IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received September 21, 2020, accepted October 12, 2020, date of publication October 20, 2020, date of current version November 12, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3032597

Plant-Wide Process Monitoring Strategy Based on Complex Network and Bayesian Inference-Based Multi-Block Principal Component Analysis

YANAN LI[1], XIN PENG[2], (Member, IEEE), AND YING TIAN[3]

[1]School of Mechanical Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China
[2]Key Laboratory of Advanced Control and Optimization for Chemical Processes, Ministry of Education, East China University of Science and Technology, Shanghai 200237, China
[3]Shanghai Key Laboratory of Modern Optical Systems, School of Optical–Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China

Corresponding author: Ying Tian (tianying@usst.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 61903251.

ABSTRACT In this work, a novel process monitoring method in a block-wised partitioning manor is proposed for plant-wide processes which can be partitioned into several sub-blocks and monitored parallelly. The focus of this method is to reduce the high complexity of global plant-wide process, while to improve the efficiency of local feature extraction. In this method, considering that not all process knowledge is available in the block division process strategy, a novel community discovery (CD) algorithms, based on the similarity of neighbor node weighted Louvain, is introduced into the framework of the multi-block Bayesian inference and principal component analysis (PCA) based plant-wide process monitoring scheme. Firstly, the complex network (CN) theory is used to establish the network topology structure for the global variables of the plant-wide process. Secondly, by analyzing the graph characteristic structure, considering the connection strength between nodes, a more reasonable sub-block division is conducted according to the improved Louvain algorithm. Then, PCA method is used to establish process monitoring model for each sub-block to obtain sub-block monitoring statistics. Finally, the total joint statistics is obtained through Bayesian inference for fault detection. The feasibility and effectiveness, in terms of the detection performance, of this method are demonstrated in a simulated plant-wide process by compared with other state-of-the-art PCA based monitoring methods.

INDEX TERMS Plant-wide process monitoring, complex networks, community discovery algorithms, Bayesian inference, principal component analysis.

I. INTRODUCTION

Due to the improvement of product quality and safety in modern industrial processes, industrial process monitoring has become a hot topic of research [1]–[5]. As the continuous development of modern data collection and storage technology, the types and quantities of industrial process data have been greatly improved, and industrial processes have entered the era of big data. Massive data contains important process variable information. Recent decades, data-driven approaches [6], especially multi-variable statistical process monitoring (MSPM), have achieved increasingly developments [7]–[9] with the help of the improvement of the computing power as well as the data analysis techniques. Among them, PCA is commonly used in MSPM [10], which is a dimension reduction algorithm, due to its efficiency and understandability. The acquired data can be projected into a relative lower-dimensional space to eliminate redundant information between variables, which can process Gaussian and linear data. Considering the possibility of linear and nonlinear problems between variables, Xu proposed [11] a nonlinear process monitoring and fault diagnosis based on kernel principal component analysis (KPCA) and multiple kernel learning support vector machines. In addition to PCA and its extension methods, other process monitoring methods are also used in different monitoring domains. Zhong proposed [12] a quality-related statistical process monitoring method based on global and local partial least squares (PLS)

The associate editor coordinating the review of this manuscript and approving it for publication was Yu Wang.

projection. Zhou proposed [13] an improved PLS algorithm to prove its robustness in process monitoring compared with traditional methods. Wang proposed a novel statistical model based on locality-preserving PLS to improve the nonlinear processing capability of the system. Chen proposed an improved canonical correlation analysis (CCA) to realize the detection of multiplication faults in industrial processes [14]. Peng proposed [15] the combination of just-in-time learning and extreme learning machine (ELM) to deal with non-Gaussian chemical processes. Tian proposed [16] to combine average multivariate accumulation and residual analysis with independent component analysis(ICA) to realize fault detection of non-Gaussian process with periodic disturbance. Du introduces [17] lazy learning (LL), support vector data description (SVDD), and modified receptor density algorithm (MRDA) to online monitor nonlinear multimode processes.

Furthermore, in modern industry, the plant-wide process is characterized by its multiple function-oriented operation units, a large number of process variables and complex variable relationships. Consequently, this kind of complexity will cause unsatisfactory monitoring performance if merely traditional MSPM methods are applied. In this case, the establishment of a global monitoring model, such as global PCA, may mask local fault information and reduce the complexity and accuracy of the plant-wide process monitoring. Therefore, in order to overcome the complexity of process variables in the plant-wide process and extract the suitable local variable information of process data, block-wised partitioning manor or distributed monitoring is an effective solution. Westerhuis *et al.* proposed the plant-wide process monitoring methods of multi-block PCA (MBPCA) or multi-block PLS [18]–[20]. Ma *et al.* proposed [21] a novel key performance indicator oriented hierarchical monitoring and propagation path identification framework for complex industrial processes. This approach uses a new gap measurement approach for monitoring key performance indicator (KPI) oriented faults in the block level. Jiang *et al.* proposed [22] to combine the randomized algorithm (RA) with evolutionary optimization-based data-driven distributed local fault detection scheme to achieve the efficient monitoring of multi-unit chemical processes. At the same time, Kohonen *et al.* demonstrated the efficiency and difference between the multi-block PLS approach and conventional PLS [23], priority regression in the plant-wide process monitoring results. Ge *et al.* proposed [24] a plant-wide process monitoring method based on distributed PCA, in which variables with high contribution rate in the same principal component(PC) direction were automatically divided into the same block. Xie *et al.* proposed [25] a process monitoring and fault isolation scheme for shrinkage PCA. Jiang *et al.* proposed sub-block division based on Hellinger distance between variables [26]. Zhang *et al.* proposed a novel plant-wide process monitoring framework [27] based on distributed Gap-support SVDD with adaptive radius, and divided the plant-wide process variables into different subblocks by using mixed similarity measure to deal with complex coupled process variables. Huang *et al.* proposed [28] a multi-block partitioning method based on mutual information (MI). Tian *et al.* proposed the use of copula correlation analysis and Bayesian inference based multi-block principal component analysis for decentralized monitoring of large-scale processes [29], [30]. Zeng *et al.* proposed [31] mutual information-based sparse multi-block dissimilarity method for incipient fault detection and diagnosis in plant-wide process. By taking advantage of the complex relationships between variables and the connections between subblocks, subblocks are generated that are easily interpreted by the process mechanism.

Nevertheless, all the aforementioned studies obtained corresponding detection results. However, considering the possibility of cascading failures in process variables, when an operating unit fails in a plant-wide process, it might affect merely this specific unit and related neighbor units. Thus, a small portion of process variables are affected by this fails. It is necessary to fully understand the structural characteristics of process variables throughout the plant-wide process during the block partitioning process. Therefore, the community discovery algorithm of complex network becomes the evaluation basis of the quality of block division. After Girvan and Newman proposed the community evaluation standard of modularity is a typical edge betweenness-based separation algorithm [32]–[34], it really opened the door to the research of complex network community discovery algorithm. Subsequently, Tatsuro *et al.* proposed the greedy algorithm to solve the degree of modularity function and obtained the approximate optimal result [35]. Gao *et al.* proposed [36] the swarm intelligence algorithm to detect community structure. Sieyum proposed [37] a clustering algorithm combining k-means algorithm and genetic algorithm, Benita *et al.* proposed [38] a clustering scheme based on Louvain. The clustering algorithm of Louvain is used to cluster the variable feature graphs, which can not only quickly process the network with hundreds of millions of nodes, but also improve the quality of the community partition. Therefore, a plant-wide process monitoring algorithm based on multi-block PCA analysis based on complex network community discovery-Louvain (CNCDL-MBPCA) is proposed in this paper.

Although the existing method has successfully implemented the fault detection of industrial processes, the following major problems still need to be solved for block division in the whole process monitoring:

(1)Although based on prior process knowledge block division methods proposed in [18]–[20] improved the plant-wide process monitoring results, they relied too much on prior process knowledge in terms of block-wised partitioning, given that the process knowledge is not always available in the process of partitioning.

(2)These data-driven [26], [28] methods can automatically partition blocks, however, only the relationship between variables is considered in the partition of blocks, and the interaction between sub-blocks is not considered.

(3)The method by [29], [30] not only the similarity between variables, but also the association pattern between sub-blocks are considered. But it takes a lot of computing time to process large networks. For the traditional Louvain algorithm [39], the default weight value of the relation between two variables is 1, and the influence of variable behavior is ignored, the variables with the same correlation degree may have different correlation pattern, resulting in inaccurate results of community partitioning and effecting the monitoring performance.

To solve these problems, the following work has been done this study:

(1)Via complex network community discovery algorithm [38], the graph feature-based clustering can not only consider the relation between variables, process variables with similar characteristics are automatically clustered through the characteristics carried by the graph itself, but consider the correlation degree between blocks. It can also effectively avoid the limitation and misdirection caused by over-reliance on process knowledge.

(2)Based on the cosine similarity of neighbor node weighted Louvain, is used for block division. Weighted network structure and attribute information fusion, not only consider the existence of the edges between nodes, and considering the close correlation between the variables, avoid missing information of process variables in the process of block partitioning to obtain better block partitioning results so as to obtain better monitoring results.

The major procedure include the following:

First, the cosine similarity between variables is calculated to build a complex network model.

Second, the variable similarity matrix is used to determine the weights of variables and to divide the sub-blocks, so as to avoid the loss of information.

Third, PCA model [40] was established for the process variables of each sub-block for monitoring and analysis. Since Gaussian or non-Gaussian distribution phenomenon existed in the process data set, the confidence limit of PCA in each sub-block was determined by kernel density estimation (KDE).

Lastly, by using Bayesian inference [41], the detection results of each sub-block are fused to obtain the final statistics to realize the comprehensive evaluation of monitoring results.

The rest of the paper is organized as follows. The second part briefly reviews the related work. The third part briefly introduces complex network for block division, principal component analysis process monitoring and Bayesian inference for statistics combination. The fourth part introduces case study. Through the analysis of TE benchmark process, the effectiveness of this method is verified. PCA, sparse PCA (SPCA), MBPCA, multi-block sparse PCA (MBSPCA) are selected to compare with the proposed method in this paper to verify the effectiveness of the method. Finally, in the fifth part, the conclusion is drawn and discussed.

II. RELATED WORK

In order to reduce the failure risk of industrial process and improve product quality requirements, process monitoring is essential in industrial operation. The plant-wide process monitoring is widely used to analyze the complex relationship of process variables. However, the global variables collected through plant-wide process are modeled, these local behaviors of variables is often ignored, leading to information loss. By turning the variables into several sub-blocks, the local behaviors [26], [42], [43] of the variable can be well demonstrated and the hidden characteristics of the local behavior can be explored.

For the partition of blocks, MacGregor *et al.* proposed [44] a multi-block projection method to detect each sub-block and the entire process. This method is realized by analyzing the prior process knowledge, but the process knowledge is not universal. Tong *et al.* proposed [45] distributed statistical process monitoring based on four-subspace construction and Bayesian inference. The subblocks are divided according to the correlation or irrelevance between process variables and principal component space (PCS) and the residual subspace. Block partitioning is limited to linear correlation. Jiang *et al.* proposed [43] a fully data-driven distributed nonlinear plant-wide process monitoring method. Sub-block partition is realized by mutual information - spectral clustering. It can consider both the linear relationship between variables and the nonlinearity of variables. However, in the process of block partition, only the relation between sub-block variables is considered, and the association mode between sub-blocks is not considered. Since the plant-wide industrial process is characterized by multiple operating units, there may be some connections between units or within units. Tong *et al.* proposed an improved MBPCA algorithm [42] for extracting block fractions with respect to the specificity of each block and the correlation between different blocks. Chen *et al.* proposed [46] a fault detection method for plant-wide process monitoring based on distributed canonical correlation analysis, and realized block partitioning by using correlation information of the neighboring nodes. The uncertainty of local process is reduced by making use of the correlation relations with other subsystems. Therefore, in this paper, the cosine similarity between variables is calculated to analyze the similarity relation of neighbor nodes to realize the partition of blocks.

III. METHODOLOGY

The proposed monitoring scheme for a multi-block plant-wide process consists of three parts: complex network for block division, principal component analysis process monitoring, Bayesian inference for statistics combination. These parts are introduced in this section.

A. COMPLEX NETWORK FOR BLOCK DIVISION

Considering the multi-operation units and the complexity between variables of the plant-wide process, the problem of ignoring local behaviors in the global model and excessively relying on process knowledge for sub-block partitioning was solved, and only considering the correlation analysis of variables in the existing block partitioning process. and set out to solve the block partition problem through graph theory. Through the network topology diagram (i.e., the plant-wide process. By connecting the variables of the plant-wide process in a specific way), the representation can not only intuitively reflect the rich resources and hidden internal relations among variables, but also present a clear network topology structure. This diagram is known as the complex network model.

In many real networks, the nodes are grouped in the form of subgraphs, and the network is made up of several ''communities'' with relatively tight connections between nodes within each community but sparse connections between communities. For the high dimensional complex data information collected in the plant-wide process, the sub-blocks are divided by the community discovery algorithm, not only deal with the linear relationship between variables, but also consider the nonlinear relationship between variables. As the structure and attribute information of the original unweighted network are fused, only the existence of edges between nodes is considered. Based on this problem, the pre-existing Louvain algorithm was optimized in this paper, and a novel community discovery algorithm based on the similarity weighting of neighbor nodes was proposed to divide the community by using a weighted network, not only consider whether or not a connection between nodes, but also consider the associated closely degree between nodes and correlation analysis between sub-blocks are also considered. By considering the weights allocated according to the similarity between nodes, the nodes are divided the node into the community with which it is most closely connected so that it is more tightly connected to the same community. According to the cosine similarity of the feature vectors of the nodes, the transition probability of the nodes joining the neighbor nodes is obtained. The cosine similarity of the two variables x_1, x_2 is calculated as follows [47]:

$$\cos\langle x_1, x_2\rangle = \frac{\sum_{i=1}^{n} x_{1i}x_{2i}}{\sqrt{\sum_{i=1}^{n} x_{1i}^2}\sqrt{\sum_{i=1}^{n} x_{2i}^2}} \tag{1}$$

where, x_{1i}, x_{2i} represents the ith sample value of variables x_1 and x_2, respectively.

For a given number of samples, cosine similarity between variables can be calculated to quantify the relationship between variables. This quantization should not just consider the linear correlation, as well as consider the nonlinear relationship between variables. Strongly correlated variables have greater cosine similarity, whereas weakly correlated variables have less cosine similarity. According to the set threshold value to determine the tightness of the variables, establish the adjacency matrix. Since cosine similarity has positive and negative values, when $|\cos\langle x_1, x_2\rangle|$ is defined to be greater than or equal to the threshold, then adjacency matrix elements is $r_{x_1x_2} = r_{x_2x_1} = |\cos\langle x_1, x_2\rangle|$, otherwise $r_{x_1x_2} = r_{x_2x_1} = 0$. Here, $r_{x_ix_i} = 0$ ignores the autocorrelation of variables. On this basis, the adjacency matrix R is generated of Equation (1) with [48], [49]:

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ r_{n1} & r_{n1} & \cdots & r_{n1} \end{bmatrix} \tag{2}$$

where, c represents the number of samples and m represents the number of variables.

If the adjacency matrix element is greater than or equal to the threshold value, then there are edges between variables; otherwise, there are no edges between variables, which are called isolated nodes. According to the adjacency matrix, the weights of the edges between the nodes are initialized, and the undirected weight network is constructed. The corresponding module degree is calculated to select the neighbor node most closely associated with the current node for community partition via the transfer probability obtained. Modularity degree is defined as follows [34], [37]:

$$Q = \frac{1}{2m}\sum_{i,j}\left[r_{ij} - \frac{K_iK_j}{2m}\right]\delta\left(c_i, c_j\right)$$
$$\delta(u, v) = \begin{cases} 1, & when\ (u == v) \\ 0, & when\ (u! = v) \end{cases} \tag{3}$$

where, r_{ij} represents the weight of edge of the connecting node i $(i = 1, 2, \cdots, n)$ an j $(j = 1, 2, \cdots, m)$, K_i represents the sum of the weight of edge connected with node i, and c_i represents the community to which node i belongs. $m = \frac{1}{2}\sum_{i,j} r_{ij}$ represents the total number of edges of the network structure.

Louvain algorithm determines whether to join another community according to the modularity gain of the node. When node i is distributed to neighborhood node j belongs to the community c, the modularity gain is [36]:

$$\Delta Q = \left[\frac{\sum in + K_{i,in}}{2m} - \left(\frac{\sum tot + K_i}{2m}\right)^2\right] - \left[\frac{\sum in}{2m} - \left(\frac{\sum tot}{2m}\right)^2 - \left(\frac{K_i}{2m}\right)^2\right] \tag{4}$$

where, $\sum in$ represents the sum of the weights of all the edges in community c; $K_{i,in}$ represents the sum of edge weights between nodes i and nodes in the community c; $\sum tot$ represents the sum of the weights of edges between external nodes of community c and community c. ΔQ is divided into two parts, the first part represents the degree of modularity after adding node i to community c, and the second part represents the degree of modularity of node i as independent community and community c.

For block division, as shown in Fig. 1 and Fig. 2, a complex network model is established for the collected normal training

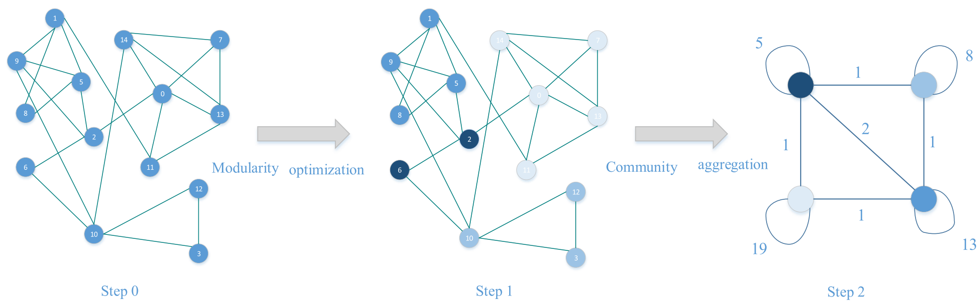

FIGURE 1. Block division step 1.

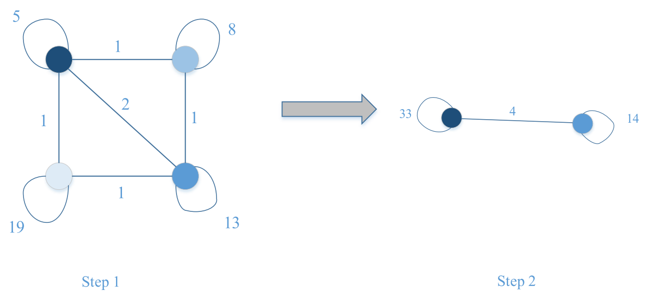

FIGURE 2. Block division step 2.

data of the plant-wide process. Step 0, each node is initialized as an independent community. The number of communities is equal to the number of nodes, also called the number of process variables; step 1, for each node i, try in turn to assign node i to the community where each of its neighbor nodes is located, and calculate the modularity changes ΔQ before and after allocation. And when ΔQ is the largest, record neighbor nodes, if max $\Delta Q > 0$, the node i is assigned to the community, where the neighbor node has the greatest modularity change, otherwise remain unchanged, this node i is defined as an independent node; until the community of all nodes no longer changes; step 2, all nodes in the same community are compressed into a new node, The weight of the edge between nodes in the community is transformed into the weight of the ring of the new node, and the weight of the edge between communities is transformed into the weight of the edge between new nodes. Repeat step 1 and step 2, when the modularity of the entire graph does not change, the iteration terminates.

B. PRINCIPAL COMPONENT ANALYSIS PROCESS MONITORING

Assuming that given a sample data set $X \in R^{n\times m}$, where n represents the sample number of each variable and m represents the observation variable. Through the division of complex network communities, the data set becomes

$$X = [X_1, X_2, \cdots X_B], \tag{5}$$

where, B represents the number of sub-blocks, and for the bth sub-block $X_b \in R^{n\times m_b}$, where m_b is the number of variables of the bth sub-block ($b = 1, 2, \cdots, B$). PCA is a fundamental MSPM method, which projects the sample data set from the high-dimensional feature space to the low-dimensional feature space to eliminate redundant information and represent the original data set information in the low-dimensional space as far as possible. PCA model is established in each subspace as follows [50]:

$$X_b = T_b P_b^T + E_b \tag{6}$$

where, $p_b \in R^{m_b\times k_b}$ represents that the bth sub-block is projected onto the load matrix of PCS, which is obtained by singular value decomposition (SVD). k_b is the number of principal elements of the bth sub-block is determined by cumulative percentage variance (CPV). T_b is the score matrix, and the score vectors are orthogonal to each other. E_b is the residual vector to extract random noise and model error information.

After the PCA model is established, based on the purpose of process monitoring, the corresponding monitoring statistics T_b^2 and SPE_b are constructed to monitor the changes of the dominant and residual subspaces respectively. Given a new observation variable $x_b \in R^{m_b\times 1}$, the two statistics are calculated as follows, respectively [42], [50]:

$$T_b^2 = x_b^T P_b (\Lambda_b)^{-1} P_b^T x_b \leq T_{b,\lim}^2 \tag{7}$$

$$SPE_b = e_b^T e_b \leq SPE_{b,\lim}$$

$$e_b = \left(I - P_b^T P_b\right) x_b \tag{8}$$

where, $\Lambda_b \in R^{k_b\times k_b}$ represents a diagonal matrix composed of eigenvalues corresponding to the previous k_b principal element; e_b is residual vector; $T_{b,\lim}^2$ and $SPE_{b,\lim}$ is confidence limits for statistics. In the traditional method, the mechanism of solving the above control limit is derived from the complex mathematical statistics, but every event has its own specific density distribution form. In the present study, KDE is often used to determine the confidence limit. This paper selects the Gaussian kernel function which is widely used. The expression of the Gaussian kernel density estimation function is as follows [29], [51], [52]:

$$\hat{f}(x) = \frac{1}{nh\sqrt{2\pi}} \sum_{i=1}^{n} \exp\left(-\frac{(x-x_i)^2}{2h^2}\right) \tag{9}$$

where, x_i is the observation variable and is n the sample number of the observation variable; h is the bandwidth, which greatly affects the accuracy of KDE. When h is small, the number of samples in the region is sparse and individual factors are amplified, which cannot better reflect the overall characteristics of the samples. When h is too large, the regional scale becomes larger, and the overall characteristics of the sample dominate, partial information of the original data is ignored.

C. BAYESIAN INFERENCE FOR STATISTICS COMBINATION

Theoretically, the corresponding control limit and statistics can be obtained by establishing PCA monitoring model for each sub-block. And each sub-block gets a different number of principal component through CPV. Therefore, for a new monitoring data set x_{new}, it is difficult to merge the different detection results that are composed of subspaces directly into the final monitoring strategy. Therefore, to solve the problem that different sub-blocks may get different number of principal component, this paper adopts the Bayesian inference strategy and combines the detection results of all subspaces with the probability of failure. The fault probability of the T^2 statistics corresponding to the subspace X_b is calculated as follows [43], [53]:

$$P_{T^2}(F|X_b) = \frac{P_{T^2}(X_b|F)\,P_{T^2}(F)}{P_{T^2}(X_b)} \tag{10}$$

$$P_{T^2}(X_b) = P_{T^2}(X_b|N)\,P_{T^2}(N) + P_{T^2}(X_b|F)\,P_{T^2}(F) \tag{11}$$

Here, the conditional probability $P_{T^2}(X_b|N)$ and $P_{T^2}(X_b|F)$ are defined as follows:

$$\begin{aligned} P_{T^2}(X_b|N) &= \exp\left(-\frac{T^2_{b,new}}{T^2_{b,\lim}}\right) \\ P_{T^2}(X_b|F) &= \exp\left(-\frac{T^2_{b,\lim}}{T^2_{b,new}}\right) \end{aligned} \tag{12}$$

where, N and F represent normal conditions and failure conditions respectively; $P_{T^2}(N)$ and $P_{T^2}(F)$ represents the prior probability under normal and failure conditions, When the confidence level α is determined to be, value of $P_{T^2}(N)$ and $P_{T^2}(F)$ are determined as α and $1-\alpha$; $T^2_{b,new}$ is the statistic of the bth sub-block of the new detection data set; finally, the total joint statistics can be calculated by combining the detection quantities of all different sub-block by Bayesian inference as follows [42], [45]:

$$BIC_{T^2} = \sum_{b=1}^{B}\left\{\frac{P_{T^2}(X_b|F)\,P_{T^2}(F|X_b)}{\sum_{b=1}^{B} P_{T^2}(X_b|F)}\right\} \tag{13}$$

Similarly, the failure probability of SPE statistics corresponding to the subspace is calculated as follows:

$$P_{SPE}(F|X_b) = \frac{P_{SPE}(X_b|F)\,P_{SPE}(F)}{P_{SPE}(X_b)} \tag{14}$$

$$P_{SPE}(X_b) = P_{SPE}(X_b|N)\,P_{SPE}(N) + P_{SPE}(X_b|F)\,P_{SPE}(F) \tag{15}$$

Here, the conditional probability $P_{SPE}(X_b|N)$ and $P_{SPE}(X_b|F)$ are defined as follows:

$$\begin{aligned} P_{SPE}(X_b|N) &= \exp\left(-\frac{SPE_{new}}{SPE_{\lim}}\right) \\ P_{SPE}(X_b|F) &= \exp\left(-\frac{SPE_{\lim}}{SPE_{new}}\right) \end{aligned} \tag{16}$$

The total joint statistics ***SPE*** are calculated as follows:

$$BIC_{SPE} = \sum_{b=1}^{B}\left\{\frac{P_{SPE}(X_b|F)\,P_{SPE}(F|X_b)}{\sum_{b=1}^{B} P_{SPE}(X_b|F)}\right\} \tag{17}$$

When the value of BIC_{T^2} and BIC_{SPE} exceeds the confidence limit of the statistic, an industrial failure is considered to have occurred; otherwise, the industrial process can run normally. Based on the performance of existing methods, the fault detection rate is used to measure. The detection rate is calculated as follows:

$$\begin{aligned} DR_{T^2} &= \frac{N_{F(T^2)}}{N^\circ} \\ DR_{SPE} &= \frac{N_{F(SPE)}}{N^\circ} \end{aligned} \tag{18}$$

where, DR_{T^2}, DR_{SPE} represents the fault detection rate of T^2 and SPC statistics respectively. N° represents the number of fault variables in the sample data set. $N_{F(T^2)}$ represents the number of BIC_{T^2} data set that exceeded the confidence limit of the $BIC_{T^2,\lim}$ statistic, $N_{F(SPE)}$ represents the number of BIC_{SPE} data set that exceeded the confidence limit of the $BIC_{SPE,\lim}$ statistic.

D. METHOD IMPLEMENTATION

In this paper, monitoring is divided into two parts: offline modeling and online monitoring. The purpose of offline modeling is to obtain the confidence limits of each sub-block statistic and the total joint indicator statistic, online monitoring is to detect the running state of the online system. The monitoring flow chart of this method is shown in Fig. 3, the implementation steps of this method are as follows:

(1) Off-line modeling

1. Carry out data standardization and normalization preprocessing for the data set under normal conditions of the plant-wide process.
2. Block division the data set after standard processing by using complex network community discovery algorithm-Louvain.
3. PCA monitoring model was established for each sub-block.
4. Calculate the confidence limit of the monitoring statistics of each sub-block PCA model by Equations (7) and (8), and then fuse them to calculate the confidence

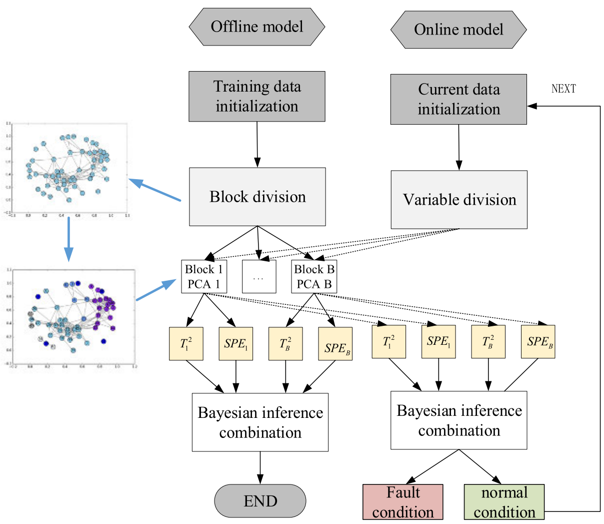

FIGURE 3. TE process flow chart.

limit of the statistics of the total joint indicators through Bayesian inference by Equations (13) and (17).

(2) On-line monitoring

1. After standard processing of the current sample set, block division is conducted according to the sub-block division results of offline modeling.
2. Calculate the statistics of each sub-block and fuse the statistics through Bayesian inference.
3. Compare the confidence limits of $BIC_{T^2,\lim}$ and $BIC_{SPE,\lim}$ calculated by off-line modeling with the values of BIC_{T^2} and BIC_{SPE} to judge whether the industrial process fails.

E. SOME REMARKS

Compared with the global process monitoring method, in this method, multiple subspace models are established, and the complex network community discovery algorithm is adopted to divide the sub-blocks. Divide closely related variables into the same sub-block, simultaneously, the variable fault information of an operation unit can be divided into the same sub-block to improve the failure probability of process monitoring. In this paper, the similarity threshold is set to 0.1 when building the complex network model. A small number of independent nodes will exist under the threshold result. In this paper, independent nodes are divided into the same sub-block. Through principal component analysis, the redundant information is deleted by selecting feature variables with CPV $\geq$ 85% rule. The subblocks monitor statistics T_b^2 and SPE_b calculated, and the two statistics were respectively used to monitor the changes of the dominant subspace and the remaining subspace. Since the variables cannot be guaranteed to be Gaussian, KDE is used to determine the confidence limit of the subspace. Since it is difficult to fuse the detection results of different subspaces, Bayesian inference is used to transform the statistical information into the problem of fault probability, where the confidence level α is 0.99. In the fourth section, the experimental results show that this method can effectively improve the monitoring results compared with the global process monitoring, and there is no substantial increase in the computational complexity. It is superior to PCA, SPCA, other MBPCA and MBSPCA.

IV. CASE STUDIES ON THE BENCHMARK PROCESS

A. TE PROCESS MODEL

In this part, based on the process monitoring scheme proposed in this paper, this paper selects the Tennessee Eastman (TE) benchmark process to analyze the whole process industrial process data. The Tennessee Eastman benchmark process is a simulation example of a chemical process first proposed by Downs and Vogel and is widely used in the field of industrial process monitoring. The simulation experiment data from http://depts.washington.edu/control/LARRY/TE/download.html.

TE process is mainly composed of five conversion units: reactor, condenser, compressor, separator and stripper. This process conforms to the characteristics of the whole process studied in this paper, so this TE process is selected for simulation study. The process flow chart is shown in Fig. 4. It simulates complex industrial processes, and its dimensions are quite rich, in which the data set consists of measurement variable and operation variable. Among them, 11 operation variables and 41 measurement variables of which the measurement variables are divided into 22 continuous process variables, and 19 component measurement variables. A detailed description of the TE process is available in the literature [18].

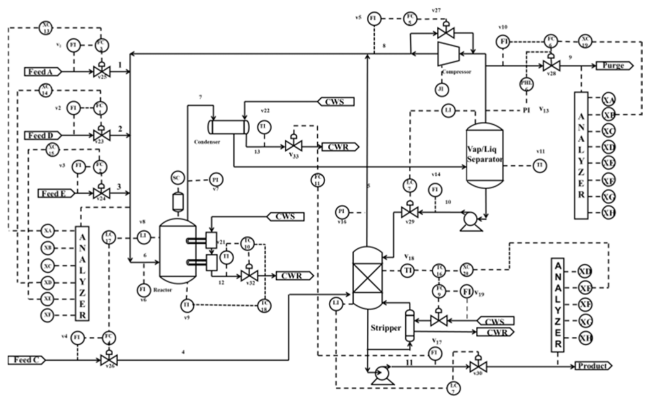

FIGURE 4. TE process flowchart.

TE process data every run for 48 hours, running a total of 22 times, the training sample and monitoring samples of the sampling interval is set to 3 min. For the first time, the normal data set was obtained under the condition of no fault. When the fault condition was simulated, the process started to run under the condition of no fault. After 8 hours of smooth operation, the fault was introduced. (i.e. the first 160 sample points for trouble-free samples). To verify the monitoring method, 21 industrial faults simulated by TE process were selected. Gaussian noise was added to 21 fault process variables. Once the industrial process fails, most variables may be interfered. As shown in Table 1. Among, faults 1-7 are step type faults, faults 8-12 are random fault types, faults 13 are caused by slow drift of reaction dynamics, faults 14, 15 and 21 are caused by viscous action of valve, and the rest are unknown faults.

TABLE 1. 21 fault STAES.

Fault label	Fault description	Fault types
0	normal	-
1	A/C feed ratio, B composition constant (stream 4)	Step
2	B composition, A/C ratio constant (stream 4)	Step
3	D feed temperature (stream 2)	Step
4	Reactor cooling water inlet temperature	Step
5	Condenser cooling water inlet temperature	Step
6	A feed loss (stream 1)	Step
7	C header pressure loss – reduced availability (stream 4)	Step
8	A,B,C feed composition (stream 4)	Random variation
9	D feed temperature (stream 2)	Random variation
10	C feed temperature (stream 4)	Random variation
11	Reactor cooling water inlet temperature	Random variation
12	Condenser cooling water inlet temperature	Random variation
13	Reaction kinetics	Slow drift
14	Reactor cooling water valve	Sticking
15	Condenser cooling water valve	Sticking
16-20	Unknown	Unknown
21	The valve for stream 4 was fixed at the steady state position	constant position

B. TE PROCESS SIMULATION BASED ON MULTI-BLOCK CNCDL-MBPCA

In order to establish the monitoring model, this paper selects the process data set as 960 × 52, where, 52 represents the number of process variables and 960 represents the number of sampling points for each variable. Firstly, as shown in Fig. 3, the whole process variables $X = [x_1, x_2, \cdots, x_{52}]$ are divided into six sub-blocks by using the community discovery algorithm of the complex network, and the results are shown in Table 2. According to the result of block division, PCA monitoring model was established. The number of principal components was determined by the CPV greater than 85%.

TABLE 2. Subblock division resultes.

Block No.	Variables
1	0, 5, 9, 10, 17, 18, 20, 23, 25, 26, 27, 29, 31, 32, 33, 34, 35, 37, 39, 43, 46, 49
2	2, 3, 4, 6, 7, 12, 15, 19, 22, 24, 28, 30, 40, 42, 44, 45
3	8, 1, 50, 21, 41
4	48, 14
5	11, 47
6	13, 16, 36, 38, 51

Based on the confidence limit of detection statistics obtained by KDE, first, T^2 and SPE values are calculated using normal operation data. Secondly, KDE is used to estimate the density function of normal T^2 and SPE values. Finally, the mean integral square error is used to select the optimal h value. In this paper, the threshold value of bandwidth h in Equation (9) is set as 1.

First, the data sets of 21 fault states are analyzed and the detection rate of each fault state is tested. The existing method and the method presented in this paper were selected for the test, and the test results were shown in Table 3. It can be seen from Table 3 that faults 3, 9 and 15 are difficult to detect in all methods. All methods for faults 1, 2, 6, 8, 12, 13 and 14 detect high failure rates. In step fault (except for failure 3), failure detection rates above 0.85 can be obtained in all multi-block process monitoring. In particular, CMBPCA and CNCBL-MBPCA obtained 0.937 and 0.904 fault detection rates respectively. In faults 4, 5, 10 and 17, compared with the traditional global model, decentralized process monitoring can detect the local behavior of process variables well and improve the monitoring performance of the whole system. In addition, PCA and SPCA detection results show that PCA fault detection rate is superior to SPCA algorithm 0.073. In multi-block process monitoring, MBSPCA detection results are also better than MBPCA algorithm. Therefore, this paper chooses the factory-wide process monitoring based on multi-block principal component analysis. Compared with other methods, the error detection rate of CNCDL-MBPCA algorithm for most monitoring results is relatively low. The following sections analyze the performance results for both faults 5 and 10.

Fault 5 in TE process reflects a step change of inlet temperature of condenser cooling water. When the temperature presents the trend of step change, it will lead to the step change of condenser cooling water flow. When the fault occurs, the flow velocity from the outlet of the condenser to the vapor/liquid separator will increase, resulting in an upward trend in the temperature of the vapor/liquid separator and increasing in the outlet temperature of the cooling water of the separator. The control loop system can easily compensate the fault state and pull the temperature of the separator back to the default temperature. Fig. 5(a) - (h) are

TABLE 3. The detection rate of different algorithm.

Fault no.	PCA T^2	PCA SPE	SPCA T^2	SPCA SPE	$PCA^{a)}$ BIC_{T^2}	$PCA^{a)}$ BIC_{SPE}	$PCA^{b)}$ BIC_{T^2}	$PCA^{b)}$ BIC_{SPE}	$PCA^{c)}$ BIC_{T^2}	$PCA^{c)}$ BIC_{SPE}	$PCA^{d)}$ BIC_{T^2}	$PCA^{d)}$ BIC_{SPE}	$PCA^{e)}$ BIC_{T^2}	$PCA^{e)}$ BIC_{SPE}
1	0.992	0.999	0.994	0.780	0.994	0.999	0.993	0.998	1	0.992	0.996	0.998	0.997	0.997
2	0.986	0.985	0.985	0.950	0.985	0.985	0.975	0.983	0.998	0.972	0.987	0.98	0.985	0.987
3	0.006	0.049	0.017	0.024	0.020	0.044	0.003	0.056	0.985	0	0.008	0.027	0.043	0.020
4	0.262	1	0.087	0.366	0.854	0.999	0.370	1	0.987	0	0.642	1	1	0.997
5	0.229	0.215	0.231	0.254	0.238	0.204	0.223	1	0.972	0.178	0.257	1	0.289	1
6	0.987	1	0.995	1	0.993	1	0.993	1	1	0.995	0.993	1	0.995	1
7	1	1	1	0.412	1	1	1	1	0.985	0.292	1	1	1	0.999
8	0.972	0.972	0.941	0.952	0.974	0.960	0.971	0.975	1	0.911	0.976	0.980	0.881	0.977
9	0.015	0.035	0.006	0.017	0.009	0.040	0.007	0.071	0.276	0	0.022	0.032	0.039	0.039
10	0.242	0.417	0.355	0.437	0.244	0.428	0.281	0.437	0.998	0	0.393	0.525	0.516	0.658
11	0.436	0.675	0.326	0.276	0.621	0.505	0.5225	0.803	0.970	0.134	0.576	0.713	0.742	0.584
12	0.982	0.950	0.951	0.971	0.983	0.944	0.987	0.997	1	0.882	0.980	0,977	0.989	0.988
13	0.945	0.950	0.937	0.944	0.943	0.948	0.945	0.953	1	0.928	0.937	0.953	0.950	0.952
14	0.985	1	0.992	0.664	1	0.821	1	1	0.998	0.006	1	1	1	0.999
15	0.005	0.065	0.031	0.054	0.023	0.053	0.063	0.068	0.023	0.001	0.021	0.047	0.069	0.025
16	0.085	0.401	0.169	0.267	0.095	0.395	0.1	0.377	0.212	0.048	0.207	0.485	0.312	0.764
17	0.762	0.945	0.721	0.855	0.825	0.931	0.865	0.972	0.998	0.556	0.888	0.973	0.946	0.936
18	0.890	0.907	0.879	0.895	0.898	0.894	0.892	0.908	0.889	0.880	0.898	0.906	0.902	0.908
19	0.064	0.162	0.007	0.004	0.084	0.090	0.108	0.108	0.279	0.003	0.081	0.182	0.019	0.472
20	0.241	0.547	0.281	0.464	0.298	0.510	0.307	0.673	0.559	0.250	0.462	0.725	0.558	0.620
21	0.380	0.507	0.384	0.329	0.356	0.428	0.391	0.462	0.783	0.253	0.431	0.521	0.461	0.485
Ave	0.601		0.528		0.609		0.639		0.600		0.655		0.693	

a) PCA: MBPCA; b) PCA: MBPCA based on Louvain (LMBPCA); c) PCA: MBSPCA; d) PCA: copula-correlation analysis and Bayesian inference-based MBPCA (CMBPCA); e) PCA: CNCDL-MBPCA

the monitoring results of PCA, SPCA, MBPCA, LMBPCA, MBSPCA, CMBPCA, and CNCDL-MBPCA for fault 5.

Where, as shown in Fig. 5(a), PCA is used to monitor the fault. It can be seen clearly, the fault signal can be obviously detected when the fault is introduced after sample 160, but after sample 350 or so, within a certain error range, it can be considered that the reaction control loop will almost pull the statistics back to the normal level. As shown in Table 3, the miss rate of statistics is 0.771, and that of SPE statistics is 0.685. As shown in Fig. 5(f), in order to visually explain the fault, the outlet temperature of the separator cooling water (reference value is 77.297) fluctuates in the reference value through the control loop compensation, but the condenser cooling water flow (reference value is 18.114) is still higher than the reference value in the 350 sample data collected. Failure results are still not detected in Fig. 5(b). When global PCA and SPCA are used for process monitoring, the local characteristics of variables are not taken into account, so the

TABLE 4. The false alarm rate of different algorithm.

Fault no.	PCA T^2	PCA SPE	SPCA T^2	SPCA SPE	$PCA^{a)}$ BIC_{T^2}	$PCA^{a)}$ BIC_{SPE}	$PCA^{b)}$ BIC_{T^2}	$PCA^{b)}$ BIC_{SPE}	$PCA^{c)}$ BIC_{T^2}	$PCA^{c)}$ BIC_{SPE}	$PCA^{d)}$ BIC_{T^2}	$PCA^{d)}$ BIC_{SPE}	$PCA^{e)}$ BIC_{T^2}	$PCA^{e)}$ BIC_{SPE}
1	0.006	0.025	0	0	0.018	0.043	0	0.075	0.614	0	0	0.025	0	0.012
2	0.006	0.043	0	0	0.006	0.037	0.006	0.068	0.735	0	0.018	0	0.006	0.006
3	0.006	0.037	0	0	0	0	0	0.037	0.721	0	0	0.025	0	0.006
4	0.006	0.043	0.012	0	0.012	0	0.006	0.081	0.850	0.006	0.012	0.012	0.018	0.006
5	0.006	0.043	0.012	0	0.012	0.006	0.006	0.081	0.850	0	0.012	0.012	0.018	006
6	0	0.037	0.006	0	0.012	0.012	0.006	0.025	0.725	0	0.018	0	0.006	0.012
7	0	0	0	0	0	0.025	0	0.050	0.736	0	0.006	0	006	0.006
8	0	0.012	0	0	0.006	0.031	0	0.025	0.901	0	0	0.006	0	0.012
9	0.006	0.043	0.018	0	0.006	0.037	0	0.068	0.276	0	0.012	0.025	0.075	0.006
10	0	0.012	0	0	0.006	0.018	0	0.031	0.800	0	0	0.018	0	0.025
11	0	0.031	0	0	0.018	0.050	0	0.031	0.887	0	0.006	0.012	0	0.018
12	0	0.056	0	0	0.025	0.037	0	0.062	0.825	0	0.006	0,012	0.006	0
13	0	0.012	0	0	0.006	0.037	0	0.018	0.762	0.006	0.012	0.006	0	0.006
14	0	0.037	0.006	0	0	0.037	0.012	0.031	0.843	0	0	0.018	0	0.018
15	0	0.006	0	0	0.006	0.025	0.006	0.043	0.701	0	0	0.006	0	0.012
16	0.018	0.018	0.018	0.025	0.006	0.025	0	0.100	0.811	0	0.025	0.018	0.068	0
17	0	0.087	0	0	0.012	0.025	0.006	0.081	0.540	0	0	0.012	0.006	0.018
18	0	0.037	0	0	0	0.050	0	0.056	0.880	0	0	0.018	0.006	0.006
19	0	0.043	0	0	0	0.037	0	0.025	0.779	0	0	0.006	0	0.031
20	0	0.031	0	0	0	0.025	0	0.050	0.559	0	0	0.012	0	0.012
21	0	0.056	0.006	0	0.006	0.056	0.006	0.062	0.937	0	0.006	0.031	0.018	0.043

failure condition that the inlet temperature of condenser cooling water is still higher than the normal value after the control circuit is compensated cannot be detected effectively. Considering the adoption of fault detection method based on multi-block division, see Fig. 5(c)-5(g). It can be intuitively seen from the detection results of statistics BIC_{SPE} that faults can still be detected after the 350th point. The failure miss rate is closely 0. MBSPCA monitoring algorithm is shown in Fig. 5(e). Although it can detect faults, it has a high false alarm rates, as shown in Table 4.

A fault is considered to have occurred as long as either of the statistics BIC_{T^2} and BIC_{SPE} exceeds the threshold values of the statistics $BIC_{T^2,\lim}$ and $BIC_{SPE,\lim}$.

Fault 10 is a typical fault of monitoring variable C feed for random temperature changes in stream 4. The inlet temperature of the reactor produces a random change, which causes the flow rate of the reactor to oscillate, resulting in a certain range of fluctuations in the reactor temperature. Fig. 6(a) - (g) are the monitoring results for fault 10. As can be seen from Table 3, the miss rate of CNCDL-MBPCA statistics BIC_{T^2} is 0.274 lower than that of PCA, and correspondingly, the detection rate of statistics BIC_{SPE} is 0.241 higher. Compared with the detection results of global PCA and SPCA, the detection performance is also greatly improved. However, MBPCA was superior to MBPCA in T^2 statistics, although it had a high false detection rate. The fault detection rate

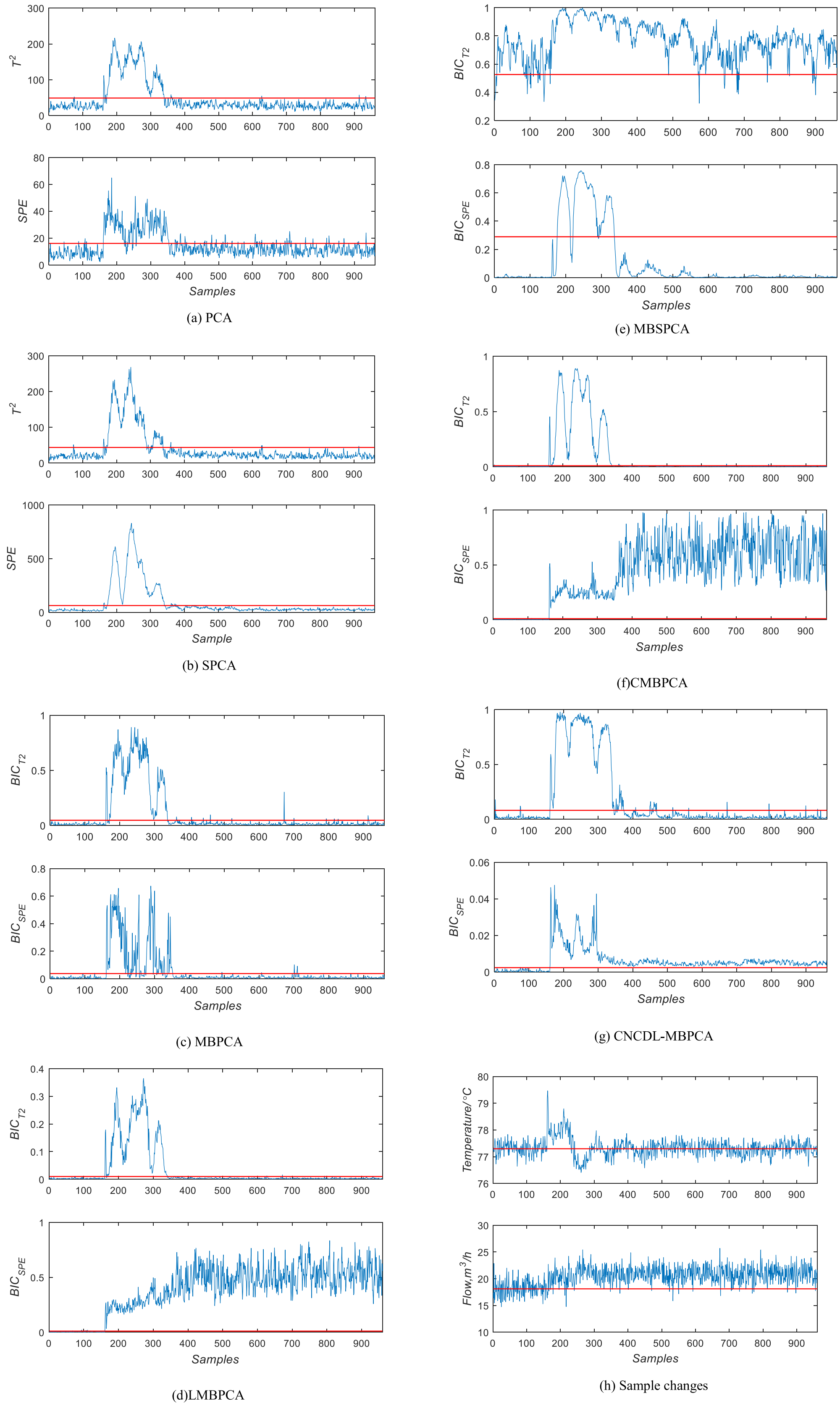

FIGURE 5. Fault 5 in TE process: (a) PCA, (b) SPCA, (c) MBPCA, (d) LMBPCA, (e) MBSPCA, (f) CMBPCA, (g) CNCDL-MBPCA (h) sample changes.

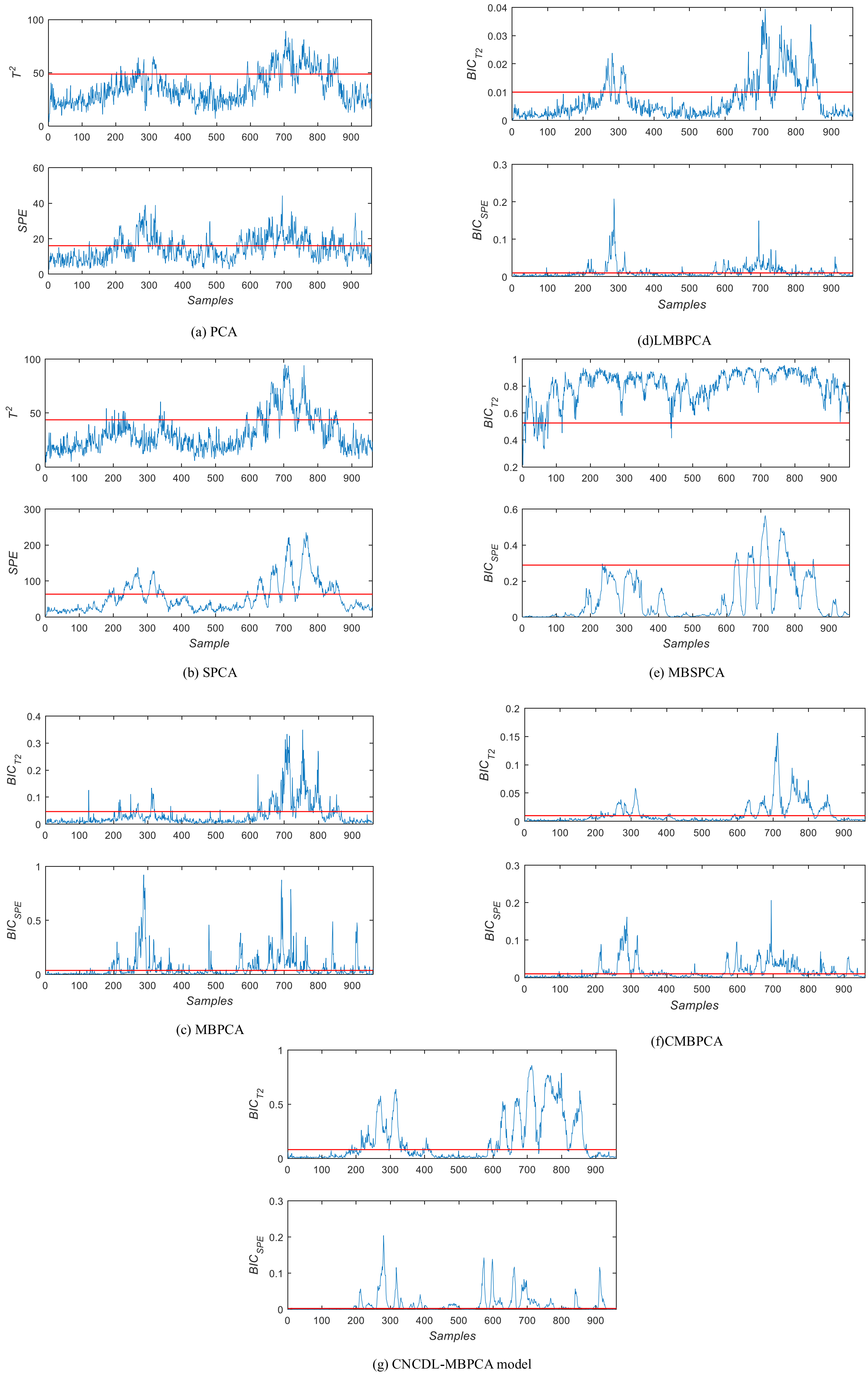

FIGURE 6. Fault 10 in TE process: (a) PCA, (b) SPCA, (c) MBPCA, (d) LMBPCA, (e) MBSPCA, (f) CMBPCA, (g) CNCDL-MBPCA.

of CNCDL-MBPCA is 0.228 higher than that of LMBPCA, which proves that the fault detection rate can be effectively improved through weighted network topology.

V. CONCLUSION

This paper proposes a novel plant-wide process monitoring algorithm based on network community discovery and Bayesian inference, which effectively reduces the complexity of process variables and improves the performance of fault detection. Louvain, a novel network community discovery algorithm, and consequently, the structural characteristics and complexity of variables are fully considered. The original data set collected in the plant-wide process is divided into blocks. After sub-blocks partition, process monitoring based on PCA model is carried out for each sub-block. Because the high-correlation variables are divided into the same sub-block by using the block division technology, the local behaviors among variables are well considered, and the disadvantages of the traditional methods are solved. Therefore, it is easier to find faults, better explain the monitoring results, and improve the detection performance. By combining the fault detection results of each module with Bayesian reasoning, the final joint index is obtained to provide intuitive monitoring results. The effectiveness of this method is verified through the analysis via the experiment in a simulated industrial TE benchmark process.

Although the variable relationship is analyzed through the complex network topology, the fault detection is realized effectively. But the method in this paper has one disadvantage: when blocks are divided among variables, not all process variables are divided by the complex network community discovery algorithm. Therefore, in the future, process monitoring based on complex network partitioning should be improved to fully consider the similarity between all variables. In the real industrial process, fault diagnosis is also very important, it can determine the cause of the fault to eliminate the fault. Therefore, fault diagnosis based on complex network can be studied in the future. By analyzing the topology of the network, it can be determined whether the fault is caused by a certain node or jointly with its neighbor nodes.

REFERENCES

[1] Z. Liu and L. Zhang, "A review of failure modes, condition monitoring and fault diagnosis methods for large-scale wind turbine bearings," *Measurement*, vol. 149, Jan. 2020, Art. no. 107002, doi: 10.1016/j.measurement.2-019.107002.

[2] M. R. Maurya, R. Rengaswamy, and V. Venkatasubramanian, "Fault diagnosis using dynamic trend analysis: A review and recent developments," *Eng. Appl. Artif. Intell.*, vol. 20, no. 2, pp. 133–146, Mar. 2007, doi: 10.1016/j.engappai.2006.06.020.

[3] L. Li, H. Luo, S. X. Ding, Y. Yang, and K. Peng, "Performance-based fault detection and fault-tolerant control for automatic control systems," *Automatica*, vol. 99, pp. 308–316, Jan. 2019, doi: 10.1016/j.automatica.2018.10.047.

[4] V. Venkatasubramanian, R. Rengaswamy, S. N. Kavuri, and K. Yin, "A review of process fault detection and diagnosis," *Comput. Chem. Eng.*, vol. 27, no. 3, pp. 313–326, 2003, doi: 10.1016/S00981354-(02)00160-6.

[5] C. Tong, T. Lan, H. Yu, and X. Peng, "Distributed partial least squares based residual generation for statistical process monitoring," *J. Process Control*, vol. 75, pp. 77–85, Mar. 2019.

[6] X. Peng, Z. Li, W. Zhong, F. Qian, and Y. Tian, "Concurrent quality-relevant canonical correlation analysis for nonlinear continuous process decomposition and monitoring," *Ind. Eng. Chem. Res.*, vol. 59, no. 18, pp. 8757–8768, 2020, doi: 10.1021/acs.iecr.0c00895.

[7] A. F. Silva, J. Vercruysse, C. Vervaet, J. P. Remon, J. A. Lopes, T. De Beer, and M. C. Sarraguça, "In-depth evaluation of data collected during a continuous pharmaceutical manufacturing process: A multivariate statistical process monitoring approach," *J. Pharmaceutical Sci.*, vol. 108, no. 1, pp. 439–450, Jan. 2019, doi: 10.1016/j.xphs.2018.07.033.

[8] M. Kano and Y. Nakagawa, "Recent developments and industrial applications of data-based process monitoring and process control," *Comput. Aided Chem. Eng.*, vol. 21, pp. 57–62, Jan. 2006, doi: 10.1016/S1570-7946(06)80026-X.

[9] S. Dražen, G. Ratko, and H. Željko, "Multivariate statistical process monitoring," *Tehnicki Vjesnik Tech. Gazette*, vol. 19, no. 1, pp. 33–41, 2012, doi: 10.1061/.(ASCE) CR.1943-5495.0000036.

[10] C. Zhao and F. Gao, "Fault-relevant principal component analysis (FPCA) method for multivariate statistical modeling and process monitoring," *Chemometric Intell. Lab. Syst.*, vol. 133, pp. 1–16, Apr. 2014, doi: 10.1016/j.chemolab.2014.01.009.

[11] J. Xu and S. Hu, "Nonlinear process monitoring and fault diagnosis based on KPCA and MKL-SVM," *Chin. J. Sci. Instrum.*, vol. 1, pp. 233–237, Nov. 2010, doi: 10.1109/AICI.2010.56.

[12] B. Zhong, J. Wang, J. Zhou, H. Wu, and Q. Jin, "Quality-related statistical process monitoring method based on global and local partial least-squares projection," *Ind. Eng. Chem. Res.*, vol. 55, no. 6, pp. 1609–1622, 2016.

[13] J. Zhou, S. Zhang, and J. Wang, "A dual robustness projection to latent structure method and its application," *IEEE Trans. Ind. Electron.*, p. 1, 2020, doi: 10.1109/TIE.2020.2970664.

[14] Z. Chen, C. Yang, T. Peng, H. Dan, C. Li, and W. Gui, "A cumulative canonical correlation analysis-based sensor precision degradation detection method," *IEEE Trans. Ind. Electron.*, vol. 66, no. 8, pp. 6321–6330, Oct. 2019, doi: 10.1109/TI-E.2018.2873100.

[15] X. Peng, Y. Tang, W. Du, and F. Qian, "Online performance monitoring and modeling paradigm based on just-in-time learning and extreme learning machine for a non-Gaussian chemical process," *Ind. Eng. Chem. Res.*, vol. 56, no. 23, pp. 6671–6684, 2017, doi: 10.1021/acs.iecr.-6b04633.

[16] Y. Tang, W. Du, and F. Qian, "Fault detection and diagnosis for non-Gaussian processes with periodic disturbance based on AMRA-ICA," *Ind. Eng. Chem. Res.*, vol. 52, no. 34, pp. 12082–12107, 2013, doi: 10.1021/ie400712h.

[17] W. Du, T. Tian, and F. Qian, "Monitoring for nonlinear multiple modes process based on LL-SVDD-MRDA," *IEEE Trans. Automat. Sci. Eng.*, vol. 11, no. 4, pp. 1133–1148, Dec. 2014, doi: 10.1109/TASE.201-3.2285571.

[18] J. A. Westerhuis, T. Kourti, and J. F. MacGregor, "Analysis of multiblock and hierarchical PCA and PLS models," *J. Chemometrics*, vol. 12, no. 5, pp. 301–321, 1998, doi: 10.1002/(SICI)1099128X(199809/10)12:5<301::AIDCEM515>3.0.CO.2-S.

[19] J. A. Lopes, J. C. Menezes, J. A. Westerhuis, and A. K. Smilde, "Multiblock PLS analysis of an industrial pharmaceutical process," *Biotechnol. Bioeng.*, vol. 80, no. 4, pp. 419–427, 2002, doi: 10.1002/bit.10-382.

[20] E. P. P. A. Derks, J. A. Westerhuis, A. K. Smilde, and B. M. King, "An introduction to multi-block component analysis by means of a flavor language case study," *Food Quality Preference*, vol. 14, nos. 5–6, pp. 497–506, 2003, doi: 10.1016/S0950-3293(03)00009-0.

[21] L. Ma, J. Dong, and K. Peng, "A novel key performance indicator oriented hierarchical monitoring and propagation path identification framework for complex industrial processes," *ISA Trans.*, vol. 96, pp. 1–13, Jan. 2020, doi: 10.1016/j.isatra.2019.06.004.

[22] Q. Jiang, Y. Wang, and X. Yan, "Optimal variable transmission for distributed local fault detection incorporating RA and evolutionary optimization," *IEEE Access*, vol. 6, pp. 3201–3211, 2018, doi: 10.1109/ACCESS.2017.2787620.

[23] J. Kohonen, S.-P. Reinikainen, K. Aaljoki, A. Perkiö, T. Väänänen, and A. Höskuldsson, "Multi-block methods in multivariate process control," *J. Chemometrics*, vol. 22, nos. 11–12, pp. 580–586, Nov. 2008, doi: 10.1002/cem.1199.

[24] Z. Q. Ge and Z. H. Song, "Distributed PCA model for plant-wide process monitoring," *Ind. Eng. Chem. Res.*, vol. 52, no. 5, pp. 1947–1957, 2013, doi: 10.1021/ie301945s.

[25] L. Xie, X. Lin, and J. Zeng, "Shrinking principal component analysis for enhanced process monitoring and fault isolation," *Ind. Eng. Chem. Res.*, vol. 52, no. 49, pp. 17475–17486, 2013, doi: 10.1021/ie401030t.

[26] Q. Jiang, B. Wang, and X. Yan, "Multi-block independent component analysis integrated with Hellinger distance and Bayesian inference for non-Gaussian plant-wide process monitoring," *Ind. Eng. Chem. Res.*, vol. 54, no. 9, pp. 2497–2508, 2015, doi: 10.1021/ie403540b.

[27] C. Zhang, K. Peng, and J. Dong, "A novel plant-wide process monitoring framework based on distributed gap-SVDD with adaptive radius," *Neurocomputing*, vol. 350, pp. 1–12, Jul. 2019, doi: 10.1016/j.neucom.2019.04.026.

[28] J. Huang and X. Yan, "Relevant and independent multi-block approach for plant-wide process and quality-related monitoring based on KPCA and SVDD," *ISA Trans.*, vol. 73, pp. 257–267, Feb. 2018, doi: 10.1016/j.isatra.2018.0-1.003.

[29] Y. Tian, T. Hu, X. Peng, W. L. Du, and H. Yao, "Decentralized monitoring for large: Cale process using copula-correlation analysis and Bayesian inference-based multi-block principal component analysis," *J. Chemometrics*, vol. 33, no. 3, 2019, Art. no. e3158, doi: 10.1002/-cem.3158.

[30] Y. Tian, H. Yao, and Z. Li, "Plant-wide process monitoring by using weighted copula-correlation based multi-block principal component analysis approach and online-horizon Bayesian method," *ISA Trans.*, vol. 96, pp. 24–36, Jan. 2020, doi: 10.1016/j.isatra.20-19.06.002.

[31] J. Zeng, W. Huang, Z. Wang, and J. Liang, "Mutual information-based sparse multiblock dissimilarity method for incipient fault detection and diagnosis in plant-wide process," *J. Process Control*, vol. 83, pp. 63–76, Nov. 2019, doi: 10.1016/j.jprocont.2019.09.004.

[32] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *Proc. Nat. Acad. Sci. USA*, vol. 99, no. 12, pp. 7821–7826, 2001, doi: 10.1073/pnas.1226537-99.

[33] S. Chauhan, M. Girvan, and E. Ott, "Spectral properties of networks with community structure," *Physreve*, vol. 80, no. 5, 2009, Art. no. 056114, doi: 10.13-71/journal.pone.0027418.

[34] M. E. J. Newman, "Modularity and community structure in networks," *Proc. Nat. Acad. Sci. USA*, vol. 103, no. 23, pp. 8577–8582, 2006, doi: 10.1073/pnas.0-601602103.

[35] I. Tatsuro, *Super-Modularity: Applications to Convex Games and to the Greedy Algorithm for LP*, vol. 25, no. 2. New York, NY, USA: Academic, 1981, doi: 10.1016/0-022-0531(81)90007-7.

[36] C. Gao, M. Liang, X. Li, Z. Zhang, Z. Wang, and Z. Zhou, "Network community detection based on the physarum-inspired computational framework," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 15, no. 6, pp. 1916–1928, Nov. 2018, doi: 10.1109/TCBB.2016.2638824.

[37] B. Sieyum, *Community Discovery Algorithm by Clustering and Optimizing Genetic*, vol. 7, no. 7. Washington, DC, USA: IEEE Comput. Soc. Press, 2020, doi: 10.21278/ART20198308.

[38] F. Benita, S. Sarica, and G. Bansal, "Testing the static and dynamic performance of statistical methods for the detection of national industrial clusters," *Papers Regional Sci.*, vol. 99, no. 4, pp. 1137–1157, Aug. 2020, doi: 10.1111/pirs.12517.

[39] H. Miao, X. Guo, and F. Wu, "Identification of the technology convergence in the field of elderly smart home based on Louvain community discovery algorithm," in *Proc. ICISDM*. New York, NY, USA: Association for Computing Machinery, Apr. 2019, pp. 47–53.

[40] L. Luo, M. Xu, S. Bao, J. Mao, and C. Tong, "Improvements to the t2 statistic for multivariate fault detection," *Ind. Eng. Chem. Res.*, vol. 58, no. 45, pp. 20692–20709, Nov. 2019, doi: 10.1021/acs.iecr.9b04112.

[41] H. Hong, C. Jiang, X. Peng, and W. Zhong, "Concurrent monitoring strategy for static and dynamic deviations based on selective ensemble learning using slow feature analysis," *Ind. Eng. Chem. Res.*, vol. 59, no. 10, pp. 4620–4635, 2020, doi: 10.1021/acs.iecr.9b05547.

[42] C. Tong and X. Yan, "A novel decentralized process monitoring scheme using a modified multi-block PCA algorithm," *IEEE Trans. Automat. Sci. Eng.*, vol. 14, no. 2, pp. 1129–1138, Apr. 2017, doi: 10.1109/-TASE.2015.2493564.

[43] Q. Jiang and X. Yan, "Nonlinear plant-wide process monitoring using MI-spectral clustering and Bayesian inference-based multiblock KPCA," *J. Process Control*, vol. 32, pp. 38–50, Aug. 2015, doi: 10.1016/j.jprocont.2015.04.014.

[44] J. F. MacGregor, C. Jaeckle, C. Kiparissides, and M. Koutoudi, "Process monitoring and diagnosis by multiblock PLS methods," *AIChE J.*, vol. 40, no. 5, pp. 826–838, May 1994, doi: 10.1002/aic.690400509.

[45] C. Tong, Y. Song, and X. Yan, "Distributed statistical process monitoring based on four-subspace construction and Bayesian inference," *Ind. Eng. Chem. Res.*, vol. 52, no. 29, pp. 9897–9907, 2013, doi: 10.1021/-ie400544q.

[46] Z. Chen, Y. Cao, S. X. Ding, K. Zhang, T. Koenings, T. Peng, C. Yang, and W. Gui, "A distributed canonical correlation analysis-based fault detection method for plant-wide process monitoring," *IEEE Trans. Ind. Informat.*, vol. 15, no. 5, pp. 2710–2720, May 2019, doi: 10.1109/TII.2019.2893125.

[47] C. H. Huang, J. Yin, and F. Hou, "A text similarity measurement combining word semantic information with TF-IDF method," *Chin. J. Comput.*, vol. 34, no. 5, pp. 856–864, 2011, doi: 10.3724/SP.J.1-016.2011.00856.

[48] Z. Lu, N. Yuan, L. Chen, and Z. Gong, "On the impacts of el Niño events: A new monitoring approach using complex network analysis," *Geophys. Res. Lett.*, vol. 47, no. 6, Mar. 2020, Art. no. e2019GL086533, doi: 10.1029/2019GL086533.

[49] R. Imbuzeiro Oliveira, "Concentration of the adjacency matrix and of the Laplacian in random graphs with independent edges," 2009, *arXiv:0911.0600*. [Online]. Available: http://arxiv.org/abs/0911.0600

[50] I. T. Jolliffe, *Principal Component Analysis*, vol. 98, no. 10, 2nd ed. Berlin, Germany: Weather, 2005, doi: 10.2307/3172953.

[51] E. Parzen, "On estimation of a probability density function and mode," *Annmathstatis*, vol. 33, no. 3, pp. 1065–1076, 1962, doi: 10.1214/aoms/1177704472.

[52] N. R. Goodman, "Annals of mathematical statistics," Amer. Stat. Assoc., Ann Arbor, MI, USA, Tech. Rep. 43, 1963, doi: hdl.handle.net/2324/3865707.

[53] Z. Ge, M. Zhang, and Z. Song, "Nonlinear process monitoring based on linear subspace and Bayesian inference," *J. Process Control*, vol. 20, no. 5, pp. 676–688, Jun. 2010.

YANAN LI received the B.S. degree in electrical engineering and automation from the Xuhai College, China University of Mining and Technology, Xuzhou, China, in 2018. She is currently pursuing the M.S. degree in electrical engineering with the University of Shanghai for Science and Technology.

Her research interests include machine learning, deep learning, industrial process monitoring, and fault diagnosis.

XIN PENG (Member, IEEE) received the B.S. and M.S. degrees in control science and engineering from the East China University of Science and Technology, Shanghai, China, in 2009 and 2012, respectively, and the Ph.D. degree from the Key Laboratory of Advanced Control and Optimization for Chemical Processes, Ministry of Education, East China University of Science and Technology, in 2017. He served as a lecturer with the university. From 2017 to 2019, he worked as a Postdoctoral Researcher with the University of Duisburg–Essen. His current research interests include process monitoring and the system modeling of chemical and biological processes, data mining, and the feature extraction of process data.

YING TIAN received the B.S. degree in electrical engineering and automation and the Ph.D. degree in control theory and control engineering from the East China University of Science and Technology, Shanghai, China, in 2010 and 2015, respectively.

She is currently a Lecturer with the School of Optical–Electrical and Computer Engineering, University of Shanghai for Science and Technology. Her research interests include process monitoring, system modeling, advanced control, and process optimization.