controllers, will contribute to the change of operation modes. Thus, the process input X can be partitioned as

$$X=\left[U\quad M\right] \tag{7}$$

where $U=\begin{pmatrix}u_1 & u_2 & \cdots & u_r\end{pmatrix}\in R^{n\times r}$ are the external variables and M are the main variables.

Assume $U_1\in R^{n_1\times r}$ and $U_2\in R^{n_2\times r}$ are different external variables in different time periods, which are normally distributed random variables with unknown mean and unknown variance. From a random sample of multiple observations $U_{1,k}(1\le k\le r)$ and $U_{2,k}$, Montgomery presented confidence interval of the true mean as

$$\bar{U}_{1,k}-t_{\alpha/2,n_1-1}\frac{s_{1,k}}{\sqrt{n_1}}\le\mu_{U_{1,k}}\le\bar{U}_{1,k}+t_{\alpha/2,n_1-1}\frac{s_{1,k}}{\sqrt{n_1}}\qquad(8)$$

$$\bar{U}_{2,k}-t_{\alpha/2,n_2-1}\frac{s_{2,k}}{\sqrt{n_2}}\le\mu_{U_{2,k}}\le\bar{U}_{2,k}+t_{\alpha/2,n_2-1}\frac{s_{2,k}}{\sqrt{n_2}}\qquad(9)$$

where $\bar{U}$ is the sample mean, s is sample variance, μ is the true mean, $t_{\alpha/2,n_1-1}$ denotes the percentage point of the t distribution with n_1-1 degrees of freedom such that $\mathrm{P}\{\mathrm{t}_{\mathrm{n}_1-1}\ge\mathrm{t}_{\alpha/2,\mathrm{n}_1-1}\}=\alpha/2$. If U_1 and U_2 belong to the same mode, their true mean will have overlapping confidence intervals, thus

$$\left|\bar{U}_{1,k}-\bar{U}_{2,k}\right|\le t_{\alpha/2,n_1-1}\frac{s_{1,k}}{\sqrt{n_1}}+t_{\alpha/2,n_2-1}\frac{s_{2,k}}{\sqrt{n_2}}$$

(10)

Base on (10) we have the following mean shift factor for external variables

$$\eta_{1,2}=\sum_{i=1}^{r}\eta_{1,2,i}$$

where for $1\le i\le r$

$$\eta_{1,2,i}=\begin{cases}\frac{1}{r} & \left|\bar{U}_{1,k}-\bar{U}_{2,k}\right|>t_{\alpha/2,n_1-1}\frac{s_{1,k}}{\sqrt{n_1}}+t_{\alpha/2,n_2-1}\frac{s_{2,k}}{\sqrt{n_2}}\\ 0 & \left|\bar{U}_{1,k}-\bar{U}_{2,k}\right|\le t_{\alpha/2,n_1-1}\frac{s_{1,k}}{\sqrt{n_1}}+t_{\alpha/2,n_2-1}\frac{s_{2,k}}{\sqrt{n_2}}\end{cases}\qquad(11)$$

2.3 The overall similarity factor

Combining PLS similarity factor (6) and the mean shift factor for external variables (11), we have the overall similarity factor as follows:

$$\phi_{1,2}=\theta_{1,2}-\lambda\eta_{1,2}\le 1\qquad(12)$$

where $\phi_{1,2}$ is overall similarity factor, λ is a discount factor which will regulate the weight of PLS model structure and the change of external variables, and generally $\lambda=1$. The overall similarity factor can be used for offline and online mode identification.

3. SIMILARITY FACTOR BASED MODE IDENTIFICATION

A moving window is used as the basic sampling unit for mode identification. PLS models are built for each window of existing process modes. The comparison between two data sets is done using the overall similarity factor (12). An adjustable parameter $\phi_{\lim}$ is introduced as the mode-switching threshold. If the similarity factor is larger than $\phi_{\lim}$, the two windows of data sets are regarded as being in a same mode. Otherwise, they belong to different operation modes. The selection of $\phi_{\lim}$ is application specific and requires process knowledge. Generally, the larger the threshold $\phi_{\lim}$ is, there are more process modes.

Suppose the process mode of the last window is the current mode. Firstly, the next window will be compared with the current process mode to determine 1) it belongs to the current mode; or 2) it doesn't belong to the current mode. For the former situation, the data points in current window will be merged into those of the current mode. For the latter situation, the similarity factor of the current window and all the previous modes should be computed one by one. The maximum similarity factor among them is found and compared with $\phi_{\lim}$. If the result is larger than $\phi_{\lim}$, the current window belongs to that existing mode with the largest similarity factor, so the data points in this window will be merged into the corresponding mode. On the other hand, if it is less than $\phi_{\lim}$, the current window belongs to a new mode which should be constructed.

The method can be applied both offline and online. For offline mode identification, the information from all the existing data can help to identify the mode correctly. Moreover, to amplify the effect of the change of the external variables, it is better to normalize all the existing data instead of only part of it. Therefore, for offline applications, all the data are scaled, whereas for online applications all the existing data are standardized.

Assume the window width is h, the current window is the i^{th} window, the current mode is the j^{th} mode. The detailed online mode identification approach for the i^{th} window is shown as below and in Fig.1.

Step 1: Scale all the current data X and Y.

Step 2: The data set $X_i\in R^{h\times p},Y_i\in R^{h\times q}$ in i^{th} window are used to construct a PLS model, and its weight W_i and component number a_i are derived.

Step 3: Assume the weight of the PLS of current mode is $W_{m,j}$. Compute the similarity factor $\phi_{i,j}$ of the data sets in i^{th} window and j^{th} mode according to (12). If $\phi_{i,j}>\phi_{\lim}$, the i^{th} window belongs to the current j^{th} mode. The data sets in i^{th} window are merged with the j^{th} mode and a new PLS is built on basis of the updated data set in j^{th} mode. Thus, the updated weight $W_{m,j}$ is retrieved, then go to Step 5. If $\phi_{i,j}\le\phi_{\lim}$, the i^{th} window belongs to a different mode.

Step 4: Compare the i^{th} window and each previous operation mode, i.e., the 1^{st}, 2^{nd},…,$(j\text{-}1)^{th}$ modes. Denote their similarity factors as $\phi_{i,1},\phi_{i,2},\ldots,\phi_{i,j-1}$. If there exists some $\phi_{i,k} > \phi_{\lim}(k=1,2,\cdots,j-1)$, choose the maximum among them and denote it as $\phi_{i,km}$ and the corresponding mode km is the mode for the i^{th} window. The data in i^{th} window are added to the km^{th} mode and a new PLS is built on basis of the updated data set in km^{th} mode. Thus, the updated weight $W_{m,km}$ is assigned as the current mode. If $\phi_{i,k} \le \phi_{\lim}(k=1,2,\cdots,j-1)$, the i^{th} window belongs to a new mode, i.e., the $(j+1)^{th}$ mode.

Step 5: Wait to process the $(i+1)^{th}$ window.

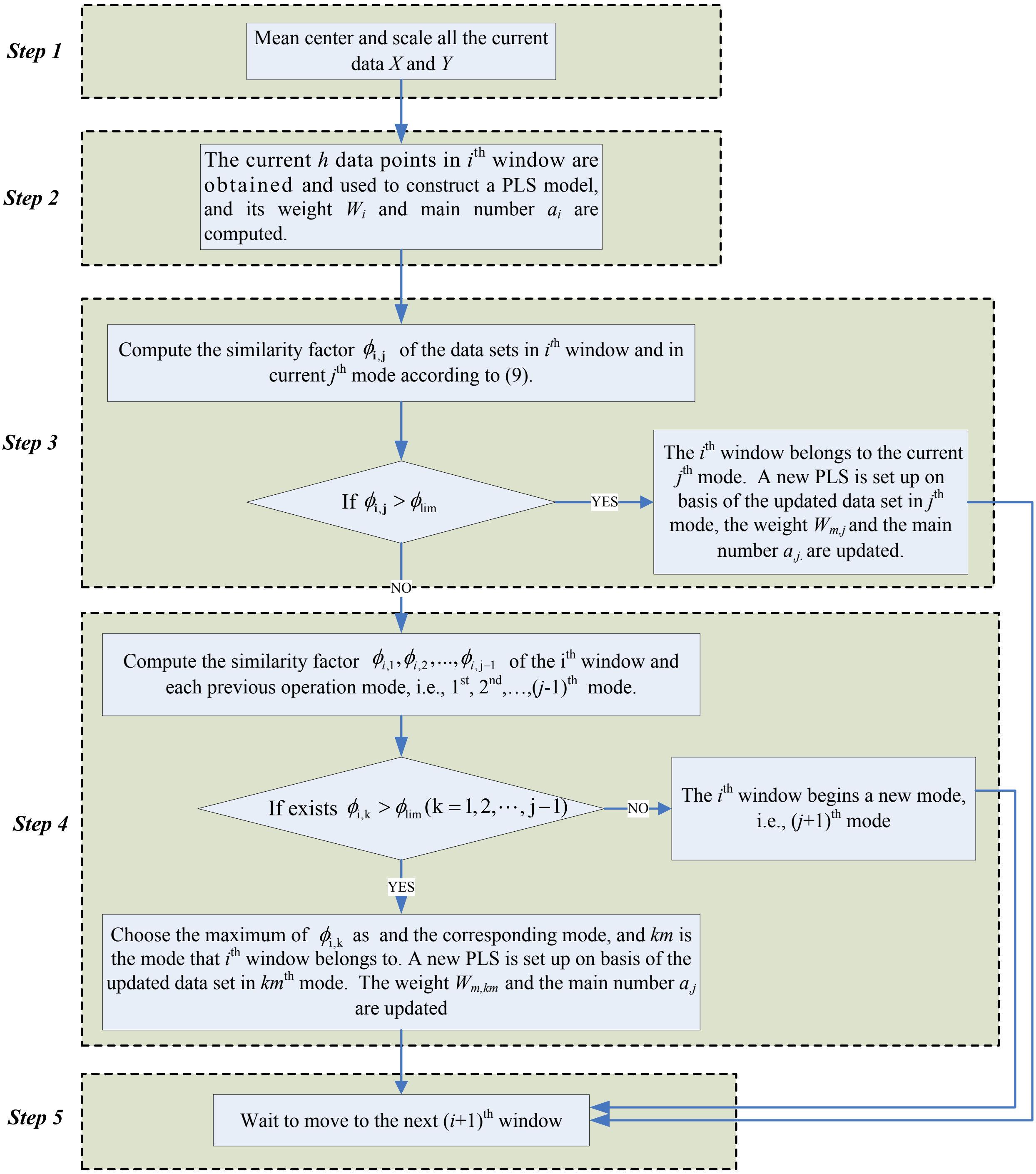

Fig.1 Online mode identification for the i^{th} data window

4. CASE STUDY OF TENNESSEE EASTMAN PROCESS

The Tennessee Eastman process is adoped to test process control and monitoring methods in process systems engineering research. As shown in Figure 2, there are five unit operations: a reactor, a condenser, a flash separator, a reboiler striper, and a recycle compressor. The process contains 12 manipulated variables and 41 measured variables. Ricker developed a decentralized control system for the Tennessee Eastman process. Downs and Vogel presented six modes of process operation, as listed in Table 1. In our study, Modes 1 and 3 are used, the simulation program is download online. The sample time is 0.01h, and 22 continuous measurements XMEAS(1-22) are chosen as the process inputs and XMEAS(37-41) are chosen as the output.

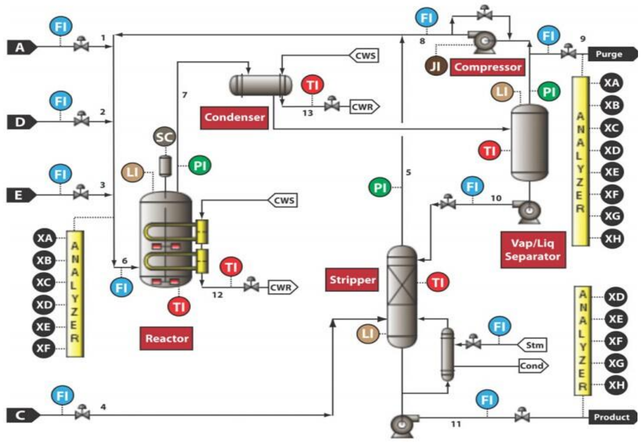

Fig.2: Schematic diagram of the Tennessee Eastman process

Table 1 Six operation modes in TE example

Mode	G/H Mass Ratio	Production Rat(Stream 11)
1	50/50	7038 kg/h G and 7038 kg/h H
2	10/90	1408 kg/h G and 1408 kg/h H
3	90/10	10000 kg/h G and 10000 kg/hH
4	50/50	Maximum
5	10/90	Maximum
6	90/10	Maximum

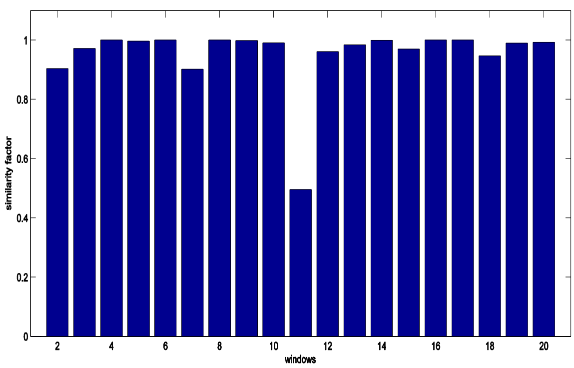

Fig.3 The similarity factor of each window and the current operation mode

The process was initially running at mode 1 for 5h and then switched to mode 3 for another 5h. The window width is 50. So there are 500 samples for Mode 1 and Mode 3 respectively. Among the process input, XMEAS(1-4) are the feed, so that they are regarded as external variables. Set $\lambda = 1$ and $\phi_{\lim} = 0.8$, the similarity factor of each window and the current operation mode is shown in Figure 4. It is obvious that the similarity factor of Windows 2-10 is above 0.9. When it comes to Window 11, the similarity factor decreases rapidly to 0.4943, which means the data in Window 11 are different from the current operation mode. Thus after window 11, the process is operated in a new mode. It can be observed the similarity factor of window 12-20 and the current new mode is returned to above 0.9, which means the window 12-20 also belongs to the new process mode. Therefore, the mode identification result is shown in Fig.4. As can be seen in Fig.4, the proposed method identifies Windows 1-10 and Windows 11-20 belong to two different modes, which indicates the proposed mode identification method effectively identifies the right mode.

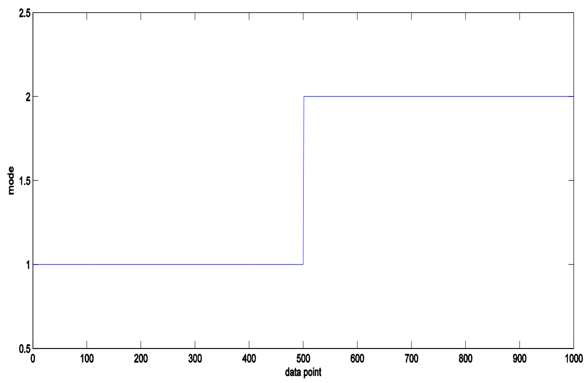

Fig.4 The mode identification result

5. CONCLUSION

Industrial processes are generally operated on multiple modes due to production rate, product grade and disturbance changes. The traditional SPM tools such as PCA and PLS are ineffective for monitoring such multimode processes. In this paper, by combing the similarity factor of PLS models and the mean shift of the external variables to form an overall similarity factor to measure the similarity, a new method is proposed with improved results. On basis of this proposed similarity factor, moving windows are used to identify the mode for multimode process monitoring. The proposed approach is demonstrated effectively on the benchmark Tennessee Eastman process.

ACKNOWLEDGEMENT

The authors are grateful for the financial support for this work from National Natural Science Foundation of China (61374139 and 61490700).

REFERENCES

Chen, T., and Zhang, J. (2010). On-line multivariate statistical monitoring of batch processes using Gaussian mixture model. *Comput Chem Eng.*, 34, 500–507.

Choi, S.W., Martin, E. B., Morris, A.J., and Lee, I. B. (2006). Adaptive multivariate statistical process control for monitoring time-varying processes. *Industrial and Engineering Chemistry Research*, 45, 3108–3118.

Downs, J.J., and Vogel, E.F. (1993). A plant-wide industrial process control problem. *Comput. Chem. Engng.*, 17, 245-255.

Feital, T., Kruger, U., Dutra, J., et al. (2013). Modeling and performance monitoring of multivariate multimodal processes. *AIChE Journal*, 59(5), 1557-1569.

Ge, Z.Q., Yang, C.J., Song, Z.H., and Wang, H.Q. (2008). Robust online monitoring for multimode processes based on nonlinear external analysis. *Industrial and Engineering Chemistry Research*, 47, 4775–4783.

Ge, Z.Q., and Song, Z. H. (2010). Mixture bayesian regularization method of PPCA for multimode process monitoring. *AICHE Journal*, 56, 2838-2849.

Ge, Z.Q., Song Z.H, and Kano M. (2014). External analysis based regression model for robust soft sensing of multimode chemical processes. *AICHE Journal*, 60, 136-147.

Haghani, A., Jeinsch, T., and Ding, S.X. (2014). Quality Related Fault Detection in Industrial Multimode Dynamic Processes. *IEEE Transactions on Industrial Electronics*, 61(11), 6446- 6453.

Kano, M., Hasebe, S., Hashimoto, I., and Ohno, H. (2004). Evolution of multivariate statistical process control: application of independent component analysis and external analysis. *Comput Chem Eng.*, 28, 1157–1166.

Lee, Y.H., Jin, H.D., and Han, C. H. (2006). On-line process state classification for adaptive monitoring. *Industrial and Engineering Chemistry Research*, 45, 3095–3107.

Ma, Y., and Shi, H. (2014). Multimode Process Monitoring Based on Aligned Mixture Factor Analysis. *Industrial and Engineering Chemistry Research*, 53, 786−799.

McClure, K. S., Gopaluni R. B., Chmelyk, T, Marshman D., Shah, S.L. (2014). Nonlinear Process Monitoring Using Supervised Locally Linear Embedding Projection, *Industrial and Engineering Chemistry Research*, 53(13), 5205–5216

Montgomery, D.C. (2013). *Introduction to Statistical Quality. 7th Edition*, Chap 4: Inferences About Process Quality, John Wiley & Sons, Inc., USA.

Ng, Y.S., and Srinivasan, R. (2009). An adjoined multimodel approach for monitoring batch and transient operations. *Comput Chem Eng.*, 33, 887–902.

Ricker, N.L. (1996). Decentralized control of the Tennessee Eastman challenge process. *Journal of Process Control*, 6, 205-221.

Singhal, A., Seborg, D. E. (2006). Evaluation of a pattern matching method for the Tennessee Eastman challenge process. *Journal of Process Control*, 16, 601–613.

Tan, S., Wang, F., Peng, J., et al. (2012). Multimode Process Monitoring Based on Mode Identification, *Industrial and Engineering Chemistry Research*, 51, 374–388.

Tong, C.D., Palazoglu, A., and Yan, X. F. (2013). An adaptive multimode process monitoring strategy based on mode clustering and mode unfolding. *Journal of Process Control*, 23, 1497-1507.

Yu, J. and Qin, S. J. (2008). Multimode process monitoring with bayesian inference-based finite gaussian mixture models. *AICHE Journal*, 54, 1811-1829.

Yu J.B. (2010). Hidden markov models combining local and global information for nonlinear and multimodal process monitoring. *Journal of Process Control*, 20, 344-359.

Zhang, Y.W, Wang, C., and Lu, R. (2013). Modeling and monitoring of multimode process based on subspace separation. *Chemical Engineering Research and Design*, 91, 831-842.

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