A Fault Detection Method Based on Horizontal Visibility Graph-Integrated Complex Networks: Application to Complex Chemical Processes

Zhiqiang Geng,[1,2] Zun Wang,[1,2] Haixia Hu,[1,2] Yongming Han,[1,2]* Xiaoyong Lin[1,2]* and Yanhua Zhong[3]

1. College of Information Science & Technology, Beijing University of Chemical Technology, Beijing 100029, China

2. Engineering Research Center of Intelligent PSE, Ministry of Education in China, Beijing 100029, China

3. Jiangmen Polytechnic, Jiangmen, Guangdong 529020, China

With the trend towards large-scale structure and complexity in modern chemical processes, it is difficult to describe the system operating status and fault conditions via traditional approaches. In addition, there are obstacles to fault detection research in complex chemical processes due to the characteristics of the equipment hardware used in chemical systems, data collection and handling, strong internal correlations in chemical systems, factors affecting transmission, and randomness and cascade in process failures. Therefore, this paper presents a fault detection method based on horizontal visibility graph (HVG) analysis-integrated complex networks. The data for each variable in the system are regarded as a time series, each time series is modelled into a network by the horizontal visibility algorithm, and each single-layer network corresponding to a time series is abstracted as a node. Meanwhile, the correlation between two single-layer networks is used to characterize the correlation between the corresponding nodes. Moreover, a complex network structure representing the chemical system can be constructed from the correlations. In addition, according to the correlation ratio matrix obtained from the fault state and the normal state, the variance of each node determines the faulty node. Finally, we verify the validity and the effectiveness of the proposed method by applying it to the Tennessee Eastman (TE) process.

Keywords: data driven, time series, HVG, fault detection, complex chemical process

INTRODUCTION

With the rapid development of the real economy and the progress of ever-changing science and technology, many industrial production equipment systems are increasing in scale and complexity. Meanwhile, the large-scale structure and complexity of the current industrial system has enhanced the strength of relationships between the devices in the system and has led to increasingly severe accidents.[1] In recent years, accidents due to failures in the industrial system have happened frequently. For example, in 2010, an accidental fire at a petrochemical plant affiliated with the Lanzhou Petrochemical Company in China occurred due to tank leakage, which caused the combustible gas concentration at the scene to reach the explosive limit; as a result of the fire, five people were missing, one was seriously injured, and five others were slightly injured. Therefore, effective protection by industrial system monitoring and fault detection is important in reducing accidents, casualties, and economic losses. Fault diagnosis technology has always been an important topic of investigation and it is widely used in industrial process systems.[2,3] In addition, with multidisciplinary development and the expansion of network theory, network-based fault diagnosis methods, such as fault diagnosis based on Bayesian networks,[4–6] fault diagnosis based on neural networks,[7,8] and fault diagnosis based on complex networks,[9,10] have also received increasing attention. Fault detection based on complex networks regards the whole industrial system as a network structure and the devices or monitoring variables in industrial systems are regarded as the nodes in the network. Meanwhile, the complex interactions are the overall topology information and local attributes in the network. Moreover, fault detection based on complex networks takes relevance, communication, and other characteristics of the industrial system into account and is highly useful in the exploration of industrial fault diagnosis. The data generated in the modern industrial process represent the state of the devices in the operating system, so that a failure of an industrial system can be understood as an abnormal change in the characteristics or device parameters in the system caused by the system itself. Therefore, data-driven fault diagnosis can find the circumstances of system failure by mining the data for potential information.[11,12] Furthermore, the time series,[13] a common and highly useful data type, is an important direction of research in the data-mining field. In 2008, Lacasa et al. introduced the method of mapping time series into network graphs based on visibility-graph theory.[14] Therefore, this paper proposes a fault detection method based on horizontal visibility graph (HVG) integrated complex networks from the perspective of time series data.

The organization of the remainder of this paper is as follows. The Related Work section presents the research status of fault detection. The details of the proposed method are introduced in The HVG Integrated Complex Networks Method section. The Experiments section presents modelling experiments based on the Tennessee Eastman (TE) process and comparisons of the results

* *Author to whom correspondence may be addressed.*
E-mail address: hanym@mail.buct.edu.cn (Y. Han); linxy@mail.buct.edu.cn (X. Lin)
Can. J. Chem. Eng. 9999:1–10, 2018

DOI 10.1002/cjce.23319
Published online in Wiley Online Library (wileyonlinelibrary.com).

with the partial-correlation method.[15] Finally, the conclusions are discussed in the Discussion and the Conclusions sections.

RELATED WORK

Fault diagnosis technology has been in development for several decades. This field is a highly comprehensive area of scientific research that has been involved in studies in various subjects. Therefore, the application of this technology is very extensive. In general, the fault diagnosis process includes the following five phases: fault modelling; fault detection; fault reasoning; fault identification; and fault processing. Fault modelling, which is the basis of the process, is designed to establish a model through the existing information in the system. Fault detection is used to determine whether the system is normal or faulty on the basis of the system's output or changes in the characteristics of parameters in the system, and this step plays the most important role in fault diagnosis. The fault reasoning stage is used to determine the location of the fault and then determine the reason after the fault is detected. Fault identification ascertains the attributes of the fault itself, including the time and ageing characteristics. Finally, the fault processing phase establishes the appropriate approach according to the scope and extent of the fault's influence.

However, as the scale of industrial processes becomes increasingly large and more complex, the need for excellence in fault diagnosis technology also grows. To keep pace with and adapt to developments over time, a good fault diagnosis method is particularly critical. The current fault diagnosis methods are numerous and have been built on one another effectively. In general, fault diagnosis methods are mainly divided into the following parts: methods based on analytical models;[16–18] methods based on qualitative knowledge;[19,20] and methods based on data.[21,22] Initially, fault diagnosis research is mostly based on an analytical model and, consequently, many achievements in the field include state estimation methods, parameter estimation methods, and equivalent space methods.[23] The most important state estimation methods are the filter methods. Zhang proposed an adaptive interactive multi-model (IMM) estimator method for fault diagnosis that uses the adaptive Kalman filter, and the IMM was applied to a random hybrid system to verify the performance of the method.[24] Bonvini et al. proposed a fault diagnosis method that is based on an infinite Kalman filter and advanced Bayesian nonlinear state estimation technology, and they validated the method's effectiveness by applying it to energy and hydraulic fault detection in chillers.[25] The parameter estimation methods, which are technical and forward-looking, can be subdivided into estimation methods based on the system parameters and estimation methods based on the fault parameters. According to the different parameters selected, these estimation methods can accurately estimate the system fault parameters while simultaneously estimating system output states.[26] Above all, these methods based on an analytical model require a precise mathematical model and explore the mechanisms underlying the system. Therefore, these methods have certain limitations because they are very difficult in practice. Methods based on qualitative knowledge mainly include directed graphs and expert systems.[27–29] Peng et al. proposed a semi-quantitative fault diagnosis approach based on a multi-logic probabilistic signed directed graph (SDG) model with Bayesian inference and solved the inability of SDG to express complicated logic relations other than the logic "OR," which usually results in spurious interpretations.[27] Yang et al. introduced an expert system based on fault tree analysis and analyzed the system's feasibility both qualitatively and quantitatively by modelling a fault in a wind turbine gearbox.[29] Such methods need adequate expert knowledge, so they are suitable only for less-variable systems. Currently, qualitative methods appear to be costly for large-scale industrial systems and massive data and, therefore, their applicability is also limited. Data-driven methods mainly include statistical analysis methods,[30–33] methods based on signal processing,[34,35] and methods based on intelligent analysis.[36–38] The methods based on statistical analysis have the following two directions: univariate statistical methods; and multivariate statistical methods. The former are mostly control graph methods;[39] the latter mainly include principal components analysis (PCA),[30] canonical variate analysis (CVA),[31] Fisher discriminant analysis (FDA),[32] and partial least squares (PLS) analysis,[33] all of which are methods of dimension reduction. In the study of signal processing, the wavelet transform and the Hilbert-Huang transform,[34,35] which are fault diagnosis methods based on signal analysis, have received attention from many scholars. In addition, these methods are based on intelligent analysis, which mainly involves neural networks,[36] Bayesian networks,[37] fuzzy logic,[38] etc. and usually use artificial intelligence technology to carry out the fault diagnosis.

Even though fault diagnosis technology has developed rapidly and its applications are highly complicated, some influencing factors always pose problems in this field. On one hand, the vast structure of the industrial system and the complexity of the production process make it increasingly difficult to describe a system's operating status and fault conditions. On the other hand, the features that exist among faults in industrial processes set barriers to the path of fault diagnosis. These factors cause problems because there is a lack of analysis of the properties of and complex interactions among device variables, as well as the local and overall system topology. However, complex network theory research can overcome these shortcomings and better show the complexity of industrial system data. Du presented a fault diagnosis strategy based on complex networks and used a network model to represent the fault data structure, thus transforming the clustering problem into the detection of the subnet structure.[40] Peng et al. used complex network theory to describe the interdependencies of various failure modes, allowing the fault diagnosis of analogue circuits to be studied from a novel perspective.[41] These examples reflect the prospects of developments in the use of complex network theory in industrial fault diagnosis. These kinds of methods consider not only local details but also the integrity of overall structural analysis in the form of a complex network. In this way, there is a class of methods that can map data with timing characteristics into a network, and then, a network analysis approach, the graph method, can be used to analyze the information contained in the original data. Visibility graph analysis has been applied in many fields because of its advantage in describing time series. Shao performed complex network modelling of the human heartbeat through the visibility graph algorithm and analyzed the dynamic characteristics of the human heartbeat.[42] The study found that the relevant network of heartbeat interval time series was always dimensionless, highly clustered, and hierarchical, among other characteristics, and that the network classification coefficient could also distinguish between healthy patients and patients with congestive heart failure. Hloupis studied a seismic time series by using the visibility graph method, analyzed the complex characteristics of the data, and found that the time period of the seismic sequence after an earthquake was consistent with the time period of the minimum value of the index.[43] The study indicated that complex network

modelling through the visibility graph algorithm was also significant in terms of seismic detection. Therefore, it is of great practical significance to perform complex network modelling by using the visibility mapping method when exploring the various aspects of system characteristics.

In relation to the above discussion, this paper proposes a fault detection method based on HVG integrated complex networks. The HVG regards the device nodes in the industrial system as a collection of multiple time series. According to the visibility algorithm, the attribute of each device node corresponding to a single-layer network can be analyzed. The correlation between any two single-layer networks is regarded as the correlation between corresponding device nodes. Meanwhile, a complex network representing the industrial system on the whole can be constructed by studying the characteristics of the network structure. Moreover, the fault can be determined from the variance of each node on the basis of the correlation ratio matrix obtained from data from the fault state and the normal state. This method avoids the error caused by the complexity of empirical knowledge acquisition and the lack of empirical knowledge and prevents human error caused by artificial threshold-setting in fault diagnosis methods based on complex networks. Finally, the proposed method is applied to analyze faults in complex chemical processes as modelled by the TE process. The experiment shows that the proposed method is feasible and effective. Furthermore, the proposed method performs better than the partial-correlation method.

THE HVG INTEGRATED COMPLEX NETWORKS METHOD

On the one hand, the HVG integrated complex networks method considers the timing of chemical system data to uncover deeper potential information about the data. On the other hand, this method can analyze the structure of the data on the basis of a complex network to describe the complexity of the chemical data.

HVG

Lacasa et al. proposed a simple and fast mathematical method for mapping time series into network graphs, which is called the visibility graph.[14] The concept is that for any two samples $[t_a, y_a]$ and $[t_b, y_b]$ in the time series, given a third sample $[t_c, y_c]$ that satisfies $t_a < t_c < t_b$ and the relationship shown in Equation (1), the two samples can be considered visible and are regarded as two connected nodes in the visibility graph:

$$y_c < y_b + (y_a - y_b)\frac{t_b - t_c}{t_b - t_a} \tag{1}$$

Then, the time series can be characterized from a new perspective.

On the basis of the visibility graph, Fioriti and Pietro modified this visibility rule in 2012 and proposed a new visibility graph algorithm called the horizontal visibility graph.[44] The horizontal visibility graph proposes that for any $i < k < j$ in the time series $X(n) = \{x(1), x(2), \ldots, x(k), \ldots, x(n)\}$ that satisfy the relationship shown in Equation (2), x_i and x_j have horizontal visibility and are two connected nodes in the horizontal visibility graph:

$$x_i > x_k, x_j > x_k \tag{2}$$

In particular, two adjacent data points, such as $x(1)$ and $x(2)$, have a direct connection because there is no other data point between them. The mapping relation between the horizontal visibility graph and timing data is shown in Figure 1.

Figure 1 shows an example of timing data for 3 variables (V1, V2, V3), which are mapped into 3 network layers (L1, L2, L3) by the HVG method. Each variable corresponds to one layer. The linking in each layer, represented by the arcs, is determined by the HVG algorithm mentioned above; for example, the partial timing data for V1 inside the dashed oval can be mapped into the structure shown in L1. The overall network and corresponding correlation matrix $\boldsymbol{R}$ are obtained by interlayer mutual information (MI). The details are introduced in the following sections.

The Degree Distribution Between Multiple Single-Layer Networks

For the HVG method, it is necessary to analyze the correlation between the single-layer networks to obtain the correlation between the device nodes after mapping the data of each device node into a single-layer network. For a system with L device nodes, there are L single-layer networks, and each layer has its own network structure. As shown in Figure 1, L1 is a single-layer network determined by V1, which is a device node in the overall network determined by the multi-layer network (L1, L2, L3). In this paper, the degree distribution of a single-layer network and the degree correlation between multiple single-layer networks are described in detail for any two device nodes r and t as an example.

Taking n samples as an example, the data sequences of the two device nodes r and t are as follows:

$$\boldsymbol{x_r} = \{x_{r1}, x_{r2}, \ldots, x_{rn}\} \tag{3}$$

$$\boldsymbol{x_t} = \{x_{t1}, x_{t2}, \ldots, x_{tn}\} \tag{4}$$

Degree distribution of a single-layer network

The degree of a node refers to the number of edges that are associated with a node in the network and is also known as the degree relevance; this value is an important measure in complex networks.

The examples given as Equations (3) and (4) correspond to a single-layer network r and a single-layer network t, respectively. Each network can be used to construct its own network model according to the horizontal visibility graph method (Equation (2)) and the corresponding degree distribution can be obtained. Taking the r-layer network as an example, the degree distribution of the layer is expressed by $P(k^r)$, where k^r is the degree of the r-layer network, and the degree distribution calculation process is as shown in Figure 2.

The degree correlation between multi-layer networks

For a single-layer network r and corresponding single-layer network t with the same n samples, the n nodes in layer r and the n nodes in layer t have degrees in each layer of the two-layer network. The degree of node p in layer r is expressed as k^r and the degree of node p in layer t is expressed as k^t. In addition, the two networks have a degree correlation between the nodes, which is expressed by $P(k^r, k^t)$:

$$P(k^r, k^t) = \frac{N_{k^r,k^t}}{n} \tag{5}$$

where N_{k^r,k^t} represents the number of nodes in which the degree in layer r is k^r and the degree in layer t is k^t and satisfies the following condition:

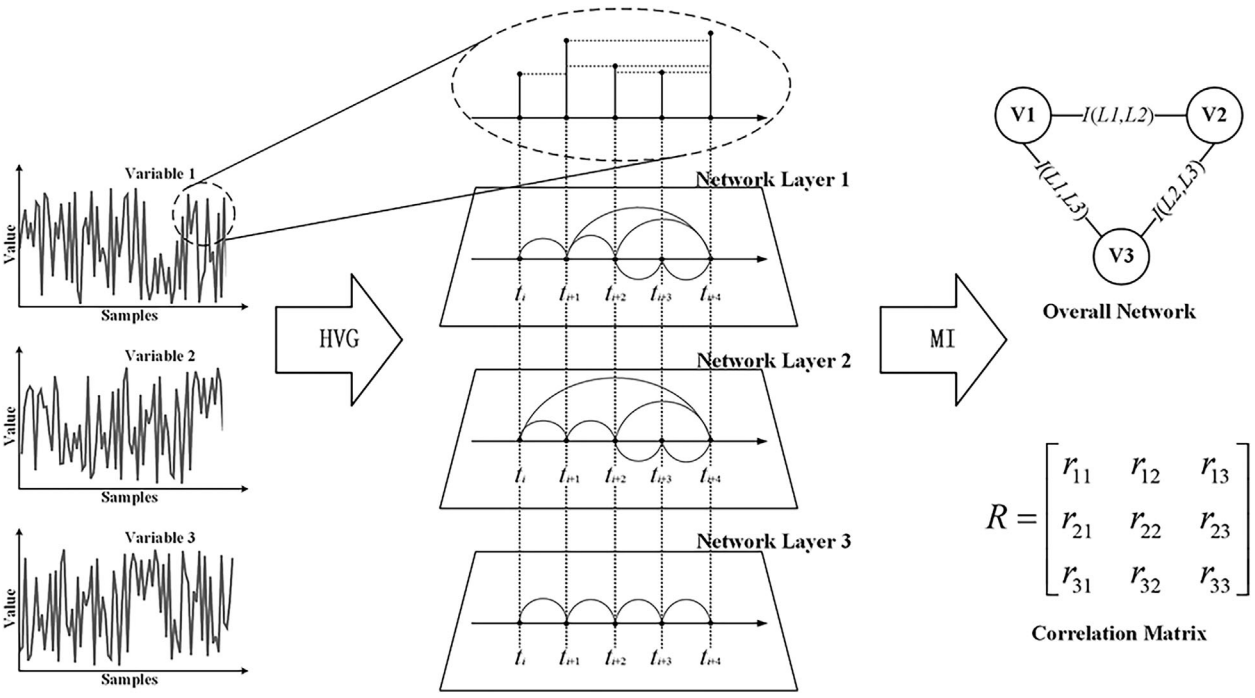

Figure 1. Horizontal visibility graph example.

$$\sum_{k^r}\sum_{k^t} N_{k^r,k^t} = n. \tag{6}$$

Build the Correlation Matrix

Data standardization

There are many different devices and forms of equipment in the industrial system, so the corresponding data have different units and orders of magnitude. In general, it is necessary to standardize data. In this paper, the z-score method is adopted; the specific process for this method is shown as follows:

$$x' = \frac{x - \bar{X}}{S_i} \tag{7}$$

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{8}$$

$$S_i = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{X})^2}{n-1}} \tag{9}$$

where $\bar{X}$ is the average of the variable X, S_i is the standard deviation of the variable X, x' is the value of the variable after standardization, and x is the original value of the variable.

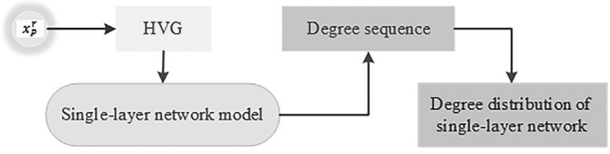

Figure 2. Calculation flow chart of a single-layer-network degree distribution.

Mutual information

Mutual information can determine whether the two variables X and Y are related and the strength of correspondence between the two variables. Mutual information is defined as the following:

$$MI(X,Y) = \int X \int Y\, P(X,Y) log \frac{P(X,Y)}{P(X)P(Y)} \tag{10}$$

where $P(X,Y)$ is the joint probability function of X and Y, and $P(X)$ and $P(Y)$ are the marginal probability distribution functions of X and Y, respectively.

It can be seen from the equation that if X and Y are independent from each other, then $P(X,Y) = P(X)P(Y)$, so that $MI(X,Y) = 0$, which demonstrates that X and Y are not related.

In the paper, after the data are standardized, the data for each device node are regarded as a time series. Each time series is mapped into a single-layer network by using the horizontal visibility graph, and then the mutual information between single-layer networks can be analyzed to characterize their correlation. For instance, the variable nodes r and t correspond to two single-layer networks, namely, layer r and layer t, respectively. The mutual information between the two single-layer networks satisfying these conditions is calculated as the following:

$$I(r,t) = \sum_{k^r}\sum_{k^t} \mathrm{P}(k^r,k^t) log \frac{P(k^r,k^t)}{P(k^r)P(k^t)} \tag{11}$$

The parameters in Equation (11) are described in the Degree Correlation Between Multi-Layer Networks section in detail.

Then, a complex network corresponding to the complex chemical system can be built for analysis and research. Every variable in the system is a node in the complex network and the degree correlation is the relationship between nodes.

The correlation matrix

The mutual information between a single-layer network and another single-layer network is used as the correlation of the corresponding device nodes. On this basis, the correlation matrix $\boldsymbol{R}$ is as follows:

$$\boldsymbol{R} = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \cdots & \cdots & \ddots & \cdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} \end{bmatrix}. \quad (12)$$

where $\boldsymbol{R}$ is equivalent to the adjacent matrix of a weighted complex network. The obtained correlation matrix is used to determine the correlation between the corresponding two devices. Therefore, fault detection can be carried out in accordance with the correlation-ratio matrix between the fault state and the normal state for the system.

Fault Detection Analysis Method

The overall process of the method is based on data and determines the fault only in terms of the analysis of data. In the measurement of data, variance is a measure of the difference between the source data and the expected value. The variance not only expresses the degree of deviation from the sample mean but also reveals the degree of fluctuation in the sample, and it can be understood that the variance represents the expectation of fluctuations in the sample.[45] Therefore, ever since Fisher proposed variance,[46] this statistic has been used to characterize the deviation of data in many practical problems.

In the case of a normal system, the data corresponding to each device variable contain the characteristics of the normal state and the interrelationship between the devices. When a system is in fault, the data corresponding to some device variables are directly changed. The changes affect the characteristics of the device variables themselves and their relationships with other device variables. For the method described here, the structure of the single-layer network corresponding to the fault node in the fault state changes, which leads to a change in the degree distribution. On the whole, the degree correlation between multi-layer networks is affected. Finally, the change in data affects the mutual information between device variable nodes and, thus, affects the complex network structure that represents the chemical system. The method detects failure according to the variance after handling the relation matrix describing system data in the normal state and faulty state. The flow chart of the proposed method is shown in Figure 3 and the specific steps are as follows:

step 1. standardize the data when the system is normal and then construct the corresponding single-layer network for each device node in the system via the horizontal visibility graph;
step 2. obtain the degree distribution of each single-layer network and the degree correlation between layers, so that the mutual information between the multi-layer networks can be obtained, which is the corresponding device node correlation when the system is under normal conditions;
step 3. collect the data when the system is operating then standardize the data and obtain the network structure associated with every device node through the horizontal visibility graph;
step 4. acquire the degree distribution of each single-layer network and the degree correlation between layers when the state of the system is to be tested and then obtain the mutual information, that is, the correlation between corresponding device nodes;
step 5. obtain the ratio matrix composed of the ratio of the correlation obtained in step 4 to the correlation obtained in step 2, which indicates the change in the correlation between device nodes for the system in an unknown state with respect to the system in a normal state; and
step 6. calculate the corresponding variance of each device node according to the ratio matrix. If there is a device node with a large variance compared to the other device nodes, then the system to be tested is in fault and the device node with the largest variance is the most likely faulty node; otherwise, return to step 3 to collect new data for further verification.

EXPERIMENTS

TE Process

To verify the feasibility and validity of the proposed method, the TE process was selected for testing. The TE process is mainly made up of 5 components, which are the reactor, condenser, compressor, separator, and stripper, and the process flow chart is shown in Figure 4. The TE process includes 41 measurement variables and 12 manipulated variables, and the 41 measurement variables contain 22 variables for continuous process measurements, such as the pressure, the temperature, and the liquid level. The other 19 variables are component variables that are not considered in this paper. The details of the 22 considered variables are shown in Table 1. A detailed description of the TE process is available in the literature.[47] The data set includes the normal condition and 21 fault conditions,[48] and we used 100 samples of the 22 considered variables under each condition.

Experimental Analysis

To verify the method, this section analyzes the fault detection results from three typical fault types, namely, a step fault, random fault, and viscous fault through an example analysis of the TE process. In addition, the results of this analysis are contrasted with the results obtained from the partial-correlation method.[15]

Normal TE process

Data are collected from a TE model that simulates zero faults. In the method, for the timing data of each variable, the corresponding single-layer network structure can be obtained through the horizontal visibility graph and the topological properties can be analysed. Then, each layer network is regarded as a node to obtain the correlation between nodes on the basis of the topological properties of each layer and the correlation between layers; correspondingly, a complex network and the correlation matrix A_0 between nodes in the normal system are obtained.

Example 1: fault 4

The TE simulation model is set to fault 4, a step fault in which the reactor cooling water inlet temperature changes and the corresponding system data are collected.

In the analysis method proposed in this paper, the data from fault 4 are used to construct multi-layer networks with 22 single-layer networks according to the algorithm given in Equation (2). Therefore, the degree order and degree distribution of each layer's network in the fault 4 state can be obtained, and the inter-layer

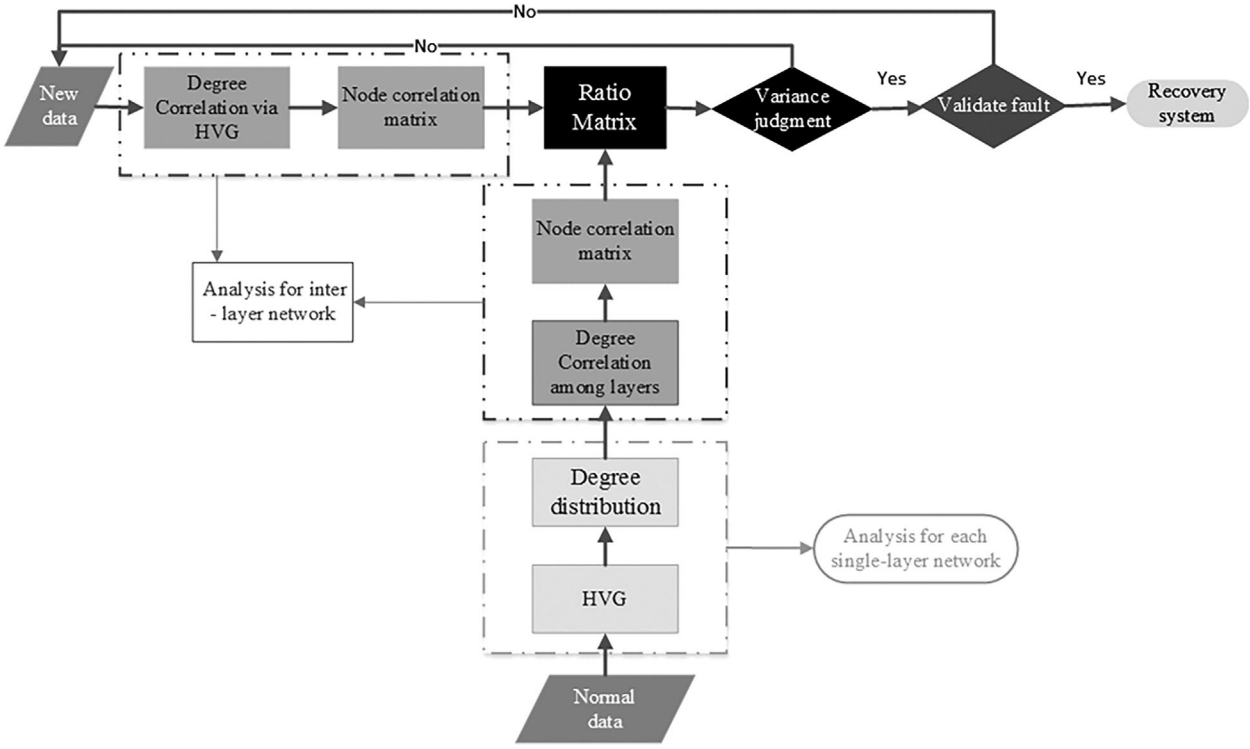

Figure 3. Flow chart of the proposed method.

mutual information between any two single-layer networks within the system is also correspondingly available. Most importantly, the correlation matrix, recorded as A_4, which represents the correlation between nodes in the complex network when system is under fault 4, is obtained.

In this paper, the variance is used to characterize the fluctuations of the correlations between the multi-layer networks between the normal state and faulty state. In the case of no fault, the system is steady, the corresponding structures of the multi-layer networks are stable, and the correlations between nodes lack

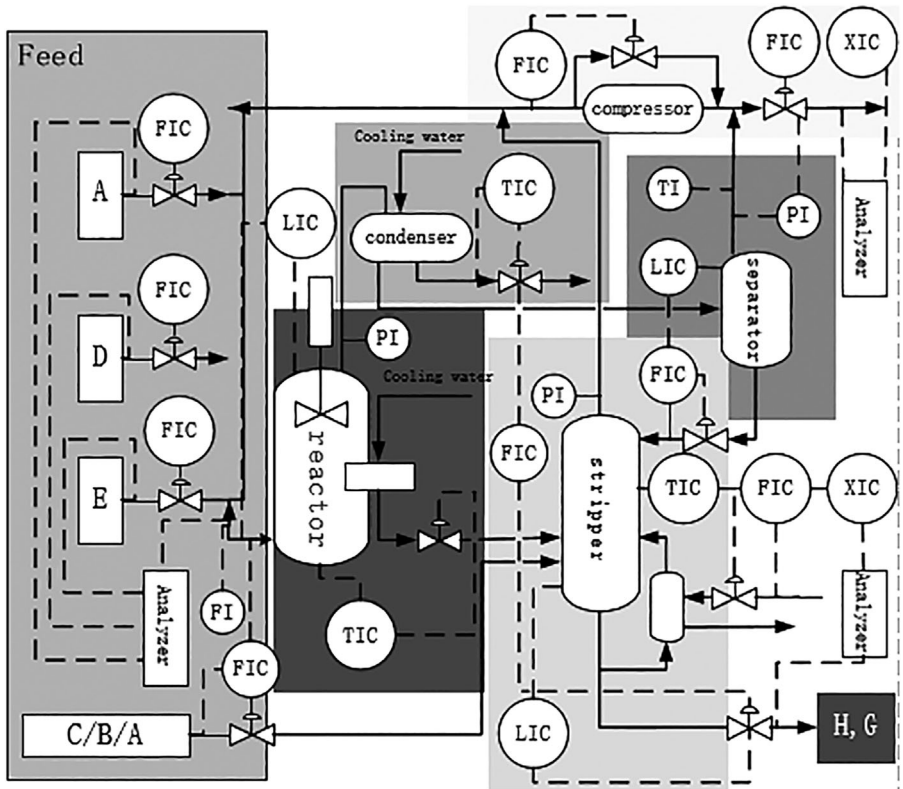

Figure 4. Flow chart of the TE process.

Table 1. Numbers and descriptions for the 22 variables of the TE process

No.	Description	No.	Description
1	A feed (stream 1)	12	Separator level
2	D feed (stream 2)	13	Separator pressure
3	E feed (stream 3)	14	Separator underflow (stream 10)
4	A and C feed (stream 4)	15	Stripper level
5	Recycle flow (stream 8)	16	Stripper pressure
6	Reactor feed rate (stream 6)	17	Stripper underflow (stream 11)
7	Reactor pressure	18	Stripper temperature
8	Reactor level	19	Stripper flow
9	Reactor temperature	20	Compressor power
10	Purge rate (stream 9)	21	Reactor cooling water outlet temperature
11	Separator temperature	22	Separator cooling water outlet temperature

obvious fluctuations; if the system fails, the intuitive impact is that the data change substantially, and the closer the fault source is, the greater the impact is. It follows that the stability of the correlations between the fault node and other variable nodes will fluctuate greatly. Figure 5 shows the variance of the correlation between each node and the other 21 nodes when comparing the system in normal conditions and in fault 4 conditions.

It can be seen from Figure 5 that the stability of the correlations between each node and the other 21 nodes fluctuates and that the nodes with large fluctuations are 7, 9, 11, 13, and 16, which are involved in the reactor, separator, and stripper, respectively, but these results do not specify the cause of fault 4. Therefore, it is not necessary to directly locate the fault source when comparing only the correlation between fault 4 and normal conditions. In this paper, the relative proportions of the correlations of each node in faulty and normal conditions are considered to represent the impact on the correlations for each node in fault 4, and then, the variance is measured to determine the most likely source of fault 4. The relative proportions of the correlation are shown in Figure 6 in detail.

Figure 6 indicates that the variance of the correlation corresponding to node 9 in the case of fault 4 compared with that corresponding to normal conditions is the largest; in other words, the correlation of topological properties throughout the network fluctuates most at node 9 due to the occurrence of fault 4. For the simulation model under fault 4, the data for variable 9 undergo a step jump, shown in Figure 7. The method used here effectively detects the node where fault 4 occurred, that is, node 9. In addition, the partial-correlation method can also detect the faulty node.[15]

Example 2: fault 11

This example considers fault 11, in which the reactor cooling water inlet temperature changes and the fault type is random and the corresponding system data are collected.

A multi-layer network with 22 single-layer networks is obtained, corresponding to the system data under fault 11. At the same time, the degree distribution and the corresponding inter-layer mutual information can be obtained. Finally, a complex network and the overall correlation matrix A_{11} are determined. The comparison of the variances of the correlations between each node and the other 21 variable nodes under no fault and under fault 11 is shown in Figure 8.

It can be seen from Figure 8 that the correlations between each node and the other 21 nodes undergo significant fluctuations and that the most obvious nodes are 7, 11, and 13, which are involved in the reactor and separator. However, the cause of the fault cannot be uniquely determined. Further, according to the method of this paper, the ratio of the correlations in fault 11 to the correlations in normal conditions is determined, and then, the variance is utilized

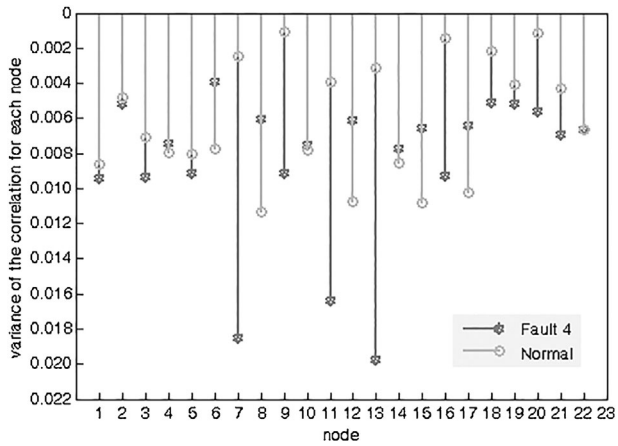

Figure 5. Comparison of the variance of the correlations between each node and the other 21 nodes.

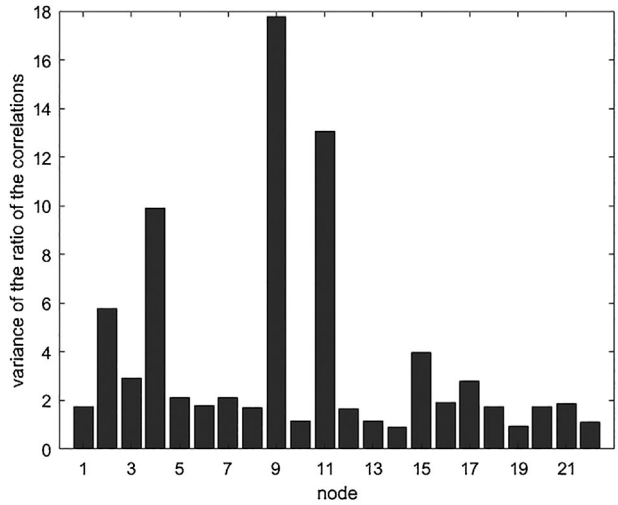

Figure 6. Variance of the ratio of the correlations in fault 4 to the correlations in normal conditions.

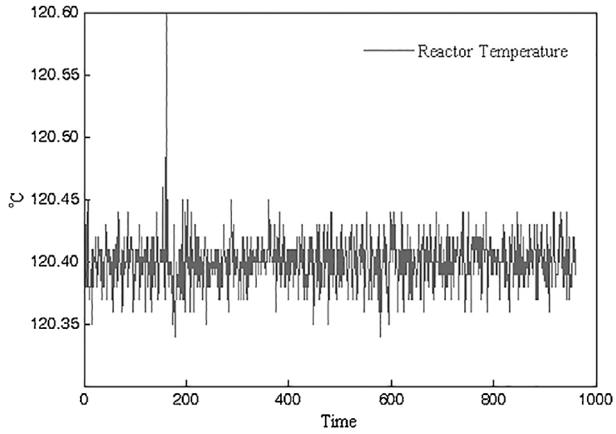

Figure 7. Data for variable 9 under fault 4.

to determine the most likely fault source. The details are shown in Figure 9.

As shown in Figure 9, the variance of node 9 is the largest after processing the correlation ratio between fault 11 and normal conditions, which means that the correlation corresponding to node 9 for the topological properties in the complex network fluctuates the most. For the simulation model under fault 11, the data for variable 9 undergo a random change. As shown in Figure 10, this method can effectively detect node 9. In addition, the partial-correlation method can also detect the responsible node.[15]

Example 3: fault 14

Fault 14, a viscous-type fault involving the reactor cooling water valve, is selected in this example and the corresponding data are obtained.

Through the proposed method, the structures of the 22 single-layer networks under fault 14 and the corresponding degree distribution can be obtained, and the inter-layer mutual information between the 22 single-layer networks can be obtained according to the inter-layer degree correlations. Ultimately, the whole complex network can be determined. Meanwhile, the correlation matrix A_{14} can be obtained. The difference in variance between the correlations is described in Figure 11.

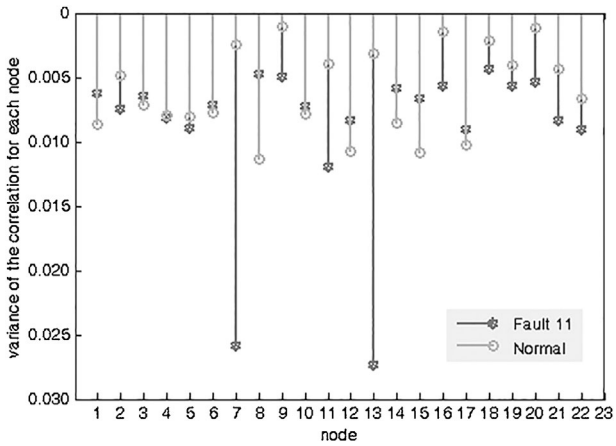

Figure 8. Comparison of the variance of the correlations between each node and the other 21 nodes.

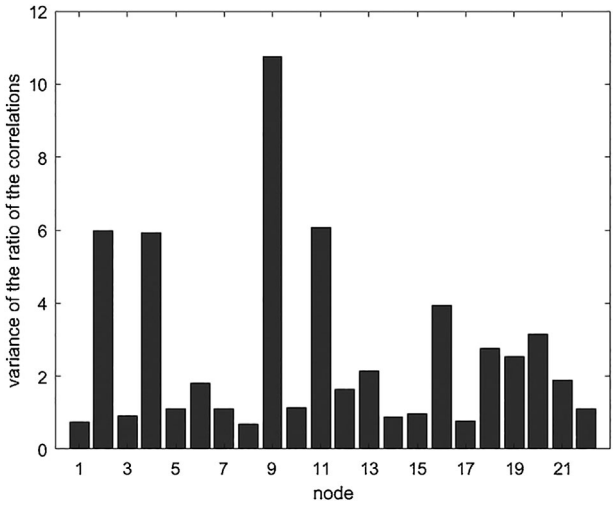

Figure 9. Variance of the ratio of the correlations in fault 11 to the correlations in normal conditions.

From Figure 11, there are significant fluctuations in the correlations of every node when the system is in fault 14 and the most obvious nodes are 7, 9, 11, 13, 16, 20, and 21, which are involved in the reactor, separator, compressor, and stripper. Therefore, this approach is useless for determining the cause of fault 14. Further, the variance of the ratio of the correlations in fault 14 to the correlations in normal conditions is shown in Figure 12.

As seen from Figure 12, the variance corresponding to node 9 is the most obvious. Therefore, within the topological nature of the overall complex network, the correlation of node 9 fluctuates the most. In the TE simulation model under fault 14, the data for variable 9 change accordingly, as shown in Figure 13. Thus, this method effectively detects node 9 as the faulty node under fault 14.

However, the partial-correlation method does not detect node 9 correctly under fault 14.[15] Figure 14 shows the degree distribution of each node under fault 14 from the partial-correlation method. The degrees of nodes 1, 7, 9, 10, 11, 20, 21, and 22 are changed, but the changes are not obvious. In addition, there are no prominent changes in these nodes, so the cause of fault 14 cannot be uniquely determined.

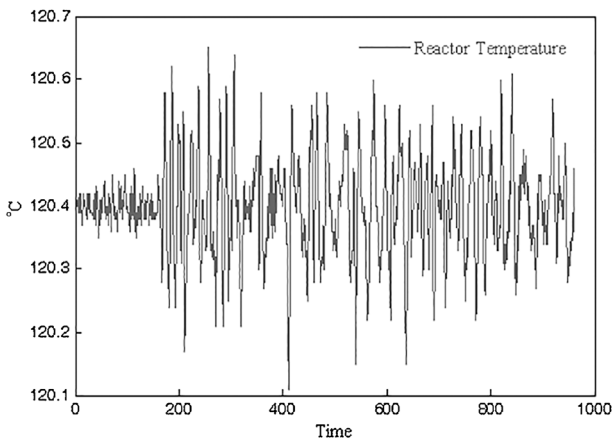

Figure 10. Data for variable 9 under fault 11.

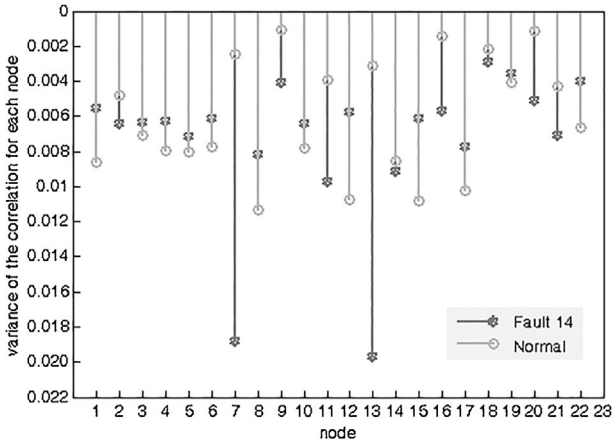

Figure 11. Variance in the correlations between each node and the other 21 nodes.

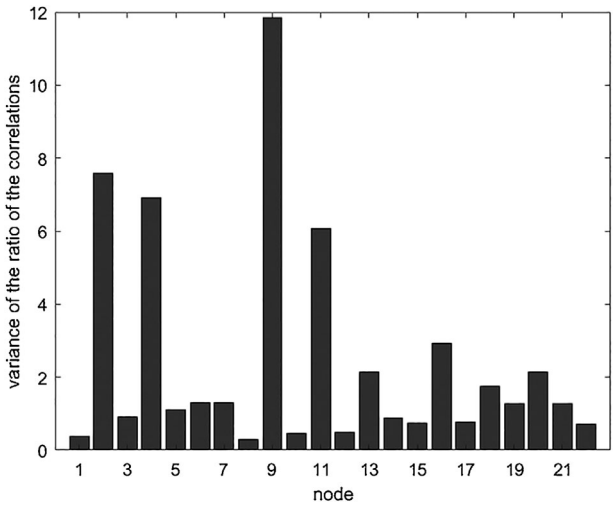

Figure 12. Variance of the ratio of the correlations in fault 14 to the correlations in normal conditions.

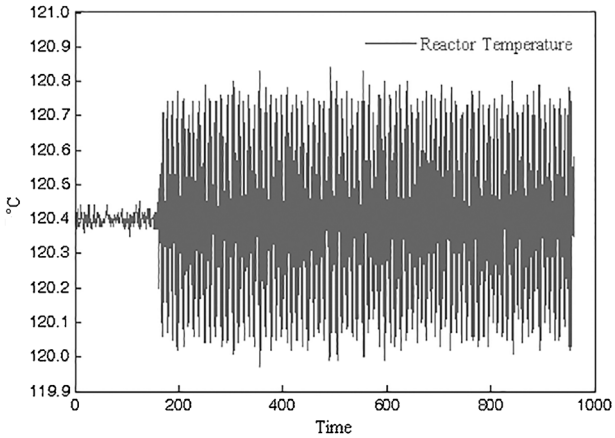

Figure 13. Data for variable 9 under fault 14.

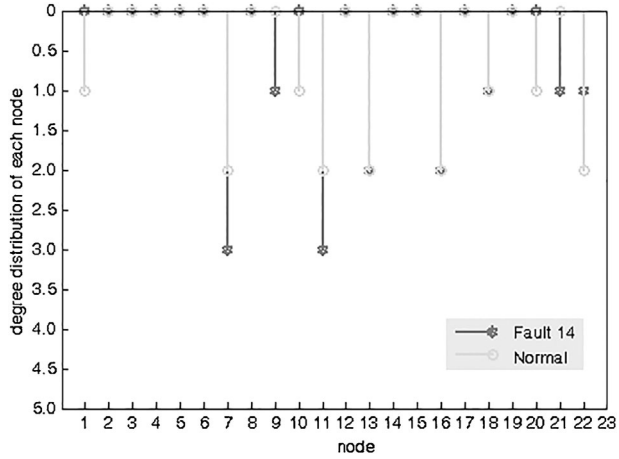

Figure 14. Contrast in node degree distribution.

DISCUSSION

First, this paper proposes an integrated HVG-complex network method to carry out the fault detection of complex chemical processes. The proposed method can not only study the network topology properties of the system integrally but also analyze the internal network structure of all the device variables in the system. Meanwhile, the experimental results show that the proposed method can describe and analyze the chemical system effectively.

Second, the proposed method is applied to detect faults in the TE process, which is representative of chemical systems. In the described experiments, this paper analyzes the data of the TE process structurally from the perspective of complex networks and studies the structural properties of the data to detect faults. From the results shown, the proposed method can effectively detect faults. Meanwhile, compared with the partial-correlation method, the proposed method has an advantage when considering viscous faults.

Third, the proposed method can mine the data in greater detail, so it is more sensitive to changes in the data. However, this method is still not perfectly comprehensive because the proposed method can only address single faults and was only applied to the TE process. Therefore, this method will be applied to more industrial systems in the future. Meanwhile, the fault detection of multi-source faults will be studied to further improve and perfect the method proposed in this paper.

CONCLUSIONS

This paper proposes a fault detection method via the utilization of HVGs in the perspective of complex networks for industrial processes. The proposed method views the data for each device variable in the chemical system as a time series and then maps them into single-layer networks corresponding to each device variable in the system. Moreover, the single-layer network is abstracted into nodes after analyzing the network structure and characteristics of the device variables themselves based on complex network theory. The proposed method takes only the collected data into account, so it ignores the physical mechanisms of the industrial system and avoids deviations due to the complexity of acquiring more comprehensive knowledge and the lack of prior knowledge. Meanwhile, this method analyzes the network structure of each variable in the system and can examine the characteristics of the data in greater detail. Finally, the

experimental results demonstrate that the proposed method can analyze a complex chemical system and detect step, random, and viscous faults effectively. Furthermore, compared with the partial-correlation method, the superiority of the proposed method is verified by experiments considering the TE process.

In future studies, the universality of the method will be validated by analyzing the faults of more complex chemical systems. Moreover, the method will be improved for application to the study of multi-fault detection, and then, the results will be compared with the existing multi-fault detection method to contrast the performance.

ACKNOWLEDGEMENTS

This research was partly funded by the National Natural Science Foundation of China (61673046, 61603025), Natural Science Foundation of Beijing, China (4162045), and National Key Technology Support Program (2015BAK36B04).

REFERENCES

[1] M. Kano, Y. Nakagawa, *Comput. Chem. Eng.* **2008**, *32*, 12.
[2] Y. Shen, S. X. Ding, A. Haghani, H. Hao, P. Zhang, *J. Process Contr.* **2012**, *22*, 1567.
[3] G. Li, S. J. Qin, T. Yuan, *Chemometr. Intell. Lab.* **2016**, *159*, 1.
[4] M. F. S. V. D'Angelo, R. M. Palhares, M. C. O. C. Filho, R. D. Maia, J. B. Mendes, P. Y. Ekel, *Appl. Soft Comput.* **2016**, *49*, 676.
[5] Z. Zhang, F. Dong, *Chemometr. Intell. Lab.* **2014**, *138*, 30.
[6] D. Peng, Z. Geng, Q. Zhu, *Ind. Eng. Chem. Res.* **2014**, *53*, 9792.
[7] Y. Qu, Z. M. Li, E. C. Li, *ISA T.* **2012**, *51*, 786.
[8] S. S. Tayarani-Bathaie, Z. N. S. Vanini, K. Khorasani, *Neurocomputing* **2014**, *125*, 153.
[9] S. Cheng, H. Yang, B. Jiang, *Neurocomputing* **2016**, *216*, 797.
[10] H. Tan, M. Peng, *Appl. Math. Comput.* **2012**, *219*, 408.
[11] J. Fan, Y. Wang, *Inform. Sciences* **2014**, *259*, 369.
[12] J. Macgregor, A. Cinar, *Comput. Chem. Eng.* **2012**, *47*, 111.
[13] E. R. Ziegel, *Technometrics* **1992**, *3*, 371.
[14] L. Lacasa, B. Luque, F. Ballesteros, J. Luque, J. C. Nuno, *P. Natl. Acad. Sci. USA* **2008**, *105*, 4972.
[15] Y. Chen, Y. M. Han, Z. Wang, Z. Q. Geng, *Journal of Chemical Industry and Engineering (China)* **2014**, *65*, 4503.
[16] N. Chatti, R. Guyonneau, L. Hardouin, S. Verron, S. Lagrange, *Eng. Appl. Artif. Intel.* **2016**, *55*, 307.
[17] Y. Cui, J. Shi, Z. Wang, *Reliab. Eng. Syst. Safe.* **2015**, *133*, 192.
[18] Z. Sun, Z. Yang, *IFAC Proceedings Volumes* **2014**, *47*, 8293.
[19] I. Yélamos, G. Escudero, M. Graells, L. Puigjaner, *Comput.-Aided Chem. En.* **2007**, *24*, 1253.
[20] I. Yélamos, G. Escudero, M. Graells, L. Puigjaner, *Comput.-Aided Chem. En.* **2006**, *21*, 1209.
[21] H. Jiang, J. J. Zhang, W. Gao, Z. Wu, *IEEE T. Smart Grid* **2014**, *5*, 2947.
[22] H. Wu, G. Gao, *Science Mosaic* **2010**, *6*, 19.
[23] Q. Jia, Y. Zhang, Y. Guan, X. Chen, *Inform. Control* **2012**, *41*, 356.
[24] Q. Zhang, *IFAC PapersOnline* **2015**, *48*, 150.
[25] M. Bonvini, M. D. Sohn, J. Granderson, M. Wetter, M. A. Piette, *Appl. Energ.* **2014**, *124*, 156.
[26] R. Sun, S. Liu, Y. Zhang, *Control and Decision* **2014**, *29*, 506.
[27] D. Peng, X. Gu, Y. Xu, Q. Zhu, *J. Loss Prevent. Proc.* **2015**, *33*, 279.
[28] Y. T. Bian, Z. H. Li, *Electronic Design Engineering* **2013**, *16*, 83.
[29] Z. L. Yang, B. Wang, X. H. Dong, H. Liu, *Syst. Eng. Proc.* **2012**, *4*, 189.
[30] S. Gajjar, A. Palazoglu, *Chemometr. Intell. Lab.* **2016**, *154*, 122.
[31] S. Stubbs, J. Zhang, J. Morris, *Comput. Chem. Eng.* **2012**, *41*, 77.
[32] Z. B. Zhu, Z. H. Song, *Expert Syst. Appl.* **2011**, *38*, 6895.
[33] G. Lee, S. O. Song, E. S. Yoon, *Ind. Eng. Chem. Res.* **2003**, *42*, 6145.
[34] J. Chen, Z. Li, J. Pan, G. Chen, Y. Zi, J. Yuan, B. Chen, Z. He, *Mech. Syst. Signal Pr.* **2016**, *70*, 1.
[35] D. Yu, Y. Yang, J. Cheng, *Measurement* **2007**, *40*, 823.
[36] R. Ata, *Renew. Sust. Energ. Rev.* **2015**, *49*, 534.
[37] J. Liu, S. J. Liu, D. S. H. Wong, *Comput.-Aided Chem. En.* **2013**, *32*, 751.
[38] S. Y. Chang, C. T. Chang, *Chem. Eng. Sci.* **2003**, *58*, 3395.
[39] R. L. Mason, *Technometrics* **1999**, *41*, 168.
[40] H. Du, *Journal of Mechanical Engineering* **2010**, *46*, 90.
[41] M. Peng, J. Wang, C. K. Tse, M. Shen, *Int. J. Bifurcat. Chaos* **2011**, *21*, 1323.
[42] Z. G. Shao, *Appl. Phys. Lett.* **2010**, *96*, 073703.
[43] G. Hloupis, *Commun. Nonlinear Sci.* **2017**, *51*, 13.
[44] V. Fioriti, A. D. Pietro, *AIP Conf. Proc.* **2012**, *21*, 3.
[45] Y. Zhang, H. Wu, L. Cheng, *Proceedings of the International Conference on Modelling, Identification & Control IEEE* **2012**, 987.
[46] R. A. Fisher, *T. RSE. Earth* **1919**, *52*, 399.
[47] J. J. Downs, E. F. Vogel, *Comput. Chem. Eng.* **1993**, *17*, 245.
[48] R. D. Braatz, "Braatz Group," *Massachusetts Institute of Technology: Department of Chemical Engineering*, **2017**, accessed on 1 May 2017, http://web.mit.edu/braatzgroup/links.html.

Manuscript received August 21, 2017; revised manuscript received May 10, 2018; accepted for publication May 28, 2018.