DOI: 10.3901/CJME.2014.0529.106, available online at www.springerlink.com; www.cjmenet.com; www.cjmenet.com.cn

ASCS Online Fault Detection and Isolation Based on an Improved MPCA

PENG Jianxin[1, 2], LIU Haiou[1, *], HU Yuhui[1], XI Junqiang[1], and CHEN Huiyan[1]

1 *Science and Technology on Vehicle Transmission Laboratory, Beijing Institute of Technology, Beijing 100081, China*
2 *The State Key Laboratory of the Vehicle Transmission, China North Vehicle Research Institute, Beijing 100072, China*

Received April 7, 2013; revised April 15, 2014; accepted May 29, 2014

Abstract: Multi-way principal component analysis (MPCA) has received considerable attention and been widely used in process monitoring. A traditional MPCA algorithm unfolds multiple batches of historical data into a two-dimensional matrix and cut the matrix along the time axis to form subspaces. However, low efficiency of subspaces and difficult fault isolation are the common disadvantages for the principal component model. This paper presents a new subspace construction method based on kernel density estimation function that can effectively reduce the storage amount of the subspace information. The MPCA model and the knowledge base are built based on the new subspace. Then, fault detection and isolation with the squared prediction error (SPE) statistic and the Hotelling (T^2) statistic are also realized in process monitoring. When a fault occurs, fault isolation based on the SPE statistic is achieved by residual contribution analysis of different variables. For fault isolation of subspace based on the T^2 statistic, the relationship between the statistic indicator and state variables is constructed, and the constraint conditions are presented to check the validity of fault isolation. Then, to improve the robustness of fault isolation to unexpected disturbances, the statistic method is adopted to set the relation between single subspace and multiple subspaces to increase the corrective rate of fault isolation. Finally fault detection and isolation based on the improved MPCA is used to monitor the automatic shift control system (ASCS) to prove the correctness and effectiveness of the algorithm. The research proposes a new subspace construction method to reduce the required storage capacity and to prove the robustness of the principal component model, and sets the relationship between the state variables and fault detection indicators for fault isolation.

Keywords: multi-way principal component analysis (MPCA), fault detection, fault isolation, automatic shift control system

1 Introduction

In many industrial processes, computerized operator systems for on-line diagnosis can play an important role in supporting operators with expert knowledge during production. An important aspect of the safe operation of an industrial process is the rapid detection of faults, process deviations, or other special events and the location and removal of the factors causing such events. However, hundreds of process variables may exist, and these variables may be recorded many times per day. Therefore, a method is required that can project a high-dimensional space into low-dimensional spaces for better analysis of data. This paper presents a new multivariate procedure for monitoring the progress of all such industrial batch processes.

Principal component analysis (PCA) is a viable option for online fault detecting, which is suitable for complex processes with a large number of highly correlated process variables[1]. Until now, PCA has been used and extended to various applications[2–4].

NOMIKOS, et al[5], presented muti-way PCA (MPCA) for monitoring batch processes in 1994. MPCA, an extension of PCA, was investigated to monitor batch processes dynamically, which has previously proven to be effective in capturing the dynamic pattern for a defined duration in batch process applications[6]. MAJID, et al[7], presented a new framework based on MPCA to detect faults in real-time in the industrial continuous aluminum electrolysis process, which incorporates the dynamic behavior of two important operations in the continuous aluminum electrolysis process, alumina feeding and anode changing. WANG, et al[8], proposed a new stage separation method for online monitoring and fault variable detection in a three-tank system based on the similarity of the loading matrices and singular value matrices in the same operation substage. MPCA can also be used to monitor the performance of batch control processes by building performance bounds based on data from batch runs operated at optimum conditions[9].

In terms of vehicle fault detection, XI, et al[10], adopted a redundancy analysis method and an action analysis method to detect failures in the automated mechanical transmission (AMT) system, and a fault diagnostic system of an air-drive AMT in a heavy commercial vehicle is developed by a model-based design that is well adapted to the development

* Corresponding author. E-mail: liuhaiou30@163.com
Supported by National Hi-tech Research and Development Program of China (863 Program, Grant No. 2011AA11A223)

of an automotive control system. AKIN, et al[11], presented a signal processing-based motor fault diagnosis scheme for electric and hybrid electric vehicle power train applications. QIN, et al[12], proposed an analytical redundancy fault diagnosis method of an automatic shift control system (ASCS). The information redundancy in the local models of the power train and the structural logic relations among the assemblages is used to detect and diagnose faults in the sensors, actuators, and unit assemblages of an ASCS. LIU, et al[13], analysed the main causes for the different types of faults and presented three fault diagnostic methods based on the construction and working principle of an AMT for a heavy-duty vehicle. PENG, et al[14], proposed a fault detection and diagnosis strategy for ASCS based on the behaviour trajectory of the ASCS's hybrid system.

In a PCA application, WEI, et al[15], proposed PCA-based fault detection methods for vertical rail vehicle suspension systems. When there is a detectable fault, the detector sends an alarm signal if the residual evaluation is larger than a predefined threshold. ROUTRAY, et al[16], proposed a data-driven method to detect anomalies in operating parameter identifiers (PIDs) and in the absence of any anomaly, to classify faults in automotive systems by analysing the PIDs collected from the freeze frame data. Ramahaleomiarantsoa[17] used PCA to address the problem of fault detection in an electric vehicle engine (EVE).

But there exist some potential problems with these applications. First, one assumption of some methods is batch process with strict periodicity, meaning that all batches have an equal duration and are synchronized[18]. Second, there is a need to reduce the storage amount required by historical data with high sampling frequency when the online detection calculator has limited storage capacity. Last, the Mahalanobis distance T^2 is often adopted to calculate the control limit for fault detection. When a fault occurs, the mapping relationship between the monitored indicators and the measured state variables must be established to realize the fault isolation function. This paper proposed a new subspace construction method to reduce the required storage capacity and to prove the robustness of the principal component model to the measurement noise. In fault isolation, the relationship between the control limit T^2 and the system state variables is built when a fault occurs, and the constraints of fault isolation algorithm are analyzed.

2 Construction of Historical Data Subspace

To understand the nature of the data available with which to monitor batch processes, consider a typical batch run in which $m = 1, 2, \cdots, M$ variables are measured at $l = 1, 2, \cdots, L$ time intervals throughout the batch. Similar batch will exist in a number of such batch runs $n = 1, 2, \cdots, N$. All of the historical data can be summarized in the three-dimensional matrix $\boldsymbol{X}(N \times M \times L)$. MPCA is equivalent to unfolding the three-dimensional matrix $\boldsymbol{X}(N \times M \times L)$ into a two dimensional matrix $\boldsymbol{X}(N \times ML)$ by cutting each of its vertical slices $(N \times M)$ side by side to the right, starting with the one corresponding to the first time interval, as shown in Fig. 1.

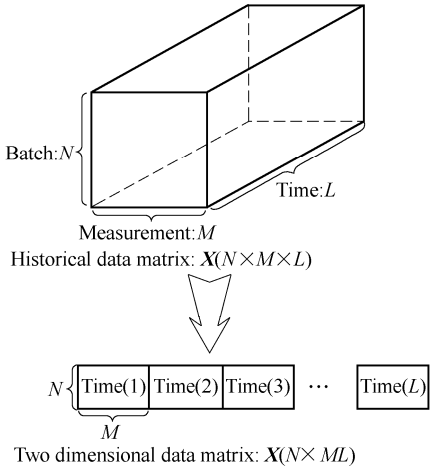


Fig. 1. Unfolding the historical data matrix

In batch processes, the number of measurement sensor is assumed to be fixed, but different batches may have different sample points. Therefore, sensor number M is invariable in the two-dimensional data matrix $\boldsymbol{X}(N \times M \times L)$. Because the same batch process is under normal operation conditions, the length of all the batches obeys a normal distribution, where the central limit theorem is effective. Batch number N is decided by the batches whose length is between $(\bar{L} - 3\delta)$ and $(\bar{L} + 3\delta)$, where $\bar{L}$ is average length of all the faultless batches and δ is the standard deviation.

The batch length L changes from batch to batch. To solve the problem of different batch lengths, the data subspace is adopted to describe the characteristics of the historical data matrix at a different sample point, which is created by moving window along the time axis step by step. In Fig. 2, if the width of moving the window includes d sample points, the subspace data $\boldsymbol{X}_i$ at time i can be described as:

$$\boldsymbol{X}_i = [\boldsymbol{x}_i, \boldsymbol{x}_{i+1}, \cdots, \boldsymbol{x}_{i+d-1}], \tag{1}$$

where $\boldsymbol{x}_i$ is the data matrix $(N \times M)$ at i sample point $(i = 1, 2, \cdots, L - d + 1)$. If a batch length is shorter than other batches at some sample point, the batch is removed from the subspace and other batches are selected as the subspace data. In this manner, the problem of different batch lengths is effectively resolved, throughout the subspace reconstruction in the moving window along the time axis.

In process fault detection, the interval between adjacent sample points decides the shortest response time of fault detection, because the system state variables are assumed to be constant during the interval. If a fault occurs in the interval, an overly large sampling interval may cause more

serious damage because the fault is not detected in a timely manner. Therefore, the process sampling frequency is typically and sufficiently high to describe dynamic process changes and guarantee the speed of fault detection. As the number of sample points increase, the three-dimensional data matrix $\boldsymbol{X}(N\times M\times L)$ becomes bigger and the degree of data redundancy increases because the process state variables do not change considerablely over the relatively short intervals. To address these issues, a new sub space construction method based on the kernel density estimation function is presented here.

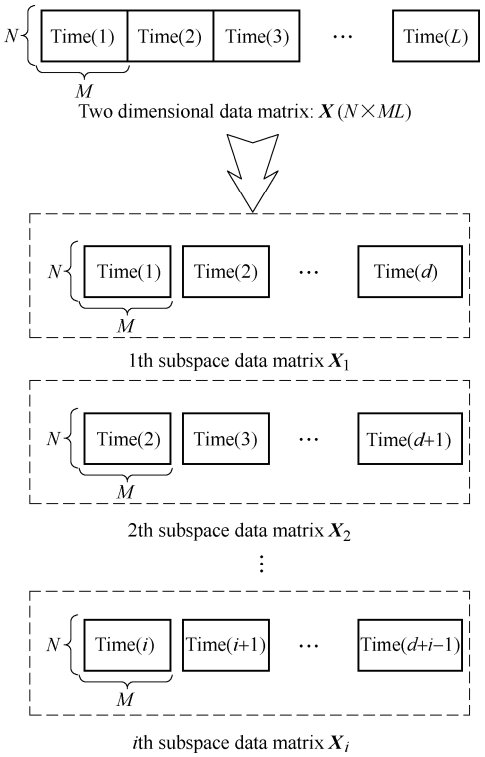


Fig. 2. Construction of the subspace data matrix

As shown in Fig. 2, every subspace matrix $\boldsymbol{X}_i$ includes d time slices, which are composed of d matrixes. At high sampling frequencies the process variables do not change over the window width. Therefore, this paper attempts to construct a new matrix $\boldsymbol{X}_i^*$ to replace the original matrix $\boldsymbol{X}_i$ with a high degree of redundancy. The new matrix $\boldsymbol{X}_i^*$ is composed of the accumulation of d time slices with corresponding weights which are produced by kernel density estimation function $K(j)$. The selection condition of function $K(j)$ can be described as:

$$\int_{-\infty}^{\infty} K(j)=1. \tag{2}$$

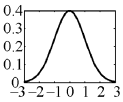
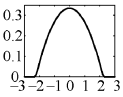
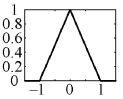
In the ith subspace, the new matrix $\boldsymbol{X}_i^*$ can be obtain by

$$\boldsymbol{X}_i^*=\frac{\sum_{i=1}^{d}K\left(i-\frac{d}{2}\right)\boldsymbol{x}_i}{\sum_{i=1}^{d}K\left(i-\frac{d}{2}\right)}, \tag{3}$$

where i is the number of d time slices.

The common kernel functions, including Gaussian, Epanechnikov, and Triangular are shown in Refs. [19–20]. The expressions and curve shapes of these kernel functions are given in Table 1.

Table 1. Expression and curve shape

Function names	Expressions	Curve shapes
Gaussian	$K(j)=\frac{1}{\sqrt{2\pi}}\exp(-0.5j^2)$	
Epanechnikov	$K(j)=\begin{cases}\frac{3}{4\sqrt{5}}\left(1-\frac{1}{5}j^2\right), & \lvert j\rvert<\sqrt{5},\\ 0, & \lvert j\rvert\geqslant\sqrt{5}.\end{cases}$	
Triangular	$K(j)=\begin{cases}1-\lvert j\rvert, & \lvert j\rvert<1,\\ 0, & \lvert j\rvert\geqslant 1.\end{cases}$	

When a batch process does not contain any faults, the sensor measurement error and noise typically follow a normal distribution. Thus, this paper employs the Gaussian kernel function to construct the new matrix $\boldsymbol{X}_i^*$ by Eq. (3) in Fig. 3. A comparison of the original matrix $\boldsymbol{X}_i$ and the new matrix $\boldsymbol{X}_i^*$ demonstrates that the dimension of every subspace is reduced from $(N\times Md)$ to $(N\times M)$, which means the storage needs of the PCA model are decreased. The PCA model produced by the new matrix $\boldsymbol{X}_i^*$ is more robust to measurement noise than the PCA model by the original matrix $\boldsymbol{X}_i$, particularly when every subspace has a large number of time slices with high data redundancy. In addition, the requirements for the PCA calculation are substantially reduced due to the reduced matrix dimensions of the subspace.

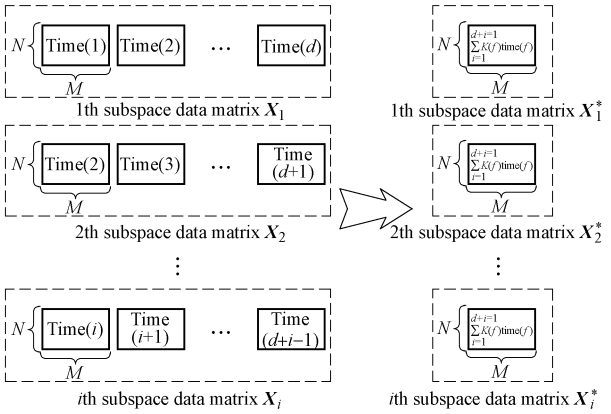


Fig. 3. Construction of new matrix $\boldsymbol{X}_i^*$

3 MPCA Fault Detection

As discussed in section 2, every subspace has a new matrix $\boldsymbol{X}_i^*$, which is used to produce the PCA model for fault detection. First, to eliminate the scaling effects from different variables, the normalisation matrix $\bar{\boldsymbol{X}}_i^*$ can be obtained by $\boldsymbol{X}_i^*$ as follows:

$$\bar{\boldsymbol{X}}_i^* = [\boldsymbol{X}_i^* - (1,1,\cdots,1)_{1\times m}^{\mathrm{T}}\boldsymbol{U}_{1\times m}]\mathrm{diag}\left(\frac{1}{s_1},\frac{1}{s_2},\cdots,\frac{1}{s_m}\right), \quad (4)$$

where $\boldsymbol{U}_{1\times m}$ is the mean vector, s_m is the standard deviation, and $\bar{\boldsymbol{X}}_i^*$ is the normalisation matrix of $\boldsymbol{X}_i^*$.

The PCA algorithm then decomposes the matrix $\bar{\boldsymbol{X}}_i^*$ as the sum of the outer product of vectors $\boldsymbol{t}_j$ and $\boldsymbol{p}_j$ plus the residual matrix $\boldsymbol{E}_i$:

$$\bar{\boldsymbol{X}}_i^* = \boldsymbol{T}\boldsymbol{P}^{\mathrm{T}} + \boldsymbol{E} = \sum_{j=\boldsymbol{1}}^{p}\boldsymbol{t}_j\boldsymbol{p}_j^{\mathrm{T}} + \sum_{j=1}^{m-p}\boldsymbol{t}_j\boldsymbol{p}_j^{\mathrm{T}} = \hat{\boldsymbol{X}}_i + \boldsymbol{E}_i, \quad (5)$$

where $\boldsymbol{p}_j$ is a loading vector containing the information of the projection space, and $\boldsymbol{t}_j$ is a score vector describing data intensity in the direction of projection $\boldsymbol{p}_j$ [21]. Both the loading vectors and score vectors are orthogonal. The dimension of the principal component matrix $\hat{\boldsymbol{X}}_i$ is typically the number of the eigenvalue when the accumulation of the contribution rate of the eigenvalue exceeds 85%.

During the fault detection process, every sampling point has a corresponding subspace that can be divided into two parts: the principal component space and the residual space. Therefore, the process state variable at some sampling point can be projected into the principal component to obtain the corresponding score vectors $\boldsymbol{t}_j$ and residual matrix $\boldsymbol{E}_i$.

According to the residual space, squared prediction error (SPE) is adopted to monitor whether the process residual matrix is consistent with the historical model. The SPE statistical indicator is given as:

$$\boldsymbol{SPE}_i = \|\boldsymbol{E}_i\|^2 = (\bar{\boldsymbol{X}}_i^* - \hat{\boldsymbol{X}}_i)(\bar{\boldsymbol{X}}_i^* - \hat{\boldsymbol{X}}_i)^{\mathrm{T}} = \bar{\boldsymbol{X}}_i^*(I - \boldsymbol{P}\boldsymbol{P}^{\mathrm{T}})\bar{\boldsymbol{X}}_i^{*\mathrm{T}}, \quad (6)$$

where $\boldsymbol{P}$ is the loading vector matrix of the principal component space $\hat{\boldsymbol{X}}_i$ and I is the identity matrix. The upper confidence limit of the SPE statistical indicator[5] can be described as:

$$SPE_\alpha = \theta_1\left[\frac{C_\alpha\sqrt{2\theta_2 h_0^{\,2}}}{\theta_1} + 1 + \frac{\theta_2 h_0(h_0-1)}{\theta_1^{\,2}}\right]^{\frac{1}{h_0}}, \quad (7)$$

$$h_0 = 1 - \frac{2\theta_1\theta_3}{3\theta_2^2}, \quad (8)$$

$$\theta_r = \sum_{j=p+1}^{n}\lambda_j^r, r = 1, 2, 3, \quad (9)$$

where C_α is the standard normal deviation corresponding to the upper $(1-\alpha)$ percentile and j is the eigenvalue number of the loading vector.

According to the principal component space, statistical indicator $\boldsymbol{T}^2$ is given to measure the variation within the PCA model. At the ith sampling point, $\boldsymbol{T}_i^2$ is given as:

$$\boldsymbol{T}_i^2 = \left\|\sum{}^{-0.5}\hat{\boldsymbol{X}}_i\right\|^2 = \sum_{j=1}^{p}\frac{\boldsymbol{t}_j^{\mathrm{T}}\boldsymbol{t}_j}{\lambda_j}, \quad (10)$$

where $\boldsymbol{t}_j$ is the score vector of the principal component matrix.

From the F-distribution, the upper confidence limit is obtained as:

$$T^2{}_{p,n,\alpha} = \frac{p(n-1)}{n-p}F_{p,n-1,\alpha}, \quad (11)$$

where n is the number of samples in the historical data, α is the confidence threshold percentile, and p is the number of principal components.

4 MPCA Fault Isolation

Fault causes an unpermitted deviation of at least one characteristic property or parameter of the system from acceptable/standard condition[22]. And fault isolation is to isolate the process state variable that causes an unacceptable deviation of the system characteristic property after fault detection.

4.1 Fault isolation based on the SPE statistic

Assuming that the SPE statistical indicator is greater than the confidence threshold at the ith sample point, it is believed that process state is abnormal and a fault occurs at that moment. According to Eq. (6), the SPE statistical of the mth state variable can be calculated by

$$\boldsymbol{SPE}_{im} = (\bar{\boldsymbol{X}}_{im}{}^* - \hat{\boldsymbol{X}}_{im})^2, m = 1, 2, \cdots, M, \quad (12)$$

where $\bar{\boldsymbol{X}}_{im}{}^*$ is the measured value of the mth state variable at the ith sample point, and $\hat{\boldsymbol{X}}_{im}$ is the estimated value of the mth state variable at the ith sample point.

When a fault occurs, the deviation caused by the fault occupies a major influence factor in the deviation. Therefore, the variable with the greatest contribution to the SPE statistical indicator is most likely the failure source (FS), which can be described as:

$$FS = \{x_m \mid \max(\boldsymbol{SPE}_{im})\}, m = 1, 2, \cdots, M, \quad (13)$$

where x_m is the mth state variable.

4.2 Fault isolation based on the $\boldsymbol{T}^2$ statistic

According to Eq. (10), if $\boldsymbol{T}_i^2$ beyond the upper limit $T^2{}_{p,n,\alpha}$ at the ith sample point, a fault is believed to occur. The statistical indicator $\boldsymbol{T}_i^2$ is calculated by the eigenvalue λ_j and score vector $\boldsymbol{t}_j$, which is the projection of process state variables in the principal component space. Therefore,

the mapping relations between the score vector $\boldsymbol{t}_j$ and the process state variables must be established for fault isolation.

Assuming that the *j*th process state variable have a fault at the *i*th sample point. Then, Eq. (8) indicates that $\boldsymbol{t}_j^{\mathrm{T}}\boldsymbol{t}_j/\lambda_j$ is the largest contribution to the statistical indicator $\boldsymbol{T}_i^2$. According to Eq. (5), this value can be expressed as

$$\frac{\boldsymbol{t}_k^{\mathrm{T}}\boldsymbol{t}_k}{\lambda_k}=\frac{(\bar{\boldsymbol{X}}_{1\times M}\bullet\boldsymbol{p}_k)^2}{\lambda_k}=\frac{(\bar{x}_{11}\bullet p_{1k}+\cdots+\bar{x}_{1m}\bullet p_{mk})^2}{\lambda_k}, \quad (14)$$

where $\bar{\boldsymbol{X}}_{1\times M}$ is the measured value of the process state variable, $\boldsymbol{p}_k$ is the loading vector, $\bar{x}_{1m}$ is the element of matrix $\bar{\boldsymbol{X}}_{1\times M}$, p_{mk} is the element of matrix $\boldsymbol{p}_k$, and λ_k is the eigenvalue $(m=1,2,\cdots,M;k=1,2,\cdots,p)$. Eq. (14) can be simplified as:

$$\frac{\boldsymbol{t}_k^{\mathrm{T}}\boldsymbol{t}_k}{\lambda_k}=\frac{(\bar{x}_{1j}\bullet p_{jk})^2+\Delta}{\lambda_k}\approx\frac{(\bar{x}_{1j}\bullet p_{jk})^2}{\lambda_k}, \quad (15)$$

where Δ is ignorable error. The deviation caused by faults is the main part of an unacceptable deviation from the process. Therefore, the contribution between the different score vectors can be described as

$$\left(\frac{\boldsymbol{t}_1^{\mathrm{T}}\boldsymbol{t}_1}{\lambda_1}\right):\left(\frac{\boldsymbol{t}_2^{\mathrm{T}}\boldsymbol{t}_2}{\lambda_2}\right):\cdots:\left(\frac{\boldsymbol{t}_k^{\mathrm{T}}\boldsymbol{t}_k}{\lambda_k}\right)\approx\frac{{p_{j1}}^2}{\lambda_1}:\frac{{p_{j2}}^2}{\lambda_2}:\cdots:\frac{{p_{jk}}^2}{\lambda_k}. \quad (16)$$

Eq. (16) can also be given as

$$(\lambda_1\boldsymbol{t}_1^{\mathrm{T}}\boldsymbol{t}_1):(\lambda_2\boldsymbol{t}_2^{\mathrm{T}}\boldsymbol{t}_2):\cdots:(\lambda_k\boldsymbol{t}_k^{\mathrm{T}}\boldsymbol{t}_k)\approx {p^2}_{j1}\lambda_1:{p^2}_{j2}\lambda_2:\cdots:{p^2}_{jk}\lambda_k. \quad (17)$$

In the principal component space, the factor loading matrix $\boldsymbol{A}$ is used to describe the fluctuations in the process state variables assigned to different score vectors $(m=1,2,\cdots,M;k=1,2,\cdots,p)$:

$$A_{mk}=p_{mk}\sqrt{\lambda_k}, \quad (18)$$

$$\sum_{i=1}^{p}A_{mk}^2=1. \quad (19)$$

By inserting Eq. (18) into Eq. (17), the equation can be obtained as

$$\lambda_1\boldsymbol{t}_1^{\mathrm{T}}\boldsymbol{t}_1:\lambda_2\boldsymbol{t}_2^{\mathrm{T}}\boldsymbol{t}_2:\cdots:\lambda_k\boldsymbol{t}_k^{\mathrm{T}}\boldsymbol{t}_k\approx {p^2}_{j1}\lambda_1:{p^2}_{j2}\lambda_2:\cdots:{p^2}_{jk}\lambda_k=A^2{}_{j1}:A^2{}_{j2}:\cdots:A^2{}_{jk}. \quad (20)$$

Because every subspace has its own loading vectors and eigenvalue, Eq. (20) proves the relationship between the score vector and the factor loading matrix $\boldsymbol{A}$. When a fault occurs at the *i*th sample point, judging by the statistical indicator $\boldsymbol{T}^2$, it first calculates a different score vector as the form shown in Eq. (17) and sorts from largest to smallest. Then, the factor loading matrix $\boldsymbol{A}$ of the *i*th subspace is calculated according to Eq. (18). Finally the state variable whose factor loading vector $\boldsymbol{A}_j$ satisfies the relation in Eq. (20) is most likely to be the fault source that caused the main deviation of the process state variables. To improve the sequence matching success in the case of noise and error, the maximum value matching method is adopted. In this method, the sequence for searching for the fault variable only includes the largest value. The largest value in the sequence carries the main deviation, which is produced by a fault.

In the fault isolation process, there are two constraints to guarantee the effectiveness of the diagnostic results:

(1) The error Δ ignored in Eq. (15) does not change the sequence shown in Eq. (16). It can also be expressed as the consistency between the sequences

$$\frac{(\bar{x}_{1j}\bullet p_{j1})^2+\Delta}{\lambda_1}:\frac{(\bar{x}_{1j}\bullet p_{j2})^2+\Delta}{\lambda_2}:\cdots:\frac{(\bar{x}_{1j}\bullet p_{jm})^2+\Delta}{\lambda_m}, \quad (21)$$

and

$$\frac{(\bar{x}_{1j}\bullet p_{j1})^2}{\lambda_1}:\frac{(\bar{x}_{1j}\bullet p_{j2})^2}{\lambda_2}:\cdots:\frac{(\bar{x}_{1j}\bullet p_{jm})^2}{\lambda_m}. \quad (22)$$

If the score value of different loading vectors are close to each other and the error Δ is ignored, the result of the fault isolation may be incorrect because of the sequence change.

(2) The fault isolation result is unique. According to the factor loading matrix $\boldsymbol{A}$, every process state variable has its own factor loading vector $\boldsymbol{A}_j$ which describes the projection relationship between the state variable and the different loading vector $\boldsymbol{p}_j$. Thus, *M* state variables have *M* sequences, and there may exist two or more sequences that can match the sequence calculated in Eq. (17). In that case, the fault isolation cannot be determined between the matched sequences.

5 MPCA Algorithm Flow

MPCA fault detection and isolation include offline and online parts. The offline part is used to preprocess the historical data and generate the principal component model in a different subspace, as shown in Fig. 4. Three dimensional historical data $\boldsymbol{X}(N\times M\times L)$ are unfolded along the time axis to obtain a two-dimensional matrix $\boldsymbol{X}(N\times ML)$ for subspace construction. Based on a moving fixed-width window, matrix $\boldsymbol{X}(N\times ML)$ is decomposed into multiple subspace matrices $\boldsymbol{X}_i$. In the *i*th subspace, a kernel function is adopted to reconstruct matrix $\boldsymbol{X}_i$ to obtain a new subspace matrix $\boldsymbol{X}_i^*$. In order to eliminate the different variable scaling effect, the new subspace matrix $\boldsymbol{X}_i^*$ is normalized for matrix $\bar{\boldsymbol{X}}_i^*$. Then, the principal component algorithm is used to extract the

eigenvector and eigenvalue from the normalized matrix $\overline{\boldsymbol{X}}_i^*$. During the loop, the eigenvector, eigenvalue, mean vector and variance vector in the *i*th subspace are stored for online fault detection.

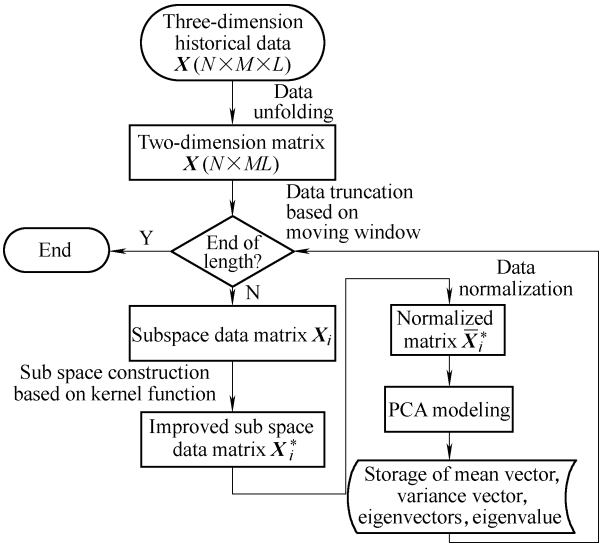


Fig. 4. MPCA offline algorithm flow

After the offline part, each subspace's principle component model can be built and saved. These models are then used for fault detection and isolation during the online monitoring of the process. In Fig. 5, the process sensor measurement $\boldsymbol{X}_i^*$ at the *i*th sampling point is handled by Eq. (4) with the mean vector and the variance vector of the corresponding subspace. Then, the SPE statistic indicator and the $\boldsymbol{T}^2$ statistic indicator are calculated separately for fault detection. If the statistic indicator is beyond the upper confidence limit, a fault occurs and the algorithm of fault isolation presented in section 4 is adopted to find the fault source in the process state variables.

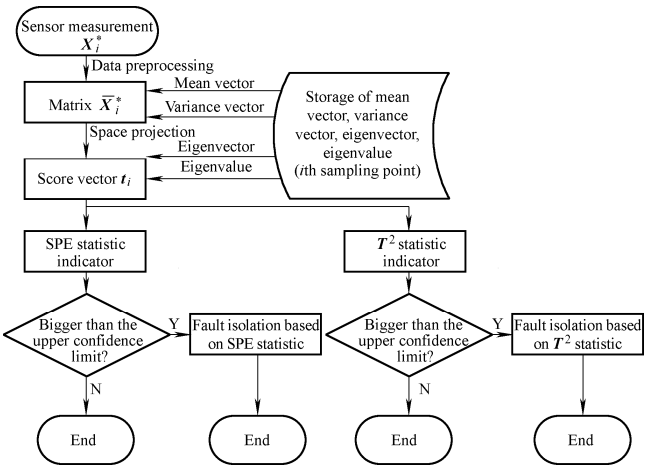


Fig. 5. MPCA online algorithm flow

6 Fault Detection and Isolation Experiments

To verify the accuracy of the algorithm, improved MPCA fault detection and isolation are adopted to monitor the automatic shift control system in an AMT vehicle. The ASCS, being the outcome of upgrading the traditional vehicular transmission system by applying high technologies, is the typical application of integrated technology of mechanism, hydraulics and electronics to wheeled vehicle, which has experienced significant growth worldwide in recent years. The ASCS of wheeled vehicle can be divided into several parts, including sensors, electronic control unit (ECU) and actuators[23–25].

In Fig. 6, the basic working principal of an AMT system is shown to have sensors to collect the system state information, an ECU to generate commands and actuators to simulate the driver control to respond and complete commands by clutch engagement and disengagement, shifting gears, and engine speed control, etc.

1) The ECU is the core of the ASCS, it controls the actuators based on the signals from sensors and can also communicate with the engine ECU through a CAN bus.

2) The sensors reflect the vehicle's running status and the driving intentions from the driver, which are arranged in different sections of the vehicle, such as the multifunctional handled gear selector, vehicle speed sensor and the clutch on-off status sensor.

3) Finally, the actuators realize automatic control of the transmission gear ratio and the clutch engaging and disengaging control.

Fig. 6 presents a composition diagram of an ASCS. The ECU computes the shifting points based on the current vehicle speed and the accelerator pedal position, and then generates the control instructions to drive the gear selecting and shifting actuator, the automatic clutch actuators and the engine coordinating control via the CAN bus to fulfill the automatic shifting control task.

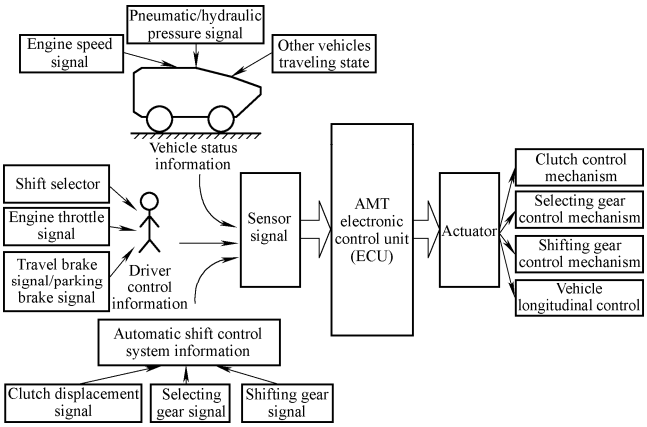


Fig. 6. AMT system structure diagram

Thus, to monitor the ASCS performance while the AMT vehicle is running, the process state variables are selected based on the AMT working principal in Table 2. The ECU temperature x_3 and electric current x_6 are important parameters to reflect the ECU working state. The Oil pressure x_4 is measured by the hydraulic oil source which provides hydraulic power for the ASCS. There are three hydraulic cylinders for selecting gear position, shifting gear

and clutch control. Then, the corresponding mechanism displacement x_1, x_2 and x_5 can be measured for process monitoring, as shown in Table 2.

Table 2. Process state variables

Variable symbol	Variable physical meaning	Physical units
x_1	Shifting gear displacement	mm
x_2	Selecting gear displacement	mm
x_3	ECU temperature	°
x_4	Oil pressure	MPa
x_5	Clutch displacement	mm
x_6	ECU electric current	A

As an example to monitor the vehicle driving in gear 1, 100 batches in gear 1 without fault are selected as the historical data to the build MPCA model. The length of these batches varies with normal variations, including an approximate duration of 300 sample points, as shown in Fig. 7.

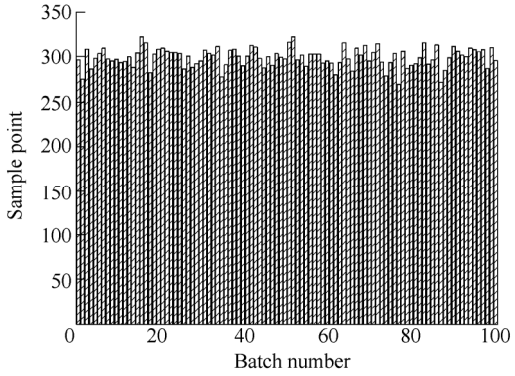


Fig. 7. Sample points in different batch

6.1 Comparison between traditional MPCA and improved MPCA

To prove the superiority of the new subspace construction method, the same historical data in Fig. 7 are used to build the MPCA model respectively based on the traditional MPCA and improved MPCA mentioned in section 2, respectively. Then, new process data without a fault are selected to test these MPCA models, and the results are shown in Figs. 8(a) and 8(b). A comparison of Figs. 8(a) and 8(b) indicates that the model based on the improved MPCA is more robust to noise than the traditional MPCA model, which effectively decreases the disturbance on fault detection and reduce the storage amount for the subspace.

Fault detection using the traditional and improved MPCA is compared in Figs. 9(a) and 9(b) for the case in which a fault occurs. The results indicate that the indicator from the improved MPCA is more sensitive to faults than the traditional MPCA.

6.2 Fault detection and isolation with SPE statistic based on the improved MPCA

Fig. 10 demonstrates that an abnormal process with a length of 285 sample points exceeds the SPE confidence limit at the 98th sampling point. Therefore, the deviation of the process state variables is not consistent with the residual model of the subspace at that moment, and a fault occurs. As mentioned in section 5, a fault detection result is sent into the fault isolation module for fault diagnosis when failure occurs. The contribution rate of the different state variables to the SPE statistical indicator is shown in Fig. 11 for fault isolation. There is no discernible rule regarding the contribution rate before the 98th sampling point. However, after the 98th sampling point, the contribution rate of the process state variable x_5 is much larger than that of the other variables, indicating that x_5 resulted in the SPE statistical indicator bigger than the confidence limit and is most likely to be the fault source. The artificial diagnostic result is that the plectrum of the clutch displacement sensor becomes rusty and deformed, which results in the error in the clutch displacement signal x_5 shown in Fig. 12.

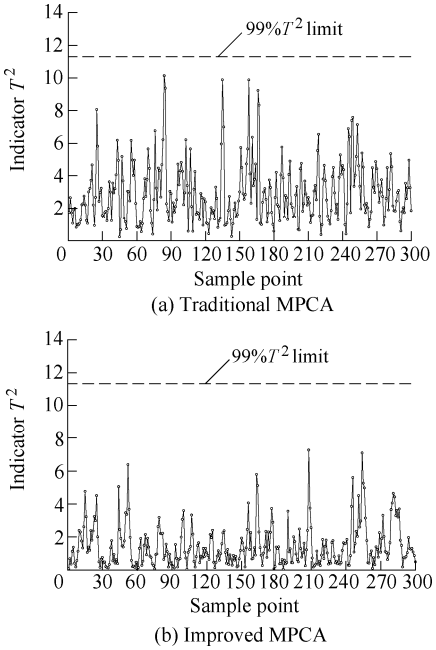


Fig. 8. Comparison of the traditional and improved MPCA with faultless process data

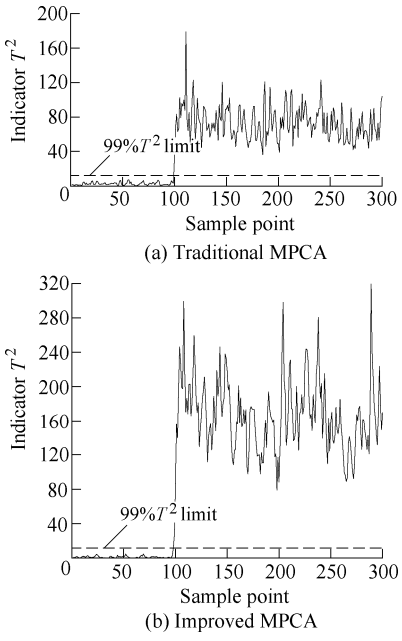


Fig. 9. Comparison of the traditional and improved MPCA with a fault in the process data

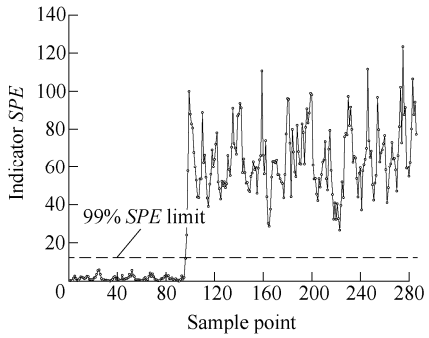


Fig. 10. SPE fault detection based on improved MPCA model

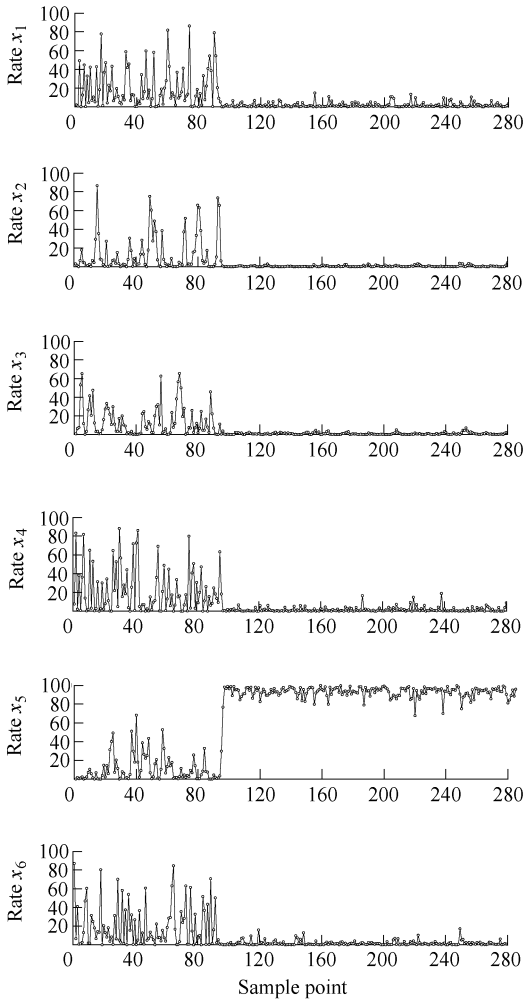


Fig. 11. Contribution rate of the process state variables to SPE

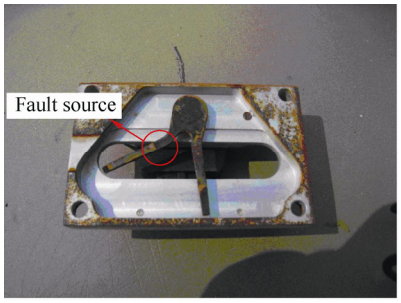


Fig. 12. Fault result

6.3 Fault detection and isolation with T^2 statistic based on the improved MPCA

To compare the fault isolation between *SPE* statistic and T^2 statistic, the abnormal process data in section 6.2 is used as the object of fault isolation for the T^2 statistic.

Fig. 13 demonstrates that the abnormal process with the length of 285 sample points exceeds the T^2 confidence limit at the 98th sampling point. Thus, a fault occurs at the 98th sampling point. After fault detection the result is sent into the fault isolation module. In the 98th subspace, it has three principal components according to the contribution rate of the eigenvalue. The fault isolation module can then calculate the relation between the different score vectors according to Eq. (16):

$$\lambda_1 \boldsymbol{t}_1^{\mathrm{T}} \boldsymbol{t}_1 : \lambda_2 \boldsymbol{t}_2^{\mathrm{T}} \boldsymbol{t}_2 : \lambda_3 \boldsymbol{t}_3^{\mathrm{T}} \boldsymbol{t}_3 = 0.22 : 0.68 : 0.08. \tag{23}$$

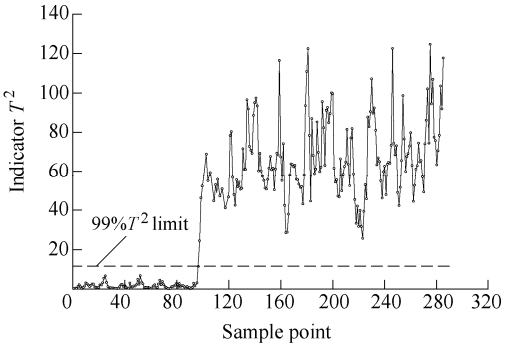


Fig. 13. Fault detection based on the T^2 static

In the 98th subspace, there are six state variables x_i and three principal component loading vectors $\boldsymbol{p}_i$. Therefore, the factor loading matrix $\boldsymbol{A}_{6\times3}$ can be calculated according to Eq. (18). This paper identifies one state variable between x_i $(i=1,2,\cdots,6)$ to satisfy the relationship shown in Eq. (20). x_5 is identified by

$$A^2_{51} : A^2_{52} : A^2_{53} = 0.24 : 0.70 : 0.05. \tag{24}$$

The consistency between Eq. (23) and Eq. (24) proves that the state variable x_5 is the fault source. However, the isolation result must be detected according to the conditions mentioned in section 4.2.

(1) Error Δ detection. According to Eq. (23) and Eq. (24), neglecting the error Δ doesn't change the following sequence:

$$\frac{(\bar{x}_{15} \bullet p_{51})^2 + \Delta}{\lambda_1} : \frac{(\bar{x}_{15} \bullet p_{52})^2 + \Delta}{\lambda_2} : \frac{(\bar{x}_{15} \bullet p_{53})^2 + \Delta}{\lambda_3} = \frac{(\bar{x}_{15} \bullet p_{51})^2}{\lambda_1} : \frac{(\bar{x}_{15} \bullet p_{52})^2}{\lambda_2} : \frac{(\bar{x}_{15} \bullet p_{53})^2}{\lambda_3}. \tag{25}$$

Therefore, the isolation result neglecting the error Δ is effective.

(2) Uniqueness of the fault isolation result. In a factor loading matrix $\boldsymbol{A}_{6\times3}$, only the sequence of the state

variable x_5 satisfies the relation in Eq. (20).

A statistical method is adopted to improve the corrective rate of fault isolation. As seen in Fig. 13, every sample point has its own principal component model, so every sample point has its own fault isolation. Therefore, after a fault occurs, the isolation results of each subspace are counted to determine the most likely fault source. Fig. 14 presents the rate at which state variable x_5 is the fault source after the 98th sampling point. Initially, the curve fluctuates wildly Because a poor fault isolation result has a significant impact on the rate when there are a small number of isolation results. However, as the number of isolation results increase, the rate curve becomes smooth and clearly exhibits an increasing trend with the rate at which the state variable x_5 is the fault source approaching 95%. As a result this statistic method effectively improves the robustness of fault isolation to unexpected disturbance.

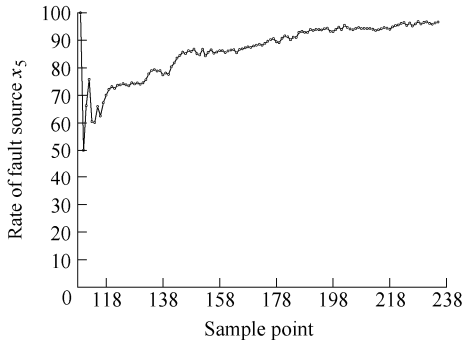


Fig. 14. Rate at which x_5 is the fault source

The result of fault isolation between SPE statistic and $\boldsymbol{T}^2$ statistic is consistent, thus validating the correctness of the improved MPCA algorithm.

7 Conclusions

(1) The structure of a new subspace based on the kernel density estimation function is presented according to traditional MPCA algorithm. The new subspace reduces the storage required and is more robust to noise, as demonstrated by the experiments in section 6.

(2) A fault detection experiment based on the SPE and $\boldsymbol{T}^2$ statistics is presented. The fault detection result demonstrates the accuracy and effectiveness of the algorithm.

(3) Fault isolation based on the SPE statistic is realized by an analysis of the different variable residual contributions when a fault occurs.

(4) The algorithm of fault isolation based on the $\boldsymbol{T}^2$ statistic is established in section 4.2. The relationship between the statistic indicator and state variables is constructed and presented in Eq. (20). When a fault occurs, the most likely fault source in the process state variables can be ensured by sequence matching according to the contribution of the different score vectors. Then, it is necessary to check whether the result of the fault isolation satisfies the two constraints in section 4.2 to guarantee the validity of the fault source.

(5) Every sample point after a fault has its own subspace and fault isolation result. A statistic method is adopted in Fig. 13 to improve the fault isolation's robustness to unexpected disturbances. As a result, it sets the relationship between a single subspace and multiple subspaces to increase the corrective rate of fault isolation.

References

[1] JACKSON J E. Principal components and factor analysis. I. Principal components[J]. *Journal of Quality Technology*, 1980, 12(4): 201–213.

[2] KOURTI T, NOMIKOS P, MACGREGOR J F. Analysis, monitoring and fault diagnosis of batch processes using multiblock and multiway PLS[J]. *Journal of Process Control*, 1995, 5(4): 277–284.

[3] WISE B M, GALLAGHER N B. The process chemometrics approach to process monitoring and fault detection[J]. *Journal of Process Control*, 1996, 6(6): 329–348.

[4] LI W H, YUE H H, VALLE-CERVANTES S, et al. Recursive PCA for adaptive process monitoring[J]. *Journal of Process Control*, 2000, 10(5): 471–486.

[5] NOMIKOS P, MACGREGOR J F. Monitoring batch processes using multiway principal component analysis[J]. *AIChE Journal*, 1994, 40(8): 1361–1375.

[6] ZHANG Y, DUDZIC M S. Industrial application of multivariate SPC to continuous caster start-up operations for breakout prevention[J]. *Control Engineering Practice*, 2006, 14(11): 1357–1375.

[7] MAJID N A A, TAYLOR M P, CHEN J J J, et al. Aluminium process fault detection by Multiway Principal Component Analysis[J]. *Control Engineering Practice*, 2011, 19(4): 367–379.

[8] WANG S, CHANG Y, YANG J, et al. Multiway principle component analysis monitoring and fault variable detection based on substage separation for batch processes[J]. *Control Theory & Amp; Applications*, 2011, 28(2): 149–156.

[9] CHEN J, WANG W. Performance monitoring of MPCA-based control for multivariable batch control processes[J]. *Journal of the Taiwan Institute of Chemical Engineers*, 2010, 41(4): 465–474.

[10] XI J Q, XU M. Research of air-drive AMT fault diagnostic system based on models[C/CD]//*2008 IEEE Vehicle Power and Propulsion Conference (VPPC)*, Piscataway, NJ, USA, Sept. 3–5, 2008.

[11] AKIN B, OZTURK S B, TOLIYAT H A, et al. DSP-based sensorless electric motor fault-diagnosis tools for electric and hybrid electric vehicle powertrain applications[J]. *IEEE Transactions on Vehicular Technology*, 2009, 58(6): 2679–2688.

[12] QIN G H, GE A L, WANG F Y. On-board fault diagnosis system of automated manual transmission control system[C]//*Proceedings of the 2003 IEEE International Conference on Intelligent Transportation Systems*, Piscataway, NJ, USA, Oct. 12–15, 2003: 932–937.

[13] LIU H, ZENG A, SONG W. Program design of on-line fault diagnosis for ASCS of AMT in a heavy-duty vehicle[J]. *Applied Mechanics and Materials*, 2011, 109: 437–440.

[14] PENG J, LIU H, CHEN H. Automated manual transmission fault diagnosis based on hybrid system characteristic[J]. *Journal of Mechanical Engineering*, 2012, 48(19): 72–79. (in Chinese)

[15] WEI X, JIA L, LIU H. Data-driven fault detection of vertical rail vehicle suspension systems[C]//*2012 UKACC International Conference on Control (CONTROL)*, Piscataway, NJ, USA, Sept. 3–5, 2012: 589–594.

[16] ROUTRAY A, RAJAGURU A, SINGH S. Data reduction and clustering techniques for fault detection and diagnosis in automotives[C]//*2010 IEEE International Conference on*

Automation Science and Engineering (CASE 2010), Piscataway, NJ, USA, Aug. 21–24, 2010: 326–331.

[17] RAMAHALEOMIARANTSOA J F, HERAUD N, BENNOUNA O, et al. Modeling fault diagnosis system for electric vehicles[C]// *38th Annual Conference on IEEE Industrial Electronics Society, IECON 2012*, Montreal, QC, Canada, October 25–28, 2012: 4127–4132.

[18] ZHAO L, CHAI T, WANG G. Double moving window MPCA for online adaptive batch monitoring[J]. *Chinese Journal of Chemical Engineering*, 2005, 13(5): 649–655.

[19] KAYRI M, ZIRHLIOGLU G. Kernel smoothing function and choosing bandwidth for non-parametric regression methods[J]. *Ozean Journal of Applied Sciences*, 2009, 6(5): 606–612.

[20] SILVERMAN B W. *Density estimation for statistics and data analysis*[M]. United States of America: Chapman & Hall, 1999.

[21] LEE J, YOO C, LEE I. Statistical process monitoring with independent component analysis[J]. *Journal of Process Control*, 2004, 14(5): 467–485.

[22] ISERMANN R, BALLE P. Trends in the application of model based fault detection and diagnosis of technical processes[C]//*Proceedings of the 13th World Congress*, Oxford, UK, June 30–July 5, 1996: 1–12.

[23] XI J Q, DING H R, CHEN H Y. The history and present status of ASCS and AMT and their development trend in China[J]. *Automotive Engineering*, 2002(2): 89–93. (in Chinese)

[24] QIN G H, GE A L. Fault self-diagnosis techniques for automotive electronic control system[J]. *Automotive Engineering*, 2000, 22(4): 286–285.

[25] PENG J X, LIU H O, HU Y H, et al. Research on fault detection and diagnosis strategy based on AMT system behavior[C]//*2012 IEEE International Conference on Vehicular Electronics and Safety (ICVES 2012)*, Piscataway, NJ, USA, July 24–27, 2012: 186–190.

Biographical notes

PENG Jianxin, born in 1986, is currently a PhD candidate at *Science and Technology on Vehicle Transmission Laboratory, Beijing Institute of Technology, China*. His research interests include vehicle transmission system control and fault diagnosis.
Tel: 13488666110; E-mail: flesher@126.com

LIU Haiou, born in 1975, is currently an associate professor at *Science and Technology on Vehicle Transmission Laboratory, Beijing Institute of Technology, China*. Her research interests include vehicle transmission system control and fault diagnosis.
E-mail:liuhaiou30@163.com

CHEN Huiyan, born in 1961, is currently professor at *Science and Technology on Vehicle Transmission Laboratory, Beijing Institute of Technology, China*. His research interest is intelligent vehicle system.
E-mail: chen_h_y@263.net