

## RESEARCH ARTICLE

# A Modeling Framework of Dynamic Risk Monitoring for Chemical Processes Based on Complex Networks

QIANLIN WANG<sup>1</sup>, JIAQI HAN<sup>1</sup>, FENG CHEN<sup>2</sup>, FENG WANG<sup>1</sup>, ZHAN DOU<sup>1</sup>, AND GUOAN YANG<sup>1</sup>

<sup>1</sup>College of Mechanical and Electrical Engineering, Beijing University of Chemical Technology, Beijing 100029, China

<sup>2</sup>College of Mechanical and Transportation Engineering, China University of Petroleum, Beijing 102249, China

Corresponding authors: Feng Chen (chenfeng@cup.edu.cn) and Guoan Yang (yangga@buct.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 52004014, and in part by the National Key Research and Development Program of China under Grant 2021YFB3301100.

**ABSTRACT** To ensure the stable and safe operations, this paper presents a modeling framework of dynamic risk monitoring for chemical processes. Multi-source process data are firstly denoised by the Wavelet Transform (WT). The Spearman's rank correlation coefficient (SRCC) of these data is calculated based on an appropriate time step and time window. An optimal correlation threshold is further applied to transform the SRCC matrix into an adjacency matrix. Accordingly, the model of complex networks (CNs) can be established for characterizing massive, disordered, and nonlinear process data. Network structure entropy is particularly introduced to transform process data into a single time series of relative risk. To illustrate its validity, a diesel hydrofining unit and Tennessee Eastman Process (TEP) are selected as test cases. Results show that the proposed modeling framework can effectively and reasonably monitor the risks of chemical processes in real time.

**INDEX TERMS** Chemical process, risk monitoring, complex network (CN), Spearman's rank correlation coefficient (SRCC), network structure entropy.

## I. INTRODUCTION

In recent years, the transformation and upgrading have been vigorously promoted for global chemical industries [1]. A trend of refining and chemical integration has become increasingly prominent to enrich product structure, enhance added value, reduce energy consumption, and save production cost. A great number of giant chemical parks are gradually established in the Gulf of Mexico (US), the Gulf of Tokyo (Japan), and the Jurong Island (Singapore). As an example of the Gulf of Mexico, the capacities of refining and ethylene are 460 million tons / year and 27 million tons / year, accounting for 52% and 95% of the total production capacity in the US, respectively [2]. With a deep integration of refining and chemical industries, production units are becoming more and more large, production processes are becoming more and

more integrated, as well as production systems are becoming more and more intensive. Thousands of monitoring points need to be set up to collect process data (e.g., temperature, pressure, flow, and liquid level) over the whole operation of chemical units [3]. For example, there are more than 1000 monitoring points in a catalytic cracking unit with the production capacity of 1 million tons / year so the process data counts in one day will be up to 860 million [4]. As a result, the distributed control system (DCS), supervisory control and data acquisition (SCADA), advanced control system (APC), industrial TV, or other monitoring systems had to display and store these massive, disordered, and nonlinear process data [5].

Risk monitoring is a common technique used to accurately and quickly master the operating condition of chemical processes. It can provide alarm information of high-probability or high-consequence incidents caused by process deviations. Generally, there are five different kinds of risk

The associate editor coordinating the review of this manuscript and approving it for publication was Chao Tong<sup>1</sup>.

monitoring methods. ① The risk monitoring method based on analytical model refers to converting the process model into state space form and comparing it with the actual operating status. If a residual is generated and its size exceeds the set threshold, the process risk can be monitored [6]. For example, Prof. J Rawlings's Group applied the state stability of stochastic input to verify the feasibility of nonlinear stochastic model of predictive control systems in process monitoring [7]. Although the analytical model-based methods are widely used and relatively mature, they require a large amount of prior knowledge and a large amount of work to obtain abnormal data and establish risk monitoring models [8], [9]. ② The risk monitoring method based on mathematical statistics refers to extracting important characteristic information of data and constructing statistical indicators to measure data attributes. This kind of method can distinguish and monitor abnormal conditions in chemical processes [10]. For example, Schaeffer and Braatz's Group developed and designed a software for multi-source data analysis of chemical processes based on principal component analysis and partial least squares [11]. González and Zavala's Group presented a new paradigm in Bayesian optimization for the analysis and monitoring of chemical process data, which allowed Bayesian optimization to effectively use composite functions [12]. Although the mathematical statistics-based methods can be combined with other optimization algorithms to achieve process risk monitoring by annotating data, they only consider the non-causality of the data and have certain application limitations [13]. ③ The risk monitoring method based on signal processing refers to spectrally transforming vibration or leakage signals and extracting risk characteristics in the frequency domain [14]. However, this kind of method is only suitable for risk monitoring and diagnosis of moving equipment in the chemical processes, and does not have the ability to monitor an entire process [15]. ④ The risk monitoring method based on process knowledge refers to establishing topology networks of chemical processes from the perspective of mechanism, and inputting data matrix to monitor the processes [16]. For example, Castaldello et al. input process data into a topological network and used optimization algorithms to mine optimal process paths, which could provide a basis for process risk monitoring [17]. However, this kind of method requires a large amount of prior knowledge, and has a limited application scope. ⑤ The risk monitoring method based on artificial intelligence refers to using the powerful nonlinear fitting ability of neural networks to extract data features to realize risk monitoring in chemical processes [18]. For example, Jiang and Yan proposed a regularized deep correlated representation method incorporating deep belief networks and canonical correlation analysis for nonlinear process monitoring [19]. However, the artificial intelligence-based methods require a large amount of labeled data, which is very complicated [20]. Although transfer learning can solve the data annotation problem of artificial intelligence algorithms to some extent, the engineering application has yet to be verified [21].

In fact, multi-source process data are coupled with each other. The process data are inevitably changed due to a unit failure or an external disturbance during the chemical processes. When a failure or disturbance occurs, chemical process risks will propagate through the material flow and information flow paths [22]. As similar as a domino effect, upstream parameter changes will propagate to the downstream in accordance with a causal relationship. Hence it is important to accurately deal with correlation rules hidden in multi-source process data, which will be conducive to quickly analyze the root causes of process risk and cut off its propagation paths [23]. However, those existing methods are not competent to efficiently deal with multi-source process data and reasonably express the correlation between them, which will easily result in serious data redundancy and information loss.

At end of the 20th century, Watts and Strogatz [24] as well as Barabasi and Albert [25] published papers in *Nature* and *Science* associated with the complex networks (CNs), respectively. They found that all the CNs had a common topological statistical property – small-world effect. It was the first time to introduce the idea of physical statistics into the graph theory. *Nature Physics* focused on the CNs once again in 2012 and Barabási clearly pointed out that “Data-based mathematical models of complex systems are offering a fresh perspective, rapidly developing into a new discipline: network science” [26]. Generally, the CNs are abstracted from complex systems in the real world. Scientific research shows that different kinds of objects can be transformed into nodes without shape and size [27]. If there are correlations within these objects, the corresponding nodes will be connected by lines, which are regarded as edges. Hence the CNs are competent to truly represent the characteristics of complex systems. In addition, the CNs have topological characteristics. The nodes which are independent of sizes, shapes, and positions only represent the studied objects. Similarly, the edges which are independent of their lengths, widths, and shapes only represent whether there are correlations between nodes or not. Nowadays, the CNs are becoming a research hotspot over the whole academic circles. They have been widely applied in many fields, such as power engineering [28], [29], transportation engineering [30], [31], and infrastructure engineering [32], [33], and so on.

Therefore, this paper aims to present a modeling framework of dynamic risk monitoring based on the CNs. During the whole period of chemical processes, the CNs can be applied to accurately describe the coupling characteristic of multi-source process data, as well as reasonably reveal the interaction relationship of energy, material, and information. This paper only focuses on multi-source process data from an aspect of time dimension, so the chemical processes are abstracted as undirected unweighted networks (UUNs) and the monitoring points are defined as network nodes. As shown below, there are two major technological contributions in this study.

First, how to establish a network model? The correlation coefficient is a common method in the modeling of CNs.

It can be used to verify whether there are correlations between multi-source process data or not. Particularly, this method not only avoids an artificial division, but also reduces human errors resulted from a large amount of process data. Furthermore, three typical correlation coefficients are generally applied in the existing studies – Partial correlation coefficient (ParCC) [34], Pearson correlation coefficient (PeaCC) [35], and Spearman's rank correlation coefficient (SRCC) [36]. The ParCC is mainly used to analyze the influence relationship between multiple variables. Because chemical process data is less affected by noise, the overall trend of the data is relatively stable under normal conditions. There is no need to use the ParCC in this paper, which can easily cause computational burden. The PeaCC measures the linear relationship between pairs of variables. Chemical process data has strong nonlinear characteristics, so the PeaCC is not applicable. The SRCC is often regarded as a nonparametric measure of rank correlation due to statistical dependence of ranking between two variables. It robustly captures the outliers and nonlinear relationships, and measures monotonic dependence relationship between two variables. Compared with the ParCC and PeaCC, the SRCC can solve nonlinear problems and have a wider application [37], [38]. The SRCC is integrated into CNs to reasonably express and efficiently deal with the nonlinear correlations of multi-source process data in the chemical processes. However, there is no unified standard for the selection of threshold interval, resulting in the effectiveness of correlation rules only depends on the expert experience; this paper will discuss it in detail.

Second, how to monitor the process risk? Existing studies show that the network structure entropy and its growth rate can effectively reveal network evolution rules and structure change characteristics under different time windows [39]. As a measure of disorder, the network structure entropy can make full use of global information of network nodes, thus it is transformed into influencing factor to reflect the trend of risk evolution. In this paper, network structure entropy is introduced to transform multi-source process data into a single time series of relative risk, which greatly reduce data dimension and accurately monitor dynamic risk during the chemical processes. However, the selection of time window of process data becomes a key problem to be solved due to the multi-dimensional, nonlinear, and dynamic characteristics; this paper will provide a solution.

The rest of the paper is organized as follows. Section II introduces a brief description of basic theories. Section III details a modeling framework of dynamic risk monitoring for chemical processes using these theories. Two case studies of a diesel hydrofining unit and Tennessee Eastman Process (TEP) are selected in Section IV. Finally, conclusions are made in Section V.

## II. PRELIMINARIES

### A. WAVELET TRANSFORM (WT)

The WT is a local transform between the time (or space) and frequency [40]. As expressed in Eq. (1), a non-orthogonal

wavelet function  $\psi(t)$  is used to develop stretching and translation operations to extract information from a signal  $f(t)$ , and the wavelet basis function  $\psi_{a,\tau}(t)$  is a cluster function generated by the stretching and translation of wavelet function.

$$\begin{aligned} \text{WT}(a, \tau) &= \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) \times \psi\left(\frac{t-\tau}{a}\right) \\ \psi_{a,\tau}(t) &= \frac{1}{\sqrt{a}} \psi\left(\frac{t-\tau}{a}\right) \end{aligned} \quad (1)$$

where  $a$  and  $\tau$  represent the frequency factor and time factor, which control the stretching and translation of wavelet function, respectively.

The basic idea of WT is to use wavelet basis function to represent a signal (i.e. data). There are five typical functions, including Haar (haar), Daubechies (dbN), Biorthogonal (bior Nr.Nd), Symlets (sym N), and Dmeyer (demy). Particularly, the index of signal-to-noise ratio  $SNR$  is selected to evaluate the applicability of different wavelet basis functions (see Eq. (2)). The higher the signal-to-noise ratio, the better the denoising effect is.

$$\begin{aligned} P_{\text{signal}} &= \frac{1}{n} \sum_{i=1}^n X_0^2 \\ P_{\text{noise}} &= \frac{1}{n} \sum_{j=1}^n (\tilde{X} - X_0)^2 \\ SNR &= 10 \times \lg\left(\frac{P_{\text{signal}}}{P_{\text{noise}}}\right) \end{aligned} \quad (2)$$

where  $P_{\text{signal}}$  and  $P_{\text{noise}}$  represent the energies of signal and noise, respectively;  $n$  represents the signal length;  $X_0$  and  $\tilde{X}$  represent the original signals and denoised signals, respectively.

### B. SPEARMAN'S RANK CORRELATION COEFFICIENT (SRCC)

Scale-independent measurement is important to the study of CNs [36]. The SRCC is a nonparametric measure of statistical dependence to assess the monotonic relation between two variables. Correlation analysis is a basis for establishing the CNs. The adjacency matrix can be obtained using a SRCC matrix and its threshold, and the connection relationship between each node in CNs can be further determined [41]. The SRCC is valued between  $-1$  and  $1$  [42]. If the value is approximate to  $-1$  or  $1$ , it indicates that the correlation between two variables is strong. If the value is approximate to  $0$ , it indicates that the correlation between two variables is weak. If the value is  $0$ , it indicates that there is no correlation between two variables. Moreover, if the value is positive (or negative), it indicates that two variables are positively (or negatively) correlated.

Suppose that there are two variables  $X$  and  $Y$ . Their element numbers are both  $n$ , and the  $i$ 'th ( $1 \leq i \leq n$ ) elements are  $X_i$  and  $Y_i$ , respectively. According to the ascending (or descending) order,  $X$  and  $Y$  are sorted to obtain two element

sets –  $x$  and  $y$ . The elements  $x_i$  and  $y_i$  are usually called as the ranks, and used to represent the position of  $X_i$  in the  $X$  and the position of  $Y_i$  in the  $Y$ , respectively. Furthermore, the elements of the sets  $x$  and  $y$  are correspondingly subtracted to obtain a sorted difference set – rank difference  $d_i = x_i - y_i$ . As expressed in Eq. (3), the SRCC can be calculated based on rank difference  $d_i$ .

$$\rho_i = 1 - \frac{6 \sum_{i=1}^m d_i^2}{n(n^2 - 1)} \quad (3)$$

### C. COMPLEX NETWORK (CN)

The CN is a statistical physics interpretation of graph theory [43]. It aims at describing and understanding the relationships between the elements of a complex system. Formally, a CN can be represented as a graph, where the entities in a complex system are abstracted as nodes and the relationships between entities are abstracted as connections [44], [45]. As shown below, this paper provides a basic review of important concepts in CNs [27].

#### 1) NETWORK

Generally, a network can be abstracted into  $G = (V, E)$ , which is composed of a node set  $V$  and an edge set  $E$ .  $e_{ij}$  represents the edge between two nodes  $i$  and  $j$ . According to the nodes and edges, the CNs can be divided into four types – the undirected unweighted networks (UUNs), the directed unweighted networks (DUNs), the undirected weighted networks (UWNs), and the directed weighted networks (DWNs). This paper only focuses on the UUNs, where the adjacent nodes have the same relationships ( $e_{ij} = e_{ji}$ ) and the edges have no weights.

#### 2) ADJACENCY MATRIX

The adjacency matrix is a common expression applied to describe a network in computer. If the adjacency matrix  $A = (a_{ij})_{N \times N}$  is a square matrix of order  $N$ , then an UUN  $G$  with  $N$  nodes can be defined using  $a_{ij}$ . In an adjacency matrix, the element “1” indicates that the nodes in its row and column have strong correlation, and there is an edge between the corresponding nodes; the element “0” indicates that the nodes in its row and column have weak correlation, and there is no edge between the corresponding nodes.

#### 3) DEGREE

The degree is one of the most simple and important concepts used to describe a single node. In the UUNs, the degree  $k_i$  of node  $i$  is defined as the number of edges which directly connect with this node. The average degree  $k$  can be further defined as the arithmetic mean of degrees over all nodes (see Eq. (4)).

$$k = \frac{1}{N} \sum_{i=1}^N k_i = \frac{1}{N} \sum_{i,j=1}^N a_{ij} \quad (4)$$

#### 4) PATH LENGTH

The path length is often applied to reflect the whole characteristics of CNs. In the UUNs, the path length  $d(i, j)$  between node  $i$  and node  $j$  is defined as the number of edges along the shortest path connecting these two nodes. The average path length  $L$  can be further defined as the arithmetic mean of path lengths between any node  $i$  and any node  $j$  (see Eq. (5)).

$$L = \frac{1}{\frac{1}{2}N(N-1)} \sum_{i \geq j}^N d(i, j) \quad (5)$$

#### 5) CLUSTERING COEFFICIENT

The clustering coefficient is a quantitative description for clustering degree of nodes used to reflect the partial characteristics of CNs. In the UUNs, the clustering coefficient  $C_i$  of node  $i$  is defined as the ratio between the number  $e_i$  of edges that actually exist between these  $k_i$  nodes and the total possible number  $k_i(k_i - 1)/2$ . The average clustering coefficient  $C$  can be further defined as the arithmetic mean of clustering coefficients over all nodes (see Eq. (6)).

$$C = \frac{1}{N} \sum_{i=1}^N C_i = \frac{1}{N} \sum_{i=1}^N \frac{2e_i}{k_i(k_i - 1)} \quad (6)$$

#### 6) NETWORK STRUCTURE ENTROPY

In the information theory, the Shannon entropy – as a measure of uncertainty in random events – can not only measure the chaos or disorder of system state, but also contains a large amount of system intrinsic information [46]. The definition of network structural entropy is based on the Shannon information entropy, which can be used to quantify the network’s topological structure complexity [47], [48]. Particularly, it considers the number of nodes in the CNs and their relationships, and abstracts the rich information contained in the CNs into specific numerical values [49]. Therefore, network structure entropy can be applied to analyze the disorder degree in chemical processes, and characterize its risk evolution characteristics and trends. There are three major definitions as follows.

As expressed in Eq. (7), the degree distribution entropy  $E$  is defined using the difference of the nodes in a network structure.

$$E = - \sum_{i=1}^N p(k_i) \log p(k_i) \quad (7)$$

where  $p(k_i)$  represents the probability distribution of the degree of node  $i$ .

As expressed in Eq. (8), the Wu structure entropy  $E$  is also defined using the difference of the nodes in a network structure.

$$I_i = \frac{k_i}{\sum_{i=1}^N k_i}$$

$$E = \sum_{i=1}^N I_i \ln I_i \quad (8)$$



where  $I_i$  represents the difference of node  $i$ .

As expressed in Eq. (9), the Cai structure entropy  $E$  not only considers the difference of the nodes in a network structure, but also integrates the connection difference over these nodes.

$$\begin{aligned} S_i &= (1 - p(k_i))N \\ D_i &= k_i(1 - p(k_i))N \\ I_i &= \alpha S_i + \beta D_i \\ E &= - \sum_{i=1}^N \frac{I_i}{\sum_{i=1}^N I_i} \log \frac{I_i}{\sum_{i=1}^N I_i} \end{aligned} \quad (9)$$

where  $S_i$  represents the difference of node  $i$ ;  $D_i$  represents the connection difference of node  $i$ ;  $I_i$  represents the comprehensive difference of node  $i$ ; the coefficients  $\alpha$  and  $\beta$  should meet the formula  $\alpha + \beta = 1$  and can be usually valued as  $\alpha + \beta = 0.5$ .

### 7) SMALL-WORLD EFFECT

The small-world effect is a typical characteristic used to describe the connectivity of CNs. Newman defined the small-world effect in the following way: “the fact that most pairs of vertices in most networks seem to be connected by a short path through the network” [50]. Budrikis coined that CNs sharing high average clustering coefficient and short average path length were one type of network displaying the small-world effect [51]. Particularly, the CNs can be considered as interpolating between one version of an ordered network and one version of a random network [25], [52]. Eq. (10) can be used to judge whether a CN has small-world effect or not.

$$\begin{aligned} C_{ord} &\approx \frac{k}{N} \\ L_{ran} &\approx \frac{\ln k}{\ln N} \\ \frac{C_{ord}}{C_{ran}} &> \frac{L_{ord}}{L_{ran}} \end{aligned} \quad (10)$$

where  $C_{ord}$  and  $L_{ord}$  represent the average clustering coefficient and average path length of an ordered network (namely the established CN), respectively;  $C_{ran}$  and  $L_{ran}$  represent the average clustering coefficient and average path length of the corresponding random network, respectively;  $N$  represents the number of all nodes in this random network;  $k$  represents the degree of this random network.

In addition, this paper defines a small-world index  $S$  using Eq. (11). A CN is provided with the small-world effect when  $S$  is greater than 1.

$$S = \frac{C_{ord}}{C_{ran}} \bigg/ \frac{L_{ord}}{L_{ran}} > 1 \quad (11)$$

## III. PROCEDURES

As shown in Fig. 1, a detailed modeling framework of dynamic risk monitoring is presented for chemical processes.

There are six main stages in this framework. First, multi-source process data should be collected from chemical processes in real time, such as temperature, pressure, flow, liquid level, and so on; then these dynamic data should be further processed and denoised based on the WT. Second, appropriate time step and time window must be selected for the modelling of CNs; specifically, the standard deviation of network structure entropy and average small-world index are applied to determine time window in this paper. Third, non-linear correlation analyses should be developed according to the SRCC; then correlation coefficient matrix can be established for these process data. Forth, the optimal correlation threshold must be determined in accordance with the following two principles – no isolated nodes and maximizing small-world indexes; then this threshold can be applied to transform correlation coefficient matrix into adjacency matrix; and then the CNs can be established and the small-world effect should be verified accordingly. Fifth, three typical network structure entropies should be calculated; specifically, the standard deviation is used to determine the optimal one in this paper. Sixth, network structure entropy should be normalized to transform the above-mentioned process data into a single time series of relative risk; as a result, dynamic risk monitoring and assessment can be effectively and reasonably realized over the whole period of chemical processes.

### A. DATA PREPROCESSING BASED ON WT

During the chemical processes, different kinds of variables (e.g., temperature, pressure, flow, and liquid level) are collected by sensors at the regular intervals; then they are converted into the electrical signals and transmitted to dynamic monitoring terminals or data storage devices (e.g., DCS, SCADA, APC, and Industrial TV). However, there is inevitably noise attributed by many factors, such as the aging of equipment, human errors, external disturbance, measurement errors of instrument or sensor, and so on [53], [54] and [55]. The noise has adverse effects on the algorithm (or model) to extract valuable information and deal with subsequent data. Therefore, it is important to filter the noise information contained in multi-source process data, which will be conducive to increase the signal-to-noise ratio and maximize the value of data.

First, multi-source process data should be decomposed by wavelet basis functions. Using Eq. (1), the optimal wavelet basis function is selected as that has the highest signal-to-noise ratio. Second, an appropriate wavelet threshold should be determined for the high-frequency coefficients after the decomposition. If the threshold is too small, a large amount of noise information will be retained, resulting in the unsuccessful denoising. On the contrary, if the threshold is too large, part of real information will be lost, resulting in the data distortion. Third, data reconstruction should be developed based on the low-frequency coefficients (after wavelet decomposition) and the high-frequency coefficients (after threshold quantization). Fourth, data normalization should be

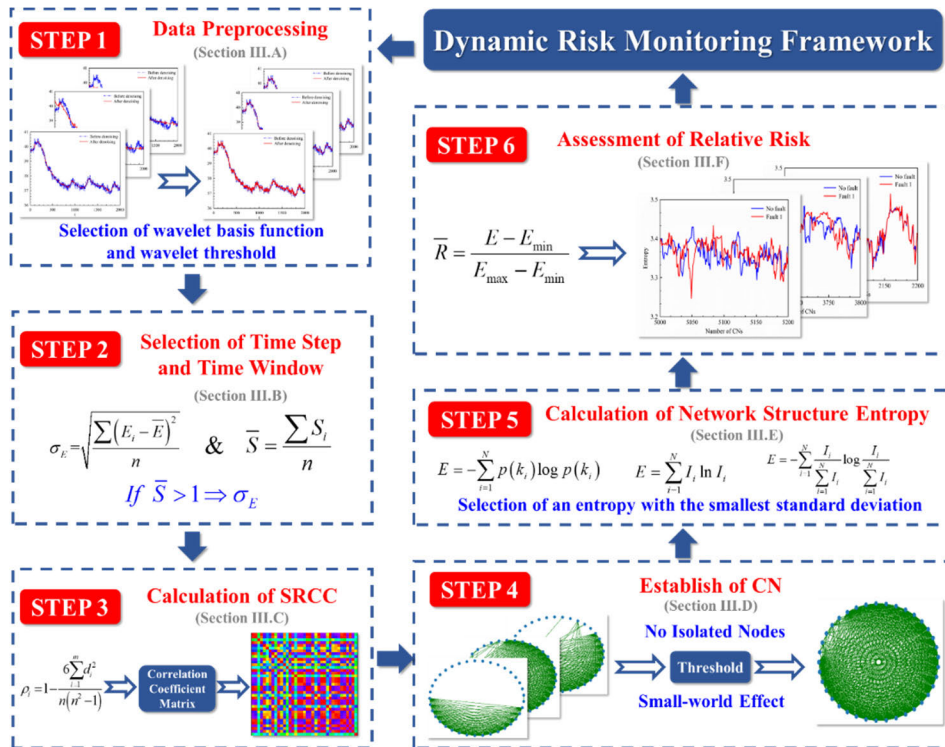


FIGURE 1. The modeling framework of dynamic risk monitoring.

applied to convert the value of process data into an interval between 0 and 1 (see Eq. (12)).

$$X' = \frac{\tilde{X} - X_{\min}}{X_{\max} - X_{\min}} \quad (12)$$

where  $X'$  represents the normalized process data;  $X_{\min}$  and  $X_{\max}$  represent the minimum and maximum values of process data, respectively.

## B. SELECTION OF TIME STEP AND TIME WINDOW

It is necessary to select appropriate time step and time window during the modelling of CNs. The selection of time step directly affects the performance of the proposed modeling framework in this paper. In accordance with chemical process data – a time series of data, the meaning of time step 1 refers to the length of a set of data, and the length of each set of data is related to its sampling time. When the time step is less than 1, it is easy to cause local optimality, reduce the generalization ability of the modeling framework, and increase the computing power [56]. As the time step increases, the modeling parameters may also jump too much on the loss function, causing the optimization process to be out of control and the modeling framework to be failed with convergence [57]. Hence this paper sets the time step as 1. Based on the time step of 1, a CN can be established when every group of data is postponed. In addition, this paper defines several groups of process data required to establish a CN as a cluster, and further defines the number of data groups in a cluster as

time window. As expressed in Eqs. (13) - (14), the standard deviation of network structure entropy  $\sigma_E$  and average small-world index  $\bar{S}$  are used to evaluate the applicability of time window, respectively.

$$\sigma_E = \sqrt{\frac{\sum (E_i - \bar{E})^2}{n}} \quad (13)$$

$$\bar{S} = \frac{\sum S_i}{n} \quad (14)$$

where  $E_i$  represents the network structure entropy corresponding to the  $i$ 'th cluster data;  $\bar{E}$  represents the average network structure entropy;  $n$  represents the number of established CNs;  $S_i$  represents the small-world index corresponding to the  $i$ 'th cluster data.

Standard deviations can quantitatively describe the data distribution and dispersion. The chemical process is stable under normal conditions, and the corresponding network structure entropy also fluctuates within a stable range [58], [59]. When a unit failure or an external disturbance occurs, the network structure entropies will change significantly and can be monitored as risk points. The greater the network structure entropy value, the higher the risk situation of chemical process is. Therefore, this paper selects the smallest standard deviation from the degree distribution entropy, Wu structure entropy, and Cai structure entropy to characterize the risk situation in CNs. In summary, the smaller the standard deviation of network structure entropy, the better the sampling effect is. Moreover, small-world indexes can quantitatively

describe the data relationship. The greater the small-world index, the closer the data relationship is and the better the sampling effect is. Particularly, two evaluation indexes –  $\sigma_E$  and  $\bar{S}$  should be normalized to obtain an optimal time window. This paper sets the time window as that has the smallest  $\sigma_E$  on the condition that  $\bar{S}$  is greater than 1.

### C. CALCULATION OF SRCC

Non-linear correlation analyses should be developed for the above-mentioned process data. In this paper, the SRCC of multi-source process data can be firstly calculated using Eq. (3). Then the correlation coefficient matrix can be further established. Lastly, this paper introduces the heat map and grey-scale map to verify whether these process data are correlated or not.

### D. ESTABLISHMENT OF CN

A correlation threshold should be determined to transform the correlation coefficient matrix into a Boolean matrix for the establishment of CNs. In this paper, the Boolean matrix with only 0 and 1 is regarded as an adjacency matrix. Note that all nodes will be adjacent to each other if the correlation coefficient matrix is directly applied to establish a CN; clearly, this supposition is not conducive to extracting the statistical characteristics of CNs, or even causes meaningless results. In sum, the correlation coefficient matrix need be modified to reduce redundant information. This paper defines that there will be strong correlation in the SRCC if its absolute value is greater than or equal to the correlation threshold; the SRCC can be transformed into 1 in this case, indicating that the corresponding nodes are connected. Similarly, this paper also defines that there will be weak correlation in the SRCC if its absolute value is less than the correlation threshold; the SRCC can be transformed into 0 in this case, indicating that the corresponding nodes are not connected. As shown below, there are two major principles in the selection of correlation threshold.

First, there are no isolated nodes in the CNs. The node degree decreases with increasing correlation threshold due to increasing isolated node number. This paper defines that the isolated node refers to a node without any adjacency nodes; namely the node degree is 0. If there are isolated nodes (also called as blind spots) in the CNs, some critical information related to chemical process risks will be lost. In addition, too many isolated nodes will cause the CNs to lose their rationality, which is not expected [60]. During the calculation process of correlation threshold, the elements with 1 on the diagonal of correlation coefficient matrix should be firstly eliminated; then the absolute values of SRCC should be also computed; lastly, the maximum values of corresponding elements in each row should be further calculated. Note that the correlation coefficient matrix is a symmetric matrix, so a sum of elements in each row is equivalent to a sum of elements in each column. In addition, the node degree associated with each cluster of multi-source process data will be not 0 and

there will be adjacency node on the condition that the correlation threshold is less than the maximum value of SRCC. As expressed in Eq. (15), the correlation threshold  $r'$  can be selected according to a minimax criterion.

$$r' = \min (\max (r_{i,j})) \quad (15)$$

where  $r_{i,j}$  represents the correlation coefficient between two elements  $i$  and  $j$ .

Second, the small-world indexes of CNs are maximized on the premise of the first principle. Even if some nodes or connections are destroyed, information can still spread in the CNs characterized by small-world effect, showing good robustness [45], [61]. Therefore, the CNs established in this paper is expected to have small-world effect to improve the robustness of the proposed modeling framework. This paper applies the enumeration method, where the correlation threshold is gradually reduced using the step of 0.01. The correlation threshold associated with the maximal small-world index is optimal. Note that if the SRCC is less than 0.05, the correlation is regarded as weak in this paper. Therefore, the minimum value of correlation threshold is defined as 0.05.

In accordance with these two principles, the optimal correlation threshold can be determined for multi-source process data. This threshold is further applied to transform correlation coefficient matrix into adjacency matrix. Accordingly, the establishment of CNs and the verification of small-world effect can be developed.

### E. CALCULATION OF NETWORK STRUCTURE ENTROPY

Three different types of network structure entropy – the degree distribution entropy, the Wu structure entropy, and the Cai structure entropy can be calculated using Eqs. (7) - (9), respectively. The standard deviation of network structure entropy  $\sigma_E$  (see Eq. (13)) is taken as an index to select the appropriate type in this paper. The smaller the standard deviation of network structure entropy, the better the reflecting effect of network evolution is.

### F. ASSESSMENT OF RELATIVE RISK

In this paper, the CNs established in different time periods can be regarded as risk dynamic evolutions over the chemical processes. Hence network structure entropies under different time windows are standardized by Max-Min dispersion and transformed as relative risks corresponding to the CNs. In similar to normalization processing, the relative risks can unify network structure entropies of different conditions into 0-1. Accordingly, the robustness and generalization capabilities of the proposed modeling framework can be improved. As expressed in Eq. (16), the normalization of network structure entropy  $E$  can be applied to transform multi-source process data into a single time series of relative risk  $\bar{R}$ ; as a result, dynamic risk monitoring and assessment can be realized during the chemical processes.

$$\bar{R} = \frac{E - E_{\min}}{E_{\max} - E_{\min}} \quad (16)$$

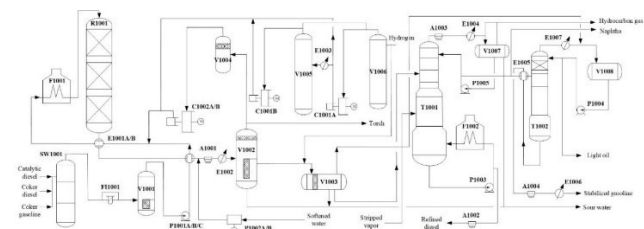
where  $E_{min}$  and  $E_{max}$  represent the minimum and maximum values in the network structure entropy sequence, respectively.

**IV. CASE STUDIES**

**A. APPLICATION OF DIESEL HYDROFINING UNIT**

1) BACKGROUND

Generally, the diesel hydrofining unit is used to remove sulfur, nitrogen, oxygen, aromatics, and other components contained in crude oil, which can improve product performance and reduce environmental pollution. However, this unit is always in the high-temperature and high-pressure conditions. In China, the diesel hydrofining is included into the first hazardous chemical processes under strict state supervision, and the transported materials belong to A in the hazardous chemical classification. An application of diesel hydrofining unit is developed to verify the effectiveness of the proposed modeling framework in this paper.



**FIGURE 2.** The PFD of a diesel hydrofining unit in one oil refinery.

As shown in Fig. 2, a diesel hydrofining unit of oil refinery mainly contains the reaction system and fractionation system. A total of 34 key devices (see Table 1) and 37 process variables (see Table 2) are involved in this unit.

2) PROCEDURES

*a: DATA PREPROCESSING BASED ON WT*

The first 2000 groups of process data associated with Tag 17 in this diesel hydrofining unit are selected as an example to compare the denoising effect of different wavelet basis functions. As shown in Fig. 3, this paper supposes the wavelet basis functions as Haar, bior3.1, sym2, dmey, db8, db16, db32, and db36, respectively. Here, the wavelet threshold is presupposed as 0.6.

The signal-to-noise ratio of every wavelet basis function can be computed (see Table 3). It is clear that the signal-to-noise ratio of db32 is the highest, and the corresponding wavelet curve is the most smoothing. Therefore, this paper selects the db32 as wavelet basis function because it has the best denoising effect.

On the basis of wavelet basis function – db32, the wavelet thresholds are supposed as 0.2, 0.4, 0.6, and 0.8 to compare the denoising effect, respectively. As shown in Fig. 4, the wavelet curve fluctuates greatly and the denoising effect is not obvious when wavelet threshold is 0.2 or 0.4; meanwhile, the wavelet curve is relatively flat when wavelet

**TABLE 1.** The key devices of a diesel hydrofining unit in one oil refinery.

No.	Item	Description
1	A1001	Air cooler of reactor effluent
2	A1002	Air cooler of diesel
3	A1003	Air cooler on top of fractionation tower
4	A1004	Air cooler of stabilized gasoline
5	C1001 A	Hydrogen make-up compressor
6	C1001 B	Hydrogen make-up compressor
7	C1002 A/B	Hydrogen recycle compressor
8	E1001 A/B	Heat exchanger of reactor effluent / reactor feed
9	E1002	Water cooler of reactor effluent
10	E1003	Inter stage cooler
11	E1004	Cooler on top of fractionation tower
12	E1005	Heat exchanger of naphtha / stabilized gasoline
13	E1006	Water cooler of stabilized gasoline
14	E1007	Cooler on top of stabilizer
15	F1001	Heating furnace of reactor effluent
16	F1002	Fractionation furnace
17	FI1001	Filter of crude oil
18	P1001 A/B/C	Charging pump
19	P1002 A/B	Charging pump
20	P1003	Pump of refined diesel
21	P1004	Reflux pump on top of stabilizer
22	P1005	Reflux pump on top of fractionation tower
23	R1001	Hydrofining reactor
24	SW1001	Dehydrator of crude oil
25	T1001	Fractionation tower
26	T1002	Stabilizer
27	V1001	Buffer tank of crude oil
28	V1002	High-pressure separator
29	V1003	Low-pressure separator
30	V1004	Inlet separator vessel of hydrogen recycle compressor
31	V1005	The 2nd inlet separator vessel of hydrogen make-up compressor
32	V1006	Inlet separator vessel of hydrogen make-up compressor
33	V1007	Reflux tank on top of fractionation tower
34	V1008	Reflux tank on top of stabilizer

threshold is 0.6 or 0.8. Moreover, it can be calculated that the signal-to-noise ratio 53.183 corresponding to wavelet threshold 0.6 is greater than 52.717 corresponding to wavelet threshold 0.8. Hence the wavelet threshold is set as 0.6 in this paper.

Note that the original process data may have invalid values. For example, all data are 0 at a certain time. Accordingly, the invalid values should be eliminated before process data denoising.

*b: SELECTION OF TIME STEP AND TIME WINDOW*

The time step is taken as 1 for this on-site diesel hydrofining unit. Accordingly, 50 CNs are established by selecting different time windows, including 50, 100, 150, 200, 250, and 300. Two indexes – standard deviation of network structure entropy and average small-world index are used to compare the applicability of time windows (see Table 4). The standard



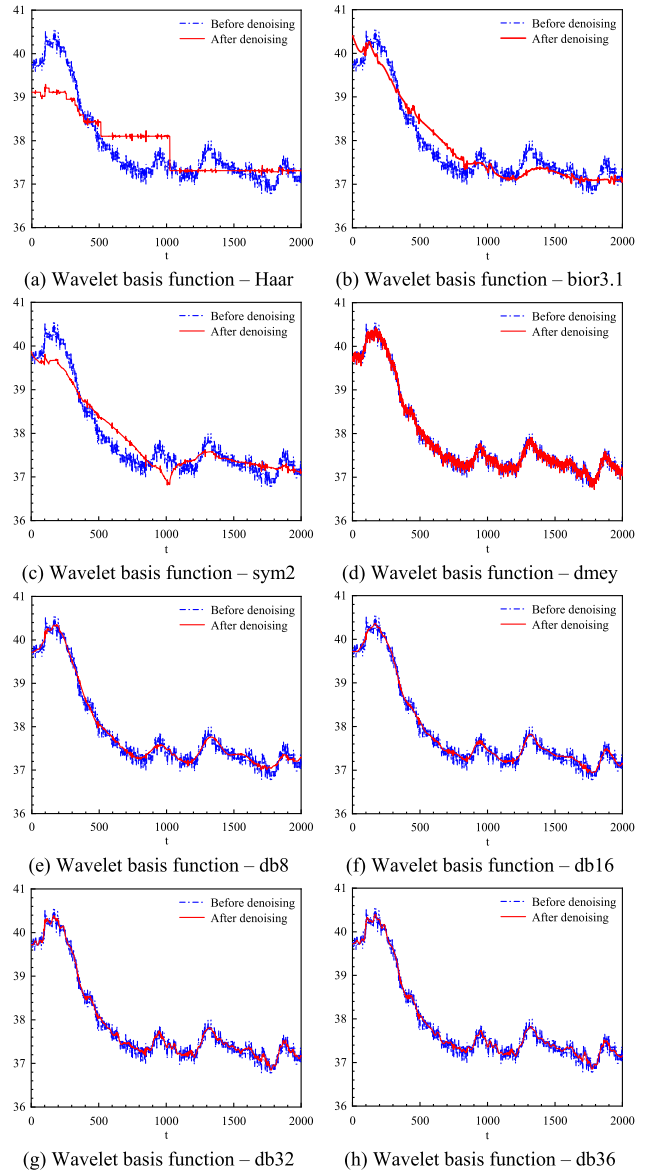
**TABLE 2.** The process variables of a diesel hydrofining unit in one oil refinery.

Variable	Description	Unit
Tag 1	Catalytic diesel flow from tank farm	kg h <sup>-1</sup>
Tag 2	Differential pressure of F11001	kPa
Tag 3	Liquid level of V1001	%
Tag 4	Outlet flow of P1001 A/B/C	t/h
Tag 5	Inlet temperature of F1001	°C
Tag 6	Bottom temperature of F1001	°C
Tag 7	Outlet temperature of F1001	°C
Tag 8	Fuel gas flow of F1001	°C
Tag 9	Hydrogen flow	Nm <sup>3</sup> h <sup>-1</sup>
Tag 10	Outlet pressure of V1005	MPa
Tag 11	Outlet pressure of V1006	MPa
Tag 12	Outlet pressure of V1007	MPa
Tag 13	Inlet temperature of R1001	°C
Tag 14	Differential pressure of R1001	kPa
Tag 15	Bed temperature 1 of R1001	°C
Tag 16	Bed temperature 2 of R1001	°C
Tag 17	Outlet temperature of E1002	°C
Tag 18	Top pressure f V1002	MPa
Tag 19	Liquid level of V1002	%
Tag 20	Top pressure of V1003	MPa
Tag 21	Liquid level of V1003	%
Tag 22	Outlet flow of V1003	kg h <sup>-1</sup>
Tag 23	Inlet temperature of T1001	°C
Tag 24	Top pressure of T1001	MPa
Tag 25	Outlet top temperature of T1001	°C
Tag 26	Liquid level of T1001	%
Tag 27	Liquid level of V1007	%
Tag 28	Top reflux flow of T1001	kg h <sup>-1</sup>
Tag 29	Firebox temperature of F1002	°C
Tag 30	Outlet temperature of F1002	°C
Tag 31	Fuel gas flow of F1002	Nm <sup>3</sup> h <sup>-1</sup>
Tag 32	Refined diesel flow	kg h <sup>-1</sup>
Tag 33	Stripped vapor flow of T1002	kg h <sup>-1</sup>
Tag 34	Feeding temperature of T1002	°C
Tag 35	Outlet top temperature of T1002	°C
Tag 36	Top pressure f T1002	MPa
Tag 37	Liquid level of T1002	%

**TABLE 3.** A comparison of the signal-to-noise ratio results from different wavelet basis functions.

No.	Wavelet basis function	SNR	No.	Wavelet basis function	SNR
1	Haar	37.372	5	db8	50.724
2	bior3.1	42.461	6	db16	51.990
3	sym2	41.662	7	db32	53.183
4	dmey	50.810	8	db36	52.913

deviation of network structure entropy is the smallest when time window is 100; meanwhile, the average small-world index is the largest when time window is 50. According to a comprehensive comparison of these two indexes, the time window is finally selected as 100. In other words, the sampling effect is the best when each cluster contains 100 groups of process data.



**FIGURE 3.** A comparison of the denoising results from different wavelet basis functions.

**TABLE 4.** A comparison of the CN modelling results from different time windows.

No.	Time window	Standard deviation of network structure entropy	Average small-world index
1	50	0.024	1.088
2	100	0.014	1.059
3	150	0.018	1.047
4	200	0.016	1.038
5	250	0.031	1.021
6	300	0.021	1.007

*c: CALCULATION OF SRCC*

The first 100 groups of data of 37 process variables are particularly applied after denoising by wavelet transform. Results from the SRCC are shown in Appendix A.

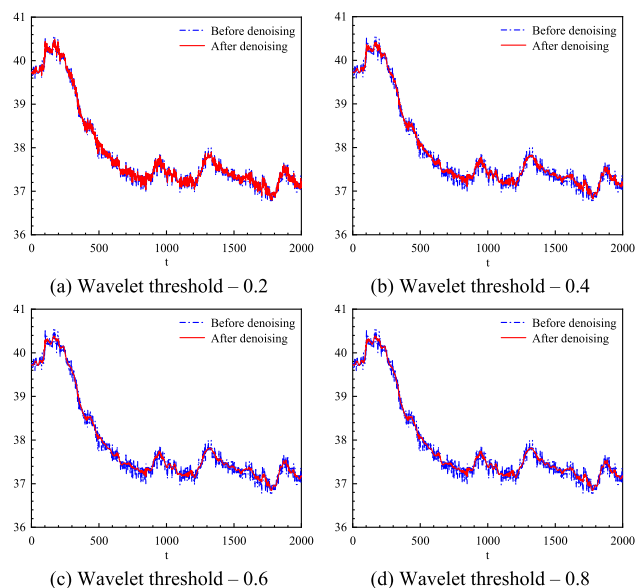


FIGURE 4. A comparison of the denoising results from different wavelet thresholds.

A SRCC matrix can be visually converted into heat map and gray-scale map, as shown in Fig. 5. The diagonal data from top left to bottom right are 1, so the above-mentioned process data are positively correlated with themselves. Moreover, other coefficients take the diagonal as symmetry axis and show a symmetrical distribution law.

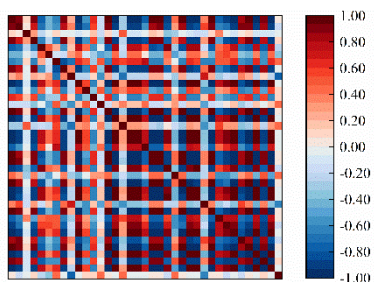


FIGURE 5. The correlation analyses based on heat map and grey-scale map.

*d: ESTABLISHMENT OF CN*

A correlation threshold should be determined to transform the SRCC matrix into adjacency matrix. In this paper, the correlation thresholds are set as 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9 respectively. Accordingly, 10 CNs can be established (see Fig. 6) and the statistical network characteristics can be also computed (see Table 5).

All the established complex networks contain 37 nodes, which represent 37 process variables. Particularly, the isolated nodes begin to appear when correlation threshold is greater than or equal to 0.5. The larger the correlation threshold is, the more the number of isolated nodes is. In other words, the number of isolated nodes is positively correlated with correlation threshold. Moreover, the maximum

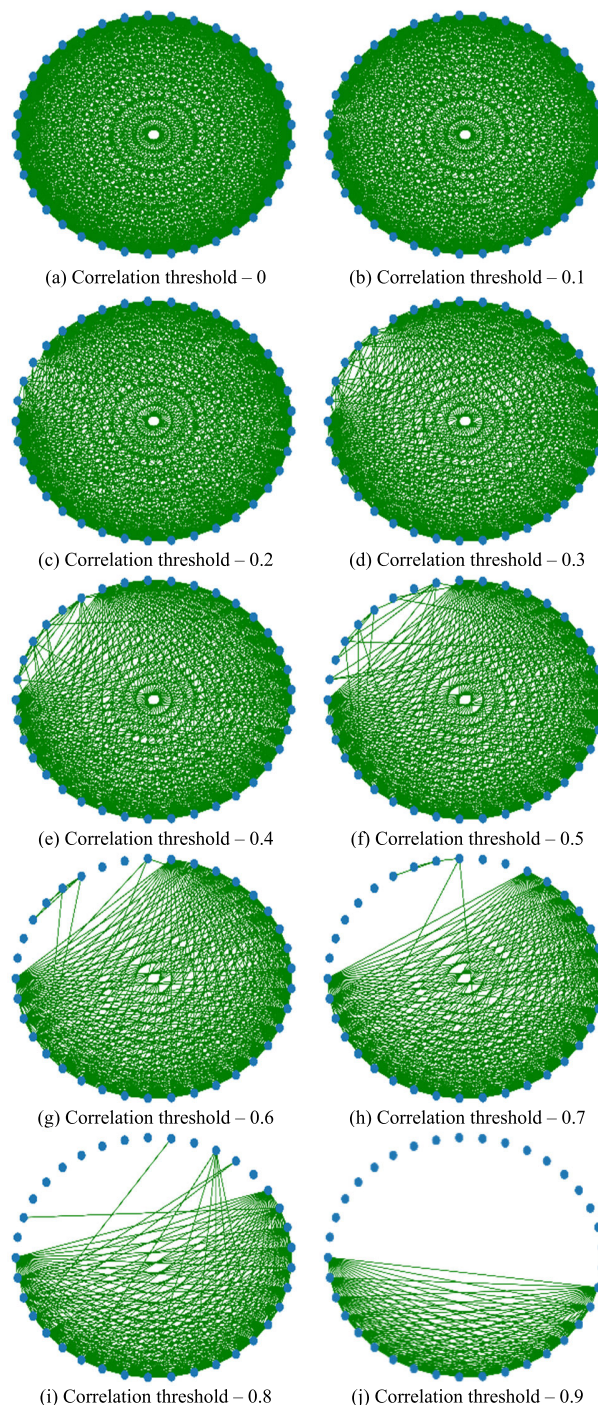


FIGURE 6. A comparison of the CN modelling results from different correlation thresholds.

correlation coefficients between 37 process variables are also computed to analyze isolated nodes (see Table 6). The minimum value between these maximum correlation coefficients is 0.490 – corresponding to Tag 3. Hence there will be no isolated nodes in the CNs as long as the correlation threshold is less than 0.490.

According to the above-mentioned results, the range of correlation threshold is 0.05~0.48. The corresponding

**TABLE 5. A comparison of the statistical characteristic results from different CNs.**

No.	Correlation threshold	Average degree	Average Path length	Average Clustering coefficient	Small-world index	Number of isolate nodes
1	0	36.000	1.000	1.000	1.036	0
2	0.1	33.351	1.074	0.951	1.011	0
3	0.2	30.649	1.149	0.923	1.024	0
4	0.3	27.784	1.228	0.885	1.042	0
5	0.4	25.243	1.351	0.859	1.043	0
6	0.5	22.944	1.413	0.895	1.137	1
7	0.6	21.758	1.504	0.882	1.010	4
8	0.7	20.800	1.686	0.836	0.852	7
9	0.8	18.444	1.330	0.860	1.070	10
10	0.9	14.900	1.029	0.874	1.264	17

**TABLE 6. The maximum correlation coefficient associated with every process variable.**

Tag	Maximum Correlation coefficient	Tag	Maximum Correlation coefficient	Tag	Maximum Correlation coefficient
Tag 1	1.000	Tag 14	0.999	Tag 27	0.724
Tag 2	0.913	Tag 15	0.994	Tag 28	1.000
Tag 3	0.490	Tag 16	0.672	Tag 29	0.802
Tag 4	0.893	Tag 17	0.999	Tag 30	0.872
Tag 5	0.886	Tag 18	0.999	Tag 31	1.000
Tag 6	0.586	Tag 19	0.900	Tag 32	0.913
Tag 7	0.811	Tag 20	0.999	Tag 33	1.000
Tag 8	0.900	Tag 21	0.999	Tag 34	0.994
Tag 9	0.786	Tag 22	0.938	Tag 35	0.936
Tag 10	0.999	Tag 23	0.571	Tag 36	0.996
Tag 11	0.669	Tag 24	0.954	Tag 37	0.786
Tag 12	0.651	Tag 25	0.997		
Tag 13	0.538	Tag 26	0.991		

small-world index can be further computed (see Table 7). When the correlation thresholds are 0.38 and 0.39, the small-world indexes are both the largest – 1.110. The correct values within four decimal places are 1.1098 and 1.1104, respectively. According to the content described in II C, the larger the small world index, the more reasonable the established CNs are. Therefore, the correlation threshold is selected as 0.39.

Based on the correlation threshold of 0.39, the adjacency matrix is calculated and its black-and-white diagram is also drawn. As shown in Fig. 7(a), white parts indicate that the two nodes are associated with each other. A CN can be finally established (see Fig. 7(b)). This network model is verified to meet the requirement of small-world effect.

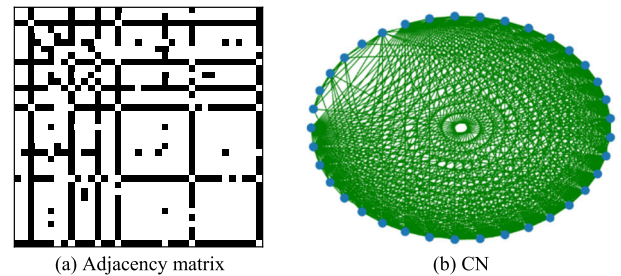
In addition, 50 complex networks can be established for this diesel hydrofining unit (note that the time step and time window are set as 1 and 100, respectively). All these networks are provided with the small-world effect.

*e: CALCULATION OF NETWORK STRUCTURE ENTROPY*

In this paper, three different network structure entropies – degree distribution entropy, Wu structure entropy, and Cai

**TABLE 7. A comparison of the small-world index results from different correlation thresholds.**

No.	Correlation threshold	Small-world index	No.	Correlation threshold	Small-world index
1	0.48	1.093	23	0.26	1.028
2	0.47	1.091	24	0.25	1.014
3	0.46	1.058	25	0.24	1.012
4	0.45	1.052	26	0.23	1.015
5	0.44	1.052	27	0.22	1.022
6	0.43	1.063	28	0.21	1.021
7	0.42	1.057	29	0.20	1.024
8	0.41	1.049	30	0.19	1.023
9	0.40	1.043	31	0.18	1.024
10	0.39	1.110	32	0.17	1.022
11	0.38	1.110	33	0.16	1.012
12	0.37	1.103	34	0.15	1.008
13	0.36	1.080	35	0.14	1.017
14	0.35	1.069	36	0.13	1.016
15	0.34	1.069	37	0.12	1.013
16	0.33	1.055	38	0.11	1.011
17	0.32	1.045	39	0.10	1.011
18	0.31	1.042	40	0.09	1.016
19	0.30	1.042	41	0.08	1.016
20	0.29	1.039	42	0.07	1.015
21	0.28	1.039	43	0.06	1.018
22	0.27	1.030			



**FIGURE 7. The adjacency matrix and CN associated with optimal correlation threshold.**

structure entropy are computed respectively. As shown in Fig. 8, the Wu structure entropy (WSE) and Cai structure entropy (CSE) almost coincide with each other, whose curves are gentle with small fluctuation up and down. However, the curve of degree distribution entropy (DDE) fluctuates violently up and down. The on-site operating conditions are stable and entropy values change little when collecting process data, so the degree distribution entropy is not applicable here. Due to the simplification of calculation, the Wu structure entropy is selected.

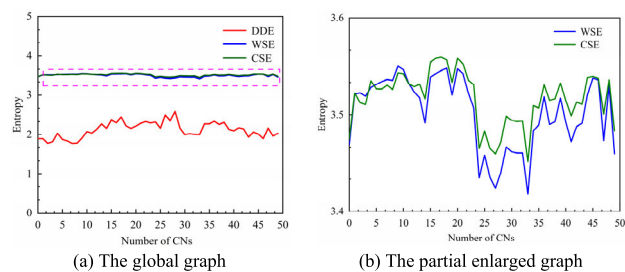
*f: ASSESSMENT OF RELATIVE RISK*

Based on the Wu structure entropy, the relative risk value can be calculated and its graph can be also drawn, as shown in Fig. 9.

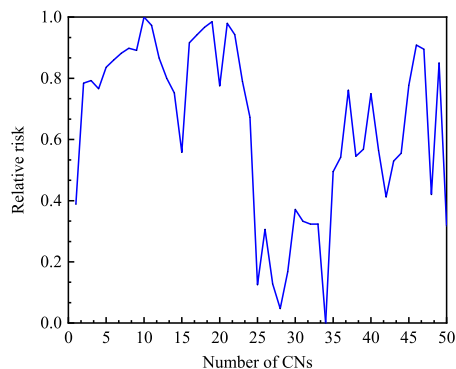
3) DISCUSSIONS

The diesel hydrofining unit involves 37 process variables and 13146 groups of effective data. The WT is applied for data





**FIGURE 8.** A comparison of different network structure entropies. (Note: The x-axis represents the number of CNs; namely, a CN corresponds to a time step).



**FIGURE 9.** The graph of relative risk.

denoising, and each group of data is normalized to  $[0, 1]$ . Based on the time step  $- 1$  and time window  $- 100$ , each cluster of data is sliced out one by one, and there is a total of 13047 clusters. Accordingly, a SRCC matrix can be established for this diesel hydrofining process (see Appendix B).

The process variables are positively correlated with themselves. This paper selects 10 groups with the largest absolute value of average correlation coefficient (see Table 8). For example, Tag 2 represents the differential pressure of FI1001, and Tag 32 represents the refined diesel flow. The greater the raw material flow – namely the greater the differential pressure of FI1001, the greater the product output – namely the greater the refined diesel flow. Therefore, Tag 2 and Tag 32 are positively and strongly correlated. For another example, Tag 7 represents the outlet temperature of F1001, and Tag 13 represents the inlet temperature of R1001. The diesel is heated by the F1001 and transferred into the R1001. Therefore, Tag 7 and Tag 13 are positively and strongly correlated. Similarly, other process variables also have strong correlations.

The process variables with weak correlations are mainly concentrated in Tag 19, Tag 24, and Tag 37 during the actual calculation. An analysis from the original data shows that there is small fluctuation between these three variables, so they are generally in a constant and weakly correlated with others. In other words, an appropriate correlation threshold is conducive to eliminate weak correlations when establishing the CNs. Accordingly, the number of adjacent nodes is the least, so there is no impact on the calculation results of statistical network characteristics.

**TABLE 8.** 10 groups of process variables with the strongest correlation in a diesel hydrofining unit.

No.	Tag	Tag	Absolute value of average correlation coefficient
1	Tag 2	Tag 32	0.949
2	Tag 7	Tag 13	0.893
3	Tag 18	Tag 20	0.855
4	Tag 30	Tag 34	0.842
5	Tag 20	Tag 25	0.838
6	Tag 34	Tag 35	0.832
7	Tag 1	Tag 2	0.832
8	Tag 14	Tag 18	0.831
9	Tag 18	Tag 25	0.830
10	Tag 3	Tag 34	0.827

In this case, the maximum, minimum, and average correlation threshold of 13047 cluster data are 0.953, 0.050, and 0.445, respectively. The reasons are explained as follows. On one hand, the sampling time lasts for a long time, and the amount of process data is very huge. There may be large fluctuations or even process faults. On the other hand, when selecting the largest small-world index, the smaller the correlation threshold is, the more edges are connected, and the evident the small-world effect is. Therefore, the minimum correlation threshold is taken as the lower limit of 0.05 in this paper, which is verified to be reasonable.

Based on the correlation threshold, the SRCC matrix is transformed into an adjacency matrix, which can be used to establish 13047 CNs in this paper. It is verified that all these CNs meet the requirements of small-world effect. Finally, the relative risk is solved and its graph is drawn for the diesel hydrofining process, as shown in Fig. 10. Based on the relative risk in Fig. 10, the risk level can be further divided to identify high-risk points in this diesel hydrofining unit, which is conducive to formulate reasonable and effective risk reduction measures for achieving accurate capture and timely pre-control of process risks. In addition, Figure 10 extracts the characteristics of multi-source process data and converts them into a single time series of relative risk in this diesel hydrofining unit, which can help on-site operators reduce works burden and manage the risk weaknesses of the entire unit promptly and efficiently.

## B. SIMULATION OF TENNESSEE EASTMAN PROCESS

### 1) BACKGROUND

Tennessee Eastman Process (TEP) is a set of chemical simulation platform developed by the Eastman Chemical Company in the US [62]. It is generally applied to simulate the operating conditions of on-site chemical enterprises. The TEP can generate process data with time-varying, coupling, and non-linear characteristics, especially covering process faults [63]. In this paper, an application of TEP is developed to verify the reasonability of the proposed modeling framework. As shown in Fig. 11, the TEP is composed of five key devices – reactor, condenser, compressor, separator, and stripper, as well as a series of instrument pipelines.



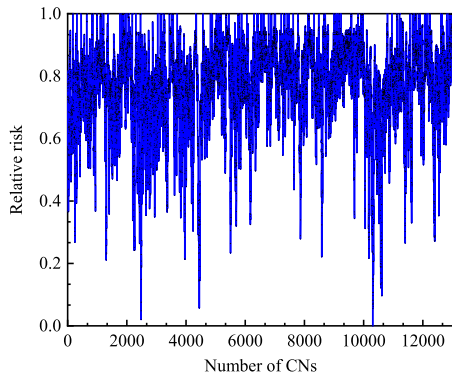


FIGURE 10. The graph of relative risk in a diesel hydrofining unit.

Gaseous materials (A, C, D, and E) and inert gas (B) are used as reactants to generate the products (G and H) and the by-products (F). Based on the Simulink 1.3.3, the TEP can cover 6 operation modes (see Table 9), and the Mode 1 is particularly applied in this paper. There are 53 main process variables, including 12 manipulated variables (see Appendix C) and 41 measured variables (see Appendix D). A total of 28 abnormal conditions are involved in the TEP, including 23 known faults and 5 unknown faults (see Appendix E).

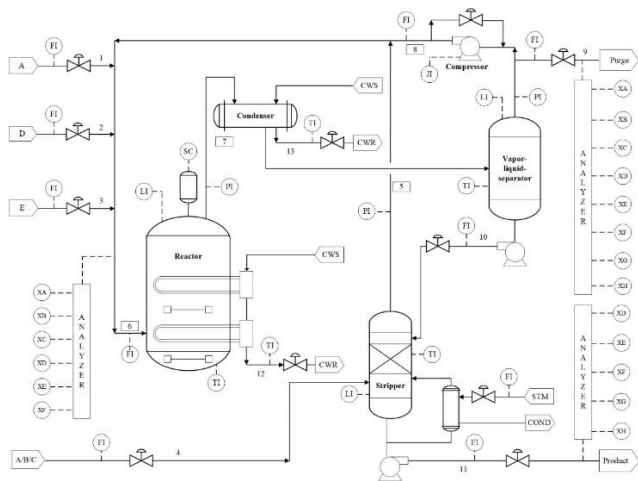


FIGURE 11. The P&ID of the TEP.

TABLE 9. 6 operation modes of the TEP.

Mode	G/H mass ratio	Production rate (stream 11)
1	50/50	7038 kg h <sup>-1</sup> G and 7038 kg h <sup>-1</sup> H
2	10/90	1408 kg h <sup>-1</sup> G and 12669 kg h <sup>-1</sup> H
3	90/10	10000 kg h <sup>-1</sup> G and 1111 kg h <sup>-1</sup> H
4	50/50	Maximum production rate
5	10/90	Maximum production rate
6	90/10	Maximum production rate

## 2) DISCUSSIONS

The TEP data of No Fault and Fault 1 are selected to compare the network structure entropies before the Fault 1 occurring,

when the Fault 1 occurring, and after the Fault 1 occurred, respectively. It is used to verify the rationality of the proposed modeling framework in this paper. Discussions are shown as below. According to the Mode 1 of Simulink 1.3.3, three manipulated variables – XMV(5), XMV(9), and XMV(12) are constant, so they are not considered here. Similarly, 19 measured variables – XMEAS(23)~XMEAS(41) show rectangular rises or falls, and 100 groups of these data are often unchanged during slicing, so they are also not considered. The on-site operating conditions can be simulated under the Model 1. Specifically, a set of TEP data is collected every one second. The simulations – two hours each time – are developed twice in this paper. One is No Fault condition, and the other is Fault 1 condition introduced at one hour. In addition, excessive signal loss may occur in the denoising process of WT, so the WT is not carried out when studying the TEP. For these simulated data, 7109 clusters can be extracted by slicing step by step with a time step of 1 and a time window of 100. Accordingly, 7109 CNs can be further established in this paper.

On the No Fault condition, a discussion is shown as follows. Over the 7109 clusters of TEP data, the maximum, minimum, and average correlation thresholds are 0.507, 0.050, and 0.236, respectively. It is reasonable to set a lower limit of 0.05. In addition, 7090 CNs are provided with the small-world effect, accounting for 99.73%. The maximum value of small-world index is 1.692, which belongs to the 4268th ~ 4367th group of TEP data. The minimum value of small-world index is 0.987, which belongs to the 2237th ~ 2336th group of TEP data. The average value of small-world index is 1.097, which has a small difference. Accordingly, the Wu structure entropies of these 7109 CNs are solved. The maximum entropy is 3.434, which belongs to the 6161st ~ 6260th group of TEP data; The minimum entropy is 3.227, which belongs to the 1943rd ~ 2042nd group of TEP data; The average entropy is 3.361, which has a small difference.

On the Fault 1 condition, a discussion is shown as follows. Over the 7109 clusters of TEP data, the maximum, minimum, and average correlation thresholds are 0.505, 0.050, and 0.234, respectively. It is reasonable to set a lower limit of 0.05. In addition, 7091 CNs are provided with the small-world effect, accounting for 99.75%. The maximum value of small-world index is 1.980, which belongs to the 3638th ~ 3737th group of TEP data. It is close to the fault time. The minimum value of small-world index is 0.987, which belongs to the 2237th ~ 2336th group of TEP data. It is consistent to the no-fault time. The average value of small-world index is 1.094, which has a small difference. Accordingly, the Wu structure entropies of these 7109 CNs are solved. The maximum entropy is 3.434, which belongs to the 689th ~ 788th group of TEP data; The minimum entropy is 3.230, which belongs to the 2132nd ~ 2231st group of TEP data; The average entropy is 3.363, which has a small difference. However, it is higher than that on the No Fault condition.

Based on the above-mentioned discussions, the 2000th ~ 2200, 3600th ~ 3800th, and 5000th ~ 5200th groups of TEP

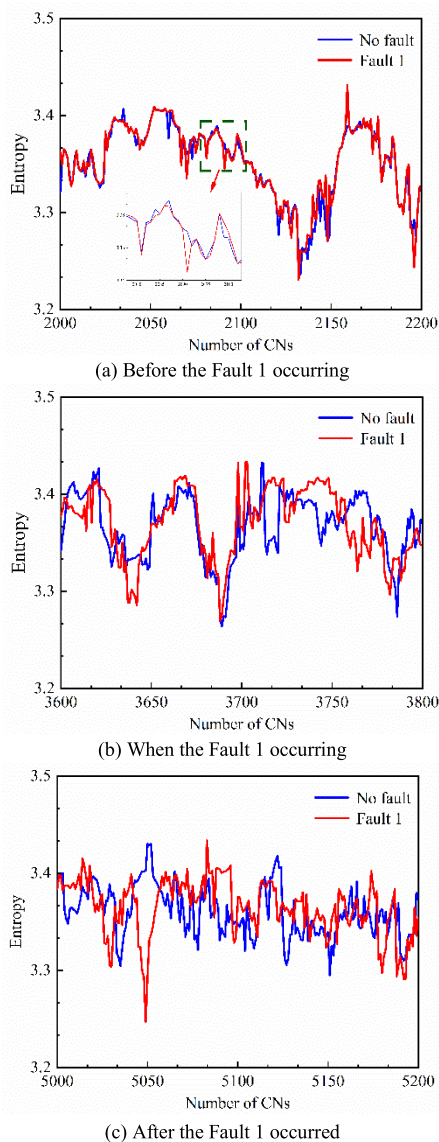


FIGURE 12. A comparison of the network structure entropies.

TABLE 10. A SRCC matrix.

	Tag 1	Tag 2	Tag 3	Tag 4	Tag 5	...	Tag 37
Tag 1	1	0.872	0.129	0.892	-0.798	...	0.042
Tag 2	0.872	1	0.064	0.715	-0.700	...	-0.167
Tag 3	0.129	0.064	1	0.320	-0.217	...	-0.030
Tag 4	0.892	0.715	0.320	1	-0.644	...	0.069
Tag 5	-0.798	-0.700	-0.217	-0.644	1	...	-0.256
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Tag 37	0.042	-0.167	-0.030	0.069	-0.256	...	1

data are regarded as the operating conditions – before the Fault 1 occurring, when the Fault 1 occurring, and after the Fault 1 occurred, respectively. The Wu structure entropies on the No Fault and Fault 1 conditions can be compared, as shown in Fig. 12. Before the Fault 1 occurring, two groups of entropies are almost the same. When the Fault 1

TABLE 11. A SRCC matrix for a diesel hydrofining unit.

	Tag 1	Tag 2	Tag 3	Tag 4	Tag 5	...	Tag 37
Tag 1	1	-0.002	-0.272	0.020	0.298	...	-0.289
Tag 2	-0.002	1	-0.031	0.051	-0.351	...	0.119
Tag 3	-0.272	-0.031	1	0.046	0.016	...	0.006
Tag 4	0.020	0.051	0.046	1	-0.204	...	0.132
Tag 5	0.298	-0.351	0.016	-0.204	1	...	-0.363
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Tag 37	-0.289	0.119	0.006	0.132	-0.363	...	1

TABLE 12. 12 manipulated variables of the TEP.

Variable	Description	Unit
XMV(1)	D feed flow (stream 2)	kg h <sup>-1</sup>
XMV(2)	E feed flow (stream 3)	kg h <sup>-1</sup>
XMV(3)	A feed flow (stream 1)	kscmh
XMV(4)	Total feed flow (stream 4)	kscmh
XMV(5)	Compressor recycle valve	%
XMV(6)	Purge valve (stream 9)	%
XMV(7)	Separator pot liquid flow (stream 10)	m <sup>3</sup> h <sup>-1</sup>
XMV(8)	Stripper liquid product flow	m <sup>3</sup> h <sup>-1</sup>
XMV(9)	Stripper steam valve	%
XMV(10)	Reactor cooling water flow	m <sup>3</sup> h <sup>-1</sup>
XMV(11)	Condenser cooling water flow	m <sup>3</sup> h <sup>-1</sup>
XMV(12)	Condenser cooling water flow	rpm

occurring, there is a small difference between two groups of entropies, but the overall trends remain the same. After the Fault 1 occurred, there is no relations between two groups of entropies intuitively. It is shown that different operation conditions have an impact on the network structure entropies. Accordingly, the network structure entropy can reflect an abnormal condition of TEP. The relative risk which is normalized from the network structure entropy can also characterize an abnormal condition.

Therefore, the proposed modeling framework using the CNs is reasonable. The multi-source process data are particularly integrated into a single time series of relative risk, which can reduce the data dimension and realize the dynamic risk monitoring of TEP.

V. CONCLUSION

This paper presents a modeling framework of dynamic risk monitoring for chemical processes based on the CNs. The WT is firstly developed for denoising – filtering and eliminating the invalid values of massive, disordered, and non-linear process data. The SRCC is further solved to analyze the relationships among these data and generate a correlation matrix. The selections of appropriate modeling parameters are explained in detail, such as time step, time window, correlation threshold, and so on. Accordingly, the CNs of chemical processes are established using a SRCC matrix and its adjacency matrix. The network structure entropy is particularly introduced to transform process data into a single time series of relative risk. Two test cases are selected to illustrate the validity of the

TABLE 13. 41 measured variables of the TEP.

Variable	Description	Unit
XMEAS(1)	A feed (stream 1)	kscmh
XMEAS(2)	D feed (stream 2)	kg h <sup>-1</sup>
XMEAS(3)	E feed (stream 3)	kg h <sup>-1</sup>
XMEAS(4)	Total feed (stream 4)	kscmh
XMEAS(5)	Recycle flow (stream 8)	kscmh
XMEAS(6)	Reactor feed rate (stream 6)	kscmh
XMEAS(7)	Reactor pressure	kPa gauge
XMEAS(8)	Reactor level	%
XMEAS(9)	Reactor temp	°C
XMEAS(10)	Purge rate (stream 9)	kscmh
XMEAS(11)	Separator temp	°C
XMEAS(12)	Separator level	%
XMEAS(13)	Separator pressure	kPa gauge
XMEAS(14)	Separator underflow (stream 10)	m <sup>3</sup> h <sup>-1</sup>
XMEAS(15)	Stripper level	%
XMEAS(16)	Stripper pressure	kPa gauge
XMEAS(17)	Stripper pressure	m <sup>3</sup> h <sup>-1</sup>
XMEAS(18)	Stripper temperature	°C
XMEAS(19)	Stripper steam flow	kg h <sup>-1</sup>
XMEAS(20)	Compressor work	kW
XMEAS(21)	Reactor cooling water outlet temp	°C
XMEAS(22)	Condenser cooling water outlet temp	°C
XMEAS(23)	Composition of A (stream 6)	mol%
XMEAS(24)	Composition of B (stream 6)	mol%
XMEAS(25)	Composition of C (stream 6)	mol%
XMEAS(26)	Composition of D (stream 6)	mol%
XMEAS(27)	Composition of E (stream 6)	mol%
XMEAS(28)	Composition of F (stream 6)	mol%
XMEAS(29)	Composition of A (stream 9)	mol%
XMEAS(30)	Composition of B (stream 9)	mol%
XMEAS(31)	Composition of C (stream 9)	mol%
XMEAS(32)	Composition of D (stream 9)	mol%
XMEAS(33)	Composition of E (stream 9)	mol%
XMEAS(34)	Composition of F (stream 9)	mol%
XMEAS(35)	Composition of G (stream 9)	mol%
XMEAS(36)	Composition of H (stream 9)	mol%
XMEAS(37)	Composition of D (stream 11)	mol%
XMEAS(38)	Composition of E (stream 11)	mol%
XMEAS(39)	Composition of F (stream 11)	mol%
XMEAS(40)	Composition of G (stream 11)	mol%
XMEAS(41)	Composition of H (stream 11)	mol%

proposed modeling framework. Results show that the risks of a diesel hydrofining unit can be effectively and timely monitored and assessed, as well as the abnormal conditions of TEP can be reasonably and accurately monitored and traced. In the future, the risk levels of chemical processes can be further divided to achieve precise risk management based on the proposed modeling framework.

## APPENDIX A

See Table 10.

TABLE 14. 28 abnormal conditions of the TEP.

No.	Description	Type
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
9	D feed temperature (stream 2)	Random
10	C feed temperature (stream 4)	Random
11	Reactor cooling water inlet temperature	Random
12	Condenser cooling water inlet temperature	Random
13	Reaction kinetics	Slow drift
14	Reactor cooling water valve	Sticking
15	Condenser cooling water valve	Sticking
16-20	Unknown	Unknown
21	A feed (stream 1) temperature	Random
22	E feed (stream 3) temperature	Random
23	A feed flow (stream 1)	Random
24	D feed flow (stream 2)	Random
25	E feed flow (stream 3)	Random
26	A and C feed flow (stream 4)	Random
27	Reactor cooling water flow	Random
28	Condenser cooling water flow	Random

TABLE 15. List of Acronyms.

Corresponding full name	Acronyms
Distributed Control System	DCS
Supervisory Control and Data Acquisition	SCADA
Advanced Control System	APC
Complex Networks	CNs
Undirected Unweighted Networks	UUNs
Partial correlation coefficient	ParCC
Pearson Correlation Coefficient	PeaCC
Spearman's Rank Correlation Coefficient	SRCC
Tennessee Eastman Process	TEP
Wavelet Transform	WT
Directed Unweighted Networks	DUNs
Undirected Weighted Networks	UWNs
Directed Weighted Networks	DWNs
Wu Structure Entropy	WSE
Cai Structure Entropy	CSE
Degree Distribution Entropy	DDE

## APPENDIX B

See Table 11.

## APPENDIX C

See Table 12.

## APPENDIX D

See Table 13.

## APPENDIX E

See Table 14.

## APPENDIX F

See Table 15.

## REFERENCES

- [1] C. Chen and G. Reniers, "Chemical industry in China: The current status, safety problems, and pathways for future sustainable development," *Saf. Sci.*, vol. 128, Aug. 2020, Art. no. 104741.
- [2] I. Filina, J. Austin, T. Doré, E. Johnson, D. Minguez, I. Norton, J. Snedden, and R. J. Stern, "Opening of the Gulf of Mexico: What we know, what questions remain, and how we might answer them," *Tectonophysics*, vol. 822, Jan. 2022, Art. no. 229150.
- [3] X. Bi, D. Wu, D. Xie, H. Ye, and J. Zhao, "Large-scale chemical process causal discovery from big data with transformer-based deep learning," *Process Saf. Environ. Protection*, vol. 173, pp. 163–177, May 2023.
- [4] S. Meng, *Chemical Process Safety Management and Technology-Systematic Approach to Preventing Major Accidents*. Beijing, China: China Petrochemical Press, 2018.
- [5] T. Niyazmand and I. Izadi, "Pattern mining in alarm flood sequences using a modified PrefixSpan algorithm," *ISA Trans.*, vol. 90, pp. 287–293, Jul. 2019.
- [6] B. Kou, Y. Jin, H. Zhang, L. Zhang, and H. Zhang, "Nonlinear analytical modeling of hybrid-excitation double-sided linear eddy-current brake," *IEEE Trans. Magn.*, vol. 51, no. 11, pp. 1–4, Nov. 2015.
- [7] R. D. McAllister and J. B. Rawlings, "Nonlinear stochastic model predictive control: Existence, measurability, and stochastic asymptotic stability," *IEEE Trans. Autom. Control*, vol. 68, no. 3, pp. 1524–1536, Mar. 2023.
- [8] W. Li, H. Li, S. Gu, and T. Chen, "Process fault diagnosis with model- and knowledge-based approaches: Advances and opportunities," *Control Eng. Pract.*, vol. 105, Dec. 2020, Art. no. 104637.
- [9] Y. Shu, L. Ming, F. Cheng, Z. Zhang, and J. Zhao, "Abnormal situation management: Challenges and opportunities in the big data era," *Comput. Chem. Eng.*, vol. 91, pp. 104–113, Aug. 2016.
- [10] Z. Li, S. Bao, X. Peng, and L. Luo, "Fault detection and diagnosis in multivariate systems using multiple correlation regression," *Control Eng. Pract.*, vol. 116, Nov. 2021, Art. no. 104916.
- [11] J. Schaeffer and R. D. Braatz, "Latent variable method demonstrator—Software for understanding multivariate data analytics algorithms," *Comput. Chem. Eng.*, vol. 167, Nov. 2022, Art. no. 108014.
- [12] L. D. González and V. M. Zavala, "BOIS: Bayesian optimization of interconnected systems," 2023, *arXiv:2311.11254*.
- [13] H. Wu and J. Zhao, "Self-adaptive deep learning for multimode process monitoring," *Comput. Chem. Eng.*, vol. 141, Oct. 2020, Art. no. 107024.
- [14] P. Gangsar and R. Tiwari, "Signal based condition monitoring techniques for fault detection and diagnosis of induction motors: A state-of-the-art review," *Mech. Syst. Signal Process.*, vol. 144, Oct. 2020, Art. no. 106908.
- [15] X. Liu, Z. Fan, Z. Cao, Y. Liu, Z. Kang, and Y. Hu, "Fault signal simulation and service condition monitoring of cracked gear system," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–14, 2023.
- [16] J. Zhu, Y. Yao, and F. Gao, "Transfer of qualitative and quantitative knowledge for similar batch process monitoring," *IEEE Access*, vol. 6, pp. 73856–73870, 2018.
- [17] C. Castaldello, P. Facco, F. Bezzo, C. Georgakis, and M. Barolo, "Data-driven tools for the optimization of a pharmaceutical process through its knowledge-driven model," *AIChE J.*, vol. 69, no. 4, Apr. 2023, Art. no. e17925.
- [18] Q. Li, S. Meng, S. Zhang, M. Wu, J. Zhang, M. T. Ahvanooy, and M. S. Aslam, "Safety risk monitoring of cyber-physical power systems based on ensemble learning algorithm," *IEEE Access*, vol. 7, pp. 24788–24805, 2019.
- [19] Q. Jiang and X. Yan, "Learning deep correlated representations for non-linear process monitoring," *IEEE Trans. Ind. Informat.*, vol. 15, no. 12, pp. 6200–6209, Dec. 2019.
- [20] W. Li, S. Gu, X. Zhang, and T. Chen, "Transfer learning for process fault diagnosis: Knowledge transfer from simulation to physical processes," *Comput. Chem. Eng.*, vol. 139, Aug. 2020, Art. no. 106904.
- [21] M. Hakim, A. A. B. Omran, A. N. Ahmed, M. Al-Waily, and A. Abdellatif, "A systematic review of rolling bearing fault diagnoses based on deep learning and transfer learning: Taxonomy, overview, application, open challenges, weaknesses and recommendations," *Ain Shams Eng. J.*, vol. 14, no. 4, Apr. 2023, Art. no. 101945.
- [22] J. Li, K. Liu, and X. Xian, "Causation-based process monitoring and diagnosis for multivariate categorical processes," *IISE Trans.*, vol. 49, no. 3, pp. 332–343, Mar. 2017.
- [23] X. Xiu, Y. Yang, L. Kong, and W. Liu, "Data-driven process monitoring using structured joint sparse canonical correlation analysis," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 68, no. 1, pp. 361–365, Jan. 2021.
- [24] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, nos. 66–84, pp. 440–442, 1998.
- [25] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, Oct. 1999.
- [26] A.-L. Barabási, "The network takeover," *Nature Phys.*, vol. 8, no. 1, pp. 14–16, Jan. 2012.
- [27] V. Latora, V. Nicosia, and G. Russo, *Complex Networks: Principles, Methods and Applications*. Cambridge, U.K.: Cambridge Univ. Press, 2017.
- [28] A. Abedi, L. Gaudard, and F. Romerio, "Review of major approaches to analyze vulnerability in power system," *Rel. Eng. Syst. Saf.*, vol. 183, pp. 153–172, Mar. 2019.
- [29] O. D. Naidu, N. George, S. Zubic, and M. Krakowski, "Time-domain-based distance protection for transmission networks: Secure and reliable solution for complex networks," *IEEE Access*, vol. 11, pp. 104656–104675, 2023.
- [30] G. Zhang, H. Jia, L. Yang, Y. Li, and J. Yang, "Research on a model of node and path selection for traffic network congestion evacuation based on complex network theory," *IEEE Access*, vol. 8, pp. 7506–7517, 2020.
- [31] G. Moutsinas and W. Guo, "Probabilistic stability of traffic load balancing on wireless complex networks," *IEEE Syst. J.*, vol. 14, no. 2, pp. 2551–2556, Jun. 2020.
- [32] W. Guan, X. Wen, L. Wang, and Z. Lu, "On-demand cooperation among multiple infrastructure networks for multi-tenant slicing: A complex network perspective," *IEEE Access*, vol. 6, pp. 78689–78699, 2018.
- [33] J. V. Milanovic and W. Zhu, "Modeling of interconnected critical infrastructure systems using complex network theory," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 4637–4648, Sep. 2018.
- [34] R. C. M. van Aert and C. Goos, "A critical reflection on computing the sampling variance of the partial correlation coefficient," *Res. Synth. Methods*, vol. 14, no. 3, pp. 520–525, May 2023.
- [35] L. S. de Pontes and L. C. Régo, "Impact of macroeconomic variables on the topological structure of the Brazilian stock market: A complex network approach," *Phys. A, Stat. Mech. Appl.*, vol. 604, Oct. 2022, Art. no. 127660.
- [36] W.-Y. Zhang, Z.-W. Wei, B.-H. Wang, and X.-P. Han, "Measuring mixing patterns in complex networks by Spearman rank correlation coefficient," *Phys. A, Stat. Mech. Appl.*, vol. 451, pp. 440–450, Jun. 2016.
- [37] W. Tian, Z. Liu, L. Li, S. Zhang, and C. Li, "Identification of abnormal conditions in high-dimensional chemical process based on feature selection and deep learning," *Chin. J. Chem. Eng.*, vol. 28, pp. 1875–1883, Jul. 2020.
- [38] W. Tian, S. Wang, S. Sun, C. Li, and Y. Lin, "Intelligent prediction and early warning of abnormal conditions for fluid catalytic cracking process," *Chem. Eng. Res. Des.*, vol. 181, pp. 304–320, May 2022.
- [39] L. Cai, W. Guo, L. He, and W. Li, "Port integrated scheduling under uncertain operation time and cascade effects: A complex network structure entropy solution," *Comput. Ind. Eng.*, vol. 182, Aug. 2023, Art. no. 109435.
- [40] H. Zhao, J. Liu, H. Chen, J. Chen, Y. Li, J. Xu, and W. Deng, "Intelligent diagnosis using continuous wavelet transform and Gauss convolutional deep belief network," *IEEE Trans. Rel.*, vol. 72, no. 2, pp. 692–702, Jun. 2023.
- [41] Y. Shimada, Y. Hirata, T. Ikeguchi, and K. Aihara, "Graph distance for complex networks," *Sci. Rep.*, vol. 6, no. 1, p. 34944, Oct. 2016.
- [42] B. B. Chaudhuri and A. Bhattacharya, "On correlation between two fuzzy sets," *Fuzzy Sets Syst.*, vol. 118, no. 3, pp. 447–456, Mar. 2001.



- [43] Y. Cheng, F. Sun, Y. Zhang, and F. Tao, "Task allocation in manufacturing: A review," *J. Ind. Inf. Integr.*, vol. 15, pp. 207–218, Sep. 2019.
- [44] S. H. Strogatz, "Exploring complex networks," *Nature*, vol. 410, no. 6825, pp. 268–276, Mar. 2001.
- [45] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. Hwang, "Complex networks: Structure and dynamics," *Phys. Rep.*, vol. 424, nos. 4–5, pp. 175–308, Feb. 2006.
- [46] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, Jul. 1948.
- [47] H. Xu, R. Luo, J. Winnink, C. Wang, and E. Elahi, "A methodology for identifying breakthrough topics using structural entropy," *Inf. Process. Manage.*, vol. 59, no. 2, Mar. 2022, Art. no. 102862.
- [48] B. Wang, F. Tan, J. Zhu, and D. Wei, "A new structure entropy of complex networks based on nonextensive statistical mechanics and similarity of nodes," *Math. Biosci. Eng.*, vol. 18, no. 4, pp. 3718–3732, 2021.
- [49] S. Cao, M. Dehmer, and Y. Shi, "Extremality of degree-based graph entropies," *Inf. Sci.*, vol. 278, pp. 22–33, Sep. 2014.
- [50] S. Schnettler, "A structured overview of 50 years of small-world research," *Social Netw.*, vol. 31, no. 3, pp. 165–178, Jul. 2009.
- [51] Z. Budrikis, "25 years of small-world network theory," *Nature Rev. Phys.*, vol. 5, no. 8, p. 440, Aug. 2023.
- [52] P. V. Paulau, C. Feenders, and B. Blasius, "Motif analysis in directed ordered networks and applications to food webs," *Sci. Rep.*, vol. 5, no. 1, Art. no. 11926, Jul. 2015.
- [53] A. Tian, Y. Zhang, C. Ma, H. Chen, W. Sheng, and S. Zhou, "Noise-robust machinery fault diagnosis based on self-attention mechanism in wavelet domain," *Measurement*, vol. 207, Feb. 2023, Art. no. 112327.
- [54] W. Zhang, G. Peng, C. Li, Y. Chen, and Z. Zhang, "A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals," *Sensors*, vol. 17, no. 2, p. 425, Feb. 2017.
- [55] C. Yang, Z. Qiao, R. Zhu, X. Xu, Z. Lai, and S. Zhou, "An intelligent fault diagnosis method enhanced by noise injection for machinery," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–11, 2023.
- [56] R. L. T. Bevan and P. Nithiarasu, "Accelerating incompressible flow calculations using a quasi-implicit scheme: Local and dual time stepping approaches," *Comput. Mech.*, vol. 50, no. 6, pp. 687–693, Sep. 2012.
- [57] T. C. N. Boekholt, T. Vaillant, and A. C. M. Correia, "Reversible time-step adaptation for the integration of few-body systems," *Monthly Notices Roy. Astronomical Soc.*, vol. 519, no. 3, pp. 3281–3291, Jan. 2023.
- [58] T. Wen and Y. Deng, "The vulnerability of communities in complex networks: An entropy approach," *Rel. Eng. Syst. Saf.*, vol. 196, Apr. 2020, Art. no. 106782.
- [59] C. G. S. Freitas, A. L. L. Aquino, H. S. Ramos, A. C. Frery, and O. A. Rosso, "A detailed characterization of complex networks using information theory," *Sci. Rep.*, vol. 9, no. 1, p. 16689, Nov. 2019.
- [60] L. Huang, Y.-C. Lai, K. Park, and J. Zhang, "Percolation and blind spots in complex networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 73, no. 6, Jun. 2006, Art. no. 066131.
- [61] M. E. J. Newman, "The structure and function of complex networks," *SIAM Rev.*, vol. 45, no. 2, pp. 167–256, 2003.
- [62] G. Li, C. F. Alcalá, S. J. Qin, and D. Zhou, "Generalized reconstruction-based contributions for output-relevant fault diagnosis with application to the Tennessee Eastman process," *IEEE Trans. Control Syst. Technol.*, vol. 19, no. 5, pp. 1114–1127, Sep. 2011.
- [63] H. Chen, P. Tiño, and X. Yao, "Cognitive fault diagnosis in Tennessee Eastman process using learning in the model space," *Comput. Chem. Eng.*, vol. 67, pp. 33–42, Aug. 2014.



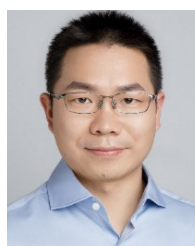
**JIAQI HAN** received the bachelor's degree in safety engineering from the Beijing University of Chemical Technology, in 2023, where he is currently pursuing the master's degree with the College of Mechanical and Electrical Engineering. His research interest includes chemical process safety.



**FENG CHEN** received the Ph.D. degree in power engineering and engineering thermophysics from the China University of Petroleum, in 2019. He is currently an Associate Professor with the College of Mechanical and Transportation Engineering, China University of Petroleum. His research interests include smart control and optimization of process equipment, multiphase flow theory, and separation technology.



**FENG WANG** received the Ph.D. degree in chemical process machinery from the Beijing University of Chemical Technology, in 2009. He is currently a Professor with the College of Mechanical and Electrical Engineering, Beijing University of Chemical Technology. His research interests include equipment fault diagnosis, early warning, chemical process safety, and intelligent diagnosis technology based on multi-parameter data fusion.



**ZHAN DOU** received the Ph.D. degree in safety science and engineering from Nanjing Tech University, in 2017. He is currently an Associate Professor with the College of Mechanical and Electrical Engineering, Beijing University of Chemical Technology. His research interests include chemical process safety, computer vision-based risk prediction, and integrated control of physical safety and cybersecurity.



**QIANLIN WANG** received the Ph.D. degree in safety science and engineering from the China University of Petroleum, in 2019. She is currently a Lecturer with the College of Mechanical and Electrical Engineering, Beijing University of Chemical Technology. Her research interests include chemical process safety, intelligent risk prediction and pre-warning, and integrated control of functional safety and cybersecurity.



**GUOAN YANG** received the Ph.D. degree in mechanical manufacturing and automation from Southeast University, in 2001. He is currently a Professor with the College of Mechanical and Electrical Engineering, Beijing University of Chemical Technology. His research interests include equipment condition monitoring, fault diagnosis, and acoustic emission testing.

...