Korean J. Chem. Eng., **31**(11), 1943-1953 (2014)
DOI: 10.1007/s11814-014-0153-9

pISSN: 0256-1115
eISSN: 1975-7220

Hyperplane distance neighbor clustering based on local discriminant analysis for complex chemical processes monitoring

Chunhong Lu, Shaoqing Xiao, and Xiaofeng Gu†

Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), Department of Electronic Engineering, Jiangnan University, Wuxi 214122, P. R. China
(*Received 18 January 2014 • accepted 2 June 2014*)

Abstract−The collected training data often include both normal and faulty samples for complex chemical processes. However, some monitoring methods, such as partial least squares (PLS), principal component analysis (PCA), independent component analysis (ICA) and Fisher discriminant analysis (FDA), require fault-free data to build the normal operation model. These techniques are applicable after the preliminary step of data clustering is applied. We here propose a novel hyperplane distance neighbor clustering (HDNC) based on the local discriminant analysis (LDA) for chemical process monitoring. First, faulty samples are separated from normal ones using the HDNC method. Then, the optimal subspace for fault detection and classification can be obtained using the LDA approach. The proposed method takes the multimodality within the faulty data into account, and thus improves the capability of process monitoring significantly. The HDNC-LDA monitoring approach is applied to two simulation processes and then compared with the conventional FDA based on the *K*-nearest neighbor (KNN-FDA) method. The results obtained in two different scenarios demonstrate the superiority of the HDNC-LDA approach in terms of fault detection and classification accuracy.

Keywords: Chemical Process Monitoring, Fault Detection and Classification, Local Discriminant Analysis, Hyperplane Distance Neighbor Clustering, Tennessee Eastman Process

INTRODUCTION

Effective and reliable monitoring plays an important role in ensuring safe plant operation, stable product quality, reduced maintenance and improved profit margin in complex chemical processes. Once abnormal events occur during production, early detection helps to avoid serious incidents [1]. Multivariate statistical process monitoring (MSPM) techniques, such as principal component analysis (PCA) and partial least squares (PLS), have been widely applied to chemical process monitoring [2-6]. These techniques can project the measurement data from the historical process space (original input space) to the low-dimensional linear subspace with the second-order statistic, covariance or correlation information. Consequently, the process faults can be identified through the derived Hotelling's T^2 and squared prediction error (SPE) statistics [3,7]. The MSPM methods based on PCA or PLS usually assume that the process data obey a Gaussian distribution. However, actual industry processes exhibit non-Gaussian characteristics due to the process nonlinearity, operating condition shift and production strategy change [8].

Independent component analysis (ICA) has been developed to deal with non-Gaussian processes [1,9-11]. By characterizing the normal operating region of statistically independent components, the ICA method extracts the inherent non-Gaussianity and reveals more hidden features. The Gaussian mixture model (GMM) has been proposed for monitoring non-Gaussian processes, which assumes that the process data follow one of the multiple Gaussian distributions with different means and covariances at a fixed prior probability [12,13]. Some machine learning models such as artificial neural network (ANN) have also been employed to address non-Gaussian processes [14-16].

The common feature of all aforementioned methods is that a set of purely normal process data are required to train the fault-free model. Nevertheless, because the plant operations contain different types of faults, the collected training data are usually mixed with both normal and faulty samples. Alternatively, the supervised monitoring methods such as Fisher discriminant analysis (FDA) [17,18] and support vector machine (SVM) [19,20] can handle such modeling problems. The SVM can divide the collected data into two classes by maximizing the margin between support vector planes, and the FDA can search the leading directions with maximized discrimination between normal and faulty data. For these two methods, the class labels of training data are assumed to be known so that their practical applications are restricted.

In this study, we propose a novel monitoring strategy for complex chemical processes. First, the hyperplane distance neighbor clustering (HDNC) is proposed to separate normal samples and different types of faulty ones without class labels. Then the local discriminative algorithm (LDA) is used to search for the characteristic direction with maximized hetero-cluster separability and minimized homo-cluster compactness as well as preserving local information. Finally, the effectiveness of the HDNC-LDA method is verified through the well-known Tennessee Eastman process (TEP) and wastewater treatment process (WWTP) and compared to that of the conventional FDA based on the *K*-nearest neighbor (KNN-FDA) method.

†To whom correspondence should be addressed.
E-mail: xgu@jiangnan.edu.cn

METHODOLOGY

1. Hyperplane Distance Neighbor Clustering

The K nearest neighbors of a monitored point can span an affine hyperplane [21], and in the HDNC the clustering is conducted based on the point-to-hyperplane distance metric. For each clustering, the local hyperplane is constructed and thus multiple local hyperplanes are determined by the different clusters. Suppose that there is a set of *n*-dimensional input measurements in the training set $X=\{x_1, \ldots, x_i, \ldots, x_m\}$, where x_i is given input sample and m is the number of training samples. Although the class labels are unavailable, HDNC has a good capability of unsupervised learning. The K nearest neighbors of a given sample x from each cluster can be denoted in a cluster *l* by x_{lk} (k=1, ..., K) and their collection is defined as $V^K(x)$. α_{lk} is defined as the reconstruction weight coefficient. The affine hyperplane of *l* cluster is written as [21,22]:

$$H_l(x)=\{h|h=\bar{x}_l+\sum_{k=1}^{K}\alpha_{lk}(x_{lk}-\bar{x}_l),\ \alpha_{lk}\in R,\ \sum_{k=1}^{K}\alpha_{lk}=1\} \quad (1)$$

where $\bar{x}_l$ is the centroid of the cluster and defined as $\bar{x}_l=\frac{1}{K}\sum_{k=1}^{K}x_{lk}$.

The square distance from x to the hyperplane $H_l(x)$ is defined as:

$$d_l(x,H_l(x))=\|x-h\|^2=\left\|x-\bar{x}_l-\sum_{k=1}^{K}\alpha_{lk}(x_{lk}-\bar{x}_l)\right\|^2=\alpha_l^T X_l^T X_l\alpha_l \quad (2)$$

where $\|\cdot\|$ denotes Euclidean distance, $X_l=[x_{l1}-\bar{x}_l, \ldots, x_{lk}-\bar{x}_l]$ and $\alpha_l=(\alpha_{l1}, \ldots, \alpha_{lk})^T$.

The optimal reconstruction weights α_l can be obtained by solving the following optimization problem:

$$\min_{\alpha_l}\alpha_l^T X_l^T X_l\alpha_l$$

$$\text{s.t.}\sum_{k=1}^{K}\alpha_{lk}=1 \quad (3)$$

Eq. (3) is equivalent to solving a linear system in α_l, which can be easily expressed in the matrix form as [23]:

$$X_l^T X_l\cdot\alpha_l=X_l^T\cdot(x-\bar{x}_l) \quad (4)$$

If $X_l^T X_l$ is nonsingular, we can obtain α_l as follows:

$$\alpha_l^*=(X_l^T X_l)^{-1}X_l^T(x-\bar{x}_l) \quad (5)$$

Once the optimal solution α_l^* is solved using the gradient search procedure, the affine hyperplane $H_l(x)$ can be computed. At the same time, the distance from x to $H_l(x)$ can be also figured out for all the clusters by Eq. (2), and the minimum distance can be determined by

$$d_m(x)=\min_{all\ l} d_l(x,H_l(x)) \quad (6)$$

Given a threshold θ, $d_m(x)$ is compared against it. If $d_m(x)>\theta$, a new cluster is generated by the point x; if $d_m(x)\le\theta$, the point is assigned to the corresponding cluster in which the minimum distance is found.

When $X_l^T X_l$ is ill-conditioned due to the collinearity, an additional diagonal term is introduced to modify Eq. (2):

$$\tilde{d}_l(x,H_l(x))=\alpha_l^T(X_l^T X_l+\varepsilon I)\alpha_l \quad (7)$$

where ε plays a role of removing the ill condition of matrix $X_l^T X_l$ by adding a positive number to all the diagonal entries [12,24] and I is the identity matrix. A small value of ε is recommended to make the optimal solution robust. Note that a careful selection of ε value is not necessary as long as it is significantly smaller than the average value of $X_l^T X_l$ [25]. The rectified reconstruction weights can be obtained by

$$\tilde{\alpha}_l^*=(X_l^T X_l+\varepsilon I)^{-1}X_l^T(x-\bar{x}_l) \quad (8)$$

Similarly, the minimum distance $\tilde{d}_m(x)$ can be determined.

The HDNC process is summarized as follows. We start the first point x_1 as the first cluster C_1. Then the distances between the new sample and all the defined local hyperplanes are computed using Eq. (7), and the minimum distance for all the clusters is determined using Eq. (6). Next we decide whether the new sample is added to the corresponding cluster or a new cluster is generated according to the threshold θ. This process is repeated until all the training samples are clustered. The flow chart of the HDNC algorithm is shown in Fig. 1.

2. LDA

In this section, a new local discriminative dimensionality reduction algorithm, termed LDA, is developed. Given a sample $x_i\in X$, its index set of the vicinity containing K nearest neighbors is denoted by $V_W^K(x_i)=\{j|x_j$ is one of the K homo-cluster nearest neighbors of $x_i\}$ and $V_B^K(x_i)=\{p|x_p$ is one of the K hetero-cluster nearest neighbors of $x_i\}$, respectively. Here, the homo-cluster and hetero-cluster represent the same cluster and different clusters, respectively. Based on the obtained reconstruction weights, the homo-cluster local distance from x_i to its nearest homo-cluster affine hyperplane is defined as follows:

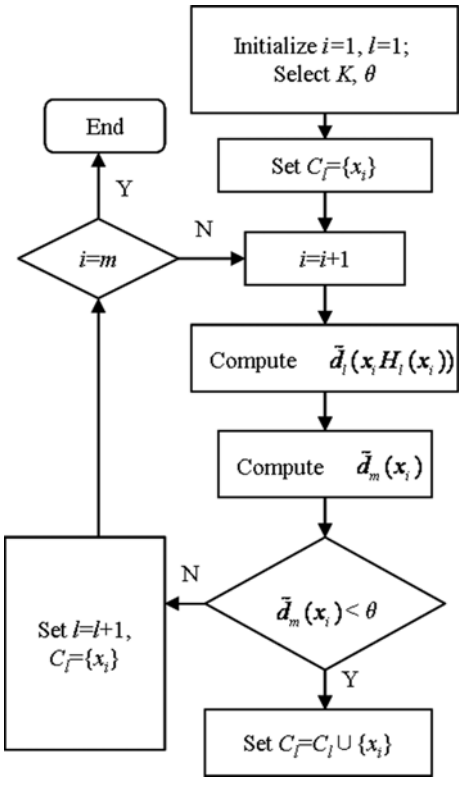

Fig. 1. The flow chart of HDNC algorithm.

$$\tilde{d}_W(x_i)=\left\|x_i-\sum_{j\in V_W^K(x_i)}\tilde{\alpha}_{ij}x_j\right\|^2 \tag{9}$$

The hetero-cluster local distance from x_i to its nearest hetero-cluster affine hyperplane is defined as follows:

$$\tilde{d}_B(x_i)=\left\|x_i-\sum_{p\in V_B^K(x_i)}\tilde{\alpha}_{ip}x_p\right\|^2 \tag{10}$$

Obviously, the distances between the close samples of the same cluster should be as small as possible, while that between the close samples of the different clusters should be as large as possible. The average homo-cluster local distance $\frac{1}{m}\sum_{i=1}^{m}\tilde{d}_W(x_i)$ is the average distance of all samples to their nearest homo-cluster affine hyperplane, representing the compactness of the close samples in the same cluster. The average hetero-cluster local distance $\frac{1}{m}\sum_{i=1}^{m}\tilde{d}_B(x_i)$ is the average distance of all samples to their nearest hetero-cluster affine hyperplane, representing the separability of the close samples in the different clusters.

A linear transformation can transfer high-dimensional data into low-dimensional transformed space by $y_i=U^Tx_i$, which can preserve local reconstruction weights. To better keep hetero-cluster separation and homo-cluster local structure discrimination, we define the objective function to satisfy the following criterion.

For y_i in the low-dimensional subspace, to make the average homo-cluster local distance as small as possible, we have

$$\arg\min_{y_i}\frac{1}{m}\sum_{i=1}^{m}\left\|y_i-\sum_{j\in V_W^K(x_i)}\tilde{\alpha}_{ij}y_j\right\|^2 \tag{11}$$

Meanwhile, to make the average hetero-cluster local distance as large as possible, we have

$$\arg\max_{y_i}\frac{1}{m}\sum_{i=1}^{m}\left\|y_i-\sum_{p\in V_B^K(x_i)}\tilde{\alpha}_{ip}y_p\right\|^2 \tag{12}$$

Since, the measurement formed by the local distance can be considered approximately linear [26]. The projection can be derived by solving the following objective function

$$\arg\min_{y_i}\left(\frac{1}{m}\sum_{i=1}^{m}\left\|y_i-\sum_{j\in V_W^K(x_i)}\tilde{\alpha}_{ij}x_j\right\|^2-\beta\frac{1}{m}\sum_{i=1}^{m}\left\|y_i-\sum_{p\in V_B^K(x_i)}\tilde{\alpha}_{ip}y_p\right\|^2\right) \tag{13}$$

where β is a multiplicative factor in [0, 1] to balance the influence of the constraints and here β=0.2. We define the coefficients vector as

$$w_i=[1,\ldots,1,-\beta,\ldots,-\beta]^T \tag{14}$$

And Eq. (13) can be reduced to

$$\begin{aligned}&\arg\min_{y_i}\left(\frac{1}{m}\sum_{i=1}^{m}\left(\left\|y_i-\sum_{j\in V_W^K(x_i)}\tilde{\alpha}_{ij}x_j\right\|^2+\left\|y_i-\sum_{p\in V_B^K(x_i)}\tilde{\alpha}_{ip}y_p\right\|^2\right)w_i\right)\\&=\arg\min_{Y_i}\mathrm{tr}\left(Y_i\begin{bmatrix}-(A_W+A_B)^T\\ I\end{bmatrix}\mathrm{diag}(w_i)[-(A_W+A_B)\ \ I]Y_i^T\right)\\&=\arg\min_{Y_i}\mathrm{tr}(Y_iG_iY_i^T)\end{aligned} \tag{15}$$

where

$$G_i=\begin{bmatrix}\sum_{i=1}^{m}(A_W+A_B)^T(A_W+A_B)w_i & -((A_W+A_B)w_i)^T\\ -(A_W+A_B)w_i & \mathrm{diag}(w_i)\end{bmatrix} \tag{16}$$

$$A_W=\begin{cases}\tilde{\alpha}_{ij}, & j\in V_W^K(y_i)\\ 0 & \text{otherwise}\end{cases} \tag{17}$$

$$A_B=\begin{cases}\tilde{\alpha}_{ip}, & p\in V_B^K(y_i)\\ 0 & \text{otherwise}\end{cases} \tag{18}$$

To make the projection matrix U linear and orthogonal, we impose a constraint to Eq. (15) by

$$\begin{aligned}&\arg\min_U\mathrm{tr}(U^TXGX^TU)\\&\text{s.t. } U^TU=I_n\end{aligned} \tag{19}$$

The above optimization problem can be solved by the standard eigenvalue decomposition

$$XGX^Tu=\lambda u \tag{20}$$

where λ is the eigenvalue and u denotes the eigenvector corresponding to a local discriminant direction. By arranging the eigenvalues $\lambda_1<\lambda_2<\ldots<\lambda_n$ in an ascending order, we can obtain the LDA projection matrix $U_{LDA}=[u_1, u_2, \ldots, u_s]$ with the s smallest eigenvalues. In process monitoring, the normal and faulty process samples are separated by projecting the raw data set onto the low-dimensional feature subspace, which is composed of the leading local discriminant vectors.

3. HDNC-LDA Based Process Monitoring Approach

Based on the above LDA algorithm, a new process monitoring approach HDNC-LDA is proposed to deal with the plant operating data mixed with both unlabeled normal and multiple types of faulty data. The historical data that we use as the training set is assumed to include both normal and faulty samples with unavailable class labels. These training samples are scaled to zero means and unit variances along all the measurement variables in the preprocessing step. Then the normalized training set is separated into multiple clusters using the HDNC approach. Each cluster corresponds to either the normal data or an individual process fault. To eliminate the requirement of a priori knowledge on the exact number of process faults existing in the data set, the geometrical validity index (GI) based on the ratio of the within-cluster density to the between-cluster separation is used to optimize the number of clusters automatically [27]. Let us suppose that *m* training samples are categorized into L clusters and the *l*th ($1\le l\le L$) cluster C_l contains m_l samples. The GI is expressed as

$$GI(L)=\min_{1\le l\le L}\left\{\frac{(2\sum_{j=1}^{n}\sqrt{\delta_{jl}})^2}{\min_{1\le q\le L}(\|v_l-v_q\|)}\right\} \tag{20}$$

where n is the dimension of the process data, the denominator is the Euclidean distance of the two closest cluster centroids and is taken as a measure of the between-cluster separation, while the numerator is a geometrical measure of the within-cluster scatter. The centroid of C_l is denoted by $v_l=\frac{1}{m_l}\sum_{x_r\in C_l}x_r$. δ_{jl} is the eigenvalue of the sample covariance matrix, whose elements are defined as

$$\sum_{ij} = \frac{1}{m_l - 1} \sum_{l=1}^{m_l} (x_{il} - \bar{x}_i)(x_{jl} - \bar{x}_j) \quad (21)$$

where $\bar{x}_i$ and $\bar{x}_j$ are the sample mean of ith and jth measurement variables, respectively. If some of the eigenvalues are negative, the index is modified by using their absolute values. The smaller the GI, the better the quality of the clustering is. The optimum combination of K and θ can be determined by computing GI for running the HDNC algorithm.

After isolating normal samples from faulty ones by the HDNC algorithm, we apply the LDA method to the training data set to obtain the optimal directions and to better classify the test data set. The LDA eigenvectors span the s-dimensional discriminant subspace. Finally, the distance from the test sample x_t to the local affine hyperplane in the transformed space can be expressed by

$$\tilde{d}_l(x_t, H_l(x_t)) = \left\| U_{LDA}^T (x_t - \tilde{x}_l) - \sum_{k=1}^{K} \tilde{\alpha}_{lk}^* U_{LDA}^T (x_{lk} - \tilde{x}_l) \right\|^2 \quad (22)$$

and then the monitored sample is categorized into the corresponding cluster by Eq. (6).

The flow chart of the HDNC-LDA based process monitoring approach is shown in Fig. 2 and the step-by-step procedure is as follows:

(1) Collect a set of historical data with unknown class labels as the training set under all possible operation conditions.

(2) Normalize the training set to scale them to zero means and unit variances along all measurement variables.

(3) Perform clustering using the HDNC approach to categorize the normal and multiple types of faulty samples.

(4) Compute the matrix GI by Eq. (16).

(5) Apply generalized eigenvalue decomposition to Eq. (20) to derive the local discriminant directions.

(6) For any monitored sample, scale it to zero mean and unit variance.

(7) Calculate the distance from the monitored sample to the local affine hyperplane of all different clusters in the reduced-dimensional discriminant space by Eq. (22).

(8) Assign the monitored sample to the corresponding cluster

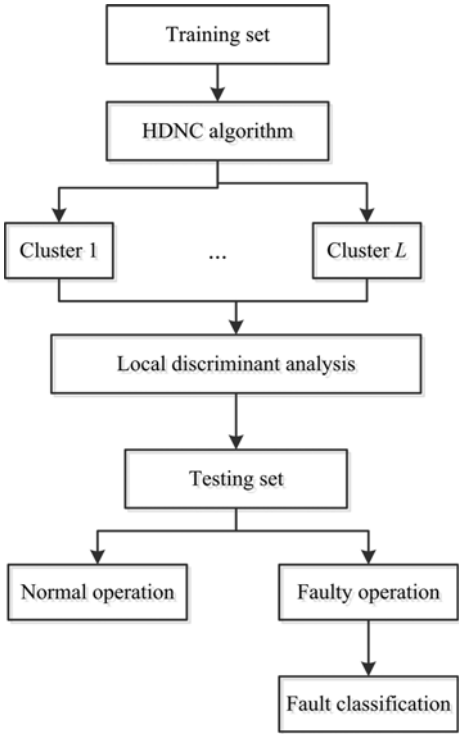

Fig. 2. Schematic diagram of the HDNC-LDA based process monitoring approach.

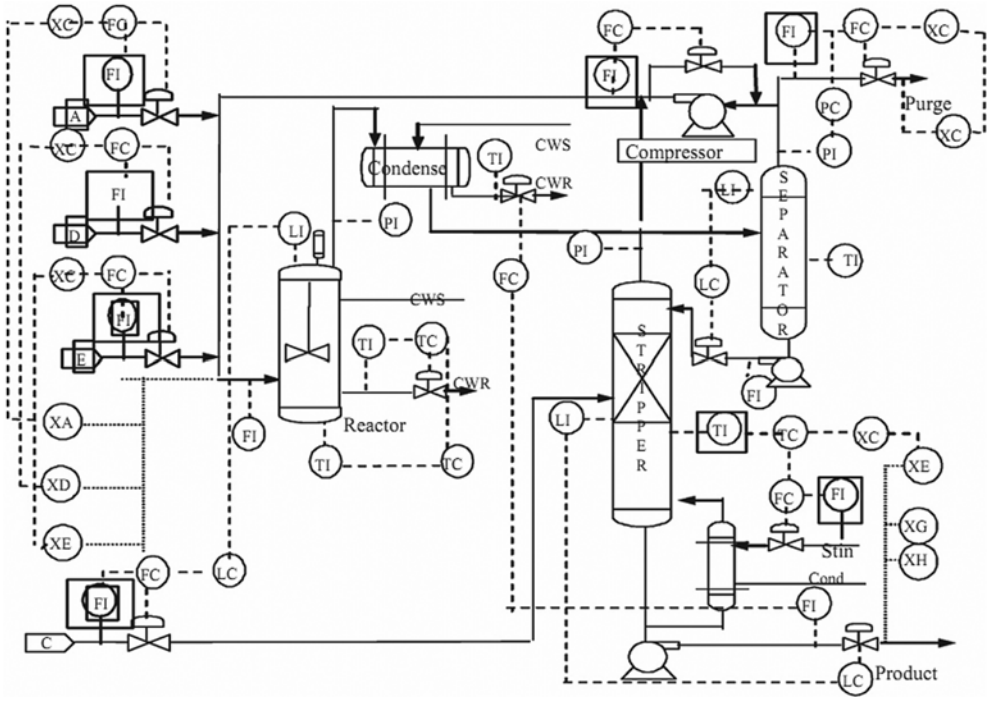

Fig. 3. Schematic diagram of the Tennessee Eastman process.

through Eq. (6).

(9) Determine whether any fault occurs in the monitored operation and which type of fault it belongs to.

APPLICATION

In this section, the proposed HDNC-LDA based process monitoring approach is applied to two case studies, the Tennessee Eastman process and wastewater treatment process, respectively.

1. Case Study on TEP

1-1. Tennessee Eastman Process

The TEP is simulated under normal and various types of faulty operations. Furthermore, the performance of the HDNC-LDA based process monitoring approach is evaluated by using the collected historical data in comparison with that of the conventional *K*-nearest neighbor Fisher discriminant analysis (KNN-FDA) method.

There are five major unit operations in the process: a reactor, a partial condenser, a recycle compressor, a stripper and a vapor/liquid separator. Four chemical reactants (A, C, D and E) are fed into the reactor to form two products (G and H) together with a byproduct (F) [28]. The schematic diagram of TEP is shown in Fig. 3. The process has 22 continuous process measurements, 12 manipulated variables and 19 composition measurement samples. As shown in Table 1, there are 22 continuous measurement variables and 12 manipulated variables in total. The decentralized control strategy has been implemented with multiple feedback or loops in this TEP [29]. In this study, only the 22 continuous measurement variables are chosen as monitored variables with sampling time of 0.05 h. Several types of process faults are mixed in two different scenarios to evaluate and compare the capability of detecting and classifying multiple faults.

Table 1. Continuous measurement and manipulated variables in the TEP

No.	Measured variable	No.	Manipulated variable
1	A Feed rate	1	D Feed flow valve
2	D Feed rate	2	E Feed flow valve
3	E Feed rate	3	A Feed flow valve
4	A+C Feed rate	4	A+C Feed flow valve
5	Recycle flow rate	5	Recycle valve
6	Reactor feed rate	6	Purge valve
7	Reactor pressure	7	Separator valve
8	Reactor level	8	Stripper valve
9	Reactor temperature	9	Steam valve
10	Purge rate	10	Reactor coolant flow
11	Separator temperature	11	Condenser coolant flow
12	Separator level	12	Agitator speed
13	Separator pressure		
14	Separator underflow		
15	Stripper level		
16	Stripper pressure		
17	Stripper underflow		
18	Stripper temperature		
19	Stem flow rate		
20	Compressor work		
21	Reactor coolant temperature		
22	Condenser coolant temperature		

Table 2. Two scenarios in the simulated TEP

Case No.	Description
1	Training set: 300 normal samples 200 faulty samples with step error in D feed temperature 100 faulty samples with random variations in reactor coolant temperature Testing set: 1^{st}-200^{th} samples: normal operation 201^{st}-300^{th} samples: random variations in reactor coolant temperature 301^{st}-400^{th} samples: normal operation 401^{st}-500^{th} samples: step error in D feed temperature
2	Training set: 400 normal samples 200 faulty samples with step error in B composition in Stream 4 200 faulty samples with drift error in reactor kinetics 200 faulty samples with condenser coolant flow valve stiction Testing set: 1^{st}-200^{th} samples: normal operation 201^{st}-300^{th} samples: random variations 301^{st}-400^{th} samples: normal operation 401^{st}-550^{th} samples: step error in B composition in Stream 4 551^{st}-700^{th} samples: normal operation 701^{st}-800^{th} samples: condenser coolant flow valve stiction

Data sets composed of both normal and faulty samples in the two scenarios are listed in Table 2. In the first case, the training set starts with 300 normal samples under the steady-state condition [30]. Then a step error in D feed temperature takes place for 200 samples in the process. Afterwards, 100 faulty samples follow with the occurrence of increased random variation in reactor cooling water inlet temperature. The testing set includes the first 200 normal operation samples and the following 100 faulty samples with the increased random variation in reactor coolant temperature. After that, the process operation shifts back to normal state with 100 samples. From the 401^{st} sample, a step error in D feed temperature occurs and maintains for 100 samples. The second case includes three categories of faults, which are step error in B composition in Stream 4, slow drift in reactor kinetics, and condenser cooling water valve stiction, respectively. The training set is initiated with 400 normal operation samples followed by 600 faults of the above-mentioned three categories. The testing set involves first 200 normal samples followed by faulty and normal segments alternating with each other. Each alternating segment is composed of 150 or 100 samples as listed in Table 2.

1-2. Process Monitoring Results

Since the training set is easily contaminated by faults, the common process monitoring methods like PCA and PLS may suffer greatly from this limitation. For both scenarios, the training sets are

Table 3. Comparison of fault detection and classification indices between KNN-FDA and HDNC-LDA methods

	Type-I error (%)		Type-II error (%)		Fault classification rate (%)	
	KNN-FDA	HDNC-LDA	KNN-FDA	HDNC-LDA	KNN-FDA	HDNC-LDA
Case 1	11.6	6.5	8.3	4.6	89.0	93.2
Case 2	12.8	7.4	10.5	5.1	86.8	92.5
Case 3	10.5	7.3	12.6	8.4	83.5	90.8

used to build the KNN-FDA or HDNC-LDA model, while the testing sets are used to validate the model in terms of fault detection and classification. Due to the unavailable class labels of data sets, the KNN and HDNC clustering procedures are used to identify normal operation and various types of faults, respectively. Then the FDA or LDA is implemented on the training set to extract the leading directions and separate the normal clusters from multiple faulty ones. Both methods adopt the quantitative performance indices, *i.e.*, Type-I error, Type-II error and faulty classification rate, as listed in Table 3.

For the first case, the fault detection results of KNN-FDA and HDNC-LDA methods are illustrated in Fig. 4(a) and (b). Note that the KNN-FDA approach is unable to identify abundant normal samples and detect abnormal events. The fault detection performance degrades because the conventional FDA algorithm sacrifices the between-cluster separability in order to minimize the within-cluster scatter [18]. In this case, the abnormal data include two types of faults, which do not always follow similar statistical distribution. The method only focuses on the global characteristics so that the multimodality within the faulty data makes the leading directions of FDA nonoptimal in separating normal and faulty clusters. Therefore, the fault detection accuracy is affected. Meanwhile, a poor fault classification is produced. In contrast, the HDNC-LDA significantly improves the fault detection accuracy. From Fig. 4(b), one can observe that there are only twelve normal samples triggering false alarm. The Type-I error is 6.5%. Such frequent alerts are acceptable during

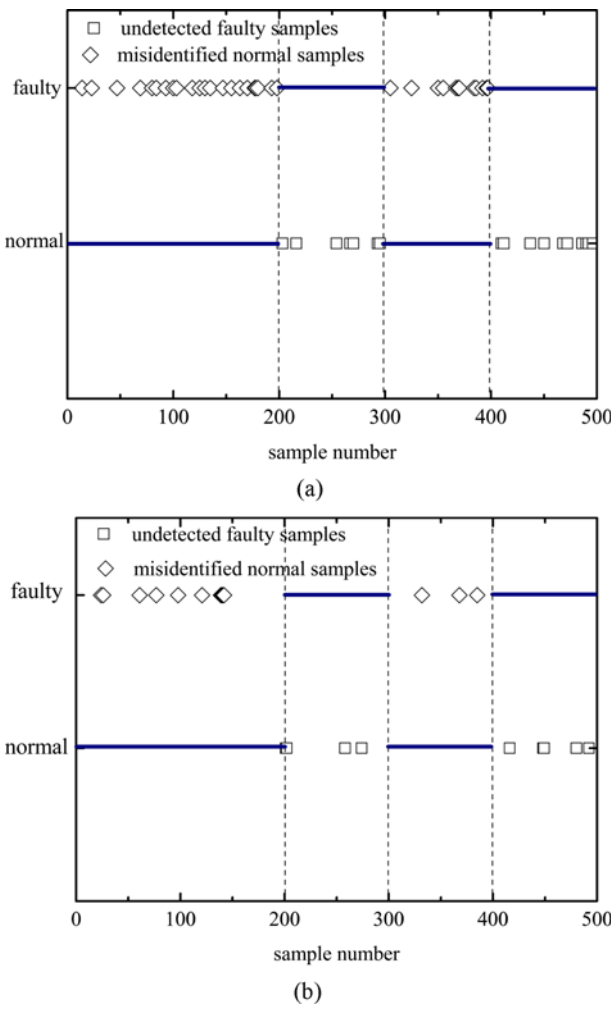

Fig. 4. Fault detection results in the first case of TEP using (a) KNN-FDA and (b) the proposed HDNC-LDA approach.

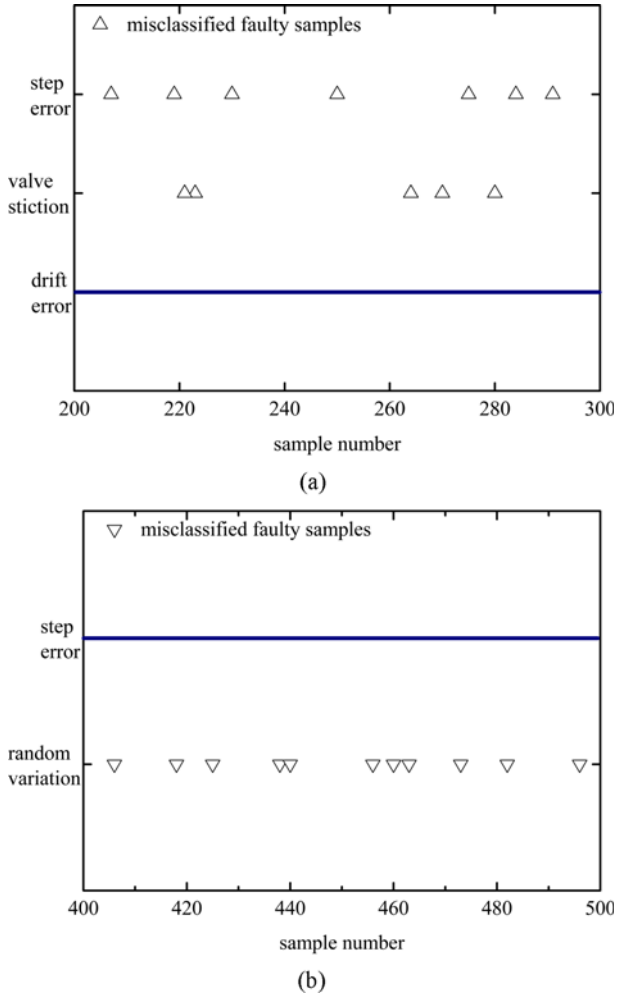

Fig. 5. Fault classification results in the first case of TEP using KNN-FDA method (a) 1st faulty segment: 201st-300th samples and (b) 2nd faulty segment: 401st-500th samples.

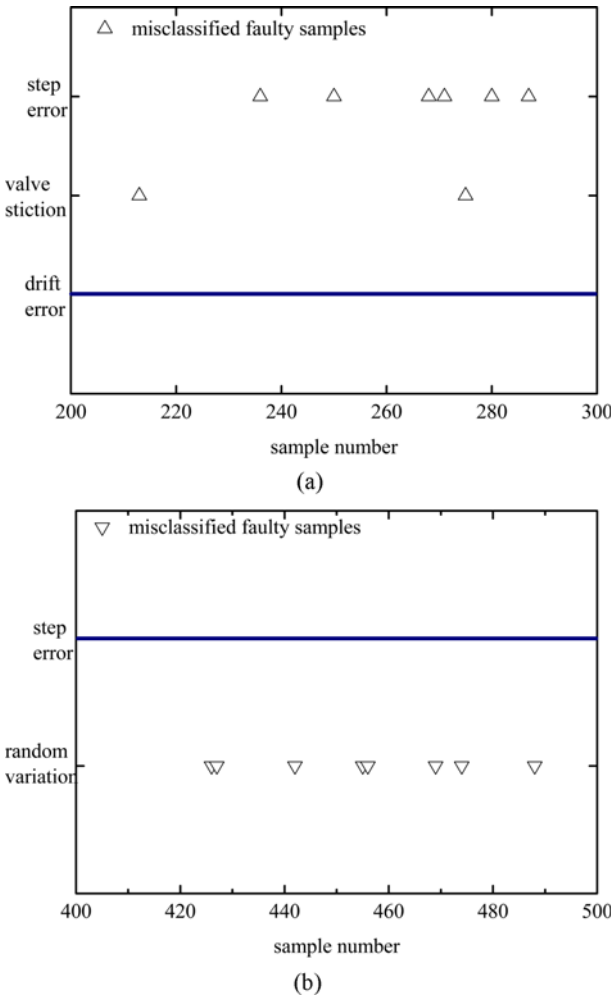

Fig. 6. Fault classification results in the first case of TEP using HDNC-LDA method (a) 1st faulty segment: 201st-300th samples and (b) 2nd faulty segment: 401st-500th samples.

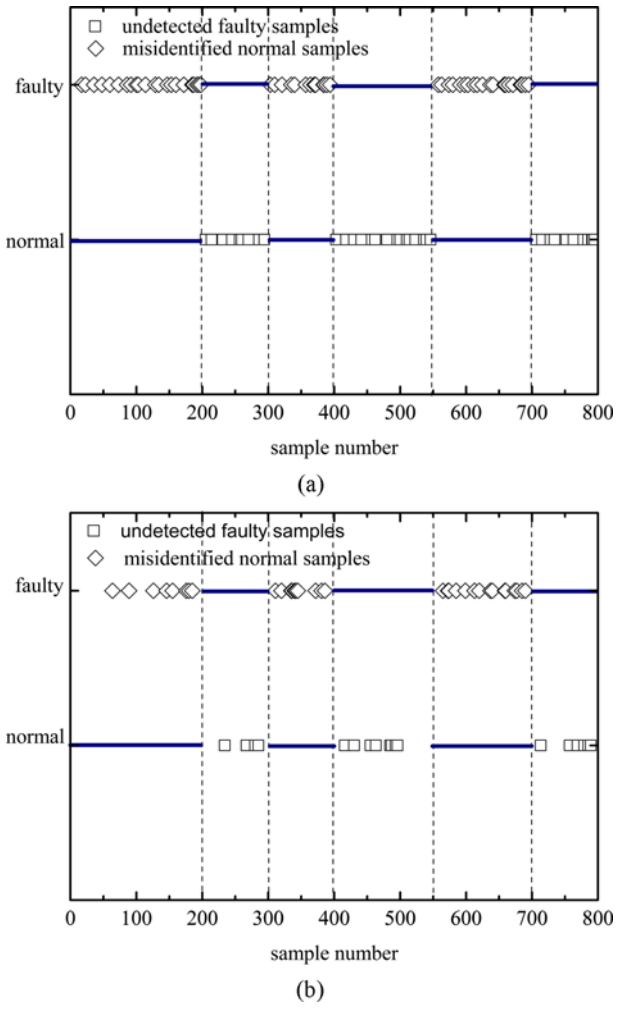

Fig. 7. Fault detection results in the second case of TEP using (a) KNN-FDA and (b) the proposed HDNC-LDA approach.

normal process operation. Moreover, the Type-II error is 4.6%, which reaches a desirable faulty detection rate. Only eight faulty samples are not captured. Few Type-I and Type-II errors occur since the LDA algorithm involves the localities of data with different class-labels and performs well in isolating the normal data from these two types of faulty data. On the one hand, the method can minimize the homo-cluster local distance to preserve the multimodality within the faulty data. On the other hand, it can maximize the hetero-cluster local distance to well separate the normal operations and multiple faulty events. Meanwhile, the classification results of different types of faulty data reveal that the HDNC-LDA approach has a stronger ability in classification than the KNN-FDA, as shown in Figs. 5 and 6. Ten samples are misclassified as step error for the KNN-FDA method when the fault of increased random variation occurs during the operating period. Another eleven faulty samples belonging to step error are wrongly separated as random variation. The method reaches an overall fault classification rate of 89.0%. In contrast, the proposed HDNC-LDA approach misclassifies only six faulty samples as step error and eight as random variation, obtaining a fault classification accuracy of 93.2%.

The second scenario designed with three types of faulty events in the process is employed to further examine the performance of these two methods. The results shown in Figs. 7, 8 and 9 demonstrate their different capabilities in detecting and classifying multiple kinds of process faults. The Type-I error, Type-II error and fault classification rate of the KNN-FDA method are 12.8%, 10.5% and 86.8%, respectively. In the normal operating period, fifty-seven false alarms are generated. Besides, a total of thirty-six faulty samples are misclassified as normal ones. Among the detected faulty samples, forty-two observations are classified into uncorrected classes. The corresponding performance indices of the HDNC-LDA method are 7.4%, 5.1% and 92.5%, indicating that the proposed method is suitable to monitor the complex processes with several kinds of abnormal events. Thirty-three normal samples trigger false alarms. Eighteen faulty samples are not detected and only twenty-five faulty ones are categorized into wrong types. Similar to the first case, the proposed method has superior ability to detect multiple types of process faults as well as to classify different faulty clusters with high accuracy.

2. Case Study on WWTP

2-1. Wastewater Treatment Process

The HDNC-LDA based process monitoring approach is also tested for its ability to detect some internal disturbances in simulated data

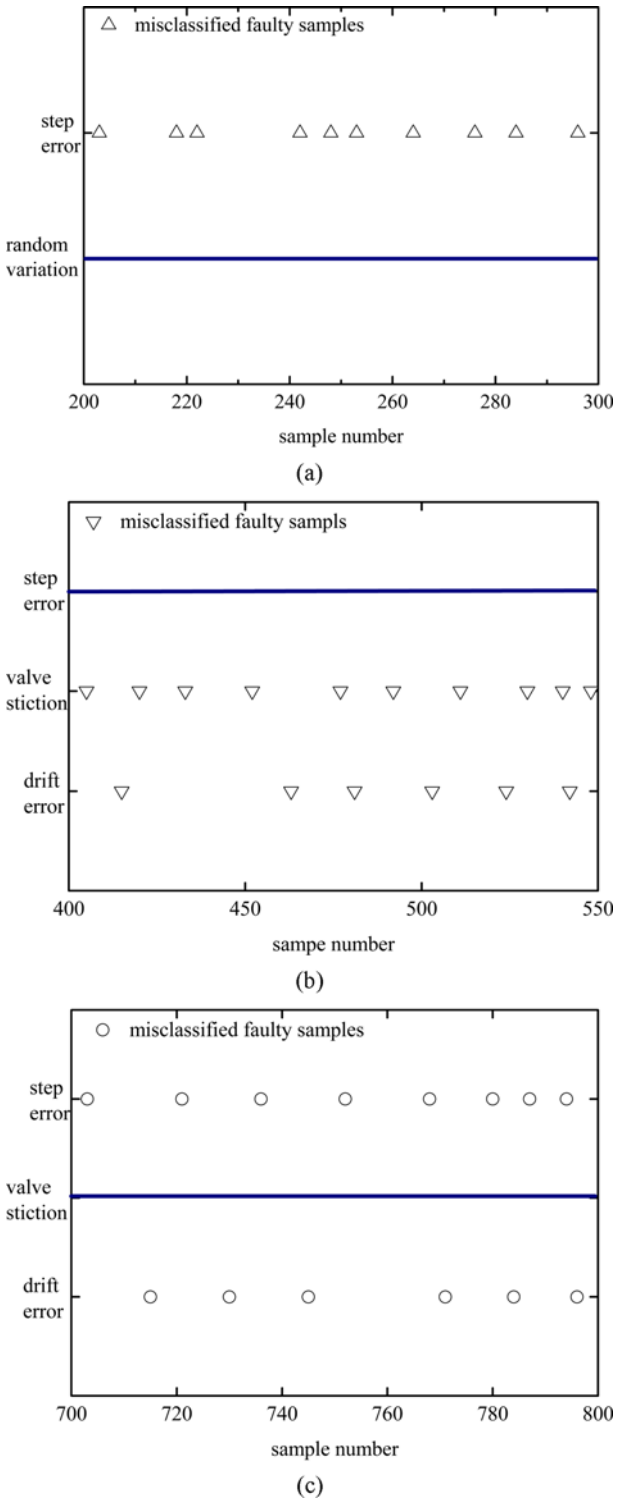

Fig. 8. Fault classification results in the second case of TEP using KNN-FDA method (a) 1st faulty segment: 201st-300th samples, (b) 2nd faulty segment: 401st-550th samples and (c) 3rd faulty segment: 701st-800th samples.

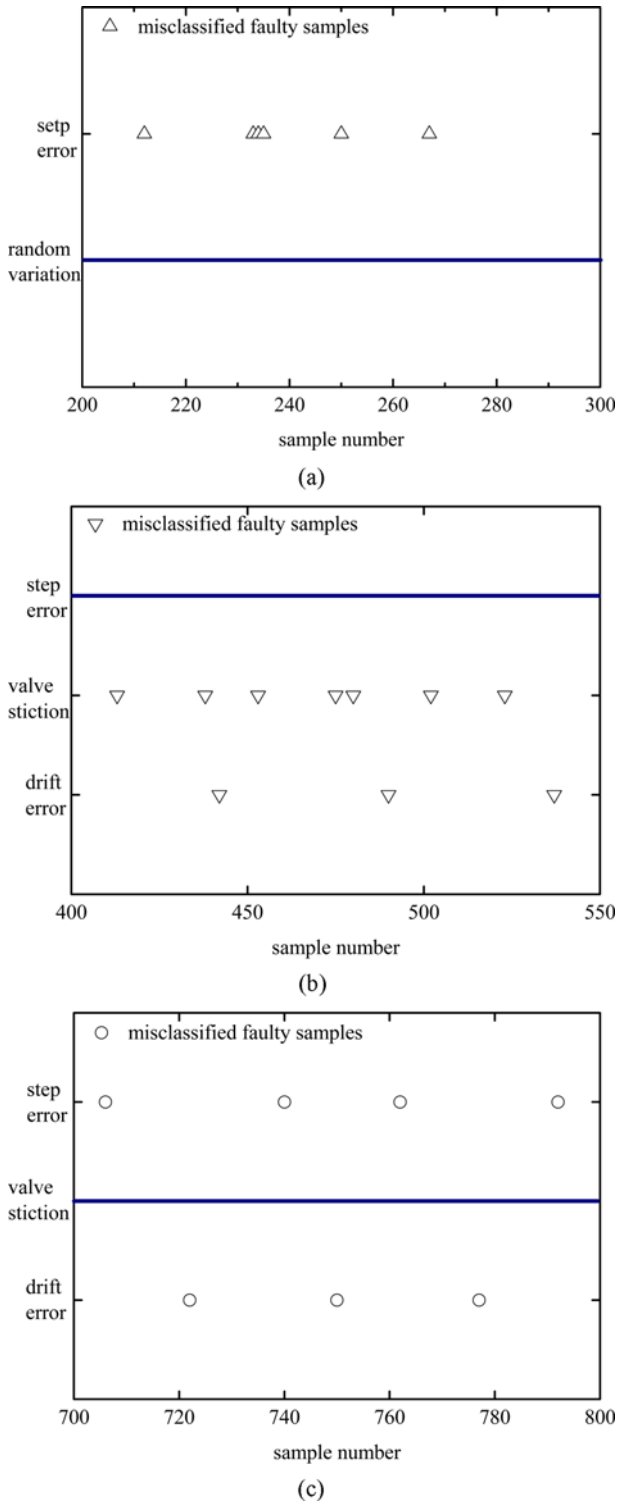

Fig. 9. Fault classification results in the second case of TEP using HDNC-LDA method (a) 1st faulty segment: 201st-300th samples, (b) 2nd faulty segment: 401st-550th samples and (c) 3rd faulty segment: 701st-800th samples.

obtained from a 'benchmark simulation' of the WWTP. The activated sludge model No. 1 (ASM 1) and a 10-layer settler model are used to simulate the biological reactions and the settling process, respectively. The plant consists of five bioreactors and a secondary settler. The first two compartments of the bioreactor are unaerated but fully mixed, whereas the other three are well aerated. As for the secondary settler, it is modeled as one-dimensional series of 10 layers. The flow diagram of the WWTP is shown in Fig. 10. More

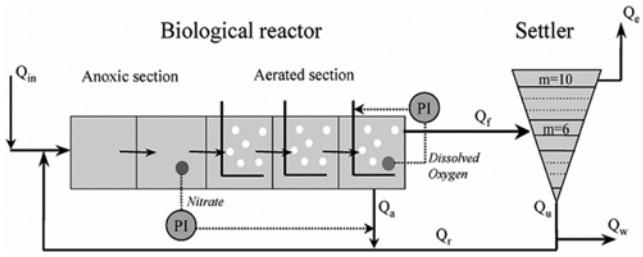

Fig. 10. Diagram of the simulated wastewater treatment process.

Table 4. Monitored variables in the WWTP

No.	Monitored variable
1	Nitrate and nitrite nitrogen S_{NO}
2	Slowly biodegradable substrate X_S
3	Soluble biodegradable organic nitrogen S_{ND}
4	Dissolved oxygen in Tank 3 $S_{O,3}$
5	Dissolved oxygen in Tank 4 $S_{O,4}$
6	Readily biodegradable substrate S_S
7	Soluble inert organic matter S_I
8	Influent flow rate Q_{in}
9	NH_4^++NH_3 nitrogen S_{NH}
10	Particulate biodegradable organic nitrogen X_{ND}

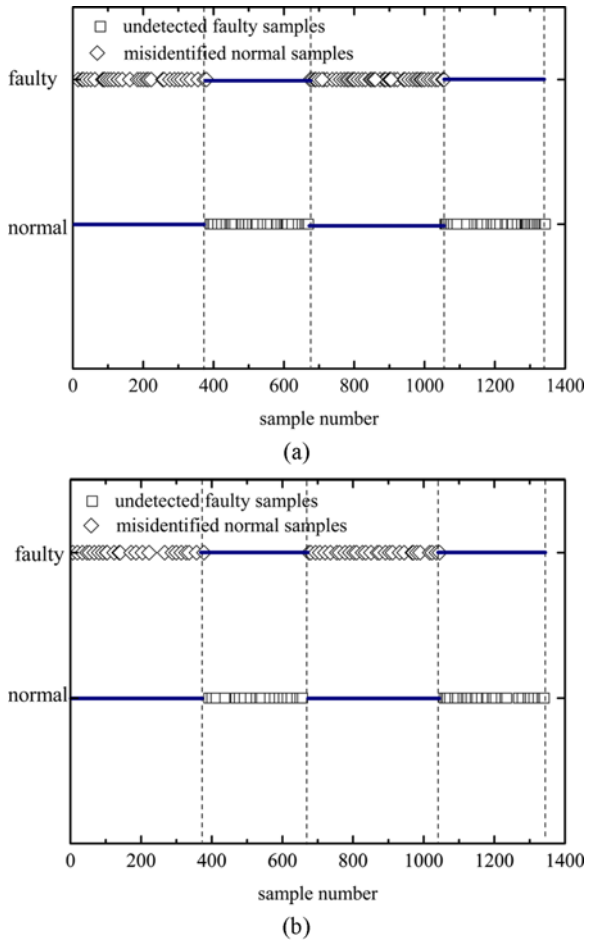

Fig. 11. Fault detection results in the third case of WWTP using (a) KNN-FDA and (b) the proposed HDNC-LDA approach.

detailed process descriptions are available in the COST website at http://www.ensic.u-nancy.fr/COSTWWTP/.

The ten monitored process variables are listed in Table 4. In the training period, the process starts with one-week normal operation under dry weather and then is shifted to a step decrease in aerobic growth rate of autotrophs from 4.0 to 3.0 per day. After three days, a linear drift in the ammonification rate of soluble nitrogen increases from 0.05 to 0.1 m^3 COD/g per day within three days. The sampling period is 15 min, and thus the training set is composed of 672 normal data and 576 abnormal process events. The test scenario (Case 3) is initiated with the normal operation for four days and then followed by a step decrease in aerobic growth rate of autotrophs from 4.0 to 3.0 per day. After three days, the process is switched to routine operation with four days duration. Finally, a linear increase in the ammonification rate of soluble nitrogen from 0.05 to 0.1 m^3 COD/g per day suddenly occurs and the faulty data in a three-day period are collected.

2-2. Process Monitoring Results

The fault detection results on the test dataset of Case 3 are illustrated in Fig. 11. The Type-I error and Type-II error of HDNC-LDA

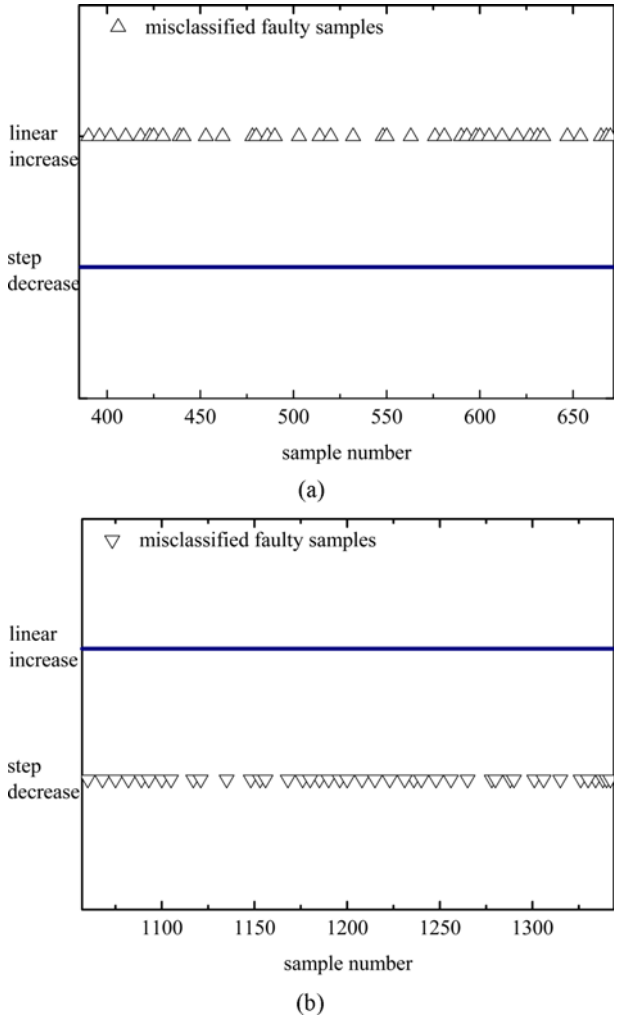

Fig. 12. Fault classification results in the third case of WWTP using KNN-FDA method (a) 1st faulty segment: 385th-672nd samples and (b) 2nd faulty segment: 1057th-1344th samples.

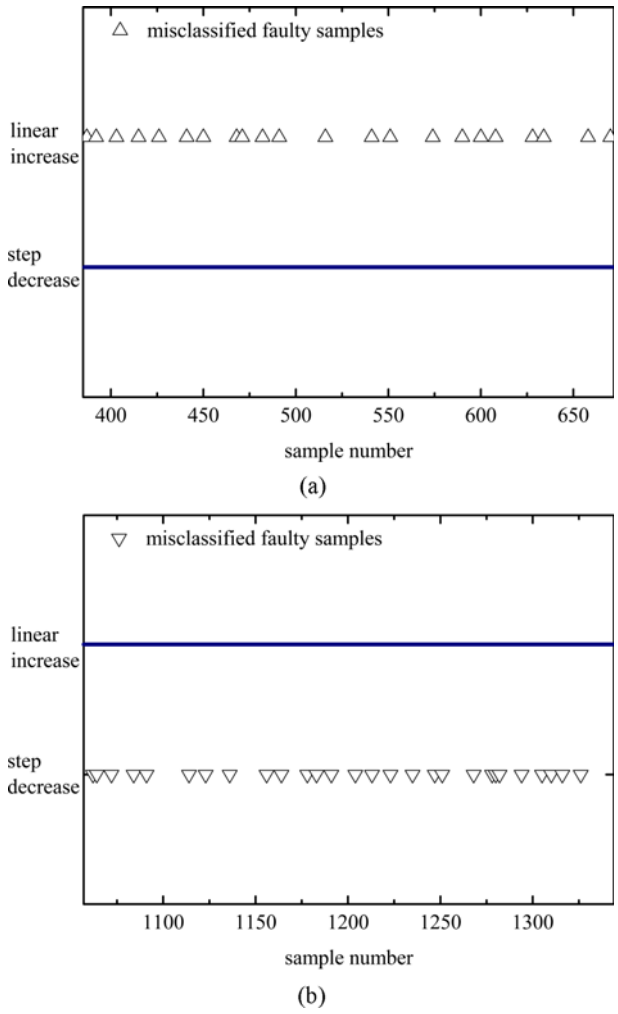

Fig. 13. Fault classification results in the third case of WWTP using HDNC-LDA method (a) 1st faulty segment: 385th-672nd samples and (b) 2nd faulty segment: 1057th-1344th samples.

method are 7.3% and 8.4%. The KNN-FDA method has higher Type-I error of 10.5% and Type-II error of 12.6%. Such results demonstrate that the proposed method has better performance for fault detection than the KNN-FDA method. The fault classification results of the two methods are compared in Fig. 12 and Fig. 13. The HDNC-LDA method leads to the better fault classification with the overall accuracy of 90.8%, while the KNN-FDA method obtains 83.5% fault classification rate. Therefore, the HDNC-LDA method can become a promising one in detecting and classifying multiple kinds of process faults.

CONCLUSIONS

We have proposed a systematic monitoring strategy for complex chemical processes with various kinds of faults. The basic idea is to first conduct the HDNC algorithm to identify the normal and multiple types of faulty clusters. The local homo-cluster and hetero-cluster distance metrics are further computed to minimize the compactness within the same cluster and maximize the separation between different clusters. Meanwhile, the local discriminative information is preserved. Hence, the HDNC-LDA scheme can be used to detect abnormal events and classify multiple types of operation faults. In comparison with the KNN-FDA method, our method has the following advantages: (1) the capability of fault classification is enhanced due to the unchanged local structure of multiple faulty clusters; (2) various types of faults are treated as individual cluster to be separated from the normal region, resulting in a higher accuracy of fault detection; and (3) the HDNC can automatically search for the optimal number of clusters, requiring no prior process knowledge on the total number of faults.

The HDNC-LDA approach is applied to deal with unlabeled process data of the TEP and WWTP in three scenarios. The results indicate that the HDNC-LDA approach exhibits better performance than the KNN-FDA method in terms of quantitative indices such as the Type-I error (false alarm rate), the Type-II error (fault detection rate) and the fault classification rate. Therefore, the HDNC-LDA method can serve a powerful tool for fault detection and classification of complex chemical process.

ACKNOWLEDGEMENTS

This work was supported by the Fundamental Research Funds for the Central Universities (JUDCF12027, JUSRP51323B), PAPD of Jiangsu Higher Education Institutions and Graduate Student Innovation Program for Universities of Jiangsu Province (CXLX12_0734).

REFERENCES

1. C. Bo, X. Qiao, G. M. Zhang, Y. J. Bai and S. Zhang, *J. Process Control*, **20**, 1133 (2010).
2. L. H. Chiang, E. Russell and R. D. Braatz, *Fault detection and diagnosis in industrial systems*, Springer Verlag (2001).
3. S. Lakshminarayanan, S. L. Shah and K. Nandakumar, *AIChE J.*, **42**, 2307 (2004).
4. M. H. Kim and C. K. Yoo, *Korean J. Chem. Eng.*, **25**, 947 (2008).
5. X. Wang, U. Kruger and B. Lennox, *Control Eng. Practice*, **11**, 613 (2003).
6. K. Han, K. J. Park, H. Chae and E. S. Yoon, *Korean J. Chem. Eng.*, **25**, 13 (2008).
7. B. R. Bakshi, *AIChE J.*, **44**, 1596 (1998).
8. J. Liang and J. X. Qian, *Chin. J. Chem. Eng.*, **11**, 191 (2003).
9. M. Kano, K. Nagao, S. Hasebe, I. Hashimoto and H. Ohno, *AIChE J.*, **49**, 969 (2003).
10. J. M. Lee, C. Yoo and I. B. Lee, *J. Process Control*, **14**, 467 (2004).
11. J. M. Lee, S. J. Qin and I. B. Lee, *AIChE J.*, **52**, 3501 (2006).
12. J. Yu and S. J. Qin, *AIChE J.*, **54**, 1811 (2008).
13. J. Yu, *Chem. Eng. Sci.*, **68**, 506 (2012).
14. R. Rengaswamy and V. Venkatasurbramanian, *Comput. Chem. Eng.*, **24**, 431 (2000).
15. H. B. Aradhye, B. R. Bakshi, J. F. Davis and S. C. Ahalt, *AIChE J.*, **50**, 2455 (2004).
16. N. Mehranbod, M. Soroush, M. Piovoso and B. A. Ogunnaike, *AIChE J.*, **49**, 1787 (2003).
17. L. Chiang, M. Kotanche and A. Kordon, *Comput. Chem. Eng.*, **28**, 1389 (2004).
18. Q. P. He, S. J. Qin and J. Wang, *AIChE J.*, **51**, 555 (2005).
19. Y. Zhang, *Chem. Eng. Sci.*, **64**, 801 (2009).
20. I. Yélamos, G. Escudero, M. Graells and L. Puigjaner, *Comput.*

Chem. Eng., **33**, 244 (2009).
21. P. Vincent and Y. Bengio, *K-local hyperplane and convex distance nearest neighbor algorithms*, in Advances in Neural Information Proc. Systems, MIT Press, Cambridge, **1**, 985 (2002).
22. C. F. Pasluosta, P. Dua and W. J. Lukiw, *Nearest hyperplane distance neighbor clustering algorithm applied to gene co-expression analysis in Alzheimer's disease*, in: Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Boston, **33**, 5559 (2011).
23. J. Xu, J. Yang and Z. H. Lai, *Information Sciences*, **232**, 11 (2013).
24. J. Mao and A. K. Jain, *IEEE Trans. Neural Networks*, **7**, 16 (1996).
25. C. Archambeau, F. Vrins and M. Verleysen, *Flexible and robust Bayesian classication by nite mixture models*, in European Symp. on Articial Neural Networks, Bruges, Belgium, 75 (2004).
26. T. Zhang, D. Tao, X. Li and J. Yang, *IEEE Trans. Knowledge and Data Eng.*, **21**, 1299 (2009).
27. B. Lam and H. Yan, *Cluster validity for DNA microarray data using a geometrical index*, in 4th International Conference on Machine Learning and Cybernetics, **6**, 3333 (2005).
28. J. J. Downs and E. F. Vogel, *Comput. Chem. Eng*., **17**, 245 (1993).
29. N. L. Ricker, *J. Process Control*, **6**, 205 (1996).
30. N. L. Ricker, *Comput. Chem. Eng*., **19**, 949 (1995).