

# Research Article

# Modified *t*-Distribution Stochastic Neighbor Embedding Using Augmented Kernel Mahalanobis-Distance for Dynamic Multimode Chemical Process Monitoring

# Haoyu Gu 💿 and Li Wang 💿

School of Electrical and Electronic Engineering, Shanghai Institute of Technology, Shanghai 201418, China

Correspondence should be addressed to Li Wang; wangli@sit.edu.cn

Received 21 August 2022; Revised 5 December 2022; Accepted 16 December 2022; Published 29 December 2022

Academic Editor: Doraiswami Ramkrishna

Copyright © 2022 Haoyu Gu and Li Wang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The traditional data-driven process monitoring methods may not be applicable for the system which has dynamic and multimode characteristics. In this paper, a novel scheme named *modified t-distribution stochastic neighbor embedding using augmented Mahalanobis-distance for dynamic multimode chemical process monitoring (AKMD-t-SNE)* is proposed to realize dynamic multimodal process monitoring. First, the augmented matrix strategy is utilized to ensure the sample contains the autocorrelation of the process. Moreover, *AKMD-t-SNE* method eliminates the scale and spatial distribution differences among multiple modes by calculating the kernel Mahalanobis distance between the samples to establish the global model. The features extracted via the proposed method are obviously non-Gaussian, and there will be a deviation in the construction of traditional statistics. Then, the *support vector data description (SVDD)* method is used to construct statistics to deal with this problem. In addition, a *hybrid correlation coefficient method (HCC)* is proposed to achieve fault isolation and improve the accuracy of isolation results. The advantages of the proposed scheme are illustrated by a numerical case and the *Multimode Tennessee Eastman Process (MTEP)* benchmark.

### 1. Introduction

Process monitoring plays a crucial role in the safe and efficient production of process industries such as the chemical industry and oil refining industry. With the increasing market demand, the complexity and the difficulty of process monitoring also raise. Massive industrial data are stored with the assistance of distributed control system (DCS). The data-driven process monitoring method is to discover potential regularities from colossal historical data for modeling and monitoring of intricate industrial processes. The most fundamental method of data-driven process monitoring is multivariate statistical process monitoring (MSPM) [1–5].

In process industries, the seasonal variation and demand fluctuation will lead to the change of the equilibrium point during production, which causes the process to have the characteristic of a multimode. In the multimode process,

once the operation mode switches, the model based on a single-mode process will not be able to achieve effective monitoring. One typical multimode process is the penicillin fermentation process, which is divided into two modes. The conventional MSPM method is only suitable for the process monitoring in the production process of a single mode. Mode one was the growth and propagation mode of Penicillium, accompanied by the rapid consumption of grape sugar base, which lasted about 43 hours. Mode two is the synthetic penicillin mode, which requires continuous glucose supplementation to the fermenter in order to maintain the high yield and high quality of penicillin [6]. The methods of multimode process monitoring mainly include the model integration method based on a rigid partition strategy and the global modeling method based on a flexible partition strategy [7, 8]. Due to the inertia of equipment operation and the hysteresis response of closed-loop systems, the output of the system has the characteristic of time delay, which is ubiquitous in process industry production [9]. Because of the autocorrelation between the successive measurements, the modeling method for static objects is no longer applicable to the dynamic system. Modeling methods for dynamic systems commonly include augmented matrix, time series analysis, subspace identification, and deep learning [10–14]. The augmented matrix method is to form the extended matrix by combining the current sample and the earlier measurements, which has been proven to be able to extract both the autocorrelation of samples and the mutual correlation of variables. Ku first proposed dynamic principal component analysis (DPCA) based on an augmented matrix to solve the modeling problem of linear dynamic systems [15]. The augmented matrix method is still an effective strategy for dynamic process modeling in the current research studies. Dynamic multimodal process monitoring has been studied in recent years.

t-distribution stochastic neighbor embedding (t-SNE) is an improved manifold learning [16, 17] algorithm based on stochastic neighbor embedding (SNE). Both global and local data structures are preserved when t-SNE is used to accomplish feature extraction. t-SNE is early used in the field of data visualization [18, 19]. Due to the high computational complexity of the t-SNE algorithm, various methods have been proposed to accelerate the t-SNE algorithm [20, 21]. The t-SNE algorithm has been used successfully in various domains due to its superior performance. However, the application of t-SNE in process monitoring is still infrequent. Zhang et al. proposed a quality-related process monitoring method for the hot rolling process based on parameter t-SNE (P-t-SNE) and modified minimum error minimax probability machine (MMEMPM) method which achieved satisfying results [22]. Tang and Yan used the t-SNE algorithm to realize process monitoring through data visualization, which can more intuitively show the trend of process data when an anomaly occurs [23]. In recent years, dynamic multimodal processes have been studied. Qin et al. proposed a mixture slow feature analysis (MSFA) method to deal with the health degradation problem. MSFA integrates the local model to establish the global model, to achieve the equipment fault detection in the multimode process [24]. Zhang et al. introduced a novel sparse dynamic inner principal component analysis (SDiPCA) based monitoring for multimode dynamic processes, which can update the model of sequential modes by memorizing the significant features of existing modes and provide outstanding performance for successive mode [25]. Deep learning is a hot topic in process monitoring and it is feasible to use deep learning to realize multimodal process monitoring. A novel self-adaptive deep learning method based on local adaptive standardization and variational auto-encoder bidirectional long short-term memory (LAS-VB) was proposed by Wu and Zhao and showed excellent performance for multimode process monitoring [26].

The t-SNE algorithm first calculates the Euclidean distance between high-dimensional data and then converts the distance into conditional probability. The dimensional reduction is realized by using the low-dimensional conditional

probability to approximate the high-dimensional conditional probability. The original t-SNE method of feature extraction is only suitable for the modeling of static singlemode process. In this article, an improved t-SNE method called *t*-distribution stochastic neighbor embedding algorithm based on augmented kernel Mahalanobis distance (AKMD-t-SNE) is proposed for dynamic multimode chemical process monitoring. The augmented matrix strategy is employed to expand the samples, which can ensure the samples contain information from the adjacent sampling instants. Moreover, the kernel Mahalanobis distance between the augmented samples is calculated to eliminate the scale differences among different modes [27]. Furthermore, samples are nonidentically distributed in different modes, the kernel Mahalanobis distance transforms the samples in different modes into independent identically distributed problems via the kernel covariance matrix. Then, the augmented kernel Mahalanobis distance is introduced into the t-SNE algorithm for dimensionality reduction. The model decomposes the high-dimensional data into lowdimensional feature subspace (FS) and residual subspace (RS). The FS reflects the main information of the process and the RS contains the redundant information of the process. Since the data is non-Gaussian, the SVDD method is adopted in this paper to construct statistics in the FS and RS, which further improve the accuracy of fault detection [28-30].

Fault detection is the fundamental step of process monitoring. After the abnormal condition is detected effectively, it is necessary to trace the variable that causes the abnormal status, i.e., to conduct a fault isolation step. The most typical fault isolation methods based on the plots model are contribution projection and reconstruction-based contribution (RBC) [31]. The contribution plots indicate the contribution value of each variable to the detection statistics, but the disturbance of one variable may affect the contribution of other uncorrelated variables to the statistics; i.e., the smearing effect, even if the fault amplitude is large, there is still the probability of misdiagnosis. The fault isolation method based on reconstruction was proposed by Dunia et al. in 1996. This approach needs to give the fault direction to realize reconstruction. By comparing the reconstructed statistics, we can judge whether the given fault direction is correct and determine the isolated variable consequently [32]. The RBC method highly relies on prior knowledge, and the computational complexity of reconstruction is high [33, 34]. Verron et al. proposed a fault isolation method based on mutual information (MI), which is computationally efficient and does not need to rely on historical data, but the smearing effect still exists [35]. To overcome the above-mentioned limitations of fault isolation, this work proposes a hybrid correlation coefficient (HCC) method based on the MI to achieve fault isolation and exclude suspicious variables, which ensures the isolated variables are more explicit and consistent with the responsible variables. In short, this paper provides a global modeling method that integrates multiple modes and dynamic characteristics.

International Journal of Chemical Engineering

#### 2. Materials and Methods

2.1. Process Feature Extraction Method. Due to the autocorrelation of process data, an augmented matrix is established to ensure the new samples contain the cross-correlation between variables and the time correlation between observations. The high-dimensional data is=  $[x_1, x_2, ..., x_N] \in \Re^{N \times D}$ , where *D* represents the number of variables and *N* represents the number of observations. Suppose the current sampling instant is *k* and the sample at *k* is x(k). x(k) and past values with orders of data dynamics *L* forms an augmented matrix:

$$x_k(L) = [x(k), x(k-1), \cdots, x(k-L)] \in R^{1 \times (L+1)D}.$$
 (1)

In equation (1), the sampling time k must be larger than the time delay L, and the form of the augmented matrix  $X_a$ formed is as follows:

$$X_{b} = \begin{bmatrix} x_{1+L} & x_{l} & \cdots & x_{1} \\ x_{2+L} & x_{1+L} & \cdots & x_{2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N} & x_{N-1} & \cdots & x_{N-L} \end{bmatrix} \in R^{(N-L)\times(L+1)D}, k > L, \quad (2)$$

when the sampling time k is smaller than the time delay L, the form of the augmented matrix  $X_b$  is as follows:

$$X_{a} = \begin{bmatrix} x_{1} & x_{1} & \cdots & x_{1} \\ x_{2} & x_{1} & \cdots & x_{1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{L} & x_{L-1} & \cdots & x_{1} \end{bmatrix} \in R^{L \times (L+1)D}, k \le L.$$
(3)

Then, the final augmented matrix formed by the training data X as  $X_L$ :

$$X_L = \left[X_a; X_b\right] \in \mathbb{R}^{N \times (L+1)D}.$$
(4)

The AKMD-t-SNE algorithm is used to map the augmented high-dimensional data  $X_L \in \Re^{N \times (l+1)D}$  to the lowdimensional feature subspace  $Y_L = [y_1, y_2, ..., y_N] \in \Re^{d \times N}$ , where  $X_L$  is a matrix with N observations and (l + 1)D dimensions.  $Y_L$  is the matrix with N observations and d dimensions.

The Mahalanobis distance could eliminate the discrepancies among different modes of training data.  $x_i, x_j$  are two samples in the high-dimensional data. The Mahalanobis distance between high-dimensional samples  $x_i$  and  $x_j$  is calculated as

$$D_M(x_i, x_j) = \sqrt{\left(x_i - x_j\right)^T \Sigma^{-1} \left(x_i - x_j\right)}, \qquad (5)$$

where  $\Sigma$  is the covariance matrix of  $X_L$ .

Since Mahalanobis distance is not suitable to deal with nonlinear problems, the kernel trick is introduced for the purpose of capturing the nonlinearity. By virtue of a nonlinear transformation  $\phi_k = F(x_k)$ , the augmented vector is mapped into a higher-dimensional feature space. Denote the mapped matrix as  $\Phi = [\phi_1, \phi_2, \dots, \phi_N]$ . The augmented kernel Mahalanobis distance is determined as where  $(\phi_i, \phi_i) \in \Phi$ ,  $\Sigma_{\phi}$  represents the covariance matrix of  $\Phi$ :

$$\Sigma_{\phi} = \frac{1}{N-1} \sum_{k=1}^{N} (\phi_k - \mu_{\phi}) (\phi_k - \mu_{\phi})^{\mathrm{T}}$$

$$= \frac{1}{N-1} \overline{\Phi} \overline{\Phi}^{\mathrm{T}},$$

$$\mu_{\phi} = \frac{1}{N} \sum_{k=1}^{N} \phi_k,$$
(8)

where  $\overline{\Phi} = [\overline{\phi}_1, \overline{\phi}_2, \dots, \overline{\phi}_N]$  satisfies  $\overline{\Phi} = \Phi H$ , *H* is a symmetric and idempotent matrix:

$$H = I_N - \frac{1}{N} I_N I_N^{\mathrm{T}}.$$
(9)

The dot product of mapped vectors is calculated by the Gaussian kernel function:

$$K(x_i, x_j) = \phi_i^{\mathrm{T}} \phi_j = \exp\left(-\frac{1}{2\sigma^2} (x_i - x_j)^{\mathrm{T}} (x_i - x_j)\right).$$
(10)

Then, the centered kernel matrix is defined as

$$\overline{K} = \overline{\Phi}^{\mathrm{T}} \overline{\Phi} = H^{\mathrm{T}} \Phi^{\mathrm{T}} \Phi H = H K H.$$
(11)

A full rank factorization via singular value decomposition (SVD) method is analyzed:

$$\overline{\Phi} = USV^{\mathrm{T}},\tag{12}$$

where U and V are both orthogonal matrices.

From (7), we have

$$\Sigma_{\phi} = \frac{1}{N-1} USS^{\mathrm{T}} U^{\mathrm{T}}.$$
 (13)

Then, its inverse matrix is calculated:

$$\Sigma_{\phi}^{-1} = (N-1)U(SS^{\mathrm{T}})^{-1}U^{\mathrm{T}}.$$
 (14)

The centered kernel matrix and its Moore-Penrose pseudoinverse matrix are given as

$$\overline{K} = SV^{\mathrm{T}}VS^{\mathrm{T}},\tag{15}$$

$$\overline{K}^{+} = V \left( S^{\mathrm{T}} S \right)^{+} V^{\mathrm{T}}.$$
(16)

Combining (15) with (17), we obtain

$$\Sigma_{\phi}^{-1} = (N-1)\overline{\Phi} \left(\overline{K}^{+}\right)^{2} \overline{\Phi}^{\mathrm{T}}.$$
(17)

The augmented kernel Mahalanobis distance can be rewritten as

$$D_{KM}(\phi_{i},\phi_{j}) = \sqrt{(\phi_{i}-\phi_{j})\Sigma_{\phi}^{-1}(\phi_{i}-\phi_{j})}$$
$$= \sqrt{(N-1)(\overline{\phi}_{i}-\overline{\phi}_{j})\overline{\Phi}(\overline{K}^{+})^{2}\overline{\Phi}^{\mathrm{T}}(\overline{\phi}_{i}-\overline{\phi}_{j})^{\mathrm{T}}}$$
$$= \sqrt{(N-1)\widehat{k}_{i,j}}^{\mathrm{T}}H(\overline{K}^{+})^{2}H\widehat{k}_{i,j},$$
(18)

where  $\hat{k}_{i,j} = \Phi \phi_i - \Phi \phi_j$ .

The conditional probability is calculated between samples in higher-dimensional space. The conditional probability of sample  $\phi_j$  as the nearest neighbor of  $\phi_i$  in higher-dimensional space is denoted as  $p_{j|i}$ :

$$p_{j|i} = \begin{cases} \frac{\exp\left(-D_{KM}^{2}(\phi_{i},\phi_{j})/2\sigma_{i}^{2}\right)}{\sum\limits_{k\neq i}\exp\left(-D_{KM}^{2}(\phi_{i},\phi_{k})/2\sigma_{i}^{2}\right)}, & i\neq j, \\ 0, & i=j, \end{cases}$$
(19)

where  $\sigma_i$  represents the variance of the Gaussian distribution with  $\phi_i$  as the center, and  $\phi_k$  represents one of the nearest neighbors of  $\phi_i$ .

The selection of  $\sigma_i$  value is related to the degree of *Perplexity*, which is defined as

$$\operatorname{Prep}\left(P_{i}\right) = 2^{H\left(P_{i}\right)},\tag{20}$$

where  $H(p_i)$  is the Shannon entropy of  $p_i$ :

$$H(P_i) = -\sum_{j} p_{(j|i)} \log_2 p_{(j|i)}.$$
 (21)

Assuming that the mapping samples of higherdimensional observations  $\phi_i$  and  $\phi_j$  in the lowdimensional space are  $y_i$  and  $y_j$ , and the variance of the Gaussian distribution with  $y_i$  as the center is set to  $1/\sqrt{2}$ , the corresponding conditional probability that  $y_j$  is the nearest sample of  $y_i$  denoted as  $q_{j|i}$ :

$$q_{(j|i)} = \begin{cases} \frac{\exp\left(-\|y_{i} - y_{j}\|^{2}\right)}{\sum\limits_{k \neq i} \exp\left(-\|y_{i} - y_{k}\|^{2}\right)}, & i \neq j, \\ 0, & i = j, \end{cases}$$
(22)

where  $y_k$  represents one of the nearest neighbors of  $y_i$ .

Kullback-Leibler Divergence is used to measure the similarity between the two individual distributions. The similarity cost function *C* as:

$$C = \sum_{i} KL(P_{i} \| Q_{i})$$

$$= \sum_{i} \sum_{j} p_{(j|i)} \log \frac{P_{(j|i)}}{q_{(j|i)}}.$$
(23)

The cost function C is minimized by gradient descent, and the gradient of the cost function C is obtained:

$$\frac{\delta C}{\delta y_i} = 2 \sum_{j} \left( p_{(j|i)} - q_{(j|i)} + p_{(i|j)} - q_{(i|j)} \right) \left( y_i - y_j \right).$$
(24)

The conditional probability is replaced by the joint probability distribution, such that for any *i* and *j*,  $q_{ji} = q_{ij}$ ,  $p_{ji} = p_{ij}$ . The following symmetric method is adopted to simplify the objective function:

$$p_{ij} = \frac{p_{(j|i)} + p_{(i|j)}}{2N}.$$
(25)

The cost function *C* is transformed into the following expression:

$$C = KL(P||Q) = \sum_{i} \sum_{j} p_{ij} \log \frac{p_{ij}}{q_{ij}}.$$
 (26)

In order to solve the crowding problem of lowdimensional data in classification, *t*-distribution is used to replace Gaussian distribution. Redefine  $q_{ij}$  with a *t*distribution of 1 degree of freedom, it follows that:

$$q_{ij} = \frac{\left(1 + \left\|y_i - y_j\right\|^2\right)^{-1}}{\sum\limits_{k \neq l} \left(1 + \left\|y_k - y_l\right\|^2\right)^{-1}}.$$
 (27)

The gradient formula is converted to the following equation:

$$\frac{\delta C}{\delta y_i} = 4 \sum_{j} (p_{ij} - q_{ij}) (y_i - y_j) (1 + ||y_i - y_j||^2)^{-1}.$$
 (28)

To avoid falling into the local optimum in the gradient descent process, a momentum term is added in the gradient update, as follows:

$$y^{(t)} = y^{(t-1)} + \eta \frac{\delta C}{\delta y} + \alpha(t) (y^{(t-1)} - y^{(t-2)}), \qquad (29)$$

where  $y^{(t)}$  represents the solution of the *t* th iteration,  $\alpha(t)$  represents the momentum of the *t* th iteration, and  $\eta$  represents the learning efficiency.

The low-dimensional feature subspace  $Y_L$  is obtained through the above steps. It is assumed that the higherdimensional data are projected to the lower-dimensional features under the action of the mapping matrix A, which can be obtained from the Moore–Penrose pseudoinverse matrix  $X_L^+$  of  $X_L$ :

$$A = Y_L X_L^{+}.$$
 (30)

The feature subspace Y is recalculated via the mapping matrix and the residual subspace is denoted as E:

$$Y = AX_L,$$
  

$$E = X_L - \widehat{X_L}$$
  

$$= X_L - A^+Y,$$
  
(31)

where  $\widehat{X_L}$  is the high-dimensional projection of lowdimensional data and  $A^+$  is the Moore–Penrose pseudoinverse of A. For testing data  $x_{on}$ , the augmented matrix  $x_{onL}$  is also constructed to ensure that the dimensions of online data are consistent with training data. The low-dimensional features of the augmented online data are calculated via the mapping matrix A:

$$y_{onL} = Ax_{onL}.$$
 (32)

The residual of  $x_{onL}$  is given as

$$e_{onL} = x_{onL} - \widehat{x_{onL}} = x_{onL} - A^{+} y_{onL}.$$
 (33)

2.2. Fault Detection Method. After dimension-reduction by the AKMD-t-SNE scheme, the low-dimensional data will be expressed as *t*-distribution, seriously deviating from the normal distribution. Whereas the effective establishment of traditional statistics needs to be carried out under the condition that the data obey the Gaussian distribution. Consequently, the establishment of statistics using  $T^2$  and SPE will affect the estimation of the control limit, leading to the unsatisfactory effect of fault detection. Hence, this section uses the SVDD method to construct statistics. The SVDD algorithm is a single classification model proposed by Tax. The target of SVDD is to seek a hypersphere with center *a* and radius *R*, which can contain all the training data. SVDD statistics were established in the FS (AKMD-t-SNE-FS-SVDD) and RS (AKMD-t-SNE-RS-SVDD), respectively.

To establish the hypersphere model for the FS of training data is to solve the following optimization problem:

$$\min_{R,a,\xi} R_{y}^{2} + C_{1} \sum_{i=1}^{n} \xi_{i},$$

$$s.t. \quad \left| \Phi(Y_{i}) - a_{y} \right\|^{2} \le R_{y}^{2} + \xi_{i}, \xi_{i} \ge 0,$$

$$(34)$$

where  $Y_i$  is the *i* th support vector of the FS;  $\xi_i$  is the slack variable to improve the fault tolerance of the SVDD model;  $C_1$  is the penalty factor, the size of the  $C_1$  reflects the extent of choosing outliers for a data set.

To achieve fault detection on the feature of online data  $(y_{onL})$  is to calculate the distance between the sample and spherical center  $a_v$  and compare it with  $R_v$ .

Distance<sub>y</sub> = 
$$\left\| \Phi\left(y_{onL}\right) - a_{y} \right\|$$
  
=  $\sqrt{1 - 2\sum_{i=1}^{n} \alpha_{i} K\left(y_{onL}, Y_{o}\right) + \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_{i} \alpha_{j} K\left(Y_{i}, Y_{j}\right)} > R_{y}$  (35)  
 $\cdot$  s.t.  $0 \leq \alpha_{i} \leq C_{1}, \sum_{i=1}^{N} \alpha_{i} = 1,$ 

where  $Y_0$  is one of the support vectors in the FS of training data,  $Y_i$  and  $Y_j$  is the *i* th and *j* th support vector in the FS of training data.

For the residual of online data  $(e_{onL})$ , SVDD statistics (Distance<sub>e</sub>) and control limit ( $R_e$ ) are calculated in the same way to determine whether a fault occurs.

2.3. Fault Isolation Method. After the abnormal condition of the system is detected, it is necessary to diagnose the cause of the fault. Generally, the location of the disturbance is analyzed by isolating the fault variable. It has been proved that the mutual information (MI) method can effectively realize fault isolation, but it still has some defects. For the purpose of eliminating the influence of suspicious variables on the isolation results, an isolation method based on the hybrid correlation coefficient (HCC) is proposed to modify the MI method. MI is a measurement of the correlation between variables. Suppose there are two random variables  $V_x$  and  $V_y$ , and the MI of  $V_x$  and  $V_y$  is defined as  $I(V_x, V_y)$ :

$$I(V_X, V_Y) = \sum_{v_x \in V_X} \sum_{v_y \in V_Y} p(v_x, v_y) \log \frac{p(v_x, v_y)}{p(v_x)p(v_y)}, \quad (36)$$

where  $p(v_x)$  and  $p(v_y)$  are the marginal distribution of  $V_x$ and  $V_y$  respectively, and  $p(v_x, v_y)$  are the joint distribution of  $V_x$  and  $V_y$ .

Once the abnormal is detected in the augmented online data, the augmented online data containing normal data and abnormal data is denoted as  $Z_{onL} = \{Z_l\}_{l=1}^{L+1} \in \Re^{(L+1)D \times N_1}$ , where  $N_1$  represents the number of samples and (L+1)D represents the number of variables. Since the augmented matrix contains the data of the previous L sampling instants,  $Z_{onL}$  is composed of submatrices  $Z_l$ , and the expression is given as

$$Z_{onL} = \begin{bmatrix} z_1^1, \cdots, z_1^D Z_1 \in \Re^{D \times N_1}, z_2^1, \cdots, z_2^D Z_2 \in \Re^{D \times N_1}, \cdots, z_{L+1}^1, \cdots, z_{L+1}^D Z_{L+1} \in \Re^{D \times N_1} \end{bmatrix}.$$
 (37)

The FS statistics of  $Z_{onL}$  constitute the vector  $Z_y$ . Compute the MI between  $Z_y$  and each vector in each submatrix as

$$MIy_l^q = I(Z_l^q, Z_y)q, = 1, 2, \cdots, D.$$
 (38)

Compute the MI between each vector in the online data and  $\boldsymbol{Z}_y$  as

$$MIy(q) = \frac{\sum_{l=1}^{L+1} MIy_l^{q}}{L+1}, \quad q = 1, 2, \cdots, D.$$
(39)

On this basis, a fault isolation method based on correlation coefficient (CC) is proposed. In order to understand the correlation level of variables expressed by *MIy*, the MI between  $Z_y$  and itself is calculated as  $I(Z_y, Z_y)$ . The correlation coefficient of FS ( $CC_y$ ) is established as

$$CC_{y}(q) = \frac{MIy^{3}(q)}{I^{2}(Z_{y}, Z_{y})}, \quad q = 1, 2, \cdots, D.$$
 (40)

The impact of each variable in the FS on the occurrence of disturbances was judged by the size of  $CC_{y}(q)$ .

Similarly, the variable composed of RS statistics is denoted as  $Z_e$ , and the CC of RS ( $CC_e$ ) is calculated in the same way to achieve fault isolation.

The hybrid mutual information (HMI) and the hybrid correlation coefficient (HCC) were defined as

$$HMI(q) = \left(\omega \frac{MIy(q)}{\sum_{q=1}^{D} MIy(q)} + (1-\omega) \frac{MIe(q)}{\sum_{q=1}^{D} MIe(q)}\right) \times 100\%,$$

$$HCC(q) = \left(\omega \frac{CC_{y}(q)}{\sum_{q=1}^{D} CC_{y}(q)} + (1-\omega) \frac{CC_{e}(q)}{\sum_{q=1}^{D} CC_{e}(q)}\right) \times 100\%.$$
(41)

Calculate the weight coefficient  $\omega$ :

$$\omega_{1} = \frac{Z_{y}}{R_{y}},$$

$$\omega_{2} = \frac{Z_{e}}{R_{e}},$$

$$\omega = \frac{\omega_{1}}{(\omega_{1} + \omega_{2})}.$$
(42)

The process monitoring flow diagram of the proposed method is shown in Figure 1.

#### 3. Results and Discussion

*3.1. Numerical Case.* To testify the effectiveness of the proposed method, a numerical dynamic multimode process is simulated. The dynamic multimode model is created as follows:

$$\begin{cases} x_1 = 0.6352s_1 + 0.1254s_2 + e_1, \\ x_2 = 0.6951s_1 + 0.1557 \ln (1 + s_2^2) + e_2, \\ x_3 = 0.5028s_1s_2 + 0.8312 \ln (2 + \cos (s)_2)) + e_3, \\ x_4 = 0.4152s_1 + 0.2617s_2^2 + e_4, \\ x_5 = 0.0569x_3 + 0.7962 \sin (x_4) + e_5, \end{cases}$$
(43)

where  $s_1$  and  $s_2$  are latent variables,  $[e_1, e_2, e_3, e_4, e_5]^T$  are independent zero-mean Gaussian white noises with and a standard deviation of 0.01. Variable  $x_5$  is the correlation variable of variable  $x_3$  and variable  $x_4$ . Set three modes for the latent variable as

$$Mode1: \frac{s_{1}(\tau): \quad Uniform (-5, -1)}{s_{2}(\tau): \quad N (-2, 2),}$$

$$Mode2: \frac{s_{1}(\tau) = 0.8 \cos^{2}(s_{1}(\tau - 1)) + 5e^{\sin(\tau - 1)}}{s_{2}(\tau) = 1.5 \cos(e^{s_{2}(\tau - 1)}) + \ln(1 + s_{2}^{-2}(\tau - 1)),}$$

$$Mode3: \frac{s_{1}(\tau) = 0.1 \sin^{2}(s_{1}(\tau - 1)) + 2.8e^{\cos^{2}(s_{1}(\tau - 1))}}{s_{2}(\tau) = 3.25e^{0.3 \sin(s_{2}(\tau - 1))} + 0.1 \cos(s_{2}(\tau - 1)),}$$

$$(44)$$

where  $\tau$  represents the index of the current sampling instant. The initial state of mode 2 and mode 3 is 0, i.e.,  $s_1(1) = s_2(1) = 0$ . Five hundred samples were generated in each mode, that is, a total of 1500 samples were generated in the three modes to form the training data. In order to show the spatial distribution characteristics of the data of the three modes intuitively, Figure 2 shows the scatter plot of variables  $x_2$ ,  $x_3$ , and  $x_4$  in three modes.

Set two types of fault:

Type 1: the system operates normally for 200 sampling instants in mode 2. At the 201st sampling instant ( $\tau = 201$ ), a step signal with an amplitude of +1.35 is added to variable  $x_2$ , and 800 sampling instants are continued in this state.

Type 2: the system operates normally for 200 sampling instants in mode. At the 201st sampling instant ( $\tau = 201$ ), the ramp signal of  $0.03(\tau - 200)$  is added to variable 3 and continues to operate for 800 sampling instants in this state.

Each of the two fault types has 1000 test samples. The augment matrix-based dynamic locally linear embedding (DLLE), t-SNE, t-SNE based on Mahaobanobis distance (M-t-SNE), and AKMD-t-SNE are used to extract the features of the established dynamic multimode numerical model.



FIGURE 1: Dynamic multimode process monitoring flow chart based on AKMD-t-SNE-SVDD and HCC.

SVDD statistics are established in all the four methods for fault detection. According to experience, the dimension reduction of the four methods is 2, the nearest neighbor of DLLE is 25, the penalty coefficient of SVDD is 0.01, and the kernel width is 300. From the perspective of accuracy and running time, the number of iterations of t-SNE, M-t-SNE, and AKMD-t-SNE were all selected as 300, and the *Perplexity* was set as 30. The time delay of DLLE and AKMDt-SNE is 10; The confidence level of the fault detection control limit is 99%.

Figures 3 and 4 display the fault detection plots for fault type 1 and fault type 2, respectively. As can be seen from Figure 3, the FS of the four methods can detect the occurrence of faults to a certain extent. The AKMD-t-SNE-FS-SVDD method can distinguish between normal data and abnormal data more obviously and exhibits a better detection effect over those of the alternative models. After the addition of disturbance, the statistics of the four methods all changed to varying degrees in Figure 4. The proposed method achieves superior detection results in both FS and RS.

Generally, it is considered that the detection of six or more continuous statistics exceeding the control limit indicates that the fault is intrinsically detected. The first



FIGURE 2: Numerical case 3D scatter plot.

sampling instant that fault is intrinsically detected is denoted as DetectionTime, and the time at which the disturbance was added is denoted as OccurTime. Delay is defined as

$$Delay = DetectionTime - OccurTime.$$
 (45)



FIGURE 3: Numerical case fault type 1 detection results of different methods: (a) DLLE-SVDD, (b) t-SNE-SVDD, (c) M-t-SNE-SVDD, and (d) AKMD-t-SNE-SVDD.

Table 1 shows the results of fault detection performance indicators for numerical case, including the miss alarm rate (MAR), detection delay, and average false alarm rate (FAR) of the two types of faults. The detection delay is defined by the earlier Delay between the FS and the RS. Under each fault type in Table 1, the position where the method proposed in this paper obtains the most excellent detection performance indexes are marked in bold. Compared with other methods, the AKMD-t-SNE-SVDD method can capture the information of anomaly more rapidly and obtains the lowest MAR.

After the system is detected in an abnormal state, fault isolation is realized through HMI and HCC, respectively, based on the four methods, and the sampling interval is  $\tau \in [1, 500]$ . Figures 5 and 6 are fault isolation bar charts for type 1 and type 2, respectively. The reason for the occurrence of fault type 1 is that variable  $x_2$  is added with step disturbance. In Figure 5, the isolation results of the DLLE-SVDD method and the AKMD-t-SNE-SVDD

method fit the fault description. Specifically, in the AKMD-t-SNE-SVDD-HCC method, the probability that  $x_2$  is judged to be the primary responsible variable exceeds 60%. Fault type 2 occurs because a ramp signal is added to variable  $x_3$ . Since variable  $x_5$  is the related variable of  $x_3$ ,  $x_3$  is the main variable causing the fault and  $x_5$  is the secondary responsible variable. The method in Figure 6(c) cannot isolate the variables which caused the change.  $x_3$  and  $x_5$  can be isolated in Figures 6(a), 6(b), and 6(d), where Figure 6(d) is more consistent with the fault description. Compared with the HMI isolation method, the ©rrelevant variables of the HCC method have less impact on the isolation results, which verifies the effectiveness of the proposed method.

3.2. Multimode Tennessee Eastman Process. TE Process (TEP) is extensively researched in process monitoring owing to its complexity. The flow chart of TEP is shown in Figure 7, which is composed of five major units, including



FIGURE 4: Numerical case fault type 2 detection results of different methods: (a) DLLE-SVDD, (b) t-SNE-SVDD, (c) M-t-SNE-SVDD, and (d) AKMD-t-SNE-SVDD.

Fault type	MAR (%) and delay in numerical case													
	DLLE-SVDD			t-SNE-SVDD			M-t-SNE-SVDD			AKMD-t-SNE-SVDD				
	FS	RS	Delay	FS	RS	Delay	FS	RS	Delay	FS	RS	Delay		
1	40.63	99.00	113	24.25	96.25	15	2.38	96.63	0	0.5	98.88	4		
2	82.13	2.38	0	68.13	43.38	540	56.63	51.25	494	21.38	2.38	0		
		Average FAR (%) in numerical case												
	11.25	41.56	_	0.25	20.75	_	9.00	19.75	_	4.75	41.50			

TABLE 1: Results of numerical case fault detection performance indicators.

the reactor, condenser, gas-liquid separator, stripper, and compressor [36]. A total of 6 working conditions are set in the multimode TE Process (MTEP), and 20 fault conditions are set for each mode. The detailed descriptions of 20 types of abnormal disturbances are shown in Table 2 [37]. There are 41 measurement variables and 11 manipulated variables in the TEP, among which the measurement variables are composed of 22 continuous measurement variables and 19 component variables. The stripper steam valve in the manipulated variables is constant at 1 under any mode, thus, 22 continuous measurement variables and 10 manipulated variables are selected to achieve modeling and monitoring. MTEP of the MATLAB simulation model can be downloaded from https://depts.washington.edu/control/ LARRY/TE/download.html.



FIGURE 5: Numerical case fault type 1 isolation bar charts of different methods based on HMI and HCC: (a) DLLE-SVDD, (b) t-SNE-SVDD, (c) M-t-SNE-SVDD, and (d) AKMD-t-SNE-SVDD.

When the system is running normally, 300 samples are collected in each mode, and a total of 1800 samples in six modes constitute the training data set. The abnormal data of mode 1 and mode 2 were selected as testing data, with 1000 samples for each disturbance, and abnormal disturbance was added from the 201st sampling instant. As shown in Figure 8, the variables  $v_1$ ,  $v_2$ , and  $v_3$  of the training data were selected to draw a three-dimensional scatter diagram.

The four models of the numerical case are also applied to fault detection of MTEP. According to historical experience, the dimensionality reduction of the four algorithms is set to 15, the iteration times of t-SNE, M-t-SNE, and AKMD-t-SNE are all selected as 500. Other parameters remain the same as in the above numerical case.

Figures 9 and 10 show the fault detection plots of IDV (3) in mode 1 and IDV (12) in mode 2, respectively. In Figure 9,

DLLE and t-SNE methods cannot detect the occurrence of the disturbance, while the M-t-SNE method can briefly identify system anomalies, only AKMD-t-SNE-SVDD can effectively detect a persistent anomaly. As shown in Figure 10, the SVDD distances of the four methods have changed to different degrees when the disturbance occurs. Compared with the other three methods, AKMD-t-SNE-FS-SVDD has obtained a higher fault detection rate.

Tables 3 and 4, respectively, show the results of the fault detection performance indicators for MTEP mode 1 and mode 2, where  $\triangle$  indicates that the fault is not truly detected. Compared with other methods, the method proposed in this paper has obvious fault detection performance advantages in multiple fault types.

After the system anomaly is detected, fault isolation is realized through HMI and HCC respectively, and the data



FIGURE 6: Numerical case fault type 2 isolation bar charts of different methods based on HMI and HCC: (a) DLLE-SVDD, (b) t-SNE-SVDD, (c) M-t-SNE-SVDD, and (d) AKMD-t-SNE-SVDD.

sampling interval is  $\tau \in [1, 500]$ . Figure 11 shows the fault isolation results of IDV (3) in mode 1. The fault of IDV (3) is described as the step disturbance added to the D feed temperature, which directly affects the change of the D feed flow valve ( $v_{23}$ ) under the regulation of the control system. The change of feed D causes the change of E feed flow valve ( $v_{24}$ ) and A feed flow valve ( $v_{25}$ ) at the same time, and the change of feed also affects the relevant variables of the upstream stripper and downstream separator ( $v_{13}$ ,  $v_{16}$ ,  $v_{29}$ ,  $v_{30}$ ). The main variables isolated by AKMD-t-SNE-SVDD-HCC conform to the fault description. Figure 12 shows the comparison of the trajectories of the three main isolation variables of IDV (3) in mode 1 under normal and abnormal conditions.

Figure 13 shows the isolation results of IDV (12) in mode 2. The reason for the occurrence of IDV (12) is that the

condenser cooling water inlet temperature adds random disturbance. As the condenser acts directly on the output product of the reactor, it will directly affect the variable of reactor pressure  $(v_7)$ . The change of the output product of the reactor will also cause the alteration of the upstream stripper pressure ( $v_{16}$ ). The gas-liquid separator pressure ( $v_{13}$ ) in the downstream unit of the condenser is also affected, which causes the fluctuation of the separator temperature  $(v_{11})$ , the purge valve  $(v_{28})$ , and the purge rate  $(v_{10})$ . Thus, the isolated responsible variables by the proposed scheme are consistent with the fault description of IDV (12). Figure 14 shows the comparison of the trajectories of the three major isolation variables of IDV (12) in mode 2 under the normal and abnormal conditions. The fault isolation results of the proposed method outperform the other methods in the above-given experiments.



FIGURE 7: TEP flowchart.

TABLE 2: Disturbances	in	the	ΤE	process.
-----------------------	----	-----	----	----------

No.	Variable description	Туре
IDV (1)	A/C feed ratio, B composition constant	Step
IDV (2)	B composition, A/C feed ratio constant	Step
IDV (3)	D feed temperature	Step
IDV (4)	Reactor cooling water inlet temperature	Step
IDV (5)	Condenser cooling water inlet temperature	Step
IDV (6)	A feed loss	Step
IDV (7)	C Header pressure loss-reduced availability	Step
IDV (8)	A, B, C feed composition	Random
IDV (9)	D feed temperature	Random
IDV (10)	C feed temperature	Random
IDV (11)	Reactor cooling water inlet temperature	Random
IDV (12)	Condenser cooling water inlet temperature	Random
IDV (13)	Reaction kinetics	Slow drift
IDV (14)	Reaction kinetics	Sticking
IDV (15)	Condenser cooling water valve	Sticking
IDV (16)	Unknown	Unknown
IDV (17)	Unknown	Unknown
IDV (18)	Unknown	Unknown
IDV (19)	Unknown	Unknown
IDV (20)	Unknown	Unknown

Table 5 shows the fault isolation results of mode 1 and mode 2 based on the HCC method. The table presents the three most important isolation variables and HCC values of each type of fault disturbance. Due to the different working conditions, there are some differences in the main variables of isolation under the two modes, and the isolated responsible variables in the two modes basically correspond to the disturbance description.



FIGURE 8: MTEP 3D scatter diagram.



FIGURE 9: MTEP mode 1 IDV (3) fault detection results of different methods: (a) DLLE-SVDD, (b) t-SNE-SVDD, (c) M-t-SNE-SVDD, and (d) AKMD-t-SNE-SVDD.



FIGURE 10: MTEP mode 2 IDV (12) fault detection results of different methods: (a) DLLE-SVDD, (b) t-SNE-SVDD, (c) M-t-SNE-SVDD, and (d) AKMD-t-SNE-SVDD.

IDV	Γ	DLLE-SVDE	)	t-	t-SNE-SVDD			M-t-SNE-SVDD			AKMD-t-SNE-SVDD		
	FS	RS	Delay	FS	RS	Delay	FS	RS	Delay	FS	RS	Delay	
1	3.63	3.25	26	1.38	2.88	13	1.63	3.00	14	2.38	3.25	19	
2	29.63	41.13	32	6.75	38.00	27	2.88	41.50	26	3.88	41.13	32	
3	100.00	100.00	$\triangle$	96.25	99.88	$\triangle$	87.38	99.88	228	2.88	100.00	1	
4	0.00	98.63	0	99.63	97.75	4	0.00	98.13	0	0.00	98.63	0	
5	100.00	100.00	$\triangle$	96.63	97.25	$\triangle$	97.63	99.75	$\triangle$	24.75	100.00	100	
6	0.00	0.75	0	0.00	0.13	0	0.00	0.50	0	0.00	0.75	0	
7	0.00	84.50	0	0.13	85.50	0	0.00	85.00	0	0.00	84.50	0	
8	15.13	26.63	114	20.50	28.38	116	12.00	27.38	107	12.38	26.63	102	
9	100.00	100.00	$\triangle$	92.13	96.00	72	84.75	99.38	86	14.38	100.00	19	
10	25.63	100.00	171	36.50	99.25	187	23.00	99.13	168	16.75	100.00	107	
11	11.88	60.50	24	68.63	65.50	25	11.00	60.75	23	3.25	60.63	21	
12	99.88	74.75	79	78.75	62.75	69	67.38	77.25	80	27.75	75.00	79	
13	16.13	12.50	100	12.50	12.13	98	11.00	12.00	93	12.50	12.50	100	
14	17.00	99.50	71	71.25	93.88	117	7.63	96.75	65	7.63	99.50	66	
15	100.00	100.00	$\triangle$	100.00	98.88	$\triangle$	88.63	99.25	228	32.75	100.00	41	

	MAR (%) and delay in MTEP mode 1													
IDV	DLLE-SVDD			t-SNE-SVDD			M-t-SNE-SVDD			AKMD-t-SNE-SVDD				
	FS	RS	Delay	FS	RS	Delay	FS	RS	Delay	FS	RS	Delay		
16	100.00	100.00	$\triangle$	100.00	97.88	$\triangle$	87.13	99.63	228	18.13	100.00	22		
17	3.88	78.25	24	13.38	78.00	40	3.13	80.13	25	2.13	78.13	17		
18	71.88	95.13	464	98.38	93.38	685	60.13	96.13	228	24.88	95.25	107		
19	12.00	93.13	64	55.25	91.63	151	4.13	92.75	43	2.75	93.13	21		
20	26.75	57.63	165	30.88	47.50	238	19.00	57.75	167	10.75	57.63	42		
					Average	FAR (%) i	in MTEP 1	node 1						
	1.05	1.53	_	3.30	1.83	_	2.53	0.68	_	10.48	1.18	_		

TABLE 3: Continued.

TABLE 4: Results of fault detection performance indicators for MTEP mode 2.

IDV	D	DLLE-SVDD			t-SNE-SVDD			M-t-SNE-SVDD			AKMD-t-SNE-SVDD		
	FS	RS	Delay	FS	RS	Delay	FS	RS	Delay	FS	RS	Delay	
1	4.00	2.13	17	3.25	2.00	23	1.25	1.50	14	1.25	2.13	10	
2	15.38	2.25	18	15.50	1.88	14	3.00	1.63	14	4.38	2.25	18	
3	100.00	98.25	123	100.00	96.50	$\triangle$	83.75	89.00	193	17.88	98.25	102	
4	0.13	92.00	1	100.00	97.38	6	0.00	86.13	0	0.00	92.13	0	
5	100.00	98.63	258	99.50	98.00	$\triangle$	87.00	90.63	194	27.25	98.75	5	
6	0.00	0.13	0	0.00	0.25	0	0.00	0.00	0	0.00	0.13	0	
7	0.00	79.25	0	0.13	77.63	0	0.00	76.00	0	0.00	79.38	0	
8	18.75	12.75	102	20.63	24.38	118	13.50	13.25	99	12.88	12.75	102	
9	100.00	98.38	123	100.00	98.88	$\triangle$	84.63	88.75	118	32.00	98.38	103	
10	26.13	96.63	123	50.00	91.13	210	22.88	86.00	162	14.13	96.63	105	
11	22.63	33.88	25	59.00	71.88	28	12.88	42.13	23	4.63	33.88	23	
12	99.63	61.00	30	87.00	84.00	28	71.38	64.75	27	15.63	61.00	18	
13	14.63	10.00	80	10.00	10.63	85	10.13	9.00	80	12.75	10.00	80	
14	3.00	43.50	21	68.88	88.88	$\triangle$	2.50	63.25	20	0.88	43.25	12	
15	100.00	96.88	122	100.00	91.25	118	86.63	89.25	118	17.88	96.88	113	
16	99.88	96.13	90	100.00	100.00	$\triangle$	94.38	87.63	228	20.50	96.00	89	
17	14.00	74.63	26	32.63	92.13	37	3.13	77.50	25	2.75	74.50	20	
18	71.50	91.25	123	90.13	96.63	645	45.75	85.38	191	16.00	91.38	114	
19	20.88	83.88	67	54.88	83.50	118	8.38	77.88	55	3.00	83.88	24	
20	26.00	22.75	122	27.88	40.38	104	19.38	27.13	118	12.50	22.75	100	
					Average	FAR (%) in	n MTEP n	node 2					
	0.00	0.03	_	2.63	2.03	_	0.33	7.03	_	5.08	0.03		





FIGURE 11: MTEP mode 1 IDV (3) fault isolation bar charts of different methods based on HMI and HCC: (a) DLLE-SVDD, (b) t-SNE-SVDD, (c) M-t-SNE-SVDD, and (d) AKMD-t-SNE-SVDD.



FIGURE 12: Isolated variable trajectories of IDV (3) in mode 1: (a)  $v_{30}$ , (b)  $v_{29}$ , and (c)  $v_{23}$ .



FIGURE 13: Continued.



FIGURE 13: MTEP mode 2 IDV (12) fault isolation bar charts of different methods based on HMI and HCC: (a) DLLE-SVDD, (b) t-SNE-SVDD, (c) M-t-SNE-SVDD, and (d) AKMD-t-SNE-SVDD.



FIGURE 14: Isolated variable trajectories of IDV (12) in mode 2: (a)  $v_{16}$ , (b)  $v_{13}$ , and (c)  $v_7$ .

IDV	Mode 1							Mode 2					
IDV	V	HCC	V	HCC	V	HCC	V	HCC	V	HCC	V	HCC	
1	11	7.52	18	7.22	29	7.06	26	8.18	28	7.15	29	6.89	
2	28	17.31	10	16.98	21	13.97	28	9.86	10	9.86	20	9.79	
3	30	21.56	29	15.73	23	12.76	28	20.78	25	17.96	1	17.02	
4	31	77.80	29	6.37	30	6.04	31	68.25	28	8.37	10	7.27	
5	23	18.68	30	13.11	24	11.25	1	18.62	25	18.24	18	14.19	
6	25	17.28	1	16.77	10	11.25	1	19.28	25	18.77	21	10.77	
7	26	39.82	18	21.06	20	15.56	26	36.00	18	15.94	20	15.59	
8	18	8.23	26	7.98	11	7.79	26	8.28	11	7.25	18	7.20	
9	23	23.93	24	11.26	13	10.54	28	16.79	25	16.36	1	15.05	
10	18	49.31	30	10.53	23	5.63	18	29.91	25	13.00	28	11.64	
11	31	62.42	21	5.82	9	5.50	31	65.88	21	4.89	7	4.73	
12	13	18.81	7	18.78	16	17.78	16	18.27	13	16.41	7	14.55	
13	28	12.79	10	12.75	21	9.95	30	12.41	29	10.19	16	7.80	
14	9	25.63	31	24.00	21	20.38	31	23.87	21	21.67	9	21.37	
15	23	17.10	24	15.16	30	11.26	25	19.44	28	18.59	1	17.49	
16	23	16.32	30	13.78	29	12.39	28	25.20	25	18.39	1	15.12	
17	31	36.07	21	30.32	9	11.45	21	45.49	31	37.46	9	3.93	
18	30	16.21	23	14.39	24	10.73	25	21.10	28	19.45	1	18.49	
19	30	52.74	29	29.00	24	4.66	30	26.02	28	22.42	29	19.39	
20	16	12.32	18	12.32	13	12.10	16	10.67	13	10.42	18	10.31	

TABLE 5: Fault isolation results of MTEP mode 1 and mode 2 based on HCC (%).

### 4. Conclusions

Process monitoring is of great significance to the safe and stable production of the chemical process industry. Most of the research studies on process monitoring mainly focus on the static single-mode process. In this paper, based on the t-SNE algorithm, an AKMD-t-SNE method is proposed to realize feature extraction for dynamic multimode chemical process, and the SVDD method is used to construct statistics for non-Gaussian data fault detection. After the system anomaly is detected, fault isolation is achieved by the HCC method. The results of the numerical case and multimode TE process demonstrate that the proposed method is effective for fault detection and fault isolation in a dynamic multimode chemical process.

# **Data Availability**

The data used to support the findings of this study are available from the corresponding author upon request.

# **Conflicts of Interest**

The authors declare that they have no conflicts of interest.

#### Acknowledgments

This work was supported in part by the National Nature Science Foundation of China (Grant no. NSFC; 61403256), and Nature Science Foundation of Shanghai (Grant no. 19ZR1455200).

#### References

- Z. Ge, Z. Song, and F. Gao, "Review of recent research on data-based process monitoring," *Industrial & Engineering Chemistry Research*, vol. 52, no. 10, pp. 3543–3562, 2013.
- [2] S. J. Qin, "Process data analytics in the era of big data," AIChE Journal, vol. 60, no. 9, pp. 3092–3100, 2014.
- [3] Z. Ge, Z. Song, S. X. Ding, and B. Huang, "Data mining and analytics in the process industry: the role of machine learning," *IEEE Access*, vol. 5, Article ID 20590, 2017.
- [4] R. Iqbal, T. Maniak, F. Doctor, and C. Karyotis, "fault detection and isolation in industrial processes using deep learning approaches," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 5, pp. 3077–3084, 2019.
- [5] A. Miao, Z. Ge, Z. Song, and L. Zhou, "Time neighborhood preserving embedding model and its application for fault detection," *Industrial & Engineering Chemistry Research*, vol. 52, no. 38, Article ID 13717, 2013.
- [6] L. Feng, C. Zhang, and Y. Li, "Study on PC- WKNN-based fault detection method in multimode batch process," *Application Research ofComputers*, vol. 35, no. 4, pp. 1130–1134, 2018.
- [7] S. Zhang, C. Zhao, and F. Gao, "Two-directional concurrent strategy of mode identification and sequential phase division for multimode and multiphase batch process monitoring with uneven lengths," *Chemical Engineering Science*, vol. 178, pp. 104–117, 2018.
- [8] X. Deng and X. Tian, "Multimode process fault detection using local neighborhood similarity analysis," *Chinese Journal* of Chemical Engineering, vol. 22, no. 12, pp. 1260–1267, 2014.

- [9] J. Yu, L. Ye, and L. Zhou, "Dynamic process monitoring based on variational bayesian canonical variate analysis," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 52, pp. 1–11, 2021.
- [10] U. Kruger, Y. Zhou, and G. W. Irwin, "Improved principal component monitoring of large-scale processes," *Journal of Process Control*, vol. 14, no. 8, pp. 879–888, 2004.
- [11] S. X. Ding, "Data-driven design of monitoring and diagnosis systems for dynamic processes: a review of subspace technique based schemes and some recent results," *Journal of Process Control*, vol. 24, no. 2, pp. 431–449, 2014.
- [12] J. Zhu, H. Shi, B. Song, S. Tan, and Y. Tao, "Deep neural network based recursive feature learning for nonlinear dynamic process monitoring," *Canadian Journal of Chemical Engineering*, vol. 98, no. 4, pp. 919–933, 2019.
- [13] K. Wang, R. B. Gopaluni, J. Chen, and Z. Song, "Deep learning of complex batch process data and its application on quality prediction," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 12, pp. 7233–7242, 2020.
- [14] G. Li, S. J. Qin, and D. Zhou, "A new method of dynamic latent-variable modeling for process monitoring," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 11, pp. 6438–6445, 2014.
- [15] W. Ku, R. H. Storer, and C. Georgakis, "Disturbance detection and isolation by dynamic principal component analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 30, no. 1, pp. 179–196, 1995.
- [16] J. B. Tenenbaum, Vd Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.
- [17] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.
- [18] L. Van der Maaten and G. Hinton, "Visualizing non-metric similarities in multiple maps," *Machine Learning*, vol. 87, no. 1, pp. 33–55, 2012.
- [19] G. C. Linderman, M. Rachh, J. G. Hoskins, S. Steinerberger, and Y. Kluger, "Fast interpolation-basedt-SNE for improved visualization of single-cellRNA-seq data," *Nature Methods*, vol. 16, no. 3, pp. 243–245, 2019.
- [20] L. Van Der Maaten, "Accelerating t-SNE using tree-based algorithms," *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 3221–3245, 2014.
- [21] D. M. Chan, R. Rao, and F. Huang, "t-Sne-Cuda: GPUaccelerated T-SNE and its Applications to Modern Data," in Proceedings of the 30th International Symposium On Computer Architecture And High Performance Computing (SBAC-PAD), pp. 330–338, Berkeley, CA, USA, June 2018.
- [22] C. Zhang, K. Peng, and J. Dong, "A P-t-SNE and MMEMPM based quality-related process monitoring method for a variety of hot rolling processes," *Control Engineering Practice*, vol. 89, pp. 1–11, 2019.
- [23] J. Tang and X. Yan, "Neural network modeling relationship between inputs and state mapping plane obtained by FDA-t-SNE for visual industrial process monitoring," *Applied Soft Computing*, vol. 60, pp. 577–590, 2017.
- [24] Y. Qin, W. T. Li, C. Yuen, W. Tushar, and T. K. Saha, "IIoTenabled health monitoring for integrated heat pump system using mixture slow feature analysis," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 7, pp. 4725–4736, 2022.
- [25] J. Zhang, D. Zhou, and M. Chen, "Continual learning for multimode dynamic process monitoring with applications to an ultra-supercritical thermal power plant," *IEEE Transactions On Automation Science And Engineering*, 2022.

- [26] H. Wu and J. Zhao, "Self-adaptive deep learning for multimode process monitoring," *Computers & Chemical Engineering*, vol. 141, Article ID 107024, 2020.
- [27] J. Shang, M. Chen, and H. Zhang, "Fault detection based on augmented kernel Mahalanobis distance for nonlinear dynamic processes," *Computers & Chemical Engineering*, vol. 109, pp. 311–321, 2018.
- [28] D. M. Tax and R. P. Duin, "Support vector data description," *Machine Learning*, vol. 54, no. 1, pp. 45–66, 2004.
- [29] J. Wang, K. Qiu, W. Liu, T. Yu, and L. Zhao, "Unsupervisedmultiscale-sequential-partitioning and multiple-SVDDmodel-basedprocess-monitoring method for multiphase batch processes," *Industrial & Engineering Chemistry Research*, vol. 57, no. 51, Article ID 17437, 2018.
- [30] Q. Jiang, X. Yan, and J. Li, "PCA-ICA integrated with bayesian method for non-Gaussian fault diagnosis," *Industrial & En*gineering Chemistry Research, vol. 55, no. 17, pp. 4979–4986, 2016.
- [31] C. F. Alcala and S. J. Qin, "Reconstruction-based contribution for process monitoring," *Automatica*, vol. 45, no. 7, pp. 1593–1600, 2009.
- [32] R. Dunia, S. Joe Qin, T. F. Edgar, and T. J. McAvoy, "Use of principal component analysis for sensor fault identification," *Computers & Chemical Engineering*, vol. 20, pp. 713–718, 1996.
- [33] H. H. Yue and S. J. Qin, "Reconstruction-based fault identification using a combined index," *Industrial & Engineering Chemistry Research*, vol. 40, no. 20, pp. 4403–4414, 2001.
- [34] C. F. Alcala and S. J. Qin, "Reconstruction-based contribution for process monitoring with kernel principal component analysis," *Industrial & Engineering Chemistry Research*, vol. 49, no. 17, pp. 7849–7857, 2010.
- [35] S. Verron, T. Tiplica, and A. Kobi, "Fault detection and identification with a new feature selection based on mutual information," *Journal of Process Control*, vol. 18, no. 5, pp. 479–490, 2008.
- [36] J. J. Downs and E. F. Vogel, "A plant-wide industrial process control problem," *Computers & Chemical Engineering*, vol. 17, no. 3, pp. 245–255, 1993.
- [37] X. Xie and H. Shi, "Dynamic multimode process modeling and monitoring using adaptive Gaussian mixture models," *Industrial & Engineering Chemistry Research*, vol. 51, no. 15, pp. 5497–5505, 2012.