Improvement on Multivariate Statistical Process Monitoring Using Multi-scale ICA

Fei Liu and Chang-Ying Wu

Institute of Automation, Southern Yangtze University,
Wuxi, 214122, P.R. China
fliu@sytu.edu.cn

Abstract. A multi-scale independent component analysis (ICA) approach is investigated for industrial process monitoring. By integrating the ability of wavelet on multi-scale analysis and that of ICA on extracting independent components for non-Gaussian process variables, the multivariate statistical monitoring techniques can obtain improved performance. Contrastive tests have been carried out on the famous benchmark chemical plant among ICA-like and PCA-like methods, which reveals that multi-scale ICA approach has lower missed detection rate of faults.

1 Introduction

Modern chemical processes, which are equipped with instrument and data collector, contain thousands of measured variables, such as temperatures, pressures and flow rates. The correlation among the process variables exists as a result of either association or causation. Instead of univariable statistical process control (SPC), it is necessary and possible to apply multivariate statistical process control (MSPC) to extract relevant information in the redundant process data, and to detect if statistically significant abnormalities occur. As a methodology, MSPC monitors whether the process is in control through the analysis of the various control charts, such as T2 and SPE. While numerous procedures of univariable SPC are available in manufacturing processes and are likely to be part of a basic industrial training program, MSPC procedures are being used to monitor chemical processes that are inherently multivariate [1].

Basically, as a mathematical tool, principal component analysis (PCA) can essentially identifies important characteristics in multivariate redundant data and has successfully been applied to performance monitoring and fault diagnosis for industrial process [2], [3]. PCA makes variables de-correlated by means of maximizing the variance within the process data, which follows a Gaussian probability distribution or independent identical distribution. Unfortunately, the process variables and its statistical information are very complex in actual industrial production, it is difficult to make certain about process variable's probability distribution [4].

As a blind source separation technique, independent component analysis (ICA) has been founded wide applications in processing of medical signals, compressing of images [5], and machine fault detection [6], etc. Compared with PCA,

J. Rosca et al. (Eds.): ICA 2006, LNCS 3889, pp. 376–383, 2006.

ICA represents a set of random variables as linear combination of statistically independent component variables, and is less sensitive to process variable's probability. Benefited from this, recently, ICA has been introduced to process monitoring [7], [8], which can be more efficacious in a non-Gaussian context. In the real-world stochastic processes, the energy or power spectrum of variables often changes with time or frequency, as a result, almost all the industrial processes are multi-scale in nature. Accordingly, while the existing ICA process monitoring has been adopted at a single scale, it may be more significant to improve the process monitoring by means of multi-scale ICA. In fact, the concept of multi-scale analysis already exists in the project of image processing [9], multi-scale PCA monitoring [10] and blind source separation [11].

In this paper, a multi-scale independent component analysis (MSICA) approach is investigated for process monitoring, which integrates the ability of wavelet on multi-scale analysis and that of ICA on extracting independent components for non-Gaussian variables. While the process is in control, the genuine model is decided by the reconstructed signals from the selected wavelet coefficients, which violate the threshold of the ICA model at the significant scales. By means of the presented MSICA, the statistical monitoring method is discussed on the famous benchmark plant, Tennessee-Eastman (TE) chemical process. Compared with traditional MSPC methods (including existing PCA and MSPCA), MSICA reveals lower missed detection rate of faults.

2 Multi-scale ICA Monitoring

2.1 Wavelet-Based MSICA Model

For an original data set $X = (x(1), x(2), \cdots, x(m)) \in R^{n\times m}$ with m measure variables and n samples, the standardization is firstly performed to make measure variables have zero mean and unit variance. By applying l steps wavelet decomposes to every measure variable $x(i)$, $i = 1, 2, \cdots, m$, there are l detail coefficient vectors $a_{k,i}$ on scale $k = 1, 2, \cdots, l$, and an approximate coefficient vector $b_{l,i}$. Assuming p is the sample length of $a_{k,i}$ and $b_{l,i}$, on the k^{th} scale, there exist a detail coefficient matrix $A_k = (a_{k,1}, a_{k,2}, \cdots, a_{k,m})^T \in R^{m\times p}$ and an approximate coefficient matrix $B_l = (b_{l,1}, b_{l,2}, \cdots, b_{l,m})^T \in R^{m\times p}$. Moreover, let $A_k = C\widetilde{S}$ or $\widetilde{S} = FA_k$, where $\widetilde{S} \in R^{p\times m}$ is source signal matrix, $C \in R^{p\times p}$ is a nonsingular constant matrix, and $F = C^{-1}$. In general, A_k may be whitened by the transformation QA_k, where $Q = \Lambda^{-1/2}U^T$, and Λ and U come from the eigen-decomposition of the covariance, i.e. $E(A_kA_k^T) = U\Lambda U^T$. To separate statistical independent source signals from A_k, the notion of entropy is used, which is a measure of uncertainty of a continuous stochastic variable. Under the condition of the same variance, the smaller the differential entropy of a random variable is, the greater non-Gaussianity it has. It is known that the Gaussian random variable has maximal differential entropy and non-Gaussian implies indepedence [12]. This gives a way to judge the independent of stochastic variable by comparing its differential entropy.

Consider a continuous stochastic variable y with probability density function $f(y)$, its differential entropy is given by $H(y) = -\int f(y)logf(y)dy$. To avoid estimating the probability density function $f(y)$, the negentropy of $H(y)$ is defined as $J(y) = H(y_{Gauss}) - H(y)$, where $H(y_{Gauss})$ is a Gaussian stochastic variable with the same variance as y. As a fast algorithm,$J(y) \approx [E\{G(y)\} - E\{G(y_{Gauss})\}]^2$ is adopted here instead of negentropy definition [13], where $G(\cdot)$ is a non-quadratic function.

For $\widetilde{S} = FA_k$, single source signal can be expressed as $\widetilde{s}(i) = f \cdot A_k$, where f is a row vector in matrix F. In the setting of entropy, the modeling of independent component is translated to following optimization problem: find an optimal f satisfying $E\{fA_k \cdot (fA_k)^T\} = \|f\|^2 = 1$ to make $J(f \cdot A_k)$ maximum. Keep in mind negentropy definition, the optimization problem is equivalent to following:

$$\min_f E\{G(f \cdot A_k)\}$$
$$s.t.E\{fA_k \cdot (fA_k)^T\} = \|f\|^2 = 1 \tag{1}$$

Based on the Kuhn-Tucker conditions, that $E\{G(f \cdot A_k)\}$ is optimal while $E\{A_k g(f \cdot A_k)\} - \beta f = 0$ where $\beta = E\{f_0 A_k g(f_0 \cdot A_k)\}$ with optimum f_0, Newton's method is borrowed to solve above optimization problem (1), and here is omitted. Obviously, all the row vectors of F can be estimated in the same way. And then constant matrix C and source signal matrix $\widetilde{S}$ can be computed via $\widetilde{S} = FA_k$.

Retaining d independent source signals, $\overline{A}_k = Q^{-1}F_d\widetilde{S}_d$ is the reconstruction of detail coefficients matrix A_k, where $F_d \in R^{m\times d}$, $\widetilde{S}_d \in R^{d\times p}$ are the corresponding retained de-mixing matrix and source signal matrix, respectively [8]. Simultaneously, a threshold is set based on the median of serial signal $\overline{a}_{k,i}$ for removing residual of signals and acquiring information of significant events [14]. Similarly, B_l is reconstructed as A_k is done.

Applying reverse wavelet transformation on the retained detail coefficient at every scale and approximate coefficient at the coarsest scale [9], the reconstructed process data $Y \in R^{n\times m}$ is obtained and then modeled in ICA form, $Y^T = TS + E$, where $S = (s(1), s(2), \cdots, s(d))^T \in R^{d\times n}$ is independent components matrix of process, $T \in R^{m\times d}$ is coefficient matrix, and $E \in R^{m\times n}$ is residual. For mathematical convenience, assume $E = 0$, $d = m$, then $S = T^{-1} \cdot Y^T = W \cdot Y^T$. Based on above multi-scale ICA model, in next subsection, the process monitoring technique is investigated.

2.2 Statistical Monitoring

Two universal statistics, I^2 or T^2 (I^2 used in ICA-based monitoring method, while T^2 in PCA-based method) and Q (i.e. Square Prediction Error, SPE), have been employed in real-time process monitoring. At the t^{th} sample, $I^2(t) = S_d^T(t)S_d(t)$, where $S_d(t)$ is the t^{th} column vector of the projection of the original data in the directions of independent components, and $SPE = e^T(t) \cdot e(t)$, where $e(t) = x(t) - \widehat{x}(t)$ is the residual between sample $x(t)$ and prediction of model $\widehat{x}(t)$.

Because independent components may not follow normal distribution, the control limit of statistic can not be decided by a special probability function, as done in PCA. Instead, kernel density estimation is introduced, in which the univariate kernel estimator is defined as $f(x) = \frac{1}{nh}\sum_{i=1}^{n} K\{(x - x_i)/h\}$, where $K(\cdot)$ is the kernel function, x is the data point under consideration, x_i is the sample data, $i = 1, 2, \cdots, n$, and h is the smoothing parameter. In practice, Gaussian function is usually chosen as the kernel function. To get the point z, which occupies the 99% area of density function $f(x)$, the following equation is used,

$$\int_{-\infty}^{z} f(x)dx = \int_{-\infty}^{z} \frac{1}{nh}\sum_{i=1}^{n} K\{(x - x_i)/h\}dx$$
$$= \int_{-\infty}^{z} \frac{1}{nh}\sum_{i=1}^{n}\{exp[(x - x_i)^2/(2h^2)]/\sqrt{2\pi}\}dx = 0.99 \quad (2)$$

The detailed selection of h may be found in [15]. The control limits of process normal operation are then easily obtained.

3 Cases Study

3.1 Missed Detection Rate Comparison

The TE process is a realistic industrial plant for evaluating process control and monitoring methods, which consists of five major units: reactor, condenser, compressor, separator, stripper. It contains eight components: reactants A, C, D, E, inert B fed to the reactor, products G, H and by-product F formed in the reactor. The process contains 41 measured variables and 11 manipulated variables, which are sampled per 3 minutes. All the process measurements involved faults are introduced at sample 301. The reader may refer to [16] for details.

For some typical statistical control methods and statistics, the missed detection rates for all 20 faults are shown in Table 1. Because the variables in Fault 3, 9 and 15 have no remarkable mean and standard deviation changing, their missed detection rates are high for all the statistics. It is conjectured that any statistic based on data-driven methods will result in high missed detection rates for these faults [2], thus in Table 1 they are marked by asterisk. Besides the three faults, the minimum missed detection rates of all faults are denoted by bold style. As a whole, it maybe true that MSICA has superiority in process monitoring and detection.

3.2 Monitoring and Detection Tests

Introduce fault 4 to TE process, which is a step change in the reactor cooling water inlet temperature. This leads a step change of the cooling water flow rate as shown in Fig. 1(Left), and a sudden jump of the reactor temperature as shown in Fig. 1(Right). The other 50 measure and manipulated variables retain steady, the variance of the mean and the standard deviation of each variable is inapparent.

Table 1. Missed detection rates for the testing set

Fault	$T^2(PCA)$	Q	$T^2(MSPCA)$	Q	$I^2(MSICA)$	Q
1	0.0076	0.0045	0.0106	0.0061	**0.0030**	**0.0030**
2	0.1515	0.0136	0.0258	0.0152	**0.0061**	**0.0045**
3*	0.9682	0.9394	0.9652	0.9788	**0.7242**	**0.8833**
4	0.7833	**0.0091**	0.9061	0.2485	**0.0515**	0.0318
5	0.7015	0.6485	0.7242	0.6227	**0**	**0**
6	0.0106	**0**	**0**	**0**	**0**	**0**
7	0.4364	**0**	**0**	**0**	**0**	**0**
8	**0.0227**	0.0258	0.0545	0.0242	0.0303	**0.0167**
9*	0.9680	0.9409	0.9621	0.9758	**0.8318**	**0.8788**
10	0.5240	0.4273	0.6277	0.4318	**0.1985**	**0.1621**
11	0.6680	0.2440	0.8076	**0.0403**	**0.1682**	0.2289
12	0.0260	0.0227	0.0394	0.0167	**0.0030**	**0.0136**
13	0.0772	0.0651	0.0924	0.6515	**0.0606**	**0.0470**
14	0.1773	0.0015	0.0015	0.0030	**0.0014**	**0.0015**
15*	0.9970	0.9030	0.9545	0.9610	**0.7575**	**0.9000**
16	0.7591	0.6455	0.7515	0.5954	**0.0742**	**0.1379**
17	0.3197	**0.0727**	0.1682	0.1848	**0.0379**	0.1000
18	0.1318	0.0969	0.0985	0.1091	**0.0818**	**0.0758**
19	0.9878	0.6091	0.7864	0.9136	**0.1303**	**0.5424**
20	0.7727	0.4379	0.7848	0.5530	**0.0970**	**0.1970**

All these make the detection and diagnosis of fault 4 more challenging than other faults. The T^2 statistic charts based on PCA, multi-scale PCA (MSPCA) are shown in the Fig. 2. It is obvious that T^2 statistics are not beyond the threshold after sample 301. However, I^2 statistic chart of MSICA gives an exciting result as shown in the Fig. 3. After sample 300, it can be seen that I^2 statistic goes beyond the control limit distinctly, its ability to detect fault 4 is better than that of other methods.

Based on the process monitoring, furthermore, the fault detection and diagnosis is carried out to identify the observation variables most closely related to the

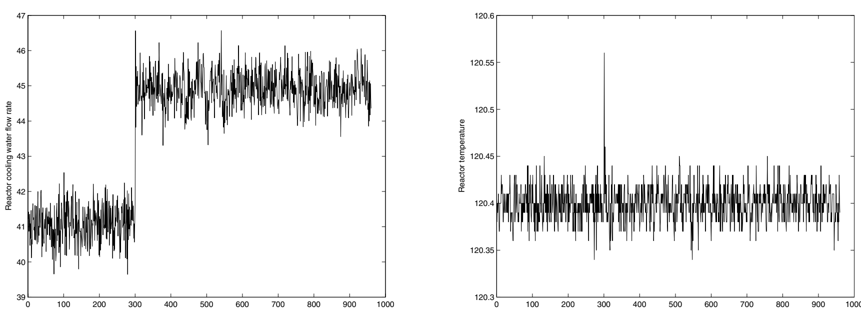

Fig. 1. Cooling water flow rate (Left) and reactor temperature (Right)

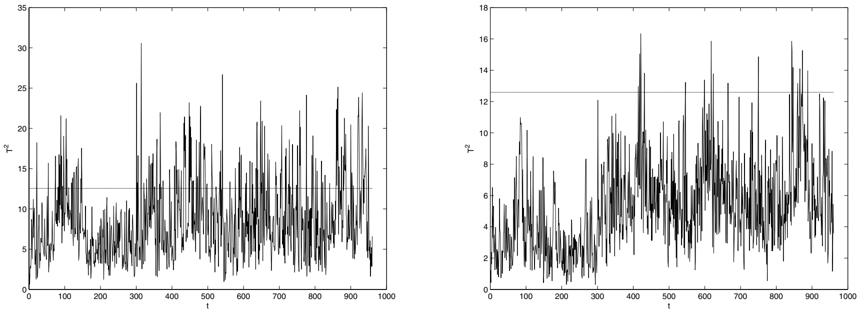

Fig. 2. The T^2 statistics of PCA (Left) and MSPCA (Right) powered by fault 4

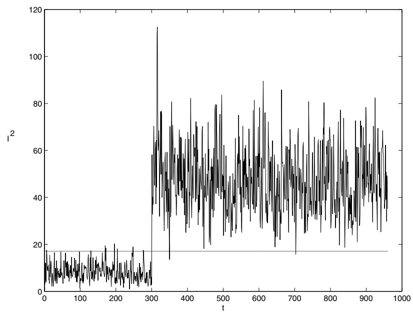

Fig. 3. The I^2 statistic of MSICA powered by fault 4

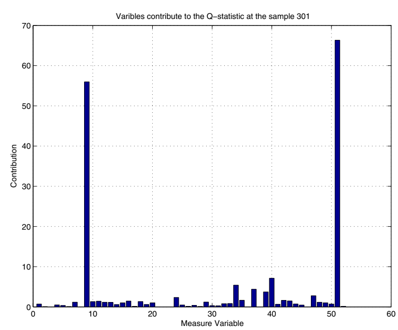

Fig. 4. The average contribution of fault 4 for the MSICA-based SPE

faults. As a basic tool, the average contribution of fault 4 at sample 301 is shown in Fig. 4. Obviously, variable 9 (Reactor temperature) and variable 51(Reactor cooling water flow rate) correlate with the fault 4 deeply. In fact, the two control loops of reactor cooling water flow rate and reactor temperature are cascade. This is consistent with the fault description.

Another interesting test is carried out by introducing fault 10, which involves a random change in the temperature of stream 4 (C feed). Even though this

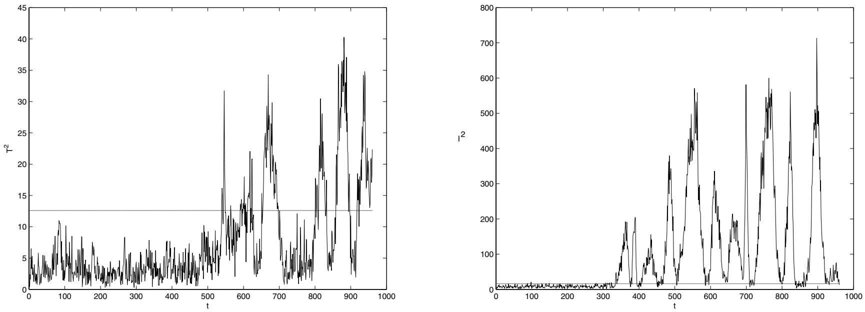

Fig. 5. The T^2 statistic of MSPCA (Left) and the I^2 statistic MSICA (Right) powered by fault 10

fault has no some effect on product quality by the titer of product in the stream 11, it is a hidden trouble on the safety of the process because of its effect on the pressure of the reactor. The T^2 statistic chart based on MSPCA and the I^2 statistic chart based on MSICA are shown in Fig. 5. It is obvious that PCA-like methods do not give an alarm in time, in which T^2 statistic is beyond the threshold after sample 500. Contrastively, the monitoring response of MSICA is acceptable.

4 Conclusion

By integrating the advantage of wavelet transform and independent component analysis, a MSICA approach is introduced to the process monitoring. The application results on TE plant illuminate some advantage over traditional MSPC methods. The ICA-based monitoring methods can give some compellent outcome in actual process operation.

References

1. Mason, R. L., Young, J. C.: Multivariate Statistical Process Control with Industrial Application. SIAM, USA (2002)
2. Chiang, L.H., Russell, E. L., Braatz, R. D.: Fault Detection and Diagnosis in Industrial Systems. Springer-Verlag, London (2001)
3. Kano, M., Nagao, K., Hasebe, S., Hashimoto, I., Bakshi, B.: Comparison of Statistical Process Monitoring Methods: Application to the Eastman Challenge Problem. Computers and Chemical Engineering 24(2000) 175-181
4. Johnson, R.A, Wichern, D.W.: Applied Multivariate Statistical Analysis. 4th edn. Englewood Cliff, Prentice Hall (1998)
5. Roberts, S., Everson, R.: Independent Component Analysis: Principles and Practice. Cambridge University (2001)
6. Ypma, A., Tax, D.M.J., Duin, R.P.W.: Robust Machine Fault Detection with Independent Component Analysis and Support Vector Data Description. Proc. IEEE Neural Networks for Signal Processing (1999) 67-76

7. Hyvarinen, A., Oja, E.: Independent Component Analysis: Algorithms and Applications. Neural Networks 13(2000) 411-430
8. Lee, J.-M., Yoo, C., Lee, I.-B.: Statistical Process Monitoring with Independent Component Analysis. Journal of Process Control 14(2004) 467-485
9. Mallat, S. G.: A Theory for Multiresolution Signal Decomposition: The Wavelet Representation. IEEE Transaction of Pattern Analysis and Machine Intelligence 11(7)(1989) 674-693
10. Bakshi, B. R.: Multiscale PCA with Application to Multivariate Statistical Process Monitoring. AICHE J. 44(7)(1998) 1596-1610
11. Kisilev, P., Zibulevsky, M., Zeevi, Y.Y.: A Multiscale Framework for Blind Source Separation of Linearly Mixed Signals. Journal of Machine Learning Research 4(2003) 1339-1364
12. Cover, T.M, Thomas, J.A.: Elements of Information theory. Wiley, New York (1991)
13. Hyvarinen, A.: Fast and Robust Fixed-Point Algorithms for Independent Component Analysis. IEEE Trans. Neural Networks 10(3)(1999) 626-634
14. Daubechies, I.: Ten Lectures on Wavelet. Capital City Press (1992)
15. Wang, M. P., Jones, M. C.: Kernel Smoothing. Chapman & Hall, London, UK (1995)
16. Downs J.J, Vogel, E. F.: A Plant-Wide Industrial Process Control Problem. Computers and Chemical Engineering 17(3)(1993) 245-255