International Journal of Innovative
Computing, Information and Control
Volume **9**, Number **4**, April **2013**

 ISSN 1349-4198
pp. 1781–1800

A MULTIVARIABLE STATISTICAL PROCESS MONITORING METHOD BASED ON MULTISCALE ANALYSIS AND PRINCIPAL CURVES

Xiangrong Shi[1,2], Yan Lv[1], Zhengshun Fei[1] and Jun Liang[1,*]

[1]State Key Lab of Industrial Control Technology
Department of Control Science and Engineering
Zhejiang University
No. 38, Zheda Road, Hangzhou 310027, P. R. China
shixr2000@163.com; {ylv; zsfei}@iipc.zju.edu.cn; *Corresponding author: jliang@iipc.zju.edu.cn

[2]School of Information
Zhejiang University of Finance and Economics
No. 18, Xueyuan Road, Xiasha Higher Education District, Hangzhou 310018, P. R. China

Received January 2012; revised June 2012

Abstract. *This study aims to develop an algorithm by integrating multi-resolution analysis (MRA) and principal curves (PC) for monitoring multivariate processes. This may pave the way for handling nonlinear data by means of principal curves in process monitoring area. We succeed in utilizing PC technique for monitoring without the assistance of neural networks, a traditional tool to deal with nonlinear model in papers, and get ideal results. The methodology proposed is tested with a mathematical example and a simulated benchmark process: the continuous stirred tank reactor (CSTR). The results demonstrate that, compared with traditional principal component analysis (PCA), PC, nonlinear PCA and multiscale PCA, the proposed approach can extract the nonlinearity and decorrelate the autocorrelated measurements effectively and is, hence, suitable for multivariate process monitoring.*

Keywords: Process monitor, Multiscale analysis, Principal curves, Nonlinear PCA

1. **Introduction.** Process monitoring and fault diagnosis are crucial tasks in industrial processes, which have attracted increasing attention from both academia and industry during the past decade. One of the most widely adopted techniques is principal component analysis. Though PCA is the optimal linear method with respect to minimizing the mean squared prediction error (SPE), sometimes traditional PCA is inadequate if two factors have to be considered in practice. One is nonlinearity. Most industrial process data are nonlinear, and if conventional PCA is used, minor components do not always consist of noise but important information. The other is autocorrelated data and wavelet decomposition is usually used to deal with such an issue.

The purpose of this paper is to propose a methodology by combining MRA of wavelet and nonlinear strategy of principal curves to handle practical process data with dynamic and nonlinear characteristics. Besides, some improvements are also made such as utilizing principal curves to establish nonlinear model without the assistance of neural networks (NN) for the training procedure of NN is tedious and the results can be instable. The results of both mathematical and simulated problems are presented to illustrate the effectiveness of this combinational method.

2. Past Related Research about Nonlinear PCA and Multiscale Analysis.

2.1. Nonlinear PCA.

To extract nonlinear features from process data, various techniques have been proposed among which two approaches are typical. One is kernel method using statistical learning theory. Data are mapped into a higher dimension space where the relationship is linear, so its essence is still linear PCA. As stated by Lee (2004), Kernel PCA (KPCA) did not consider reconstruction steps of the data in feature space; thus, it raised a problem while applied directly to process monitoring since the value of SPE can not be obtained [1].

The other approach is to find the best representation of data's inner structure directly; for example, principal curves proposed by Hastie (1984) are smooth 1-D curves that pass through the middle of a p-D data set [2]. They minimize the distance from itself to the points, and provide a nonlinear summary of the data. Figure 1 illustrates the difference between linear PCA and PC.

Based on PC technique, Dong and McAvoy (1996) proposed a method for process monitoring [3]. Neural networks were adopted in his approach to calculate the predicted

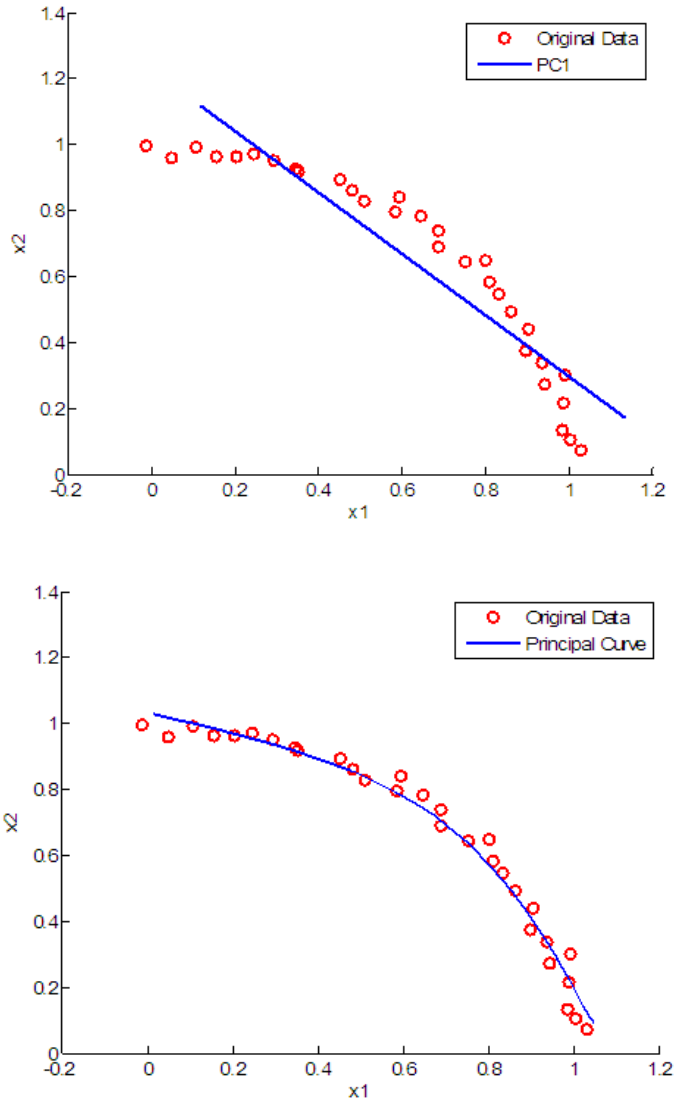

FIGURE 1. The first linear principal component and principal curve

values. In fact, neural network can be utilized independently to model nonlinear principal components [4]. A feedforward structure with a bottleneck layer is employed typically in NN model to extract the nonlinearity. Zhao and Xu (2005) found that such networks were so sensitive that the obtained results significantly differed from the underlying system [5], and the same conclusion can be drawn in Section 4 of this paper.

2.2. **Multiscale analysis based on wavelets.** Wavelet based multi-resolution analysis is usually utilized to extract features from industrial data while multiscale PCA (MSPCA) is considered the framework of later MRA based process monitoring [6]. It uses wavelet decomposition to decorrelate the autocorrelated data, and also uses linear PCA to reduce the cross-correlation among the multivariate data. Besides, the methods of how to combine wavelet decomposition and PCA are various. Ganesan et al. presented a summary of these methodologies as shown in Figure 2 [7]. In this study, method 1 is adopted to illustrate the effectiveness of multiscale principal curves (MSPC).

3. **Main Methods.** As mentioned above, conventional PCA is best suited for the analysis of steady state data with uncorrelated measurements. However, in practice, the dynamics of typical chemical or manufacturing processes cause the measurements to be autocorrelated and linear PCA is no longer appropriate. Consequently, a two-step procedure is proposed in this study to deal with such a problem as shown in Figure 3.

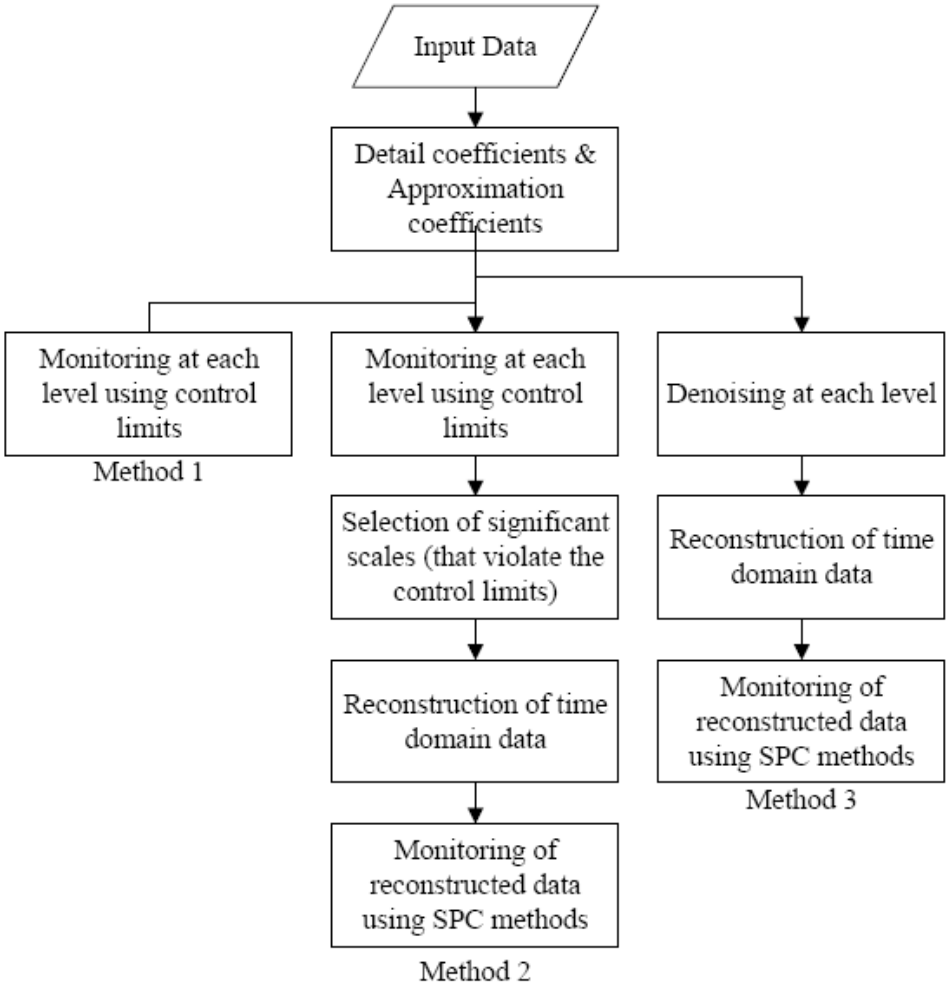

FIGURE 2. Methods of combining wavelet decomposition and PCA

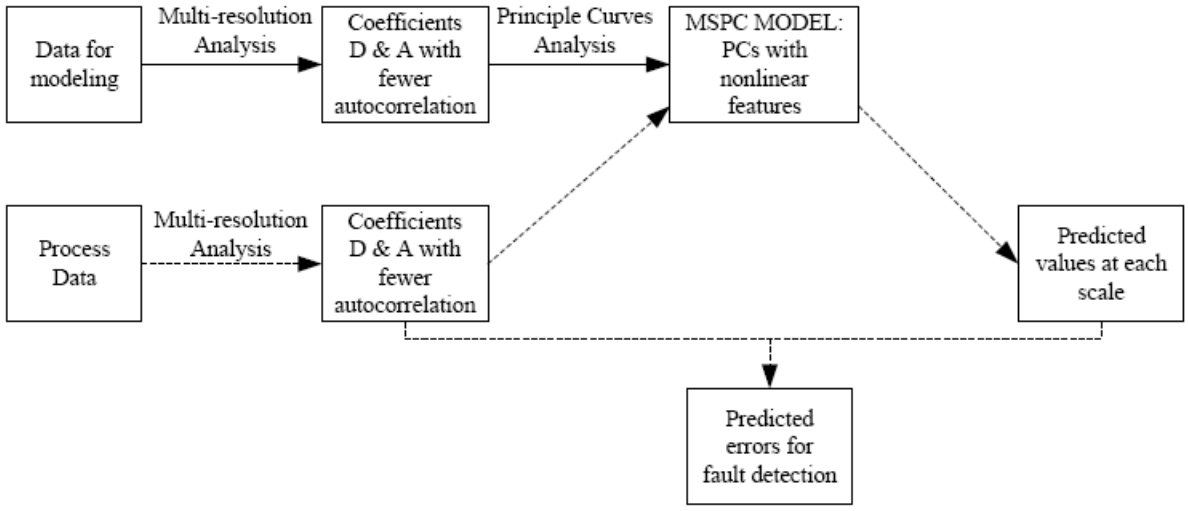

FIGURE 3. Two steps to deal with autocorrelation and nonlinearity

First, multi-resolution analysis is used to decorrelate the autocorrelation in modeling data by which the coefficients $D\&A$ are obtained. Then principal curves analysis step is employed to extract nonlinear features of $D\&A$ respectively. So PCs of each scale are obtained and these are the final process monitoring models. As far as online process data are concerned, we can acquire coefficients $D\&A$ in the same manner. Then, by interpolating (this paper) or NN method of Dong and McAvoy [3], the predicted values of coefficients and fault detection variable (such as SPE) at each scale can be calculated for process monitoring.

3.1. **Decorrelating the autocorrelation by multi-resolution analysis.** According to wavelet theory, any function can be described by a wavelet set which provides a mapping of the original function from time domain to frequency domain. The following is a wavelet mother function Ψ:

$$\Psi_{s,u}(t) = \frac{1}{\sqrt{s}}\Psi\left(\frac{t-u}{s}\right) \tag{1}$$

where s, u are the dilation and translation parameters. To deal with such situations that the measurements are discretely available, parameters s, u can be dyadically discretized as: $s = 2^j$, $u = 2^j k$, $j, k \in Z$ and Equation (1) can be rewritten as:

$$\Psi_{j,k}(t) = 2^{-j/2}\Psi(2^{-j}t - k) \tag{2}$$

Multivariate signal S can be decomposed successively onto scaling functions Φ and wavelet functions Ψ respectively. The projections onto Φ are the coarse approximation (A) of the original signal known as scaling coefficients. On the other hand, the projections onto Ψ are called wavelet coefficients, capturing the details (D) of the signal lost when moving from an approximation at one scale to the next. The decomposition step can be carried out repeatedly as in Figure 4.

Still, there exist nonlinearities in industry data and now they are in wavelet coefficients, for example, in $cA3$, $cD3$, $cD2$, $cD1$ of Figure 4. Hence, adopting linear PCA directly is improper, i.e., nonlinear PCA methods should be considered.

3.2. **Extracting nonlinear features by principal curves.** As mentioned, KPCA is regarded as a typical strategy to extract nonlinearities; however, Maulud et al. (2006) found that KPCA required a large matrix computation if the input observations were large [8]. In addition, it is not trivial to find a good kernel for a given problem under

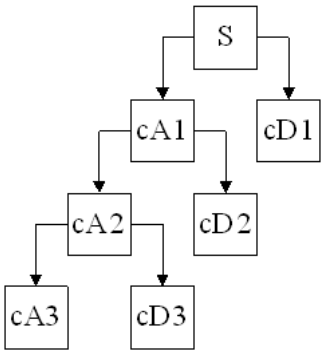

FIGURE 4. Successive wavelet decomposition

most situations, so it does not yield good results always. In this study, PC is chosen as PCA's nonlinear extension due to its excellent performance of representing data's inner structure. Dong's experiments proved the effectiveness of such a method except that PCs of original data rather than wavelet coefficients were calculated in his study [3].

Principal curves are nonparametric, nonlinear generalizations of the first principal components. A 1-D curve f is a vector of functions of a single variable λ which provides an ordering along f, and it is usually the arc length along the curve.

Let $X_c \in R^m$ be a continuous random vector, then $f(\lambda)$ is a principal curve of X_c if it does not intersect and is self consistent. That is:

$$E[X_c|\lambda_f(X_c) = \lambda] = f(\lambda) \tag{3}$$

where λ_f is defined as a projection index of $R^m \to R$:

$$\lambda_f = \sup_{\lambda} \left\{ \lambda : \|X_c - f(\lambda)\| = \inf_{\mu} \|X_c - f(\mu)\| \right\} \tag{4}$$

Self consistency means that each point on the curve is the conditional mean of the points that project there. Next, let $d(x, f)$ denote the Euclidean distance from a point x to its projection on f:

$$d(x, f) = \|x - f(\lambda_f(x))\| \tag{5}$$

And the expectation of the sum of the squared error is defined as:

$$D^2(X_c, f) = Ed^2(X_c, f) \tag{6}$$

In practice, the distribution of X_c is unknown. For example, a finite data set $X_d \in R^{m\times n}$ should be considered which is a matrix of m observations on n variables. Hastie regarded the data set as a sample from an underlying probability distribution, and used it to estimate the principal curves of that distribution. The steps to obtain the Hastie and Stuetzle's (HS) principal curve are as follows:

Step 1: Use linear principal components and their corresponding direction vectors as initial estimates of λ_f and f^0.

Step 2: Set $f^{i+1}(\lambda) = E[X_d|\lambda_{f^i}(X) = \lambda]$, then calculate the distance $D^2(X_d, f^{i+1})$. Iterate until the relative change in the distance is below some threshold.

Figure 5 is an example to calculate PC by iterating. The original data are obtained by generating points with independent noise $N(0, 0.01)$ around an ellipse with major radius 2 and minor radius 1. The principal curves are estimated using HS's algorithm with the smoother cubic spline and span value 0.7. After 35 iterations, the final PC is reached as shown in the last diagram of Figure 5.

While using HS's method, the only control parameter that needs to be specified is the span value which is the fraction of all the data points that are considered to be in

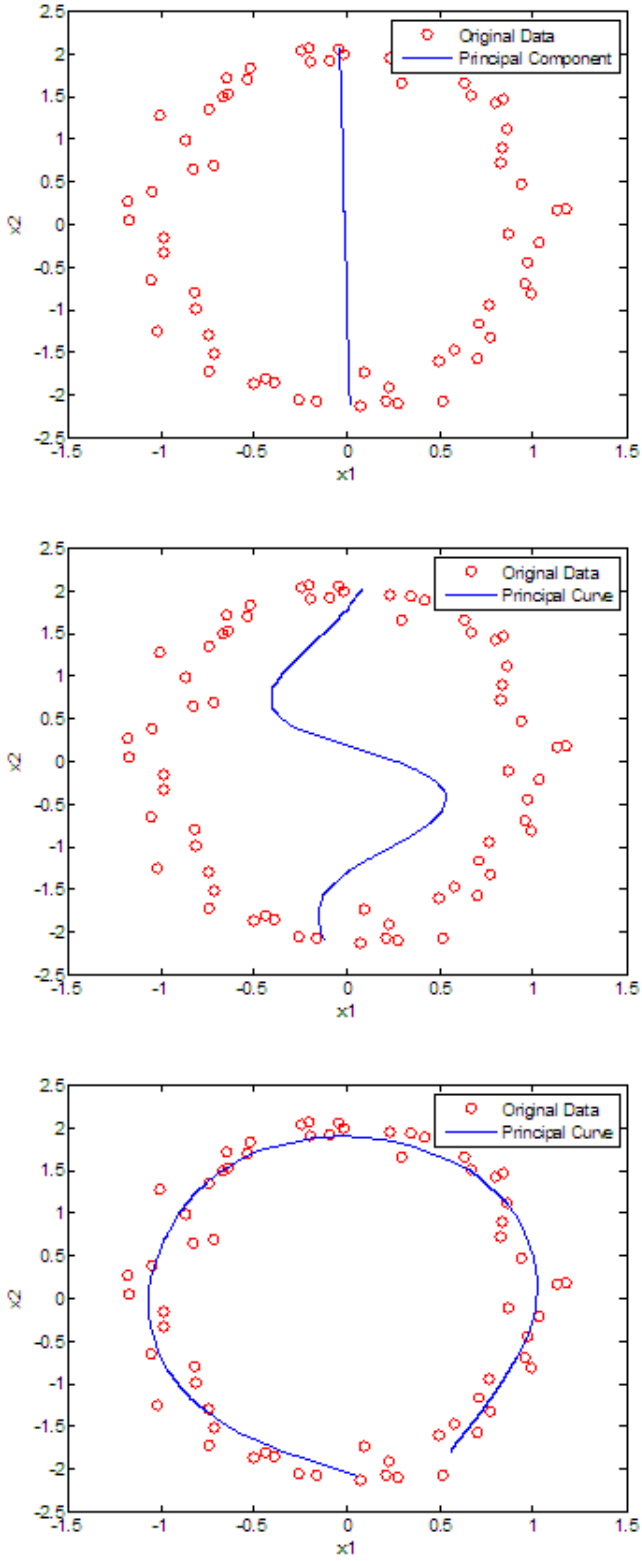

FIGURE 5. The estimated principal curves for simulated data of step 1, 5 and 35

the neighborhood. The shape of final curves can be influenced by this value more or less corresponding to varying data, so determining the value involves a trade off between smoothness and high fidelity to original data.

3.3. **Multiscale principal curves method for process monitoring.** Multiscale principal curve integrates the ability of PC to extract the nonlinear relationship of variables with that of wavelets to decorrelate the autocorrelation among the measurements.

When the PC approach is introduced to process monitoring, neural network is thought to be indispensable. This is partly due to its powerful approximation character and majorly because of the fact that people need not only the principal curve but also the projection points of new measurement on the curve. Plenty of research has been done mainly focusing on modification or improvement of the neural network. For instance, Harkat et al. (2003) proposed an enhanced HS's algorithm together with radial basis function (RBF) network [9]. In this study, we abandon the NN method based on three facts:

(1) The neural networks are sensitive and sometimes unreasonable results can be obtained.
(2) The neural network training process can be time-consuming and tedious, i.e., the time is unpredictable.
(3) In our strategy, principal curve algorithm is used following multiscale analysis, so if NN is incorporated, the whole procedure of modeling will be: MS-PC-NN, and it is excessively time expensive.

The method proposed is similar to interpolating during model calculating procedure as shown in Figure 6. To be exact, if data set X_d is finite, the principal curve can be regarded as a polyline l and the target of monitoring is to search for a point (predicted value) in it which is the nearest one to the measured data Y. Complexity of this procedure is $O(N)$ where N denotes the number of endpoints in l. That is to say, the complexity depends on the scale of original data set for establishing model.

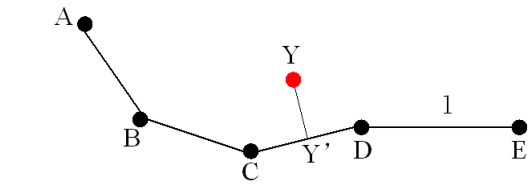

FIGURE 6. Data point Y and the predicted point of Y

In Figure 6, endpoints $\{A, B, C, D, E\}$ form the polyline l (principal curve), and Y is the measurement data point, we can find that Y' is the most nearest point in l to Y. In most cases, Y' is the vertical projection point of Y. The method of searching for Y' can be first determining the distances from Y to each sub-line and then selecting the minimum one. This method seems cumbersome but its complexity is inflexible and the computing time can be effectively guaranteed if the number of endpoints is proper. Experiment results show that this approach exhibits nearly same performance with that of NN as illustrated in Figure 7, and the related models are based on data in Figure 5. PC's span value is also 0.7. For NN, a back propagation (BP) network with the hidden layer size 13 is employed and the transfer function (TF) of layer 2 is logarithmic sigmoid.

Further, the steps of MSPC methodology are illustrated in Figure 8 and the following algorithm.

Building process monitoring model:

Step 1: Normalize the normal data to zero mean and unit variance. As these data will be used for decomposition, the number of measurements should be the integer power of 2.

Step 2: For each column (variable) in data matrix obtained in step 1, compute wavelet decomposition.

Step 3: For each scale, compute the 1st principal curve, and the 2nd, ... principal curves if necessary.

Step 4: For each scale, calculate the SPE control limit. The detailed calculating method is introduced in next section.

Calculation of process monitoring:

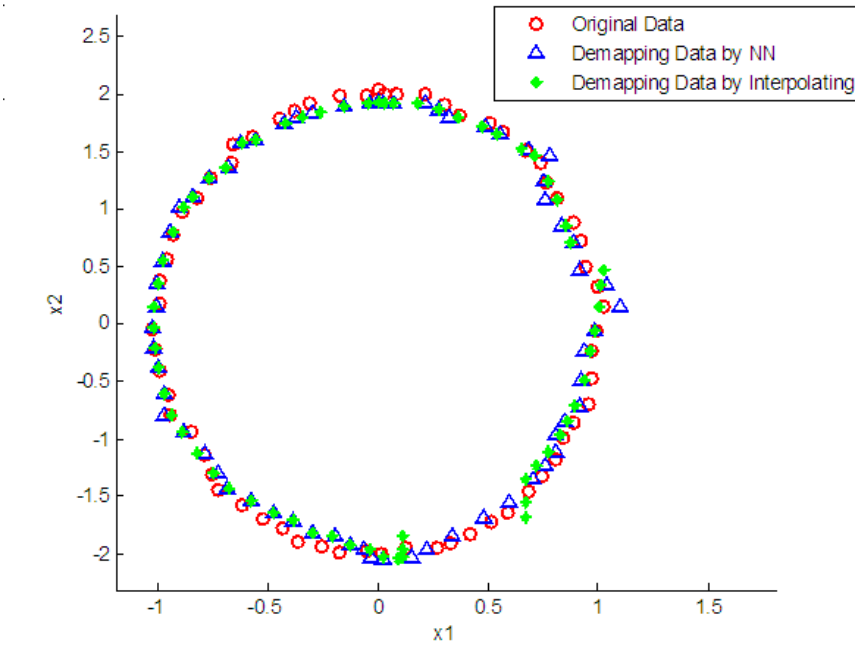

FIGURE 7. The predicted values for elliptic data via NN model & interpolating method

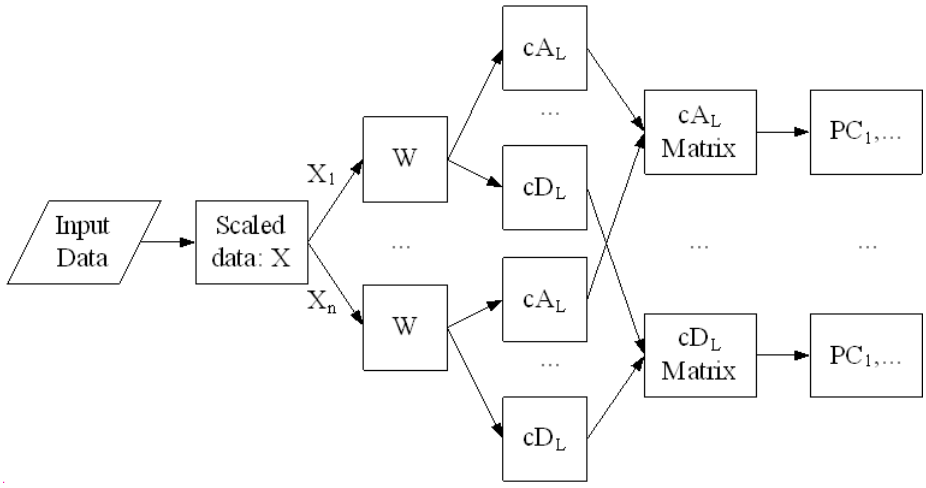

FIGURE 8. Integration of the wavelet and principal curves

Step 1: Standardize the process data of dyadic length 2^n using means and standard deviations obtained in above step 1.

Step 2: For each column in data matrix got in step 1, compute wavelet decomposition.

Step 3: For each scale, compute the projection points in PCs acquired in above step 3.

Step 4: For each scale, calculate the SPE values.

For real time monitoring, a moving window technique should be introduced. It updates the process data by including the most recent sample and abandoning the most distant one. After wavelet decomposition, the last coefficients at each level should be picked out; therefore, SPE values can be figured out.

Still, there are some important decisions to make in practice. One is the number of wavelet decomposition level and this topic has been studied abundantly during the past decades [8]. In this paper we select it mainly according to experience. The second is the number of PCs at each scale. Ideally, the number should be selected to provide maximum feature representation of a signal. If the number is too small, the predicted signal will probably have much noise, but if it becomes large, the computation will be exhaustive. A good choice is to determine the number by increasing it gradually until the cumulative percent variance (CPV) rises slowly. Besides, how to regulate the parameter span is of interest. A suggestion given in Dong's study is to decrease it gradually from a higher value to a lower one [3]. In this study, we use a fixed value strategy in all examples.

4. **Experiments.**

4.1. **A mathematical example.**

This example consists of three variables with an underlying dimensionality of one.

$$\begin{aligned} &x_1 = \sin(t) + e_1; \\ &x_2 = -2t^2 + 4t + e_2; \\ &x_3 = t^3 - 2t^2 + e_3; \\ &t = [t_1, t_2, \ldots, t_{512}] \in [0, 1.022] \end{aligned} \tag{7}$$

where e_1, e_2, e_3 are independent noise $N(0, 0.0002)$. The sample number 512 equals 2 to the 9th power so it can be wavelet decomposed conveniently. These data are used as normal condition. The fault conditions come by making small changes at x_3 in two places:

$$\begin{aligned} &\text{Fault 1: } x_3 = x_{3,normal} * 1.3;\ t = t_{257} \\ &\text{Fault 2: } x_3 = 1.1t^3 - 2.15t^2 + e_3;\ t = [t_{385}, t_{386}, \ldots, t_{512}] \end{aligned} \tag{8}$$

t_{257} is in the middle of whole time and there exists a sudden change. $t_{385} \sim t_{512}$ are at the last quarter of the whole samples which deviate from normal values. Figure 9 shows the normal and fault samples.

In result of linear PCA, both faults are not obvious as in Figure 10 due to the nonlinear nature of the experiment data though two principal components capture 99.91% of the variation. If one or three principal components are selected, even worse results can be obtained.

We calculate the 95% control limit using Box's algorithm introduced in Nomikos's study [10]. Because the prediction errors (PE) are normally distributed, the SPE (squared PE) can be regarded as a chi-squared distribution of $g.\chi_h^2$ where parameters g, h depend on the distribution parameters μ, σ as follows:

$$g = \frac{\sigma^2}{2\mu} \tag{9}$$

$$h = 2\left(\frac{\mu}{\sigma}\right)^2 \tag{10}$$

μ, σ are the estimated mean and standard deviation of the chi-squared distributed SPE values, and h is the free degree of χ^2 distribution. In this paper, we calculate all the confidence limits for SPE using the same method. SPE charts such as Figure 10 are presented in this study for it is an effective statistical method for detecting process faults.

When PC is adopted, it outperforms traditional PCA remarkably (Figure 11). We can observe a spike at point 257 (Fault 1) and many SPE values exceed the 95% limitation during the last quarter of the time axis (Fault 2). Here Verbeek's (2002) algorithm is used to find the 1st PC with the segment number 14 and the smoothing parameter 0.0018 [11]. Because it explains 99.98% of the variation, no more PCs are needed.

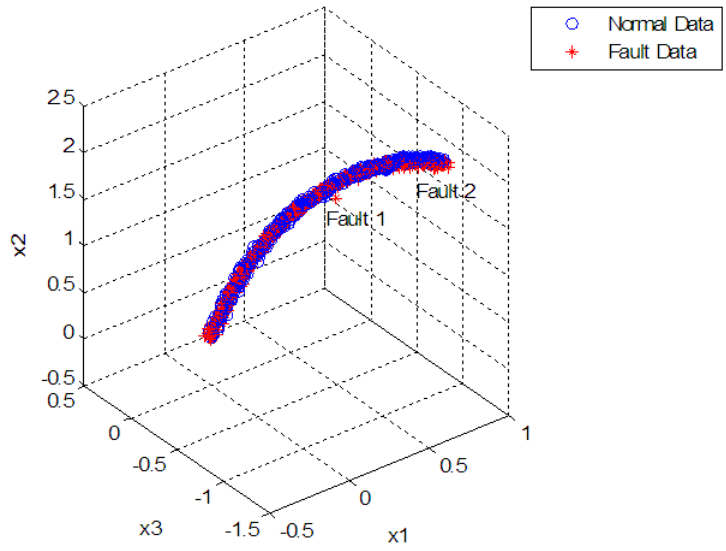

FIGURE 9. Data of normal condition and fault conditions

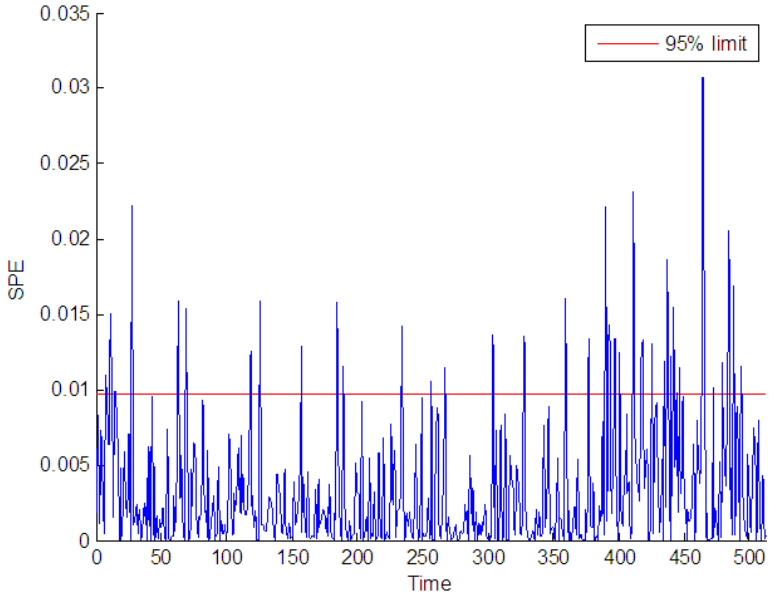

FIGURE 10. The SPE of linear PCA model

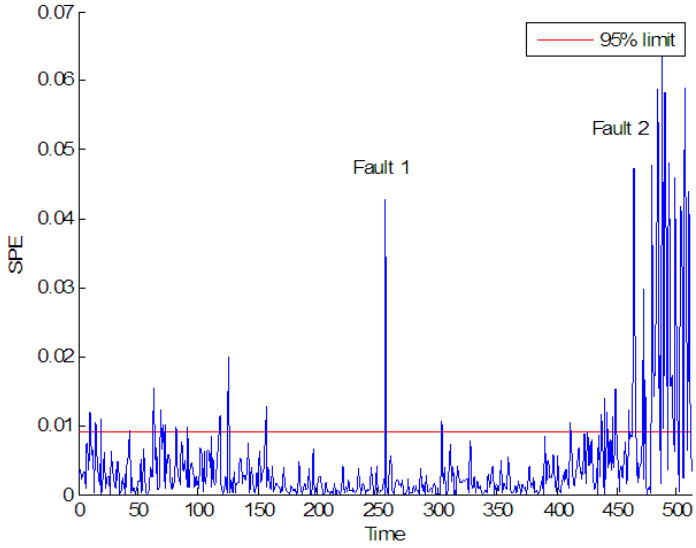

FIGURE 11. The SPE of PC model

As can be seen in upper subfigure of Figure 12, Kramer's method is adopted where a BP network is involved and the sizes of layer 2-4 are 7, 1 and 7. The TFs of layer 2, 4 are logarithmic sigmoid and TFs of layer 3, 5 are linear. After 153 times training, the network converges and the result is fairly good compared with PCA and PC. However, we got two kinds of results in experiments which illustrated the low reliability of this method: The training times can range from dozens to more than 2000, and sometimes unreasonable results can be observed as in lower subfigure of Figure 12.

Using Haar wavelet for three-level decomposition, we find that MSPCA can detect Fault 2 in approximation level as in A3 subfigure of Figure 13. The false alarms at first several points in A3 subfigure are caused by boundary effect of wavelet decomposition. We addressed this problem in Liu (2009) recently [12]. MSPCA is not so sensitive to Fault 1, a 30% sudden fluctuation, and it is because the artificially constructed data have not the autocorrelation nature, i.e., the performance of MSPCA is no better than conventional PCA with this kind of fault.

Figure 14 presents the results of MSPC. We choose wavelet type and the related parameters such as number of decomposition level just as in Figure 13. MSPC can detect Fault 1, a sudden change, in the highest scale D1 and Fault 2, a mean bias, in approximation scale A3. The results agree with Misra's [13] conclusion very well. Further, the performances of methods mentioned above are summarized in Table 1.

As shown in Table 1, if relevant parameters such as wavelet type, decomposition levels and PC numbers are fixed, the experiment results of PCA and MSPCA keep unchanged. Kramer's method tends to have low robustness as in Figure 12. As for PC and MSPC,

TABLE 1. Fault detection performance of various methods

Methods	Fault 1	Fault 2	Robustness of results
PCA	Not obvious	Partly (17.19%)	Absolutely no change
PC	Can detect	Partly (28.13%)	Pretty good
NLPCA (Kramer)	Can detect	Partly (34.38%)	Inferior
MSPCA (Bakshi)	Not obvious	Partly (68.75%)	Absolutely no change
MSPC (this paper)	Can detect	Can detect (100%)	Pretty good

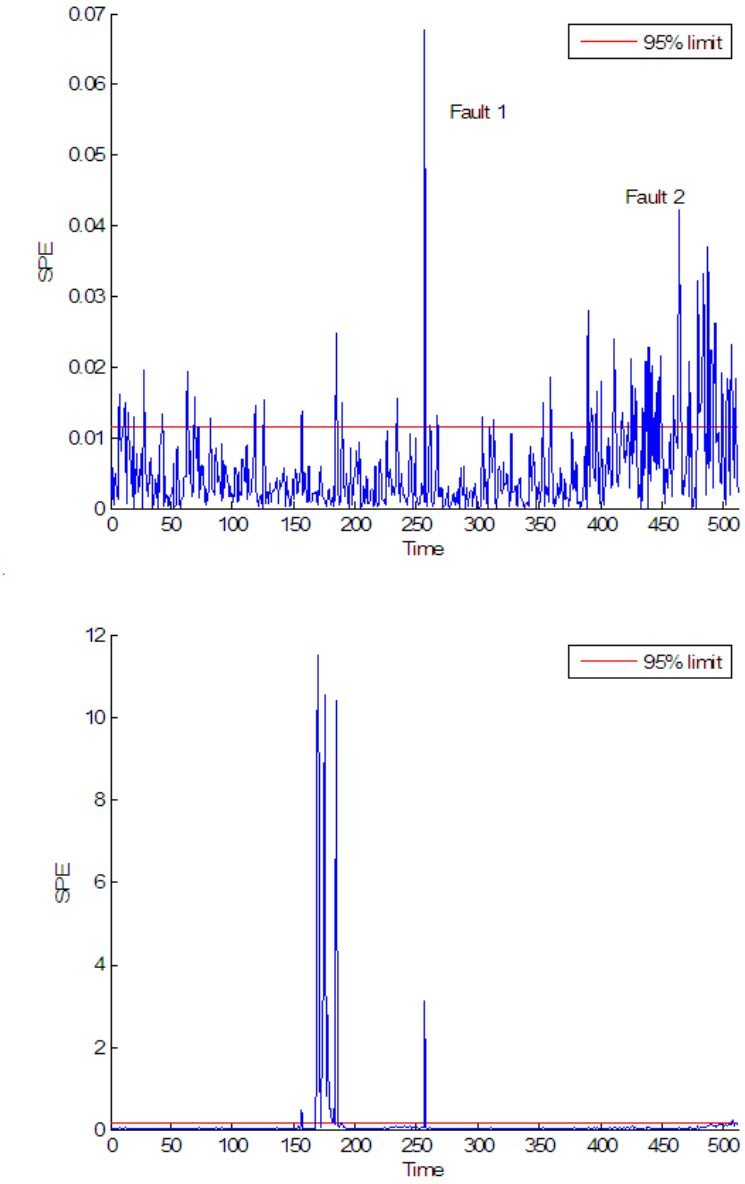

FIGURE 12. The SPE of Kramer's NLPCA model

TABLE 2. Detection rate of Fault 2 with MSPC method

Detection rate	Times (100 in all)
100% (16/16)	62
93.75% (15/16)	26
87.5% (14/16)	8
81.25% (13/16)	3
75% (12/16)	1

experiments show that the reliability is pretty good. For example, we tested MSPC method 100 times, and the detection rate of Fault 2 had 5 different values: 100%, 93.75%, 87.5%, 81.25% and 75%. Most of the values (88%) are the top two: 100% and 93.57%. 100% in Table 1 is the best result and the detailed detection rate of Fault 2 can be seen in

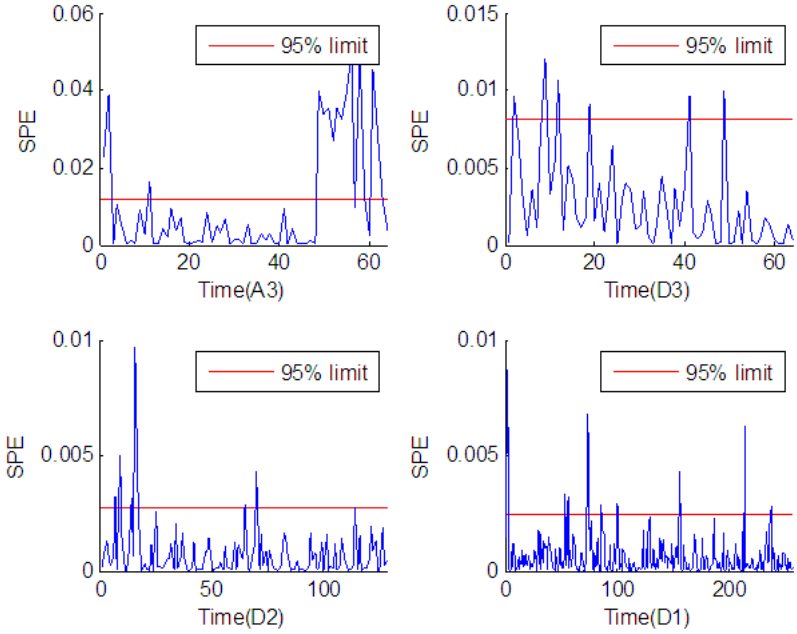

FIGURE 13. The SPE of MSPCA model

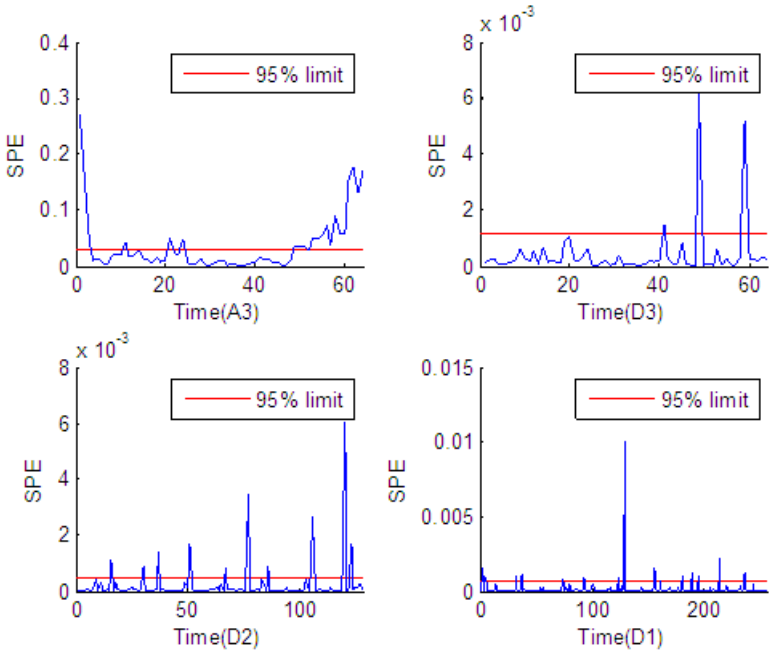

FIGURE 14. The SPE of MSPC model

Table 2. The average rate obtained from Table 2 is 96.56% and it surpasses all the other methods evidently.

Experiments also demonstrate that if the segment number and smoothing parameter are regulated slightly while using Verbeek's algorithm, different curves can be acquired; therefore, different monitoring model and results are obtained. That is to say, some model determining (parameter selection) should be done to achieve ideal results. In this experiment, parameters are set according to Verbeek's theory [11].

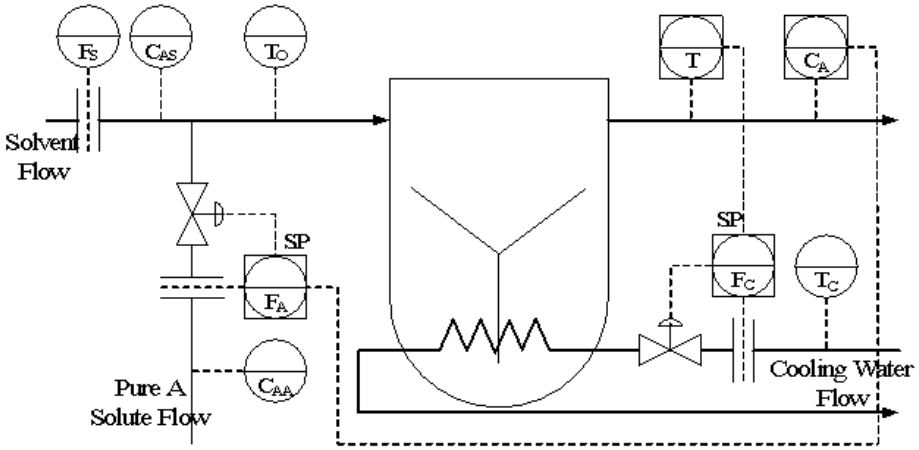

FIGURE 15. Process flow diagram of CSTR system

4.2. **CSTR process.** The nonisothermal CSTR process has a first order reaction and the reaction temperature is under feedback control (Figure 15). It was first proposed by Marlin (1995) and has been widely employed as a benchmark in fault detection and diagnosis [14]. In this paper, two kinds of faults, step changing and drift, are studied to test MSPC methodology. The dynamic behavior of the process can be observed in appendix of Yoon's paper [15].

4.2.1. *Fault of step change.* The sampling period is one minute and the fault condition is set by inserting a mean bias of 1.6 degree at T_O from point 257 to the end point 512 (Figure 16).

Nine variables are selected to construct detection model as shown in Table 3.

Figure 17 is the SPE chart obtained using linear PCA. In the first half (256 points) of whole samples, 14 points are falsely alarmed with the false alarm rate (FAR) 5.47% while

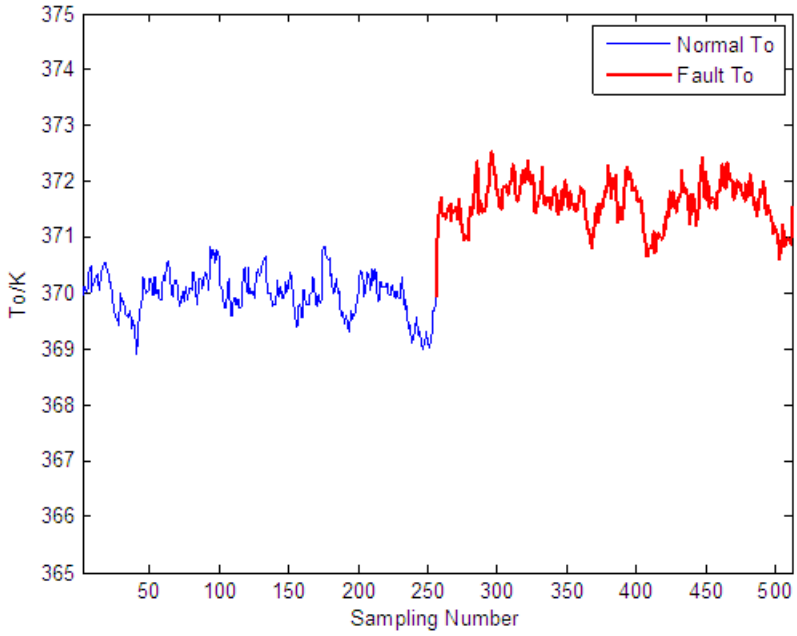

FIGURE 16. The normal and fault condition of T_O

TABLE 3. Nine involved variables

Variables	Description	Unit
F_A	Flow rate of reactant A	m^3/min
F_C	Flow rate of cooling water	m^3/min
C_A	Outlet concentration of the product	$kmole/m^3$
T	Outlet temperature	K
C_{AS}	Inlet concentration of solvent flow	$kmole/m^3$
C_{AA}	Inlet concentration of solute A	$kmole/m^3$
T_O	Inlet temperature	K
F_S	Flow rate of the solvent	m^3/min
T_C	Cooling water temperature	K

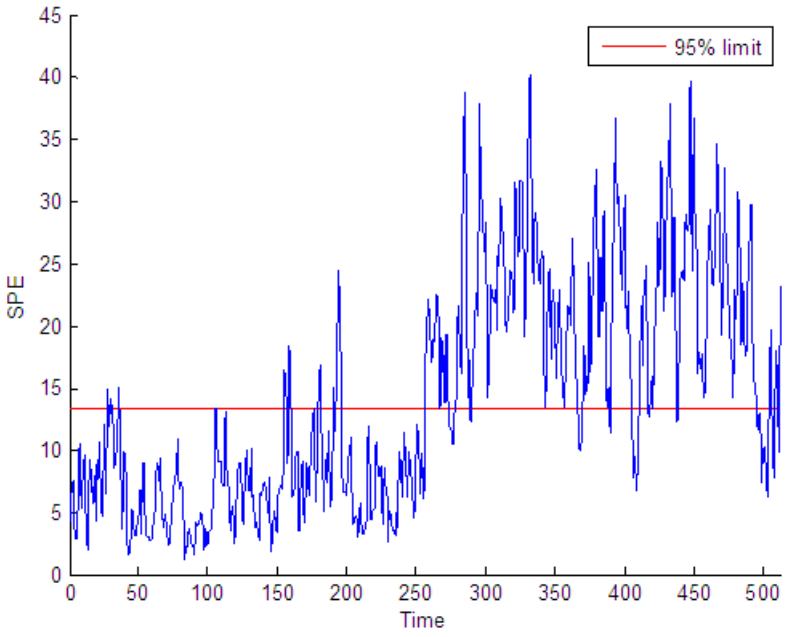

FIGURE 17. The SPE of linear PCA model

in the second 256 points, 33 fault samples fail to be detected, i.e., the miss alarm rate (MAR) is 12.89%. One principal component is chosen according to cross validation and it captures 11.57% of the variation. One principal component in such a system is proper for linear PCA. That is to say, if the number of components is increased, worse results can be found.

As shown in Figure 18, the result of PC, 30 points in the former half are falsely alarmed and 236 samples of the latter half violate the 95% warning limit. The FAR and MAR are listed in Table 4. Here, HS's method is utilized with the smoother cubic spline and span value 0.1 to find the 1^{st} PC and it explains 15.84% of the variation, while the 1^{st} linear principal component captures only 11.57%. Experiments show that if the number of principal curves is increased, the CPV value will become larger gradually but the MAR and FAR will first remain almost unchanged and then increase.

In the MSPC's SPE chart, Figure 19, 31 of 32 points at A3 scale violate 95% limitation and 1 false alarm takes place. Haar wavelet together with the same parameters as in Figure 13 is adopted. HS's method is used to find PCs and the related parameters such as span value and PC number are determined as in experiment of common PC. Further,

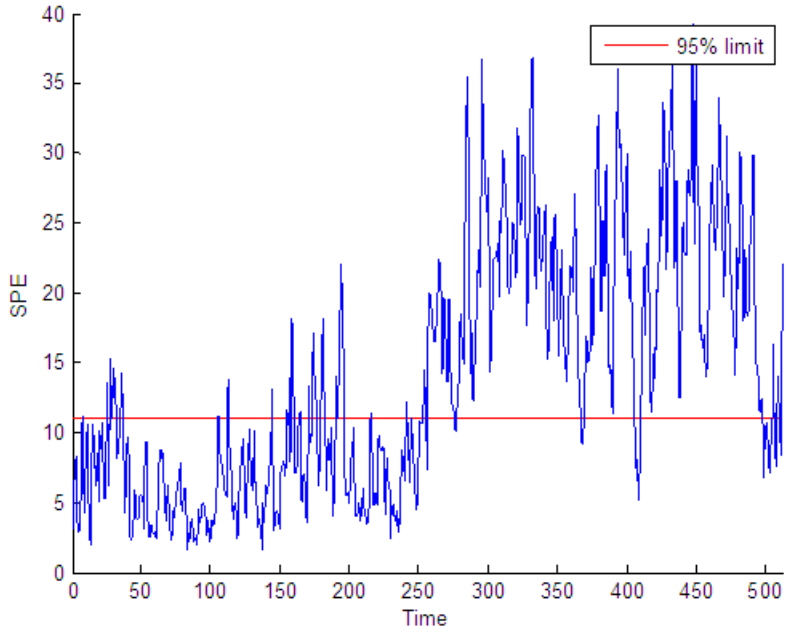

FIGURE 18. The SPE of PC model

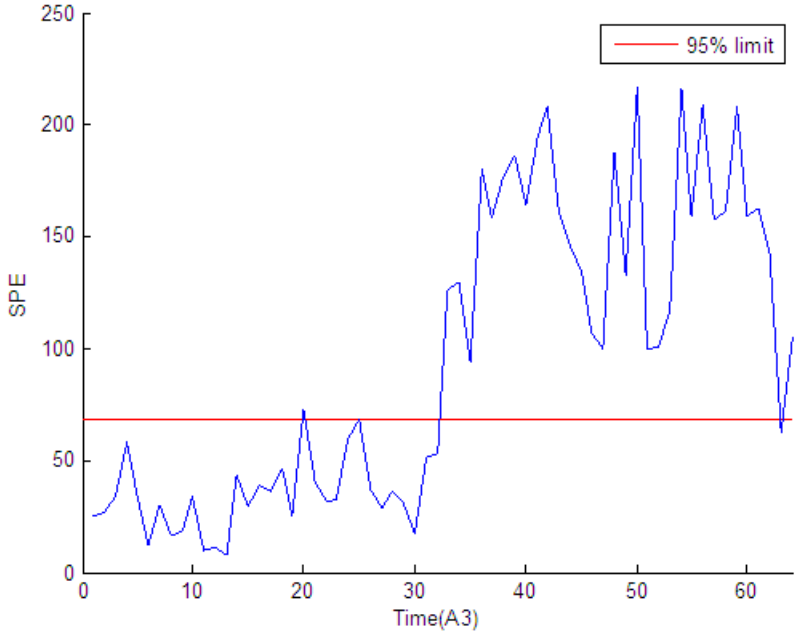

FIGURE 19. The SPE of MSPC model (A3 scale)

the performances of methods mentioned above are summarized in Table 4. We can see MSPC performs fairly well compared with conventional PCA and PC.

As MSPC decorrelates the autocorrelated measurements on the one hand, and extracts the essential feature of model data on the other hand, it can capture 34.12% of the variation at A3 scale and has the best detection performance as in Table 4.

TABLE 4. Fault detection performance of various methods

Methods	False Alarm Rate	Missed Alarm Rate	CPV
PCA	5.47%	12.89 %	11.57%
PC	11.72%	7.81 %	15.84%
MSPC	3.13%	3.13%	34.12%

4.2.2. *Fault of drift type.* Now we test the fault condition caused by increasing concentration of solute A (C_{AA}) gradually with step size 0.012 kmole/m^3 from point 257 to the end point 512 (Figure 20).

Also, nine variables in Table 3 are selected to build detection model and we can get the following results (Figure 21 and Table 5). MSPC can detect the ramp fault earlier than traditional PC and PCA, and also PC exhibits better performance than that of PCA (for example: $257 < 335 < 492$). As pointed out by Hsu, early detection of process faults is important for ensuring plant safety and retaining high yield of product [16].

FAR and MAR of MSPC method are 12.5% and 9.37% respectively. They are both the lowest one of three tested methods. As stated above, the result of PCA model remains unchanged in repeated experiments if the principal component number is fixed. For PC and MSPC method, we choose the same span value and smoother function as in step change experiment.

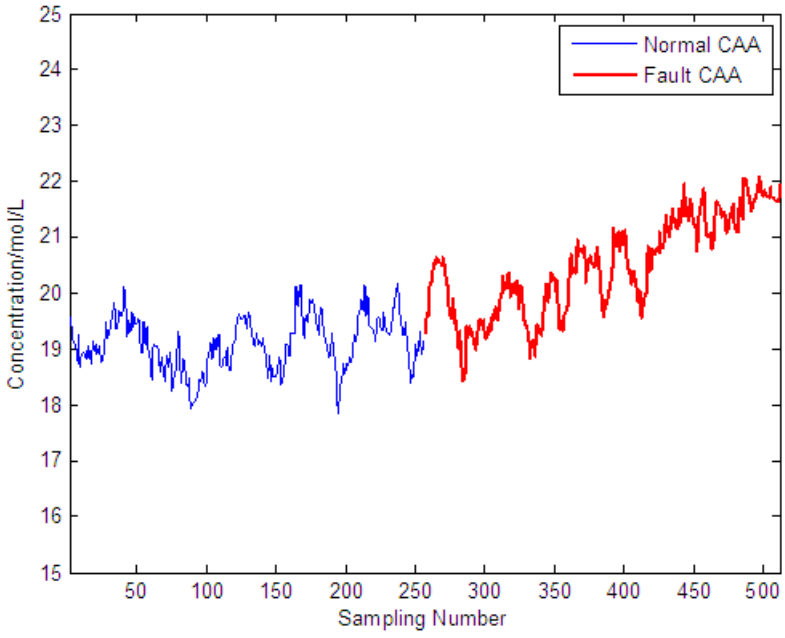

FIGURE 20. The normal and fault condition of C_{AA}

TABLE 5. Fault detection performance of various methods

Methods	False Alarm Rate	Missed Alarm Rate	Point (after it MAR < 15%)	Point (after it MAR < 10%)	Point (after it MAR < 5%)
PCA	12.11%	33.2%	492	495	496
PC	13.28%	21.48%	335	385	412
MSPC	12.5%	9.37%	257	297	305

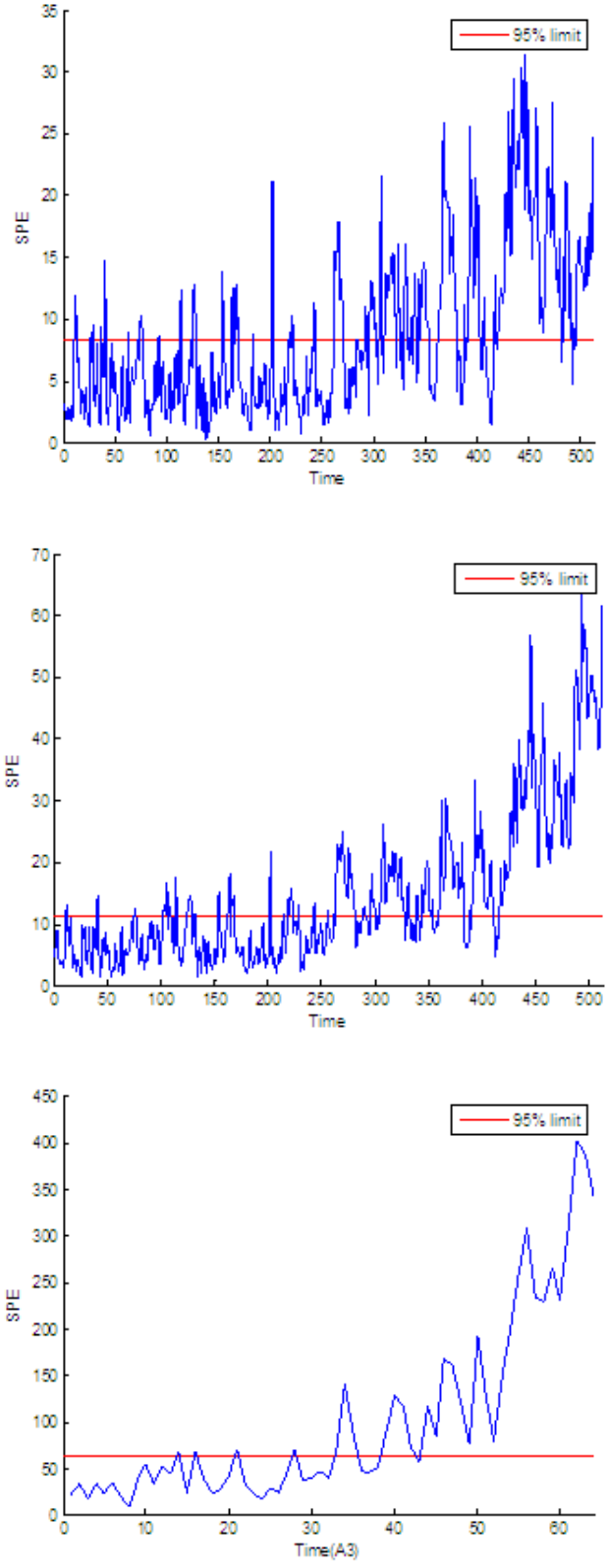

FIGURE 21. The SPE of PCA, PC & MSPC

While using HS's method, different curves can be obtained if the span value deviates from the optimal one. The value should be decided deliberately for increasing it tends to increase the smoothness of the fits, while decreasing it causes interpolating the data. A good balance should be found to get ideal monitoring performance. We decide the value mainly based on experience. Nevertheless, principal curves are representation of data's inner structure and unreasonable results are either not possible or will rarely occur in experiments. On the other hand, once the span and smoother is determined, satisfactory results can always be obtained in practice.

5. **Conclusions.** This study presents an approach by combining the attractive properties of PC and multiscale analysis. Thus the two typical issues of industrial data, nonlinearity and autocorrelation, are solved by the resulting MSPC methodology. Because PC-NN method proposed by Dong has a complicated net structure and a long time is needed to train such a network to reach the convergence, we acquire the predicted value by searching for the nearest point in PCs instead of incorporating neural networks.

The proposed approach is applied to artificial data and CSTR process. Faults of step change, sudden fluctuation and ramp are tested. The experiment results are consistent with theory analysis. The advantage of current MSPC method is verified compared with traditional approaches.

Future directions of this work include the improvement of procedure of finding PCs. For example, span is an important parameter which influences the converging rate remarkably. How to regulate it dynamically may be an issue of interest. In addition, we should widen the fault types to include such as precision degradation to inspect this method thoroughly and we think this may be of great value though the procedure is more time consuming.

Acknowledgment. This research work is supported by the National Natural Science Foundation of China (61174114), the Research Fund for the Doctoral Program of Higher Education in China (20120101130016), and the National High Technology Research and Development Program of China (2007AA04Z168).

REFERENCES

[1] J. M. Lee, C. K. Yoo and I. B. Lee, Fault detection of batch processes using multiway kernel principal component analysis, *Computers and Chemical Engineering*, vol.28, no.9, pp.1837-1847, 2004.
[2] T. Hastie, *Principal Curves and Surfaces*, Ph.D. Thesis, Stanford University, 1984.
[3] D. Dong and T. J. McAvoy, Nonlinear principal component analysis-based on principal curves and neural networks, *Computers and Chemical Engineering*, vol.20, no.1, pp.65-78, 1996.
[4] M. A. Kramer, Nonlinear principal component analysis using autoassociative neural networks, *AIChE Journal*, vol.37, no.2, pp.233-243, 1991.
[5] S. Zhao and Y. Xu, Multivariate statistical process monitoring using robust nonlinear principal component analysis, *Tsinghua Science & Technology*, vol.10, no.5, pp.582-586, 2005.
[6] B. R. Bakshi, Multiscale PCA with application to multivariate statistical process monitoring, *AIChE Journal*, vol.44, no.7, pp.1596-1610, 1998.
[7] R. Ganesan, T. K. Das and V. Venkataraman, Wavelet-based multiscale statistical process monitoring: A literature review, *IIE Transactions*, vol.36, no.9, pp.787-806, 2004.
[8] A. Maulud, D. Wang and J. A. Romagnoli, A multi-scale orthogonal nonlinear strategy for multivariate statistical process monitoring, *Journal of Process Control*, vol.16, no.7, pp.671-683, 2006.
[9] M. F. Harkat, G. Mourot and J. Ragot, Nonlinear PCA combining principal curves and RBF-networks for process monitoring, *WeA*, vol.14, no.3, pp.1956-1961, 2003.
[10] P. Nomikos and J. F. MacGregor, Multivariate SPC charts for monitoring batch processes, *Technometrics*, vol.37, no.1, pp.41-59, 1995.
[11] J. J. Verbeek, N. Vlassis and B. Krose, A K-segments algorithm for finding principal curves, *Pattern Recognition Letters*, vol.23, pp.1009-1017, 2002.

[12] Y. M. Liu, J. Liang, B. Hu, L. B. Ye and X. R. Shi, A multivariable statistical process monitoring method based on multiscale analysis, *CIESC Journal*, vol.60, no.4, pp.952-958, 2009.
[13] M. Misra, H. H. Yue, S. J. Qin and C. Ling, Multivariate process monitoring and fault diagnosis by multi-scale PCA, *Computers and Chemical Engineering*, vol.26, pp.1281-1293, 2002.
[14] T. E. Marlin, *Process Control*, McGraw-Hill, New York, 1995.
[15] S. Yoon and J. F. MacGregor, Fault diagnosis with multivariate statistical model part I: Using steady state fault signatures, *Journal of Process Control*, vol.11, pp.387-400, 2001.