Improved Relative-transformation Principal Component Analysis Based on Mahalanobis Distance and Its Application for Fault Detection

SHI Huai-Tao[1,2] LIU Jian-Chang[2] XUE Peng[3] ZHANG Ke[1] WU Yu-Hou[1] ZHANG Li-Xiu[1] TAN Shuai[2]

Abstract Principal component analysis (PCA) has been widely used in process industries, which could maintain the maximum fault detection rate. Although many issues have been addressed in PCA, some essential problems remain unresolved. This study improves PCA for fault detection performance in the following ways. Firstly, a relative transformation scheme based on Mahalanobis distance (MD) is introduced to eliminate the effect of dimension of data instead of dimensionless standardization, and improve the accuracy and real-time performance of fault detection. The theoretical derivation proves that relative transformation based on MD can directly eliminate the effect of dimension and give reasonable explanation of PCA in the relative space, the analysis and simulation results show its superiority and effectiveness. Secondly, an improved squared prediction error (SPE) statistic is given to improve the fault detection performance of standardized PCA, which can make the standardized PCA-based fault detection method more suitable for the actual industrial process. Finally, two improved methods are combined to detect the fault more effectively. The proposed methods are applied to detect single fault and multi-fault of looper system in hot continuous rolling process, simulation results demonstrate the effectiveness of these improvements for fault detection performance in terms of sensitiveness, accuracy and real-time performance of fault detection.

Key words Relative-transformation PCA based on Mahalanobis distance, improved squared prediction error (SPE), fault detection, single fault, multi-fault, looper system

Citation Shi Huai-Tao, Liu Jian-Chang, Xue Peng, Zhang Ke, Wu Yu-Hou, Zhang Li-Xiu, Tan Shuai. Improved relative-transformation principal component analysis based on Mahalanobis distance and its application for fault detection. *Acta Automatica Sinica*, 2013, **39**(9): 1533−1542

DOI 10.3724/SP.J.1004.2013.01533

Over the last few decades, multivariate statistical approaches, such as principal component analysis (PCA), partial least squares (PLS) and independent component analysis (ICA) have aroused extreme interest in the fields of fault detection. Of all these approaches, PCA is one of the most widely used fault detection method. The core idea of PCA is to determine a set of orthogonal vectors, i.e., loading vectors, ordered by the amount of variance described in the loading vector directions. PCA has been heavily researched and applied to industrial processes for fault detection[1−3]. In order to adapt various characteristics of an industrial process, a lot of extensions of PCA have been developed such as nonlinear PCA[4], dynamic PCA[5] and multi-block PCA[6], 2DPCA[7], etc.

Although many issues have been discovered, two essential problems remain unresolved in PCA.

On one hand, in order to eliminate the influence of dimensions on the system, process data often need to be standardized so that each variable is given equal weight before the application of PCA[8], but this is accompanied with new problems. From the standpoint of feature extraction of PCA, extracting principal components (PCs) from standardized data is equal to extracting the PCs from the comparative matrix, but comparative matrix only contains interacting information without differential information among variables. If we use PCA to extract feature information from the standardized data, accurate principal component model will not be obtained; from the standpoint of dimensionality reduction of PCA, when the data is standardized, each variable is allocated with the same weights, which will cause the eigenvalues of covariance matrix to be approximately equal, i.e., the feature information of each PC is so approximately equal that the representative PCs can not be extracted effectively; from the standpoint of energy conservation, the standardized data cannot obey the criterion of energy conservation. For this problem, Wen et al.[9] propose a relative PCA-based fault detection method, in which each variable is allocated corresponding weight via determining important levels of different variables according to prior knowledge. However, the method needs a large amount of system′s prior knowledge to select the relative transforming factor and the weight coefficients. Unfortunately, the method has no extensive applications, The main reason was because it is pretty difficult to obtain enough prior knowledge in the actual process.

On the other hand, traditional monitoring indices, Hotelling′s T^2 and squared prediction error (SPE), are generally used in PCA-based fault detection methods to determine whether the fault has occurred or not. However, measures of Hotelling′s T^2 statistics and SPE statistics are different because the former is Mahalanobis distance (MD) measure and the other is Euclidean distance (ED) measure. Two statistics cannot act together to estimate process condition accurately. The traditional SPE ignores the compactness between the current prediction error and normal prediction error, which will result in data information loss and low detection rate. Some researchers have committed themselves to improve the monitoring index. Dunia et al. improved the SPE index by introducing an important tuning parameter, which can reduce the false alarms[10], but it is very difficult to choose appropriate tuning parameter. Zhao et al. proposed an MD-based SPE index and an overall index to improve the monitoring performance[11].

Manuscript received January 6, 2012; accepted March 19, 2013
Supported by National Natural Science Foundation of China (51375317, 61374137), National 12th Five-Year Science and Technology Based Plan (2011BAJ02B07), the Ministry of Education Innovation Team Support Projects (IRT1160), and Science and Technology Plan, Shenyang City (F12-036-2-00)
Recommended by Associate Editor JIANG Bin
1. College of Transportation and Mechanical Engineering, Shenyang Jianzhu University, Shenyang 110168, China 2. College of Information Science and Engineering, Northeastern University, Shenyang 110819, China 3. Baoshan Iron and Steel Co., Ltd., Shanghai 200941, China

However, how to improve effectively the monitoring index and detection rate are still open problems in PCA-based fault detection methods.

For dealing with the drawbacks above, PCA for fault detection is improved in two ways. Firstly, a relative transformation based Mahalanobis distance (MD)[12] is proposed to handle the problem that eigenvalues are equal and energy is non-conservative. The idea is to transfer the original space into the relative space by computing MD between original sample data, and the theoretical derivation proves that relative transformation based on MD can directly the effect of dimension, effectively extract the representative principal components and obey energy conservation law. Secondly, an improved SPE based on MD is provided, which can effectively enhance the fault detection rate and make the PCA-based fault detection method more suitable for actual industrial process.

1 PCA-based fault detection

PCA is a multivariable statistical technique which is commonly used to monitor process operation. It produces a lower-dimensional representation in a way that preserves the correlation structure between the process variables, and is optimal in terms of capturing the variability of the data[8]. PCA involves the computation of loadings and scores matrix by computing the covariance matrix of X, $X \in \mathbf{R}^{n\times m}$, where n is the total number of samples and m is the number of variables. Covariance matrix of X is expressed as $C_X = \mathrm{E}[(X-\mathrm{E}(X))(X-\mathrm{E}(X))^{\mathrm{T}}]$, and the eigenvalues and corresponding eigenvector can be solved by (1) and (2).

$$|\lambda I - C_X| = 0 \tag{1}$$

$$(\lambda_i I - C_X)p_i = 0,\ \ i = 1, 2, \cdots, m \tag{2}$$

which p_i is eigenvector of eigenvalues λ_i, and $P = \{p_1, p_2, \cdots, p_n\}$, hence, score matrix is expressed as $T = X \cdot P$.

Considering that the process variables are often correlated, m variables can be expressed by a uncorrelated fewer variables, where a represents the number of the principal components that retains to explain the majority of the variability in the data[13]. Data matrix can be represented by a PCA model in a lower dimension as follows:

$$X = T_a \cdot P_a^{\mathrm{T}} + E = \sum_{i=1}^{a} t_i p_i^{\mathrm{T}} + E \tag{3}$$

T_a, P_a, E are score, loading and error matrices, respectively. $T_a \cdot P_a^{\mathrm{T}}$ is an a-dimensional PC subspace and represents $(m-a)$-dimensional residual subspace that cannot be expressed by the PCA model.

For a new testing observed vector x_{new}, two statistics are usually used to monitor the process. One of the indices used for fault detection is the SPE statistics:

$$SPE = \sum_{j=1}^{m} (x_{\mathrm{new}} - \hat{x}_{\mathrm{new}})^2 = \left\| x_{\mathrm{new}} - x_{\mathrm{new}} P_a P_a^{\mathrm{T}} \right\|^2 = \left\| x_{\mathrm{new}} \tilde{P}_a \tilde{P}_a^{\mathrm{T}} \right\|^2 \tag{4}$$

where $\tilde{P}_a$ is an $m \times (m-a)$ matrix that consists of the remaining columns of eigenvectors. Therefore, SPE can be viewed to measure the deviation of the observations to the lower-dimensional PCA representation. The distribution for SPE statistic is defined as follows by [14]:

$$SPE_\alpha = \theta_1 \left[\frac{C_\alpha \sqrt{2\theta_2 h_0^2}}{\theta_1} + 1 + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right]^{\frac{1}{h_0}} \tag{5}$$

where $\sum_{j=a+1}^{m} \lambda_j^i$ $(i = 1, 2, 3)$, $h_0 = 1 - 2\theta_1\theta_3/3\theta_2^2$, and C_α is normal deviate corresponding to the $(1-\alpha)\times 100\,\%$ percentile. A violation of the threshold would indicate that the systematic variations are out of control, namely, the fault occurs.

The other index used for fault detection is Hotelling′s T^2 statistic:

$$T^2 = t D_{\lambda a}^{-1} t^{\mathrm{T}} = (x_{\mathrm{new}}^{\mathrm{T}} P_a) D_{\lambda a}^{-1} (x_{\mathrm{new}}^{\mathrm{T}} P_a)^{\mathrm{T}} \tag{6}$$

where $D_{\lambda a}$ is a diagonal matrix, which is the estimated covariance matrix of the principal component scores and t is the scores vector which indicates deviation from normal values inside a-dimensional PC subspace. T^2 statistic measures the variations in the score space only, which can be interpreted as measuring the systematic variations of the process, and a violation of the threshold would indicate that the systematic variations are out of control. The threshold for T^2 can be computed by the following equation:

$$T_\alpha^2 = \frac{a(n^2-1)}{n(n-a)} F_\alpha(a, n-a) \tag{7}$$

where $F_\alpha(a, n-a)$ is the upper $100\alpha\,\%$ critical point of the F-distribution with m and $(n-m)$ degrees of freedom.

2 Relative-transformation PCA based on Mahalanobis distance

To handle the problem that eigenvalues are equal and energy is non-conservative, a relative transformation based Mahalanobis distance is given, on this basis, a relative-transformation PCA method based on Mahalanobis distance is proposed.

2.1 Relative transformation based on Mahalanobis distance (MRTPCA)

We define a relative transformation on the original space for building a new space whose dimensions are composed of all points in the original space. The newly-created space is called the relative space and generated through relative transformation, for $X \in \mathbf{R}^{n\times m}$,

$$\begin{aligned} &\Gamma : X \to Y \subset \mathbf{R}^{|X|} \\ &\Gamma_X(\boldsymbol{x}_i) = (d_{i1}, d_{i2}, \cdots, d_{i|X|}) = y_i \in Y \end{aligned} \tag{8}$$

where $X = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_{|X|}\}$, $|X|$ is the element number of matrix, $d_{ij} = d(\boldsymbol{x}_i, \boldsymbol{x}_1) = (\boldsymbol{x}_i - \boldsymbol{x}_1)^{\mathrm{T}} \sum^{-1} (\boldsymbol{x}_i - \boldsymbol{x}_1)$ is the Mahalanobis distance between sample i and j of X, here $\sum$ represents covariance of matrix X[12].

Definition 1. Relative transformation based on Mahalanobis distance is denoted by:

$$\boldsymbol{X}^R = \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1n} \\ d_{21} & d_{22} & \cdots & d_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ d_{n1} & d_{n2} & \cdots & d_{nn} \end{bmatrix} \tag{9}$$

Relative transformation based on Mahalanobis distance has some advantages: 1) It is independent of dimension; 2) It can remove the correlations among variables; 3) Mahalanobis distance which is calculated by both standardized data and centralized data are equal.

Theorem 1. Relative transformation is a kind of distance enlarging transformation, i.e.,

$$d(\boldsymbol{x}_i^r, \boldsymbol{x}_j^r) \geq d(\boldsymbol{x}_i, \boldsymbol{x}_j), \ \ \forall \boldsymbol{x}_i, \boldsymbol{x}_j \in X \tag{10}$$

where $\boldsymbol{x}_i^r$ represents relative transformation space.

Proof. $x_i^r = (d(\boldsymbol{x}_i, \boldsymbol{x}_1), d(\boldsymbol{x}_i, \boldsymbol{x}_j), \cdots, d(\boldsymbol{x}_i, \boldsymbol{x}_n))$, $j = 1, \cdots, n$, $d(\boldsymbol{x}_i, \boldsymbol{x}_j) = (\boldsymbol{x}_i - \boldsymbol{x}_j)^{\mathrm{T}} \sum^{-1} (\boldsymbol{x}_i - \boldsymbol{x}_j)$, then,

$$\begin{aligned}
d(\boldsymbol{x}_i^r, \boldsymbol{x}_j^r)^2 = & \sum_{k=1}^{n} \left(d(\boldsymbol{x}_i, \boldsymbol{x}_k) - d(\boldsymbol{x}_j, \boldsymbol{x}_k)\right)^2 = \\
& \sum_{k=1}^{n} ((\boldsymbol{x}_i - \boldsymbol{x}_k)^{\mathrm{T}} \sum^{-1} (\boldsymbol{x}_i - \boldsymbol{x}_k) - \\
& (\boldsymbol{x}_j - \boldsymbol{x}_k)^{\mathrm{T}} \sum^{-1} (\boldsymbol{x}_j - \boldsymbol{x}_k))^2 = \\
& \sum_{k=1, k\neq j}^{n} ((\boldsymbol{x}_i - \boldsymbol{x}_k)^{\mathrm{T}} \sum^{-1} (\boldsymbol{x}_i - \boldsymbol{x}_k) - \\
& (\boldsymbol{x}_j - \boldsymbol{x}_k)^{\mathrm{T}} \sum^{-1} (\boldsymbol{x}_j - \boldsymbol{x}_k))^2 + \\
& ((\boldsymbol{x}_i - \boldsymbol{x}_k)^{\mathrm{T}} \sum^{-1} (\boldsymbol{x}_i - \boldsymbol{x}_k) - \\
& (\boldsymbol{x}_j - \boldsymbol{x}_k)^{\mathrm{T}} \sum^{-1} (\boldsymbol{x}_j - \boldsymbol{x}_k))^2 \mid_{k=j} \ = \\
& \sum_{k=1, k\neq j}^{n} ((\boldsymbol{x}_i - \boldsymbol{x}_k)^{\mathrm{T}} \sum^{-1} (\boldsymbol{x}_i - \boldsymbol{x}_k) - \\
& (\boldsymbol{x}_j - \boldsymbol{x}_k)^{\mathrm{T}} \sum^{-1} (\boldsymbol{x}_j - \boldsymbol{x}_k))^2 + \\
& ((\boldsymbol{x}_i - \boldsymbol{x}_j)^{\mathrm{T}} \sum^{-1} (\boldsymbol{x}_i - \boldsymbol{x}_j))^2 \ \geq \\
& ((\boldsymbol{x}_i - \boldsymbol{x}_j)^{\mathrm{T}} \sum^{-1} (\boldsymbol{x}_i - \boldsymbol{x}_j))^2
\end{aligned}$$

here i, j, $k = 1, 2, \cdots, n$, because $d(\boldsymbol{x}_i, \boldsymbol{x}_j) = (\boldsymbol{x}_i - \boldsymbol{x}_j)^{\mathrm{T}} \times \sum^{-1} (\boldsymbol{x}_i - \boldsymbol{x}_j)$, we can obtain that: $d(\boldsymbol{x}_i^r, \boldsymbol{x}_j^r)^2 \geq d(\boldsymbol{x}_i, \boldsymbol{x}_j)^2$, and $d(\boldsymbol{x}_i, \boldsymbol{x}_j) \geq 0$, $d(\boldsymbol{x}_i^r, \boldsymbol{x}_j^r) \geq 0$, $d(\boldsymbol{x}_i, \boldsymbol{x}_j) \leq d(\boldsymbol{x}_i^r, \boldsymbol{x}_j^r)$. □

Therefore, relative transformation can make the data more distinguishable[12].

Theorem 2. Mahalanobis distance of standardized data is equal to that of centralized data, i.e.,

$$d_m(\tilde{\boldsymbol{x}}_i, \tilde{\boldsymbol{x}}_j) = d_m(\bar{\boldsymbol{x}}_i, \bar{\boldsymbol{x}}_j) \tag{11}$$

here, $d_m(\tilde{\boldsymbol{x}}_i, \tilde{\boldsymbol{x}}_j)$ represents Mahalanobis distance of standardized data and $d_m(\bar{\boldsymbol{x}}_i, \bar{\boldsymbol{x}}_j)$ represents MD of centralized data.

Proof. Given a training set of n observations and m process variables: $X \in \mathbf{R}^{n\times m}$, assuming the data is centralized, so the standardized formula is as follows:

$\tilde{\boldsymbol{x}}_{i,j} = \bar{\boldsymbol{x}}_{i,j}/s_j$ $(i = 1, \cdots, n;\ j = 1, \cdots, m)$, s_j is the standard deviation of jth observation.

$$\begin{aligned}
& d_m(\tilde{\boldsymbol{x}}_i, \tilde{\boldsymbol{x}}_j) = (\tilde{\boldsymbol{x}}_i - \tilde{\boldsymbol{x}}_j)^{\mathrm{T}} \sum^{-1} (\tilde{\boldsymbol{x}}_i - \tilde{\boldsymbol{x}}_j) = \\
& \quad [S^{-1} \cdot (\boldsymbol{x}_i - \boldsymbol{x}_j)]^{\mathrm{T}} (S^{-1} \cdot \sum \cdot (S^{-1})^{\mathrm{T}})^{-1} [S^{-1} \cdot (\boldsymbol{x}_i - \boldsymbol{x}_j)] \times \\
& \quad (\boldsymbol{x}_i - \boldsymbol{x}_j)^{\mathrm{T}} \cdot S^{\mathrm{T}} \cdot (S^{\mathrm{T}})^{-1} \cdot \sum^{-1} \cdot S^{-1} \cdot S \cdot (\boldsymbol{x}_i - \boldsymbol{x}_j) = \\
& \quad (\boldsymbol{x}_i - \boldsymbol{x}_j)^{\mathrm{T}} \sum^{-1} (\boldsymbol{x}_i - \boldsymbol{x}_j) = d_m(\boldsymbol{x}_i, \boldsymbol{x}_j)
\end{aligned}$$

here $\sum$ represents covariance of centralized data matrix. □

2.2 MRTPCA for fault detection

MRTPCA consists of three main steps. The first step is devoted to design the relative space matrix, in which the relative-transformation principal component statistical model is built. The second step demonstrates the derivation of the control limits of T^2 and SPE statistics. The third step explains how to monitor new observations by SPE statistics and T^2 statistics.

Step 1. Build relative principal components statistical model

1) Acquire a training data set $X = [\boldsymbol{x}_1^{\mathrm{T}}, \boldsymbol{x}_2^{\mathrm{T}}, \cdots, \boldsymbol{x}_n^{\mathrm{T}}]$, $\boldsymbol{x}_i \in \mathbf{R}^m$, and then relative space is built by computing Mahalanobis distance between samples, which is described in the following: $\Gamma_X(x_i^*) = (d(x)_{i1}, d(x)_{i2}, \cdots, d(x)_{in}) = \boldsymbol{x}_i^R \in X^R$.

2) Build relative space matrix, $X^R = [d_1, \cdots, d_n]$, $d_i \in \mathbf{R}^n$.

3) Compute the covariance matrix of X^R, eigenvalues, corresponding eigenvector and relative-transformation principal components $t_i^R = (x^R)^{\mathrm{T}} p_i$, here, p_i represents the ith loading vector.

Step 2. Determine the new T^2 and SPE control limits

The data in relative space does not obey Gaussian distribution. Therefore, the control limits of T^2 and SPE can not be obtained by (5) and (7). The kernel-density method[15] does not require any assumption on data distribution, and the monitoring performance is superior to the T^2 of the control limits, in this paper, the control limits are designed by using kernel-density method. Kernel density estimation can be expressed as $f(x, \sigma) = \frac{1}{n\sigma} \sum_{i=1}^{n} \varphi^{-1/2}(\sigma(x - x_i))$ which is random variable of m-dimensional matrix, and $\varphi(x)$ is a kernel function.

Step 3. New process monitoring for observations

1) Given a test data set $X_{\text{test}} \in \mathbf{R}^{N*m}$, N is the number of sample, and m is the number of process variables. Compute Mahalanobis distance between test data and train data, $d_t(x) = D_m(X, X_{\text{test}})$. And the test relative space matrix is built as: $X_t^R = d_t \in \mathbf{R}^{m*n}$.

2) Compute the test relative-transformation Principal Components, $t_{ti}^R = (X_t^R)^{\mathrm{T}} p_i$.

3) T^2 and SPE statistic can be computed by (4) and (6).

2.3 Effectiveness analysis of MRTPCA method

Firstly, relative transformation based on MD is independent of dimension, MD of standardized data is equal to that of centralized data (which is proved in Theorem 2). Therefore, it is indicated that relative transformation based on MD can directly eliminate the effect of dimension and solve the following problems: the loss of information of useful variables, and the difficulty to extract representative PCs, etc.

Secondly, relative transformation based on MD complies with the law of energy conservation.

Theorem 3. Relative transformation based on Mahalanobis distance complies with law of energy conservation[9], i.e.,

$$\|X\|_2^2 = \lambda \left\| X^R \right\|_2^2 \tag{12}$$

Proof. Data set is $X = \{\boldsymbol{x}_i, i = 1, 2, \cdots, n\}$, where $x_i \in \mathbf{R}^m$ is the ith sample, and the weight of data set is represented as $\bar{X} = (\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_n)$. Let, $d_m^2(x_i, x_j) = (d_m(x_i, \bar{X}) + d_m(x_j, \bar{X}))^2 = d_m^2(\boldsymbol{x}_i, \bar{X}) + d_m^2(x_j, \bar{X}) + 2d_m(x_i, \bar{X})d_m(\boldsymbol{x}_j, \bar{X})$, here $d_m(\boldsymbol{x}_i, x_j)$ is the Mahalanobis distance of vector $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$.

$$
\begin{aligned}
& d_m(\boldsymbol{x}_i, \bar{X})d_m(\boldsymbol{x}_j, \bar{X}) = \\
& \quad \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\boldsymbol{x}_i - \bar{\boldsymbol{X}})(\boldsymbol{x}_j - \bar{X})^{\mathrm{T}} = \\
& \quad \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i - \bar{X})\sum_{j=1}^{n}(\boldsymbol{x}_j - \bar{X})^{\mathrm{T}} = \\
& \quad \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i - \bar{X})(n\bar{X} - n\bar{X})^{\mathrm{T}} = 0
\end{aligned}
$$

Then,

$$
\begin{aligned}
& \left\|X^R\right\|_2^2 = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} d_m^2(\boldsymbol{x}_i, \boldsymbol{x}_j) = \\
& \quad \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} d_m^2(\boldsymbol{x}_i, \bar{X}) + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} d_m^2(\boldsymbol{x}_j, \bar{X}) = \\
& \quad \frac{1}{n}\sum_{i=1}^{n} d_m^2(\boldsymbol{x}_i, \bar{X}) + \sum_{j=1}^{n} d_m^2(\boldsymbol{x}_j, \bar{X}) = \frac{2}{n}\sum_{i=1}^{n} d_m^2(\boldsymbol{x}_i, \bar{X}) = \\
& \quad \frac{2}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left\{(\boldsymbol{x}_i - \bar{X})^{\mathrm{T}}\sum^{-1}(\boldsymbol{x}_i - \bar{X})\right\}^2 = \\
& \quad \sum_{j=1}^{n}\left[\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i - \bar{X})^{\mathrm{T}}\sum^{-1}(\boldsymbol{x}_i - \bar{X})\right]^2 = \\
& \quad 2n(\sum^{-1})^2(\sum_{j=1}^{n} s_j^2)^2 = 2n(\sum^{-1})^2(\sum_{j=1}^{n} var(\boldsymbol{x}_j))^2
\end{aligned}
$$

here $\sum$ represents covariance of centralized data matrix. Thanks to the facts that

$$
\begin{aligned}
& (\sum_{j=1}^{n} var(\boldsymbol{x}_j))^2 = \sum_{j=1}^{n}\mathrm{E}\{\boldsymbol{x}_j^2\} = \\
& \quad \mathrm{E}\{\mathrm{tr}(XX^{\mathrm{T}})\} = \mathrm{E}\{\mathrm{tr}(XX^{\mathrm{T}})\} = \mathrm{E}\{\mathrm{tr}(X^{\mathrm{T}}X)\} = \\
& \quad \mathrm{E}\{\mathrm{tr}(X^{\mathrm{T}}PP^{\mathrm{T}}X)\} = \mathrm{E}\{\mathrm{tr}(TT^{\mathrm{T}})\} = \\
& \quad \sum_{j=1}^{n}\mathrm{E}\{\boldsymbol{T}_j^2\} = \|T\|_2^2
\end{aligned}
$$

and due to normal PCA obey conservation of system energy, i.e., $\|T\|_2^2 = \|X\|_2^2$.

So $\left\|X^R\right\| = 2n(\sum^{-1})^2\|X\|_2^2 = \lambda\|X\|_2^2$, here, $\lambda = 2n(\sum^{-1})^2$. □

Thirdly, relative transformation is a kind of distance enlarging transformation, which has been proved in Theorem 1. Relative transformation can make the data more distinguishable. Therefore, extracted principal components using MRTPCA method are more representative.

Example 1. To validate the extracting performance of relative-transformation PCA, Tennessee Eastman process (TEP) data is used. The details on the process description can be found in [16]. Normal data is considered, which consists of 23 process variables. In the same operation condition of modeling data, Fig. 1 and Fig. 2 respectively show t_1 versus t_2 graph (t_1 refers to the first principle component, t_2 refers to the second principle component) using standardized PCA method and MRTPCA method.

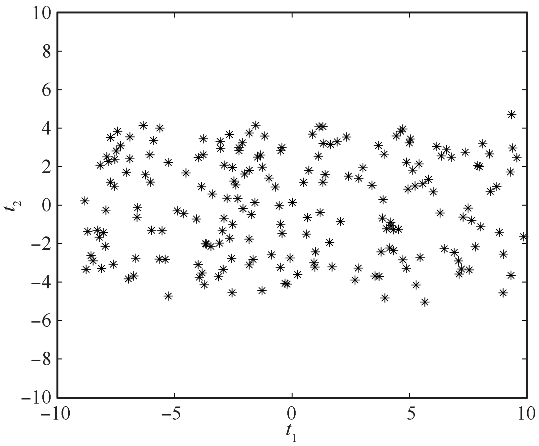


Fig. 1 t_1 versus t_2 graph by standardized PCA method

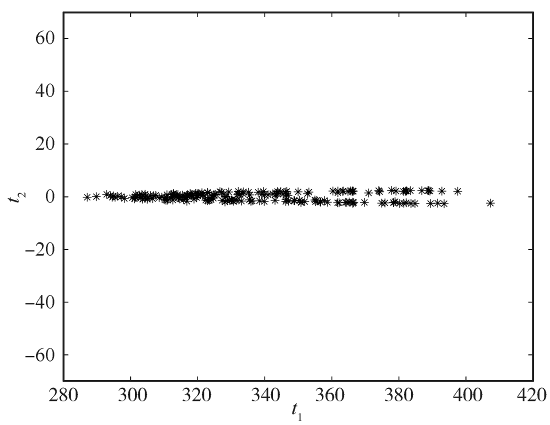


Fig. 2 t_1 versus t_2 graph by MRTPCA method

It is easy to observe that the variation by PCA method along the horizontal axis equals to the vertical axis approximately in Fig. 1, it means that the eigenvalues of covariance matrix are approximately equal. In Fig. 2, the variation by MRTPCA based method along the horizontal axis is much greater than the vertical axis. Moreover, the contribution rate of the first and second principle component based on standardized PCA is 22.5 % and 19.6 %, respectively. The contribution rate of the first principle component based on MRTPCA is 91.5 %, which has obvious representativeness. Therefore, extractive principal components of the proposed method are more representative.

2.4 Complexity analysis of MRTPCA method

In the training stage of MRTPCA, the matrix $X \in \mathbf{R}^{m\times n}$ is transformed into the relative space matrix by relative transformation based on MD. The size of relative space matrix is changed to the square of the number of samples from $(m\times n)$. When the sample number becomes large, the calculation of eigenvalues and eigenvectors in off-line modeling will be time consuming. Therefore, the off-line modeling complexity and training time of the proposed method is more than SPCA method. However, in on-line fault detection process, only 1 principal component can represent change of data in that separability between normal data and fault data is enlarged by relative transformation based on MD. Thus, from the perspectives of on-line detection process, the MRTPCA can dramatically improve real-time performance of on-line detection, especially when the process variables are very large.

3 Improved *SPE* based on Mahalanobis distance

In standardized PCA, SPE and T^2 statistics can be computed by (4) and (6), It can be seen above that T^2 statistics monitors the change of principal component subspace by measuring the controlled condition of process; SPE statistics monitors the change of error subspace by measuring the correlation among each variable of sample data. In terms of distance measure, T^2 statistics represents the distance between relative principle components and clustering center of normal data, namely, $T^2 = (M - \mathbf{0})\Sigma^{-1}(M - \mathbf{0})^{\mathrm{T}}$, where $M = [t_1, t_2, \cdots, t_p]$ is the principal components, $\mathbf{0}$ represents clustering center of normal data. On the other hand, $SPE_i = (I - t_i)(I - t_i)^{\mathrm{T}}$ can be represented by the distance between prediction error and clustering center of normal data, then $SPE = (E - 0)(E - 0)^{\mathrm{T}}$, where E is the prediction error. Measure of SPE and T^2 statistics are different because one is MD measure and the other is Euclidean distance (ED) measure. Therefore, two statistics can not act together to estimate process condition accurately. Furthermore, ED only suits for measuring evenly distributed sample data, but prediction error can hardly obey evenly distribution. The SPE statistics ignores the compactness between the current prediction error and normal prediction error, which will result in data information loss, poor effect of monitoring and low detection rate[11].

MD measure can take full account of the sparsity between the current prediction error and normal prediction error, and MD of the residual vector can measure totally feature of residual subspace. Thus, this paper proposes a new process monitoring for observations which uses the MD of the residual vector instead of SPE statistics,

$$MSPE = (E-0)\Sigma^{-1}(E-0)^{\mathrm{T}} = (I-t_i)\Sigma^{-1}(I-t_i)^{\mathrm{T}} \quad (13)$$

where Σ is the covariance of residual error.

4 MRTPCA based improved *SPE* for fault detection

To further improve the fault detection performance, the current paper combines the MRTPCA and improved SPE to detect the fault more effectively, namely, improved MRTPCA (IMRTPCA), which consists of three main steps. The first step is devoted to the design of relative space matrix using MRTPCA in which the relative-transformation principal component statistical model is built. The second step demonstrates the derivation of the control limits of T^2 and $MSPE$ statistics. The third step explains how to monitor new observations by T^2 and improved $MSPE$. The flowchart of fault detection process based on IMRTPCA is shown in Fig. 3.

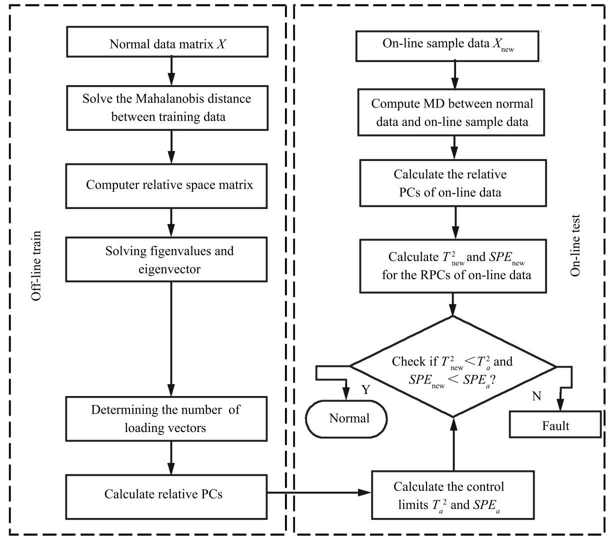


Fig. 3 The flowchart of fault detection and identification process based on IMRTPCA

5 Fault detection for looper system

In this section, the proposed methods are used to detect looper faults of rolling process. the looper system and its faults are first discussed, and then the proposed methods are applied to detect looper faults. The simulation data used in this paper, contain training data and testing data, are from the same production condition in rolling process, and the time interval of training and testing data sampling is very short. As is known, 10 000 sample data are collected, and the sampling period of rolling process data is 10 ms, and hence the time interval of training and testing data sampling is less than 100 s. In conclusion, it is definitely feasible that the single fault data and multi-fault data of looper system in rolling process can be used to verify the effectiveness of the proposed method without considering the slow time-variant characteristic of rolling process.

5.1 Looper system description

Looper system plays a significant role in rolling mill systems, the schematic of the looper system is shown in Fig. 4[17].

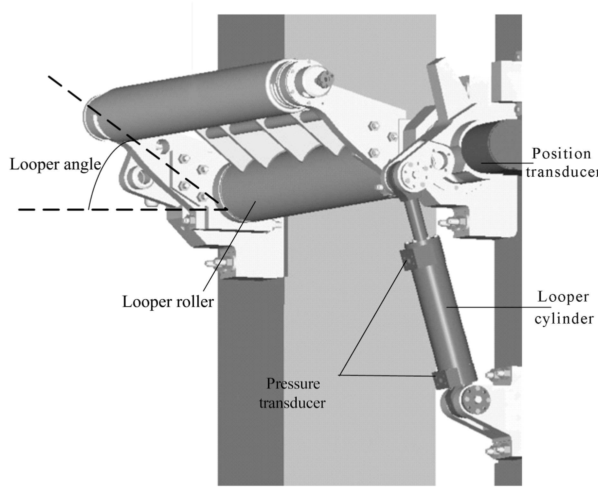


Fig. 4 The structure schematic of the looper system

If the looper fault occurs, continuous rolling process has to be interrupted and a lot of time and manpower is wasted to deal with it for restoring production. The change in physical or structural parameters will give birth to the faults for looper system, in other words, abnormality of looper angle may lead to looper faults, besides, abnormality of rolling force and rolling velocity are also major causes of looper faults[18]. To monitor the looper faults, MRTPCA method, PCA based improved SPE method and a combination based on two improved methods are used. There are 77 variables related with looper system. The process variables and units are shown as Table l.

Table 1 The process variables and units in rolling process

No.	Process variables	Unit
$1\sim6$	Looper torque of $L1\sim L6$	kN·m
$7\sim12$	Looper hydraulic cylinder pressure of $L1\sim L6$	kN
$13\sim18$	Looper angle of $L1\sim L6$	Deg
$19\sim24$	Tensile force of $L1\sim L6$	kN
$25\sim31$	Error force of $F1\sim F7$	kN
$32\sim38$	Up work roll position of $F1\sim F7$	Mm
$39\sim45$	Down work roll position of $F1\sim F7$	Mm
$46\sim52$	Work roll velocity of $F1\sim F7$	m/s
$53\sim59$	Electrical machine torque of $F1\sim F7$	kN·m
$60\sim66$	Force of $F1\sim F7$	kN
$67\sim73$	Bending force of $F1\sim F7$	kN
74	Strip steel velocity of finish rolling entry	m/s
75	Strip steel velocity of finish rolling exit	m/s
76	Strip steel temperature of finish rolling exit	°C
77	Strip steel temperature of finish rolling entry	°C

5.2 Simulation for single fault

Looper shaft deadlocking fault is the common fault of looper system, which is caused by abnormal looper hydraulic cylinder pressure and looper angle. The looper shaft deadlocking fault will lead to shape wave of strip surface and even the strip folded fault and strip breaking fault, and then affect the normal operation of looper system and strip quality. In this section, looper shaft deadlocking fault is considered. There are 200 samples collected in normal condition for modeling, also, there are 200 samples for test, testing data were on-line collected when the looper shaft deadlocking fault occurs, and the fault occurs in 101th sample. All data, normal data and fault data, are collected from actual rolling process.

In this section, we carry out IMRTPCA-based fault detection for looper system, firstly, acquire training data set of looper system: $X=[\boldsymbol{x}_1^{\rm T},\boldsymbol{x}_2^{\rm T},\cdots,\boldsymbol{x}_{200}^{\rm T}]$, $\boldsymbol{x}_i\in\mathbf{R}^{77}$, and then build relative space by computing Mahalanobis distance between samples, $X^R=[d_1,\cdots,d_{200}]$, where $d_i\in\mathbf{R}^{200}$, and compute the covariance matrix of X^R, eigenvalues, corresponding eigenvector and relative-transformation principal components $t_i^R=(x^R)^{\rm T}p_i$, here, p_i represents the ith loading vector. Secondly, obtain the control limits of T^2 and improved SPE statistic by (5) and (7). At last, for on-line collected looper system data, compute Mahalanobis distance between it and training data, and obtain T^2 and improved SPE statistic by (4) and (6). For comparison purpose, the MRTPCA, IMRTPCA, standardized PCA (SPCA) and relative PCA (RPCA)[9] are applied to detect looper shaft deadlocking fault. First, we use a data set of 200 normal samples for training purpose. Second, a test data set of 200 samples is used to test the fault detection ability of MRTPCA, IMRTPCA, SPCA and RPCA. To draw a significative comparison between proposed method, SPCA and RPCA for fault detection, respectively, the 99 % control limits are calculated in each simulation.

For RPCA, we first standardize every variable′s dimension in looper system. Secondly, according to prior information, and analyzes and determines the different important levels of different variables. Then it allocates weights for each variable under the criterion of conservation of system energy. Finally, we use the relative-principal-component model to analyze the system. In this simulation, the relative weight $W=\text{diag}\{[2, 2, 2, 2, 2, 2, 10, 10, 10, 10, 10, 10, 8, 8, 8, 8, 8, 8, 3, 3, 3, 3, 3, 3, 2, 2, 2, 2, 2, 2, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 1, 1, 1, 1, 1, 1, 1, 5, 5, 5, 5, 5, 5, 5, 3, 3, 3, 3, 3, 3, 3, 1, 1, 1, 1]\}$.

The fault detection result obtained from the proposed method (MRTPCA and IMRTPCA), SPCA and RPCA are shown in Figs. $5\sim8$, respectively. The black solid line in each figure is the threshold for the T^2 and improved SPE statistics, the statistics above its threshold indicates that the fault is detected.

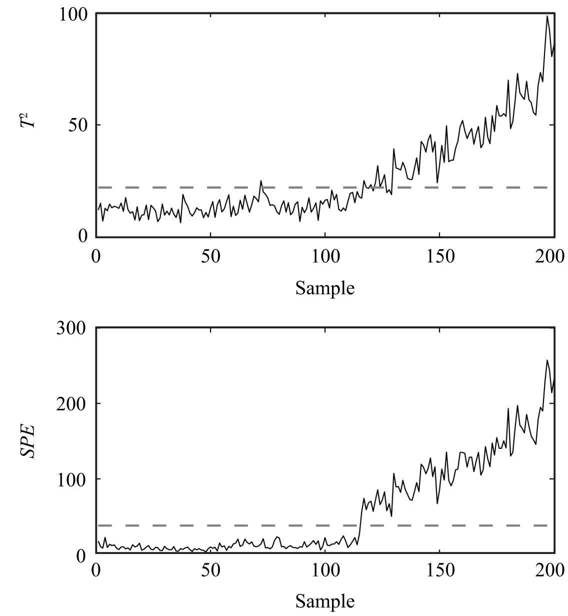


Fig. 5 The SPCA-based T^2 and SPE statistics for fault detection for looper shaft deadlocking fault

It can be seen from Fig. 5, the fault is detected at 126th and 118th sample, and the persistence of the SPCA-based T^2 and SPE statistic is not very satisfactory. It is obvious that the RPCA, MRTPCA and IMRTPCA are the effective methods from Figs. $6\sim8$.

However, MRTPCA and IMRTPCA-based statistics are more sensitive than the RPCA-based statistics, RPCA can detect fault at 105th and 107th sample. And MRTPCA has the similar detection performance as RPCA. The simulation result obtained using improved SPE ($MSPE$) statistic has been compared with traditional SPE statistic, and is shown in Fig. 7 and Fig. 8.

It is concluded that $MSPE$ statistic has shown superior performance of fault detection in terms of detection rate and persistence compared to traditional T^2 statistic and SPE statistic. IMRTPCA-based MSPE statistics is the most sensitive since the SPE statistics is improved by using MD, and $MSPE$ statistics can detect fault in sample 102. And MRTPCA and IMRTPCA-based statistic outperform the RPCA and SPCA-based statistic in terms of exceeding

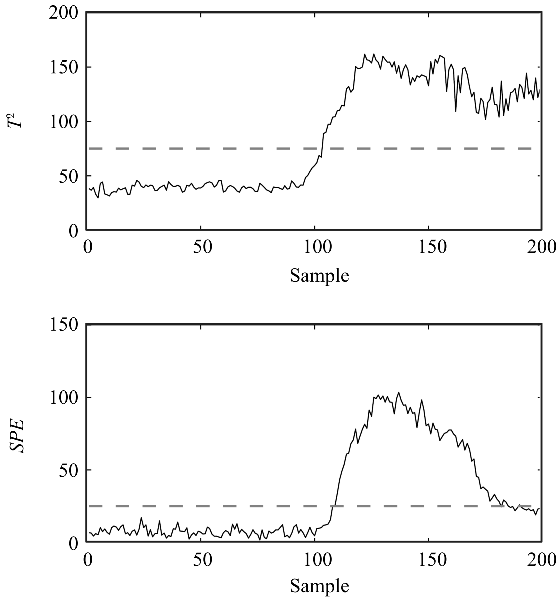


Fig. 6 The RPCA-based T^2 and SPE statistics for fault detection for looper shaft deadlocking fault

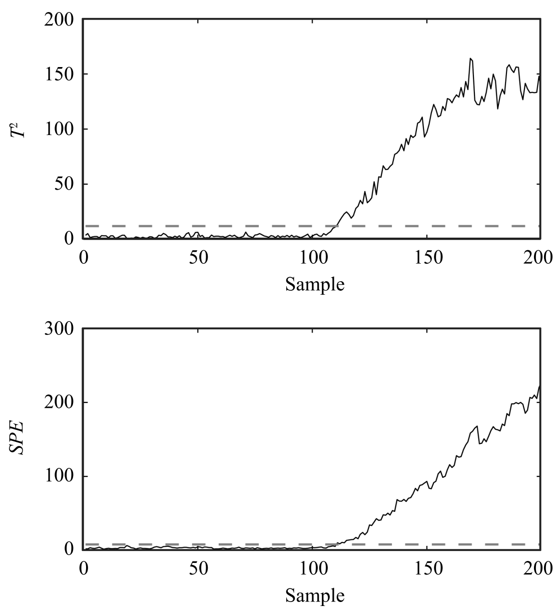


Fig. 7 The MRTPCA-based T^2 and SPE statistics for fault detection for looper shaft deadlocking fault

the threshold by a greater degree.

Moreover, MRTPCA and IMRTPCA have significant effect on computational cost for fault detection. From the perspectives of off-line training model, the scale of relative matrix for MRTPCA and IMRTPCA is 200×200 matrix, and the original matrix is 200×77, it is obvious that the eigen decomposition calculation of MRTPCA and IMRTPCA will be time-consuming. However, in this simulation, under the same fault detection rate, the number of principal components used for MRTPCA and IMRTPCA are only 1, which for SPCA and RPCA are 28 and 17, respectively. Thus, from the perspectives of on-line detection process, the MRTPCA and IMRTPCA can dramatically improve real-time performance of on-line detection, especially when the process variables are very large.

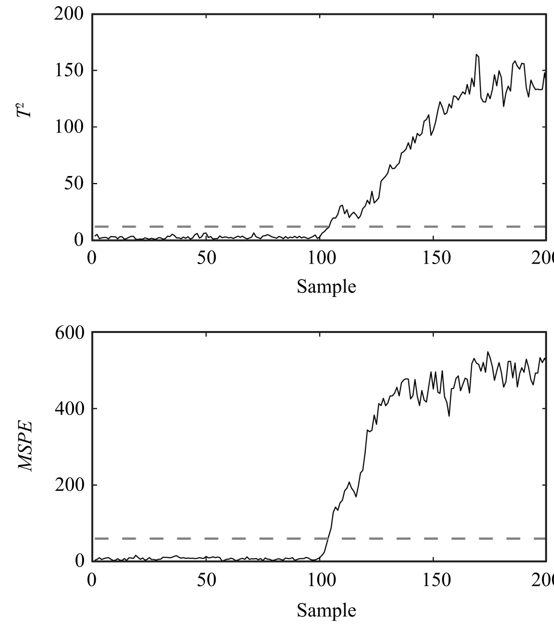


Fig. 8 The IMRTPCA-based T^2 and $MSPE$ statistics for fault detection for looper shaft deadlocking fault

Table 2 Comparisons of fault detection performance for 4 types of methods for single fault

Method	Statistic	Fault detection rate (%)	Off-line modeling time (s)	On-line detection time (s)
SPCA	T^2	87.0	0.330	0.140
SPCA	SPE	91.0	0.330	0.140
RPCA	T^2	97.5	0.300	0.140
RPCA	SPE	96.5	0.300	0.140
MRTPCA	T^2	97.0	1.250	0.080
MRTPCA	SPE	98.0	1.250	0.080
MRTPCA	T^2	97.0	1.250	0.080
MRTPCA	$MSPE$	99.0	1.250	0.080

Fault detection results and performance of 4 types of methods are shown in Table 2 for single fault. The results show that the proposed approach with improved SPE has better detection performance than other methods, although training model is more time-consuming.

In fact, the objective of fault detection is to detect fault as soon and accurate as possible, rather than the off-line modeling time. So off-line modeling time can be ignored. Form Table 2, the potential advantage of applying relative transformation based on MD to remove dimension and improve performance of PCA is clearly shown.

5.3 Simulation for multi-fault

In looper system, there are several faults that often occur simultaneously or within a short amount of time. Multi-faults will lead to coupling of fault feature, which requires more effective fault detection methods. In this section, multi-faults testing data are considered. Training samples and testing samples used in SPCA, RPCA, MRTPCA, and IMRTPCA for comparison are as follows:

1) The training data consists of 400 samples obtained

from actual rolling process.

2) The testing data consists of 380 samples obtained by introducing some different fault signal at different sample instant, the introduced faults are considered as representative, based on experience with the process. The testing data is constructed by the following way, Fault 1 and Fault 2 are introduced between 81th and 160th sample, Fault 3 and Fault 4 are introduced between 201th and 280th sample, Fault 5 is introduced from 351th sample to 380th sample, where Fault 1 stands for increasing gain (by 10 %) of $L2$ angle transducer; Fault 2 represents increasing gain (by 10 %) of $F7$ pressure transducer on piston side; Fault 3 stands for reducing (by 70 bar) system pressure; Fault 4 stands for adding offset (of 5 bar) to pressure transducer on rod side; Fault 5 represents wear of control edges (overlap decreased by -1 %) in servo-valve[16].

This subsection will show that the proposed MRTPCA and IMRTPCA method can extract more useful information from the sampling data and are more sensitive to multi-fault comparison with SPCA and RPCA method.

Fig. 9 is the T^2 and SPE statistics charts of SPCA method for multi-fault case, and Fig. 10 is the T^2 and SPE statistic charts of RPCA for the same case. For RPCA, we allocate weights for each variable under the criterion of conservation of system energy and use the relative-principal-component model established to analyze system. In this simulation, the relative weight is chosen to be the same as that in Subsection 5.2. $W = \mathrm{diag}\{[2, 2, 2, 2, 2, 2, 10, 10, 10, 10, 10, 10, 8, 8, 8, 8, 8, 8, 3, 3, 3, 3, 3, 3, 2, 2, 2, 2, 2, 2, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 1, 1, 1, 1, 1, 1, 1, 5, 5, 5, 5, 5, 5, 5, 3, 3, 3, 3, 3, 3, 3, 1, 1, 1, 1]\}$.

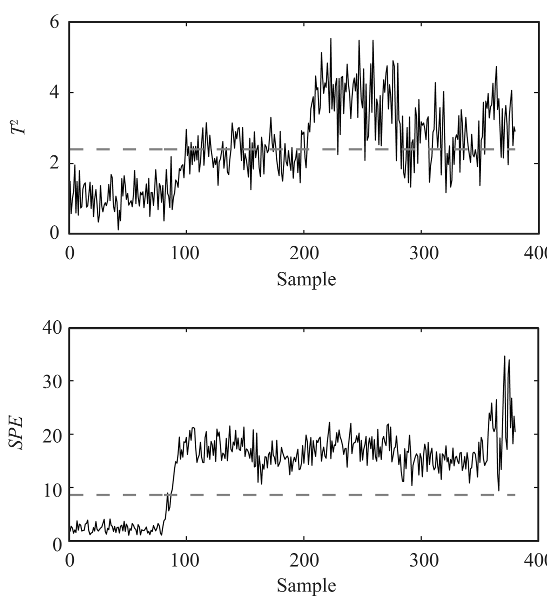


Fig. 9 The PCA-based T^2 and SPE statistics for fault detection for multi-fault

From the two figs, we see that the SPCA method cannot effectively detect fault, whose T^2 and SPE statistics only can detect fault at the beginning, and both can hardly detect fault in other conditions. In Fig. 10, however, the simulation results show that RPCA can extract more fault information and has better fault detection performance than SPCA method, the main cause is that it allocates weights for each variable by analyzing and determining the different important levels of different variables according to prior information. Unfortunately, the RPCA method cannot effectively and directly eliminate effect of dimension and relative weights of RPCA are mainly dependent on prior information, and thus more false alarms will arise in the T^2 and SPE statistic charts.

Fig. 11 is the T^2 and SPE statistic chart of proposed MRTPCA for multi-fault, and Fig. 12 is the T^2 and $MSPE$ chart of proposed IMRTPCA using improved SPE for the same fault. It is exciting to see from Fig. 12 that the IMRTPCA-based statistics is able to detect faults between sample 86 and 160, between sample 205 and 280, and between sample 357 and 370, respectively. And $MSPE$ statistics can detect faults between sample 84 and 160, between sample 201 and 280, and between sample 353 and 374, respectively.

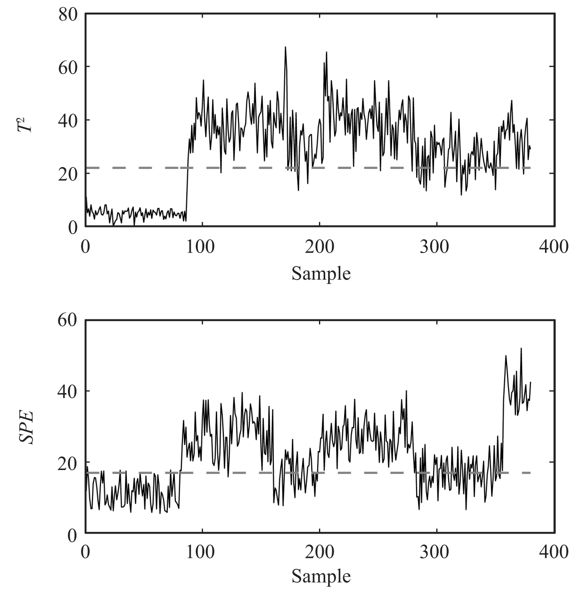


Fig. 10 The RPCA-based T^2 and SPE statistics for fault detection for multi-fault

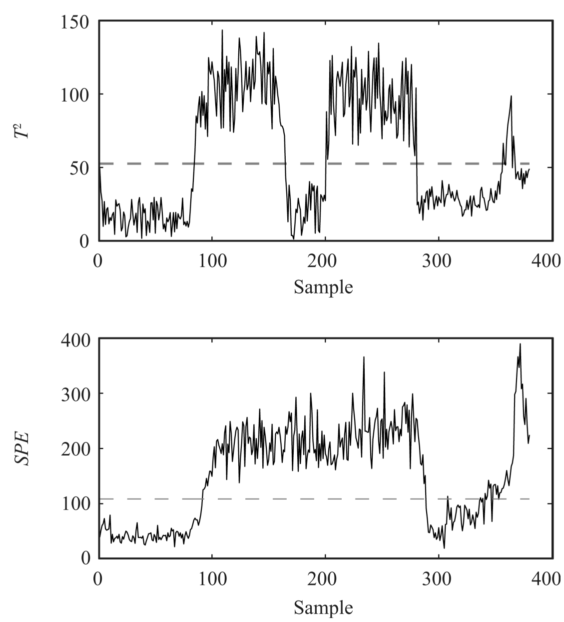


Fig. 11 The MRTPCA-based T^2 and SPE statistics for fault detection for multi-fault

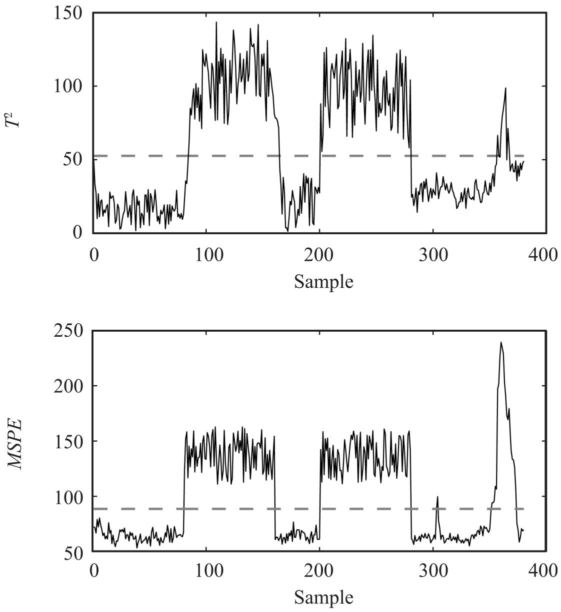


Fig. 12 The IMRTPCA-based T^2 and $MSPE$ statistics for fault detection for Multi-fault

In Fig. 11, however, traditional SPE statistic of MRTPCA false alarms arise and it detects fault only in the beginning of fault and between sample 358 and 380, and can hardly detect fault in other conditions, since traditional SPE statistic is ED measure which only suits for measuring sample data evenly distributed and ignores the compactness between the current prediction error and normal prediction error. Therefore, it results in data information loss, poor effect of monitoring and low detection rate.

Overall, the IMRTPCA-based T^2 and $MSPE$ statistics give much higher detection rate. Therefore, the IMRTPCA statistics are more sensitive and perform better in case of multi-faults than the SPCA, RPCA and MRTPCA statistics.

On the other hand, the SPCA-based cumulative contribution rate of first PC is only 26.79 %, and that of the first 32 principle components is more than 85 %, the number of principal components used for RPCA are 21 as the selected relative principal components are representative by allocating relative weights. Moreover, the IMRTPCA-based cumulative contribution rate of the first relative principle component is more than 85 %, which have meet reached the requirement of feature extraction. So IMRTPCA-based principle component model is simpler and real-time performance of fault detection is higher.

For multi-faults, the IMRTPCA-based statistics are more sensitive than the other methods, and have superior performance for fault detection. The applications in industrial looper system show that the proposed method in this paper can correctly find incipient multi-fault, and can greatly reduce false alarms in SPCA and other improved PCA methods.

Fault detection results and performance of 4 types of methods are shown in Table 3 for the case of multi-fault. The results show that the proposed approach with improved SPE has the best detection performance in all of the fault detection methods, although training model is more time-consuming.

Table 3 Comparisons of fault detection performance for 4 types of methods for multi-fault

Method	Statistic	Fault detection rate (%)	Off-line modeling time (s)	On-line detection time (s)
SPCA	T^2	52.0	0.460	0.210
SPCA	SPE	63.0	0.460	0.210
RPCA	T^2	75.5	0.430	0.210
RPCA	SPE	80.5	0.430	0.210
MRTPCA	T^2	93.5	1.75	0.120
MRTPCA	SPE	86.0	1.75	0.120
MRTPCA	T^2	93.5	1.75	0.12
MRTPCA	$MSPE$	95.5	1.75	0.12

5.4 Conclusion and further work

Over the last few decades, PCA has been heavily researched and widely applied to industrial processes for fault detection, and in this paper, an improved relative-transformation principal component analysis based on Mahalanobis distance method is developed for fault detection. This proposed method focuses on the further improvement of PCA for fault detection from two aspects, i.e. data preprocessing and monitoring index, which are suitable for the detection of single fault and multi-fault. The method in the paper can effectively extract fault information and remove the effect of dimension. The false alarms which exist in conventional PCA and other improved PCA are greatly reduced by the proposed technique. These improved schemes are applied to detect the looper faults in hot continuous rolling process. Some significant contributions are summarized as follows:

1) A relative transformation scheme based on Mahalanobis distance is adopted to remove the dimension of sample data instead of dimensionless standardization, the proposed improved method is applied to detect looper roller distortion fault, the result has shown that, relative-transformation PCA based on Mahalanobis distance can improve the sensitivity performance of fault detection. Furthermore, relative-transformation PCA based on Mahalanobis distance can enhance the on-line detection efficiency, especially when the process variables are very large.

2) An improved SPE ($MSPE$) statistic is given to improve the fault detection performance of standardized PCA, the proposed improved method is applied to detect looper shaft deadlocking fault, the result indicates that $MSPE$ statistic is more suitable for actual industrial process.

3) Two improved methods are combined to improve better performance of fault detection for multi-faults. The combined method, i.e., IMRTPCA, is proposed to detect the multi-faults which is constructed based on experience with the process, the result illustrates that the one relative principal component extracted by combined method is enough to represent most of the feature information of original data, which is much less than that of standardized PCA method. The proposed method is more sensitive than SPCA and other improved PCA method, and has superior performance for fault detection.

References

1 Garcia-Alvarez D, Fuente M J, Sainz G I. Fault detection and isolation in transient states using principal component analysis. *Journal of Process Control*, 2012, **22**(3): 551−563

2 Shi Huai-Tao, Liu Jian-Chang, Ding Xiao-Di, Tan Shuai, Wang Xue-Mei. Fault detection based on hybrid dynamic

principal component analysis. *Control Engineering of China*, 2012, **19**(1): 148−151 (in Chinese)

3 Wen Cheng-Lin, Hu Yu-Cheng. Fault diagnosis based on information incremental matrix. *Acta Automatica Sinica*, 2012, **38**(5): 832−840 (in Chinese)

4 Wu Feng, Zhong Yan, Wu Quan-Yuan. Online classification framework for data stream based on incremental kernel principal component analysis. *Acta Automatica Sinica*, 2010, **36**(4): 534−542 (in Chinese)

5 Dobos L, Abonyi J. On-line detection of homogeneous operation ranges by dynamic principal component analysis based time-series segmentation. *Chemical Engineering Science*, 2012, **75**: 96−105

6 Zhang Y W, Zhou H, Qin S J. Decentralized fault diagnosis of large-scale processes using multiblock kernel principal component analysis. *Acta Automatica Sinica*, 2010, **36**(4): 593−597

7 Luan Fang-Jun, Guo Hong-Mei, Lin Lan, Wang Yong-Hui. Algorithm of face authentication based on wavelet transform and combining two different 2-dimensional participial component analyses. *Journal of Shenyang Jianzhu University* (*Natural Science*), 2010, **26**(5): 1001−1005 (in Chinese)

8 Chiang L H, Russell E L, Braatz R D. *Fault Detection and Diagnosis in Industrial Systems*. Berlin: Springer, 2001

9 Wen Cheng-Lin, Hu Jing, Wang Tian-Zhen, Chen Zhi-Guo. Relative PCA with applications of data compression and fault diagnosis. *Acta Automatica Sinica*, 2008, **34**(9): 1128−1139 (in Chinese)

10 Dunia R, Qin S J, Edgar T F, McAvoy T J. Identification of faulty sensors using principal component analysis. *American Institute of Chemical Engineers Journal*, 1996, **42**(10): 2797−2812

11 Zhao Zhong-Gai, Liu Fei. Application research of statistical monitoring index based on Mahalanobis distance. *Acta Automatica Sinica*, 2008, **34**(4): 493−495 (in Chinese)

12 Wen G H, Jiang L J, Wen J. Local relative transformation with application to isometric embedding. *Pattern Recognition Letters*, 2009, **30**(3): 203−211

13 Li Yuan, Tang Xiao-Chu. Improved performance of fault detection based on selection of the optimal number of principal components. *Acta Automatica Sinica*, 2009, **35**(2): 1550−1557 (in Chinese)

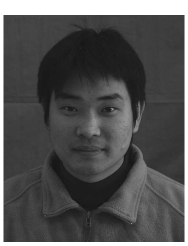
14 Tamura M, Tsujita S. A study on the number of principal components and sensitivity of fault detection using PCA. *Computers and Chemical Engineering*, 2007, **31**(9): 1035−1046

15 Banerjee A, Burlina P. Efficient particle filtering via sparse kernel density estimation. *IEEE Transactions on Image Processing*, 2010, **19**(9): 2480−2490

16 Shi H T, Liu J C, Zhang Y, Li L. Fault detection method based on relative-transformation partial least squares. *Chinese Journal of Scientific Instrument*, 2012, **33**(4): 816−822

17 Marcu T, Köppen-Seliger B, Stücher R. Design of fault detection for a hydraulic looper using dynamic neural networks. *Control Engineering Practice*, 2008, **16**(2): 192−213

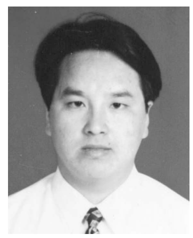
18 Yue Yu-Mei, Ma Ya-Li, Wang Jiao-Qing. Research about fault intelligent diagnosis of high speed rotating machinery. *Journal of Shenyang Jianzhu University (Natural Science)*, 2005, **21**(6): 770−773 (in Chinese)



SHI Huai-Tao Received his bachelor degree from Northeastern University in 2005, and his Ph. D. degree in control theory and control engineering at Northeastern University in 2012. Since 2012, he has been a lecturer. His research interest covers fault diagnosis in complex industrial processes and their applications.
E-mail: shihuaitao_0408@126.com

LIU Jian-Chang Received his bachelor degree in automation from Northeastern University in 1983, and his Ph. D. degree in control theory and control engineering at Northeastern University in 1999. Since 2004, he has been a professor. His research interest covers fault detection and diagnosis in complicated industrial processes, intelligent control theory and applications.
E-mail: liujianchang@ise.neu.edu.cn

XUE Peng Received his bachelor and master degrees from Northeastern University in 2005, and 2008, respectively. Since 2008, he has been an engineer in Baoshan Iron & Steel Co., Ltd.. His research interest covers automation modeling and fault diagnosis of rolling process.
E-mail: xuepeng@baosteel.com



ZHANG Ke Received his bachelor and master degrees in mechanical engineering from Shenyang Jianzhu University in 1992 and 1999, respectively, and his Ph. D. degree in mechanical engineering at Northeastern University in 2007. Since 2006, he has been a professor at Shenyang Jianzhu University. His research interest covers intelligent control and diagnosis of industrial process, high-accuracy processing technology and equipment.
E-mail: zhangke@sjzu.edu.cn

WU Yu-Hou Received his bachelor degree in mechanical engineering from Shenyang Jianzhu University in 1982, and his Ph. D. degree in mechanical engineering at Northeastern University in 1994. Since 1994, he has been a professor in the Institute of Mechanical Engineering and Automation, Shenyang Jianzhu University. His research interest covers intelligent control and fault detection, ceramic motorized spindle development and their applications. Corresponding author of this paper.
E-mail: wuyh@sjzu.edu.cn

ZHANG Li-Xiu Received her master degree in mechanical engineering from Shenyang Jianzhu University in 2002, and her Ph. D. degree in mechanical engineering at Dalian University of Technology in 2012. Her research interest covers fault diagnosis in complicated industrial processes, control and optimization of ceramic motorized spindle system.
E-mail: sanyoe@sina.com

TAN Shuai Received her bachelor degree from Northeastern University in 2005, and her Ph. D. degree in control theory and control engineering from Northeastern University in 2012. Her research interest covers fault diagnosis in batch industrial processes and their applications.
E-mail: tanshuai@ise.neu.edu.cn