A Monitoring Method Based on Modified Dynamic Factor Analysis and Its Application*

Xueyan Yin and Fei Liu

Institute of Automation, Jiangnan University, 214122, wuxi, China

Abstract. Aimed at the dynamic characteristic of the process, the auto-correlation and the cross-correlation among process variables, a monitoring method based on Dynamic Factor Analysis was investigated which was widely concerned in academia and industry. However, the traditional Dynamic Factor Analysis utilizes the same time lags currently among all variables, but the computing work is complex and the auto-correlation isn't adequately exemplified. The paper proposes a new monitoring method based on modified dynamic factor analysis that determines different time lags according to different variables by correlation analysis, and constructs several statistics as the monitoring indices to monitor the operation condition of the dynamic process. The improved monitoring method is applied in the Tennessee-Eastman process and compared with DFA, the validity and superiority are proved.

Keywords: Modified Dynamic Factor Analysis, correlation analysis, time-lags, EM algorithm, dynamic process monitoring.

1 Introduction

With the promotion of technology, the size of the industrial system expands, the manufacture technique is getting more complex, and the improvement and the innovation of process monitoring technology are increasingly focused on. However, industrial process is so complex that it is very difficult to obtain a mechanism model. On the other hand, the capability of data collection, storage and analysis is also upgraded significantly, which makes Multivariate Statistical Process Control (MSPC) methods based on the data-driven technique be concerned [1-3].

The Principal Component Analysis method is a classical method of dimension-reduction and features extraction, which has been widely applied in process monitoring field [2-4]. However, the current sample has not only static correlations among variables but time-serial correlations to its historical samples in the actual process. In order to solve this problem, some scholars have proposed Dynamic PCA method that utilizes auto-regressive model to extend the sample matrix [5]. However, PCA contains many constraints and restrictions, such as the minimum variance and isotropic of noise, etc, so it is a special form of Factor analysis method which has been widely used in many fields such as economics and sociology, and introduced into process

* Supported by basis study program of Jiangsu province (BK2009068) and six talent summit program of Jiangsu province.

K. Li et al. (Eds.): LSMS/ICSEE 2010, Part II, LNCS 6329, pp. 442–449, 2010.

monitoring field recently [6].This method overcomes the limitation of the traditional MSPC method and can relatively reflect the essential characteristics of data, so the monitoring results is more ideal. Reference [7] proposes a new process monitoring method based on Dynamic Factor Analysis which combines the idea of Dynamically PCA with factor analysis method, and utilizes the traditional data extension method, without a detailed analysis for autocorrelation between industrial process variables. However, there are different correlations between variables. So the same time-lags length sometimes can't reflect the dynamic relationship but may increase calculation error because of introducing redundant dynamic relationship.

This paper proposes an Modified Dynamic Factor Analysis method to overcomes the shortage of DPCA and DFA by eliminating unnecessary dynamic relationship and expanding the time-lag of variables with strong auto-correlation [8]. The improved monitoring method is applied in the Tennessee-Eastman process and compared with the traditional DFA, the validity and superiority are proved.

2 Modified Dynamic Factor Analysis Model and Process Monitoring Method

2.1 Factor Analysis

The Factor Analysis model can be built as :

$$X = AF + e \tag{1}$$

Let $X = [x_1^T, x_2^T, ..., x_n^T] \in R^{m\times n}$ be the normalized process data matrix, m is the number of variables, n is the number of samples and $x \sim N(0,\Sigma)$. Factor loading matrix is $A \in R^{m\times k}$ and k is the number of factors. The factor is $F \in R^{k\times n}$, and $F \sim N(0,I)$. The noise is $e \sim N(0,\phi)$ and ϕ is a random diagonal matrix.

The joint probability density function of the factor is:

$$p(f) = (2\pi)^{-k/2} \exp\{-\frac{1}{2} f^T f\} \tag{2}$$

The conditional probability density function of the data is :

$$p(x \mid f) = (2\pi)^{-2/m} \left|\varphi\right|^{-1/2} \exp\left\{-\frac{1}{2}(x - Af)^T \varphi^{-1}(x - Af)\right\} \tag{3}$$

The probability density function of data is derived:

$$\begin{aligned} p(x) = \int p(x \mid f)p(f)df =&(2\pi)^{-m/2}\left|\varphi + AA^T\right|^{-1/2} \times \\ \exp\{-\frac{1}{2}x^T(\varphi + AA^T)^{-1}x\}& \end{aligned} \tag{4}$$

Factor posteriori probability density function is derived by the Bayesian theorem:

$$p(f \mid x) = (2\pi)^{-k/2}\left|M\right|^{1/2}\exp\{-\frac{1}{2}(f-\beta x)^T(M)(f-\beta x)\}$$
$$\beta = A^T(\varphi + AA^T)^{-1} \quad M^{-1} = I - A^T(\varphi + AA^T)^{-1}A \tag{5}$$

The generation model is $x \sim \mathrm{N}(0, AA^T + \phi)$, including parameters A and ϕ .The EM algorithm for parameters A and ϕ are given by

$$A^{new} = (\sum_{i=1}^{n} x_i E(f_i \mid x_i)^T)(\sum_{i=1}^{n} E(f_i f_i^T \mid x_i))^{-1} \tag{6}$$

$$\phi^{new} = (1/n)diag(\sum_{i=1}^{n} x_i x_i^T - A^{new}E(f_i \mid x_i)x_i) \tag{7}$$

The model will be established by iteratively calculating (6) and (7) until convergence [9 – 10].

2.2 Modified Dynamic Factor Analysis Model

In normal conditions, process sample data is matrix X which is in formula (1). Current Dynamic Factor Analysis expands the variables to $[x_i, x_{i-1}, \ldots\ldots x_{i-d}]$ with the time-lag d at the i-th sample time, based on auto-regression method .

Reference [11] uses iterative algorithm as follows to determine the time-lag. Firstly, suppose d=0, without considering the dynamic conditions, the number of factors is still k and the number of static relationship is n-k; Secondly, suppose d=1, Calculate again and the number of the dynamic relationship becomes n-k-r. Lastly, the time-lag d is increased gradually, and r is calculated according to the formula (1). The time-lag will be d until r≤0, and the number of samples reduce from m to m-d, the number of variables becomes n×(d+1) .

$$r_{new}(d) = r_r(d) - \sum_{i=0}^{d-1}(d-i+1)r_{new}(i) \tag{8}$$

This paper has improved the traditional method which determines the time-lag, and overcame the deficiency that all variables take the same time-lag. This paper adopts correlation analysis method to calculate the related coefficient $\rho_i(\tau)$ of variables including variable i and time-lag τ and sets up a reasonable threshold ρ .

① Let $\tau = 0$, if $\rho_j(\tau) \le \rho$,then the i-th variables should be fixed ;

② If a variable does not satisfy the conditions ①,then calculate $\rho_j(\tau)$ again ,if $\rho_j(\tau) \le \rho$,then the delay of this variable is 1;

③ Calculate until all variables have the appropriate time-lag. After the expansion, process data sample matrix is shown as follows:

$$\widetilde{X} = [x_1^T(t-d_1), x_1^T(t-d_2), \ldots\ldots x_1^T(t-d_i), \ldots\ldots$$
$$x_n^T(t-d_1), x_n^T(t-d_2), \ldots\ldots, x_n^T(t-d_j)]^T \in R^{(m-D)\times N} \tag{9}$$

D is the maximum of $(d_i,\ldots,d_j)$, that is the longest time-lag, and N is the total number of variables. $\widetilde{X}$ will be instead of X in formula (1) and the model is established.

2.3 Process Monitoring Method Based on MDFA

In the factor analysis model, the factor is assumed to source signal and the sample is formed by source signal .So the number of factors is very important. There are many methods to select factor number and this paper uses the method based on Kaiser rule [11].

Factor is a comprehensive index which can be extracted from the process. Factor f_i is instead by the estimate formula $E(f_i \mid x_i) = \beta x_i$ because the f_i is difficult to compute and the monitoring indicators of factor is established.

$$GT^2 = \left\| \hat{f}_i \right\|^2 = x_i^T \beta^T \beta x_i \le \chi^2_{1-\alpha}(k) \tag{10}$$

Noise reflects degree of fitting between the process variables and the model. According to the formula (1), we have:

$$x = AE(f \mid x) + E(e \mid x) \tag{11}$$

$E(f \mid x)$ is the estimate $\hat{f}$ of factor, and $E(e \mid x)$ represents the estimate $\hat{e}$ of noise。 we have:

$$x = AA^T(AA^T + \varphi)^{-1}x + E(e \mid x) \tag{12}$$

Formula (11) is equivalent to:

$$\hat{e} = E(e \mid x) = (I - AA^T(AA^T + \varphi)^{-1})x \tag{13}$$

Then the monitoring index of noise space is established.

$$GSPE = \left\| \phi^{-0.5} \hat{e} \right\|^2 \le \chi^2_{1-\alpha}(m) \tag{14}$$

In order to reduce the workload of monitoring, and ensure the accurate degree of process monitoring, the integrated monitoring index ST is constructed which considers both the factors and the noise using the Mahalanobis distance measurement [6].

$$ST = \left\| (AA^T + \phi)^{-0.5} x \right\|^2 \le \chi^2_{1-\alpha}(m) \tag{15}$$

The three kinds of monitoring indexes meet the distribution of χ^2, whose confidence bounds are $1-\alpha$ with different freedoms.

3 Simulation

Tennessee-Eastman process is created by Eastman Chemical Company of the USA, which in order to provide a realistic industrial process for evaluating process control and monitoring methods, and its data have time-varying, strong coupling characteristics and so on. The reaction consists of five major units: a reactor, condenser, compressor, separator and stripper; eight components: A, B, C, D, E, F, G and H, where A, C, D, E is the reactant, B as the catalyst, F is a by-product, G, H is the final product. The whole process includes 12 control variables, 41 process variables. The data sample once per 3 minutes for each simulation time of 48 hours. Through 22 different operating conditions of the simulation constitutes-including normal and 21 different fault conditions- all the faults are introduced in the 160th sample points. The detailed description of the process and the introduction of various forms of faults is introduced in Reference [3], the flow sheet is shown in Figure 1:

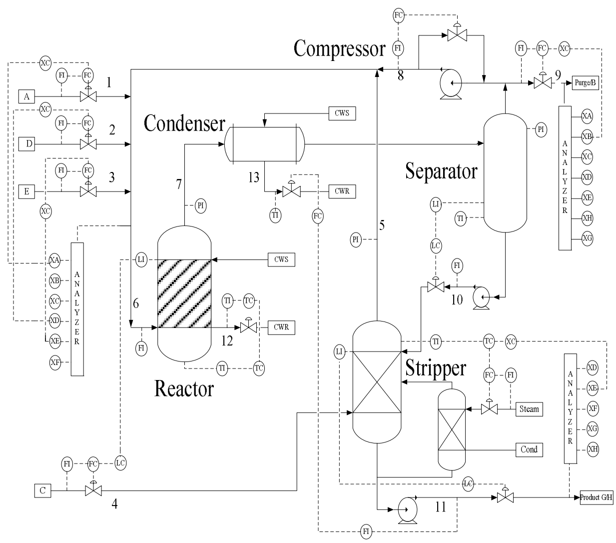

Fig. 1. Process flow sheet for the Tennessee Eastman Process

This paper compares the monitoring method based on MDFA with DFA taking fault 10 for example.

The traditional DFA model is established in which the time-lag is 2, the extended data matrix is $R^{958\times156}$and the number of factors is 45.The control chart is shown in Reference [7].

At last, the monitoring method based on MDFA is utilized in this article. The time-lags of the variables range from 0 to 6 based on the analysis results of different variables and the expanded data matrix R958×117. Compared with the traditional dynamic methods, a major advantage of the improved method is to expand the number of lower-dimensional matrix and effectively reduce the computational. The option for the number of factors is 29, building the DFA model, the control chart is shown in Figure 2:

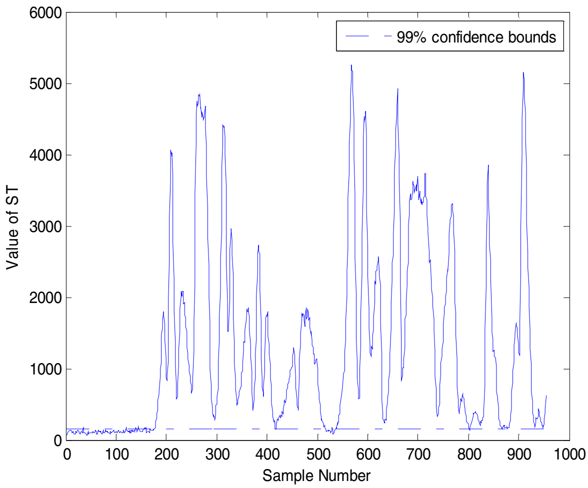

Fig. 2(a). Data monitoring chart based on MDFA

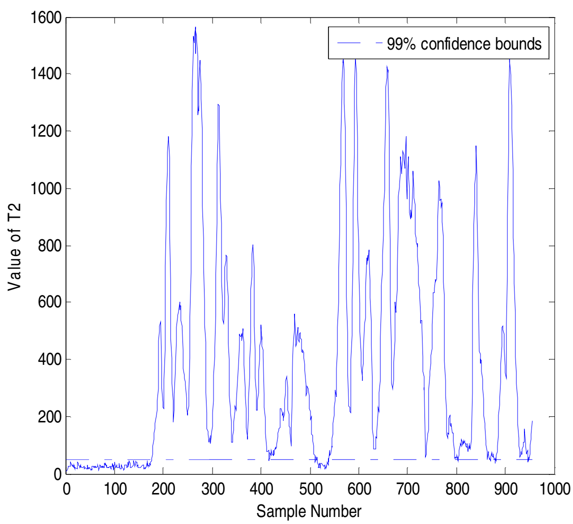

Fig. 2(b). Factor monitoring chart based on MDFA

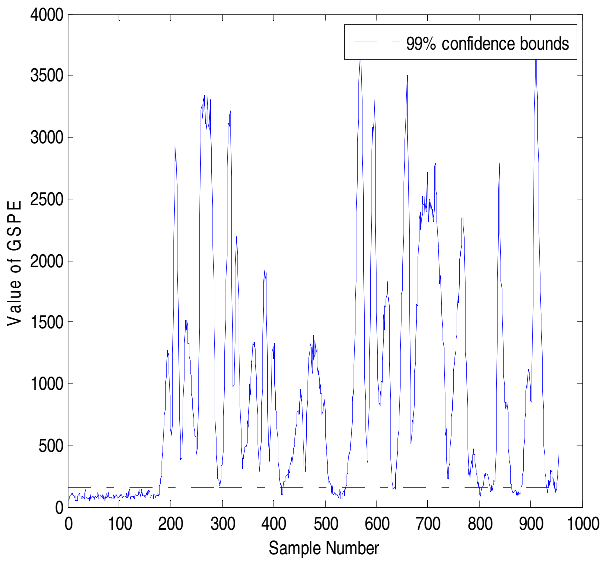

Fig. 2(c). Noise monitoring chart based on MDFA

The two methods of missing detection rates and detection delays are shown in the following table 2:

Table 1. (a) Missing detection rates

Method	ST	GT^2	GSPE
DFA	0.054	0.227	0.226
MDFA	0.067	0.062	0.102

Table 2.(b) Detection delays

Method	ST	GT^2	GSPE
DFA	60	102	66
MDFA	54	89	57

By comparing the two different methods of control chart, we can find both of the two monitoring methods could detect the faults, but the method based on MDFA is better than DFA, because when the faults occurs, index value out of control is even more obvious. The advantage will be more outstanding in the case of the faults are not so obvious. Also the MDFA is more advantage than the DFA in the index values of the missing detecting rates and detection delays.

4 Conclusion

This paper presents a new method to overcome the shortage of the traditional MSPC by determining different time lags according to different variables through correlation analysis, which avoid taking unnecessary dynamic relationship, and also ensure the embodiment of strong correlation. Combine this method with FA and DFA proposed in reference [7], the validity and superiority are proved by the simulation results of TE process, because the modified method succeed to reduce the dimension of data, the missing detection rates and the detection delays.

References

1. Zhao, Z.G.: Several issues of statistical process monitoring based on PCA.: School of Communication and Control Engineering. Jiangnan University, Wuxi (2007)
2. Changl, H., Russell, E.L., Braatzr, D.: Fault detection and diagnosis in industrial systems. Springer, London (2001)
3. Liang, J., Qian, J.X.: Multivariate Statistical Process Monitoring and Control: Recent Development and Applications to Chemical Industry. Chem. Eng. 11(2), 191–203 (2003)
4. Zhang, H., Xu, W.B., Liu, F.: The Multivariate Statistical Theory Based on PCA Analysis is Applied in Process Monitoring. Control and Automation 22(3-1), 74–76 (2006)
5. Liu, Y.M., Liang, J., Qian, J.X.: Dynamic Principal Component Analysis Method For Chunk Monitoring of Industrial Fluidized-bed Reactor. Journal of Chemical Industry and Engineering (China) 55(9), 1546–1549 (2004)
6. Zhao, Z.G., Liu, F.: Factor analysis and its application in process monitorin. Journal of Chemical Industry and Engineering (China) 58(4), 970–974 (2007)
7. Zhao, Z.G., Liu, F.: Modeling using dynamic factor analysis and its application in process monitoring. Journal of Chemical Industry and Engineering (China) 60(1), 183–186 (2009)
8. Ku, W., Storer, R.H., Georgakis, C.: Disturbance detection and isolation by dynamic principal component analysis. Chemometrics and Intelligent Laboratory Systems 30(1), 179–196 (1995)
9. Rubin, D., Thayer, D.: EM algorithms for ML factor analysis. Psychometrika 47(1), 69–76 (1982)
10. Ghahramani, Z.B., Hinton, G.E.: The EM Algorithm for Mixtures of Factor Analyzers. Technical Report RG-TR-96-1, Dept of Comp. Sci., Uiv. of Toronto (1996)
11. Shen, Q., Liu, Y.M., Liang, J.: Fault detection and diagnosis based on DPCA. Manufacturing Automation 27(6), 51–54 (2005)
12. Yin, X.Y., Liu, F.: A Improved Monitoring Method Based On Factor Analysis And Its Application In Chemical Separation Process. In: China Process System Engineering and MES Annual Meeting Memoir, pp. 131–134 (2009)