Chemical Process Alarm Root Cause Diagnosis Method Based on the Combination of Data-Knowledge-Driven Method and Time Retrospective Reasoning

Xiaomiao Song, Qinglong Liu, Mingxin Dong, Yifei Meng, Chuanrui Qin, Dongfeng Zhao,* Fabo Yin, and Jiangbo Jiu

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Due to the abrupt nature of the chemical process, a large number of alarms are often generated at the same time. As a result of the flood of alarms, it largely hinders the operator from making accurate judgments and correct actions for the root cause of the alarm. The existing diagnosis methods for the root cause of alarms are relatively single, and their ability to accurately find out complex accident chains and assist decision making is weak. This paper introduces a method that integrates the knowledge-driven method and the data-driven method to establish an alarm causal network model and then traces the source to realize the alarm root cause diagnosis, and develops the related system modules. The knowledge-driven method uses the hidden causality in the optimized hazard and operability analysis (HAZOP) report, while the data-driven method combines the autoregressive integrated moving average model (ARIMA) and Granger causality test, and the traceability mechanism uses the time-based retrospective reasoning method. In the case study, the practical application of the method is compared with the experimental application in a real petrochemical plant. The results show that this method helps to improve the accuracy of correct diagnosis of the root cause of the alarm and can assist the operators in decision making. Using this method, the root cause diagnosis of alarm can be realized quickly and scientifically, and the probability of misjudgment by operators can be reduced, which has a certain degree of scientificity.
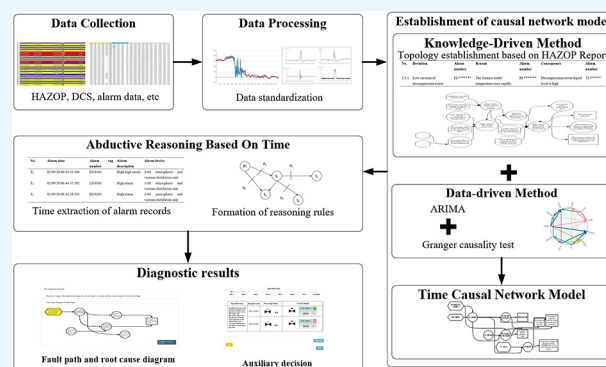
1. INTRODUCTION

Due to the complexity, reaction coupling, mutation, and uncertainty of the chemical process, whether the process is complex or not, the scale of the installation is large or small, process alarms are everywhere. The significance of the alarm is that when the abnormal working condition occurs, the alarm system will send out a warning signal with clear guiding significance to the engineering operator and supports the response.[1] However, with the advent of the era of big data, the intelligence level of the chemical process industry has been greatly improved. The application scope of intelligent systems, such as basic process control system (BPCS) and safety instrumentation system (SIS), has been gradually deepened, and the collection of chemical process data has changed. While the collection of chemical process data is getting easier, the number of alarm variable configurations has also increased exponentially.[2] At the same time, the phenomenon called "alarm flooding" has aroused widespread concern.[3] In the actual production process, because of alarm flooding, the operator was tired to confirm the authenticity of each information and chose to ignore most of the alarms according to his experience, which

also led to the occurrence of some safety accidents and heavy losses for the enterprise. For example, in the Three Mile Island incident in the United States, the main reason was that when a mechanical failure occurred, the operator did not take effective measures in time due to alarm flooding. Operation errors led to the most serious nuclear leakage accident in the history of the United States.[4]

The reason for the alarm flooding lies mostly in five aspects:[5] (a) the process correlation features are complex, and the abnormal spreading range is wide; (b) the alarm threshold is designed randomly, and the rate of false alarms and missed alarms is out of balance; (c) the alarm priority is fuzzy classification and the processing sequence is improper; (d) there

are many types of alarms, which makes it difficult to distinguish effective alarms; (e) the alarm performance evaluation is not timely, and redesign has not been realized. To suppress and eliminate some unnecessary alarm generation and alleviate the phenomenon of alarm flooding, researchers have taken a series of measures, such as introducing filters to reduce tremor alarms,[6] using delay timer,[7] and adjusting alarm threshold.[8] These methods can reduce nuisance alarms to a certain extent, but still cannot solve the problem of alarm flooding from the root cause. Therefore, when a large number of alarms emerge at the same time, how to identify the root cause of the alarm, find out the key alarm variables, and guide the operator to make an accurate response have become the key to solving the problem of alarm flooding.

According to the diagnosis method of alarm root cause diagnosis proposed by existing experts and scholars, it can be divided into two categories: pattern-matching method and the method based on causal network models. The pattern-matching method can mine historical data to determine the data category of the known alarm source. Then, the new unknown data and the known data are matched by similarity, cluster analysis, and other methods to complete the pattern matching, so as to directly determine the alarm based on the data source. For example, Cheng[9] proposed an improved Smith−Waterman (SW) algorithm, but the calculation amount increased during the process. To improve computational efficiency, Hu[10] proposed a basic local alignment search tool (BLAST) local comparison algorithm in consideration of priority and other factors. Inspired from pattern matching, Bouillard[11] proposed a new methodology to find alarm correlations with or without prior knowledge about the monitored system. For fault diagnosis, Xu et al.[12] proposed a novel and effective pattern matching method using kernel canonical variate analysis (KCVA) integrated with an adaptive rank-order morphological filter (ARMF). Furthermore, Li et al.[13] report a new fault diagnosis method that exploits dynamic process simulation and pattern matching techniques. Although the method based on pattern matching can find the root cause of the alarm, it can only give a simple diagnosis result and cannot provide more detailed reference for the operation of the operator or the optimization of the alarm management. Moreover, this method can only realize the analysis of the alarm root cause of the existing abnormal mode, and the correct rate of the analysis results will be greatly reduced in the face of abnormal conditions that have never occurred.

The method based on the causal network model consists of two steps. First, the causal network model of alarm variables is established (mainly using data-driven and knowledge-driven methods), and then the traceability analysis mechanism is formulated according to the established causal network model to determine the root cause of alarm. The data-driven modeling method is mainly based on historical data generated in the production process and uses data mining algorithms to identify the causal relationship between variables.[14] Among them, more mainstream algorithms include algorithms that are more widely used, including system identification approach,[15] cross-correlation analysis,[16] Granger causality analysis,[17] directed transfer function/partial directed coherence analysis,[18] transfer entropy analysis,[19] Bayesian network learning,[20,21] and so on. Among them, Granger causality analysis has been applied to many fields and worked well. Chen[22] added the Gaussian process regression (GPR) approach into the framework of the multivariate G-causality test to better indicate the causal relationships between the candidate process variables. Lindner[23]

made a comparative analysis of Granger causality and transfer entropy to present a decision flow for the application of oscillation diagnosis. To highlight the root cause variable and facilitate the diagnosis, Liu[24] proposed a simplified Granger causality map for root cause diagnosis. Moreover, Ghosh's[25] work demonstrated the suitability and applicability of process-accident models in capturing temporal dependence using process data. Zhang et al.[26] proposed a novel data-driven framework for root cause alarm localization, combining both causal inference and network embedding techniques. The improvement of the above methods mainly focuses on improving the learning efficiency, and the improvement of the accuracy mainly depends on the process knowledge.

The main principle of the knowledge-driven method is through analyzing the known system information (such as system flowchart, etc.), expert knowledge (mechanism model, material flow involved in the production process, energy flow, etc.) and other process knowledge; the causal relationship between alarm variables can be obtained; and then a causal relationship model can be established. Among them, the methods favored by many experts and scholars are structural equation modeling,[27] graphical models,[28] rule-based models,[29] extracting plant topology from Web language,[30] and so on. For instance, considering the analysis of consequential alarms, Wang[31] introduced a weighted fuzzy association rules mining approach to discover correlated alarm sequences. Aziz[32] presented a dynamic hazard identification methodology founded on an ontology-based knowledge modeling framework coupled with probabilistic assessment. Only using the knowledge-driven method will not be able to meet the modeling requirements of this type of system, so scholars have begun to study the hybrid-driven method combined with the data-driven method and the knowledge-driven method. In the study of practical problems, combining data-driven method and knowledge-driven method will help improve the overall performance of the method, enhance the application effect of the method, and achieve full utilization of chemical process data and knowledge. For building a good causal network model, Zhu[33] integrated process knowledge with modified transfer entropy. Jeon[34] coined the term entity normalization model with a novel edge weight updating neural network to realize knowledge-driven graph and data-driven graph.

The mechanism of alarm traceability analysis refers to the method of mining alarm propagation path based on a causal network model, which relies on backtracking,[35] reasoning,[36] hypothesis testing,[37] etc. The common methods include methods based on expert systems and methods based on depth-first traversal.[38] The common limitation of the above two methods is that they are qualitative analysis methods, which rely on system knowledge and expert experience, and cannot give quantitative conclusions. Realizing the importance of time in fault diagnosis, Shang[39] introduced a method of finite state machine model of fault reasoning for distribution system under time sequence constraints. And aiming at the problem of fault diagnosis when there are only a few labeled samples in the large amount of data collected during the operation of rotating machinery, Ye et al.[40] proposed a fault diagnosis method based on knowledge transfer in deep learning.

To solve the defects of the above methods and realize the accurate and timely diagnosis of the alarm root cause in the chemical process, a novel method of alarm root cause diagnosis is developed. To overcome the inherent weaknesses of the traditional data-driven method and knowledge-driven method, a
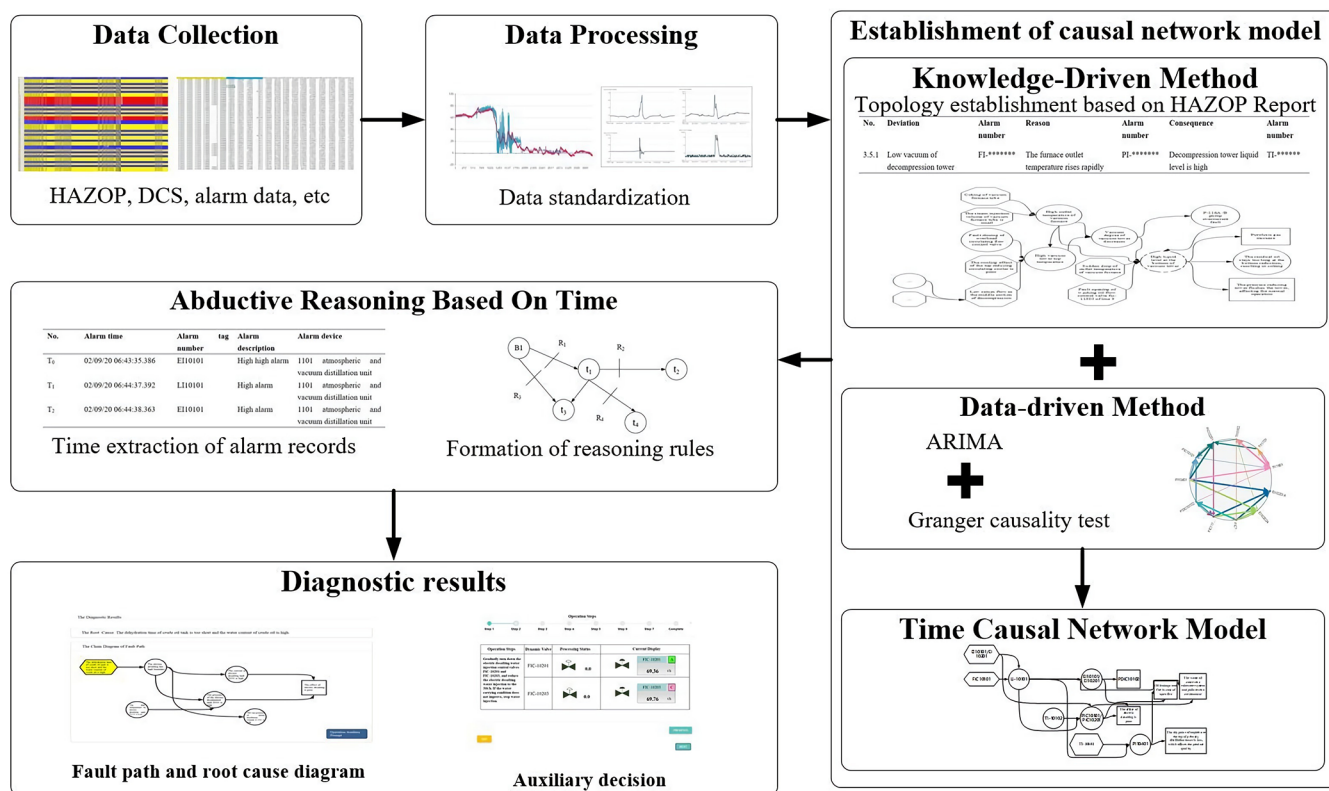
Figure 1. Flow of the chemical process alarm root cause diagnosis method.

causal network model is established using a hybrid-driven method. The combination of hazard and operability analysis (HAZOP) topology and time causal network obtained by the autoregressive integrated moving average model (ARIMA) improved the results of the Granger causality test, which makes the model more reliable and effective. In the traceability mechanism, time-based abductive reasoning methods are used, which are introduced in section 2. In section 3, the corresponding system was developed and the method was verified with an example of an oil refinery to prove the practicability of the method. Finally, a concluding statement is made in section 4.

2. DIAGNOSIS METHOD OF ALARM ROOT CAUSE

The literature pointed out that before the alarm occurs, the DCS data will show a different trend from the normal state, and there is a certain rule between the data changes of the causal related parameters. Therefore, before an alarm occurs, an appropriate data-driven method can be used to extract the causal relationship between data, thereby constructing a time causal network model, using the autoregressive integrated moving average model (ARIMA) and the Granger causality test. Furthermore, according to the standard, the HAZOP analysis must be carried out within the specified time in the chemical plant. The HAZOP reports contain knowledge of the chemical process, such as PID diagrams and expert knowledge. Therefore, the knowledge-driven method can be used to extract the topology diagram in the HAZOP report. After the hybrid-driven method is used to construct a causal relationship network, the time relationship in the alarm record and the operation record is used for retrospective reasoning as a traceability mechanism to complete the diagnosis of the alarm root cause. In view of the problem that it is difficult to quickly judge the root cause of alarm flooding in

the chemical process, in this paper, we developed a method of diagnosis of chemical process alarm root cause based on the data-knowledge-driven method as well as a corresponding system, which can effectively improve the accuracy of the alarm root cause diagnosis and assist the operator in making decisions and taking actions. The flowchart of the method of diagnosing the root cause in the chemical process alarm based on the hybrid drive is shown in Figure 1.

2.1. Establishment of the Causal Network Model. *2.1.1. Establishment of Topology Based on HAZOP.* *2.1.1.1. Introduction to HAZOP.* Hazard and operability analysis (HAZOP) is a structured analysis method used to identify design defects, process hazards, and operability problems. The essence of this method is that an analysis group composed of professionals systematically studies each unit (i.e., analysis node) in a prescribed way and analyzes the hazards and operability problems caused by deviations from the design process conditions.

The main dangerous items analyzed by HAZOP and the discussion results should be recorded in the standard operation form in time. The contents of the form include numbers, elements, guiding words, deviations, causes, consequences, measures, etc., as shown in Table 1.

2.1.1.2. Causal Transmission in the HAZOP Analysis Report. The process deviation in the cause and consequence of HAZOP analysis and the current deviation constitute the transfer relationship between the process parameters before and after. Therefore, the variable data monitored by some instruments can influence each other and there is a correlation relationship. If the current process parameters are configured with alarms in the system, then these alarms will also form a certain correlation. On the contrary, the correlation can be used

Table 1. HAZOP Analysis Record Sheet

| no.	deviation	specific deviation	reason	consequence	initial event probability	enabling events and correction factors	serious ness	possibility	risk level	control level	whether to perform LOPA	independent protective layer	other protective measures	TMEL	MEL	RRF	SIL level	severity	possibility	risk level	is it acceptable	suggested measures	MEL	severity	possibility	risk level
								initial risk analysis						residual risk analysis									risks after implementation of recommended measures			
1	no/little liquid level	low liquid level of flash tower C101	Lic3002 control failure (valve closed)	the evacuation of flash bottom oil pump P102 was damaged, causing a fire	0.1		2	4	24 (8)	Operation Department	yes	(1) flash tower liquid level alarm LI3001; (2) emergency shutoff valve EBV3010 at the bottom outlet of the flash tower	1.0.1; 2.0.1	1.0; 0E-02	1.0; 0E-03			2	2	2; 2 (4)	yes		1			
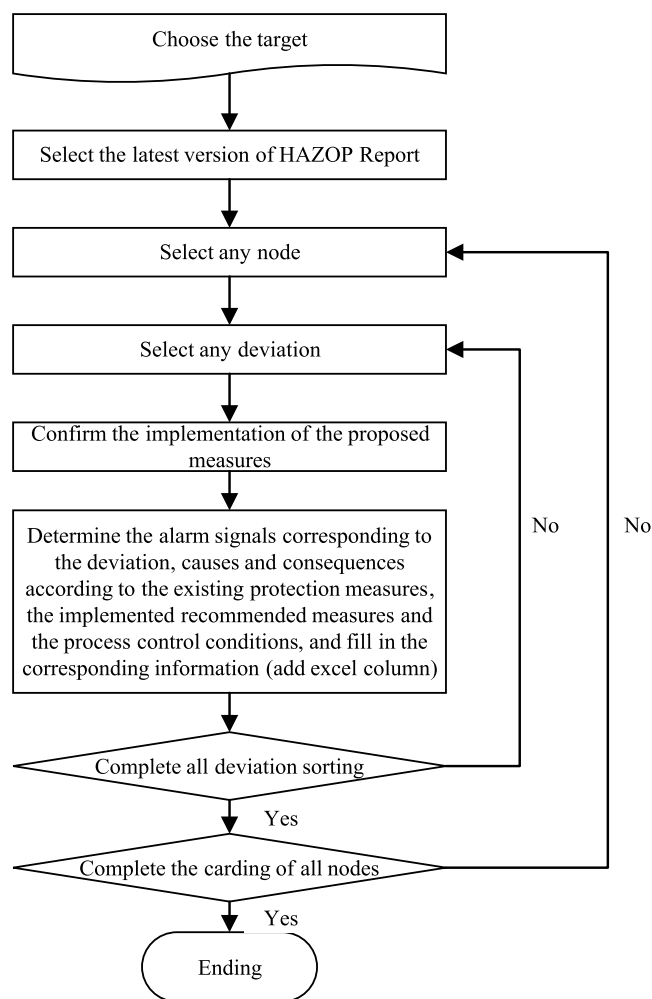
Figure 2. Adapting process of the HAZOP report.

to analyze the alarm signals and further optimize the alarm system.

It should be noted that during the analysis of causes or consequences, although some causes or consequences are expressed in the form of failures or accident scenarios, they may also be related to certain process deviations (alarms). This situation shall be considered as much as possible during correlation to identify potential process deviation correlation.

The HAZOP analysis report contains the topological relationship of the process, but in actual applications, it is necessary to standardize the data processing of the HAZOP analysis report to make it directly analyzed and utilized by the computer. This process can be called the deviation-alarm signal mapping process, and the specific process is shown in Figure 2.

First, according to the selected research object, obtain the latest version of the HAZOP analysis report of the object, and analyze and process the nodes and the deviations one by one. The main process of processing includes three steps. The first step is to confirm whether all of the "recommended measures" corresponding to the current deviation have been implemented. The main significance of the process is that the recommended measures may contain some alarm settings. If they have been implemented, the existence of the alarm should be considered in the analysis. The second step is to determine the corresponding alarm signal of the deviations, causes, and consequences in combination with the existing protection measures, imple-

Table 2. HAZOP Analysis Form after Processing

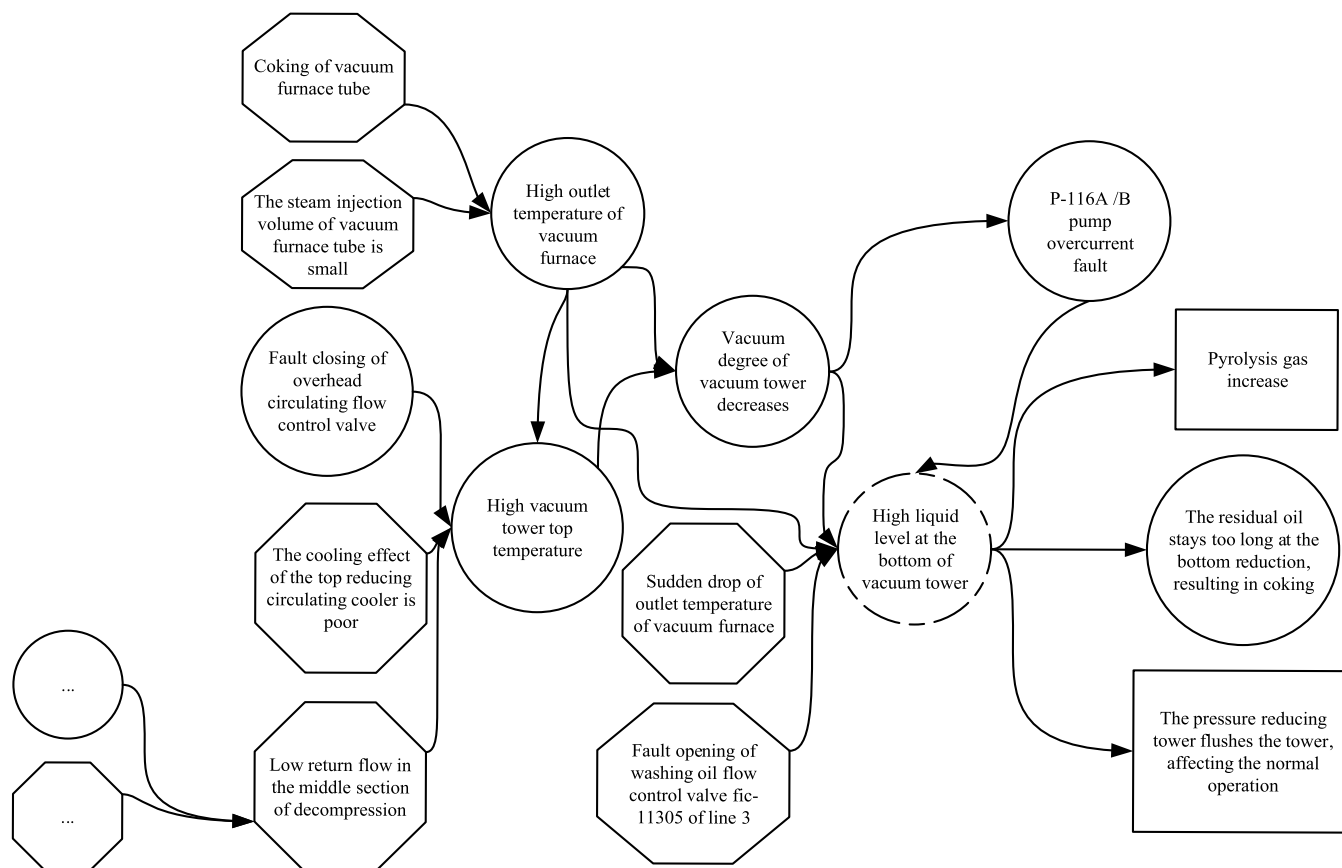| no.	deviation	alarm number	reason	alarm number	consequence	alarm number
3.5.1	low vacuum of decompression tower	FI-*******	the furnace outlet temperature rises rapidly	PI-*******	decompression tower liquid level is high	TI-******

Figure 3. HAZOP topology diagram.

mented recommended measures, and process control. In this step, it should be noted that some deviation causes and deviation consequences in the original report are not expressed in the form of deviation, but there are actually corresponding alarms due to the inconsistent standard and other reasons. We try to identify them. The third step is to add a column after the process deviation column, cause column, and consequence column, respectively, and write out the alarm signal corresponding to the deviation (standard format). If there is no corresponding alarm signal for the current deviation cause or deviation consequence, it is unnecessary to fill in. Through the implementation of the above process, the HAZOP analysis table shown in Table 2 can be obtained.

2.1.1.3. Topology Diagram and Related Parameter Extraction. After the HAZOP form is processed, the correlation relationship between the relevant process deviations (part of the corresponding alarm signal) can be obtained, and the implementation of the alarm data can be processed and analyzed using the correlation relationship to realize the optimal management of the alarm signal. As shown in Figure 3, the relationship topology diagram and related parameter topology diagram are obtained. The related parameters extracted at this time also provide the basis for the establishment of the time causal network model.

2.1.2. Establishment of Causal Network Models for Time Series. Because of the fixed time interval of the actual data collected in the factory, the time series causality network model is constructed using the ARIMA model to analyze the historical data, and then the Granger causality test is used to construct a time series causal network model.

2.1.2.1. Time Series Correlation Analysis Model. The basic idea of autoregressive integrated moving average model (ARIMA) is to treat the data sequence formed by the research object over time as a set of random variables depending on time t. The correlation of this set of random variables can be described by combining past observations with random disturbance factors to establish a random time series model, revealing the rules that exist between the target variables and time changes, and after predicting and estimating them, inferring the state of something at a certain time in the future based on the rules and the past and present historical data. AR^1 is a p-order autoregressive model, and MA (a) is an a-order moving average model. The ARIMA (p, d, q) model is a combination of the AR^1 model and the MA^2 model, where d is the number of differences made to convert the nonstationary time series into stationary time sequences. Before building a model, it is necessary to check whether the original data series are stationary. If the original sequence is not stationary, the nonstationary time data sequence
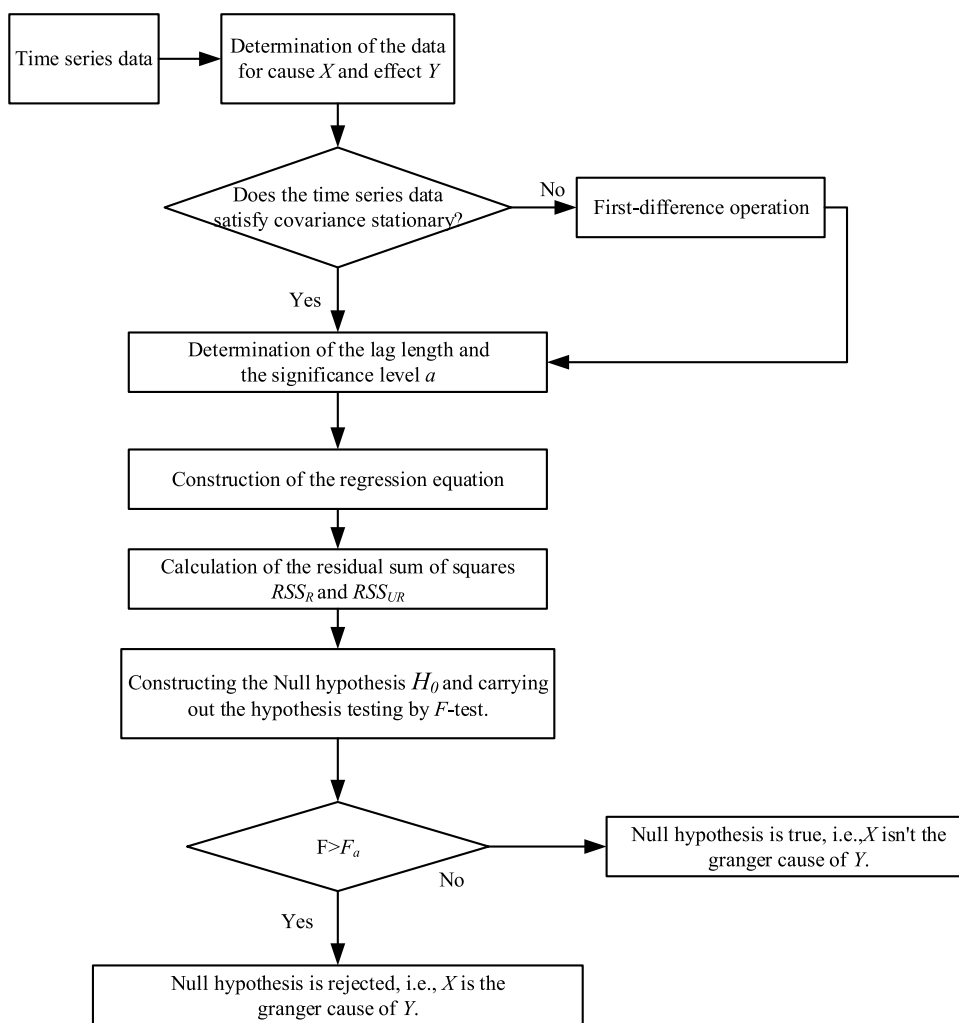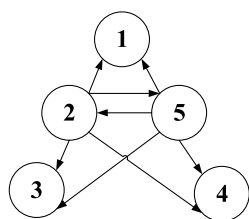
Figure 4. Granger causality test process.

Figure 5. Granger causality graph. Note: This Granger causality diagram contains five time series, namely, time series 1, 2, 3, 4, and 5. Directed edges point from time series 2 to time series 1, 3, 4, and 5, and directed edges point from time series 5 to time series 1, 2, 3, and 4, which means that time series 2 and time series 5 Granger cause all other time series.
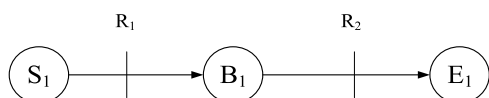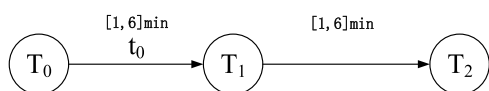

Figure 6. Example of rules and constraints

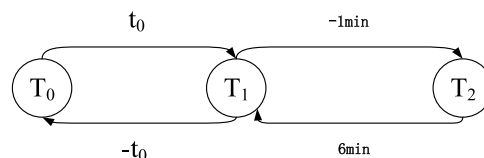**Figure 7.** Example of time-constrained networks.

Figure 8. Distance graph of the example.

needs to be converted into a stationary time data sequence, and then the dependent variable is used to recheck the present value and lag value of its lag value and random error term before establishing the model.

The general expression of the ARMA (p, d, q) model is

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + ... + \varphi_p X_{t-p} + \mu_t + \theta_1 \mu_{t-1}$$
$$+ \theta_2 \mu_{t-2} + ... + \theta_q \mu_{t-q} \quad (1)$$

In eq 1, X_t represents a time data series with stationarity; μ_t represents a white noise data sequence that conforms to a normal distribution; φ_a, θ_b ($a = 1, 2, 3, ..., p; b = 1, 2, 3, ..., q$) are the parameters of the data sequence X_t and μ_t, respectively; and p and q represent the autoregressive orders.

2.1.2.2. Granger Causality Test. Granger causality began as a measurement method generally accepted by economists. Because of the strong comparability between the complex
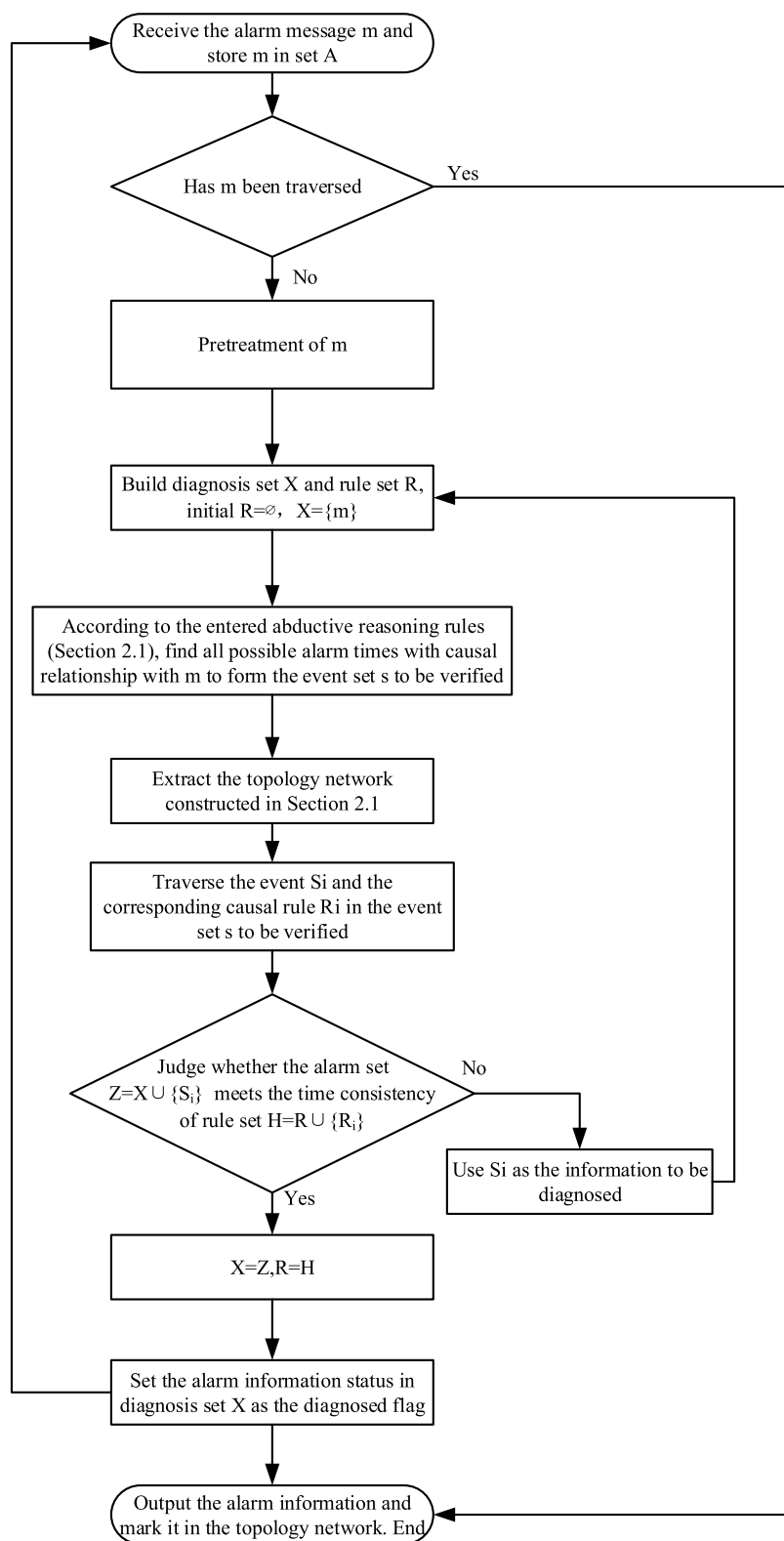
Figure 9. Flowchart of time abduction reasoning.

causality and propagation characteristics of chemical process parameters and the complex correlation between variables in the economic system, both of them are complex nonlinear large systems. Therefore, Granger causality can be introduced into the study of the relationship between chemical process faults and symptoms. In the field of time series processing, Granger

causality refers to a kind of predictive causality. The basic idea is: Given two time series X and Y, compared to only using historical information of X for prediction, the injection of Y historical information is more helpful to predict X, then it is said that Y Granger causes X. Time series Y is time series X Granger cause,
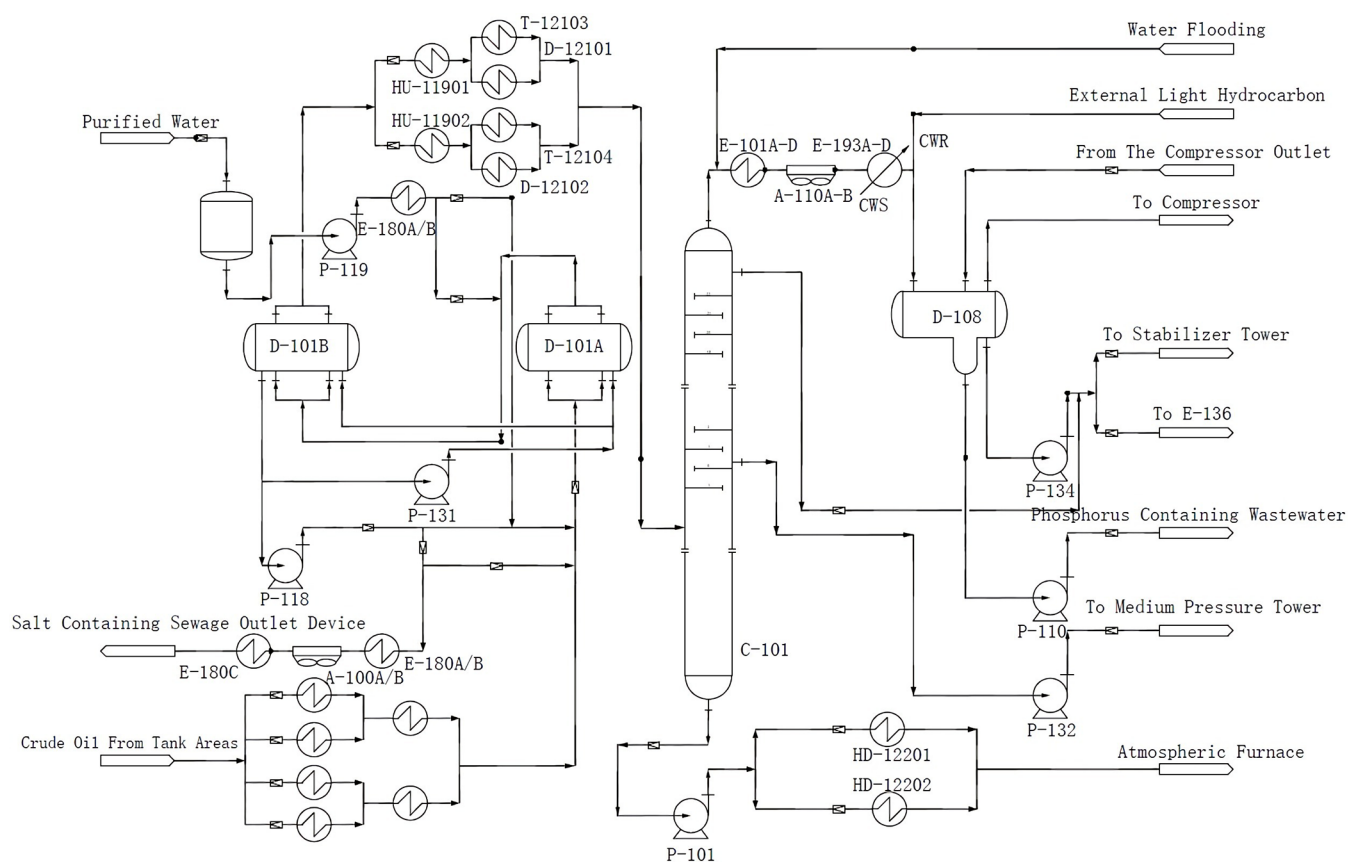
Figure 10. Electric desalination process.

time series X is time series Y Granger result. In addition, the dependency between X and Y is Granger causality.

The method of determining Granger causality between process variables is called Granger causality test. According to the definition of Granger causality, judging whether there is Granger causality between X and Y means establishing two regression equations and comparing the explanatory power of the two regression equations.

To test the Granger causality between two variables X and Y, it is necessary to construct a regression equation containing the lag (x_t and y_t) of X and Y, as shown in eqs 2 and 3:

$$y_t = \sum_{i=1}^{q} \alpha_i x_{t-i} + \sum_{i=1}^{q} \beta_j y_{t-j} + u_{1t} \tag{2}$$

$$x_t = \sum_{i=1}^{s} \lambda_i x_{t-i} + \sum_{j=1}^{s} \delta_j y_{t-j} + u_{2t} \tag{3}$$

In eq 2, x_{t-i} is the lag term of x_t; y_{t-i} is the lag term of y_t; q is the lag length in the regression equation of the variable Y; i and j are the number of lag terms; s is the subsequent length in the regression equation of the variable X and the maximum of lag period length q and s is the order of regression model; u_{1t} and u_{2t} are the white noise; α_i and λ_i are the estimated values of the coefficient of x; and β_j and δ_j are the estimated values of the coefficient of y. If δ_i ($i = 1, ..., q$) is statistically significant as a whole not 0, then X is the Granger cause of Y. Similarly, if δ_j ($j = 1, ..., s$) is statistically significant as a whole not 0, then Y is the Granger cause of X.

2.1.2.3. Establishment of the Time Causal Network Model.

The steps to build a temporal causal network model are as follows:

1. *Extract time series data of relevant parameters.* In section 2.1.1.3, the relevant process parameters have been extracted according to the processed HAZOP report, and then the time series data are extracted for the selected process parameters and the process parameters where the alarm occurs. Time series data are the historical data within a time range of 30 min forward from the moment of alarm. Assuming that there are m process parameters selected by HAZOP that may cause alarms, their time series are, respectively, set as $\{x_{1t}\}, \{x_{1t}\}, ..., \{x_{rt}\}, ..., \{x_{mt}\}$. At the same time, the time series of the alarm process parameters is set to $\{y_t\}$.

2. *Construct an ARIMA model for each relevant parameter.* For each time series established in step 1, an ARIMA model must be constructed separately. In the process of establishing the model, there are the following three points to note: (1) whether the time data series is stable must be checked; (2) the values of the autoregressive order p and the moving average order q must be determined by observing whether the autocorrelation function graph (ACF) and partial autocorrelation function graph (PACF) of the original data series show a tailing phenomenon; and (3) after the modeling is completed, an error test of the model prediction results is required, and the qualified model can be used for prediction.

Table 3. Excerpts from the HAZOP Report[a]

no.	parameter Indicator	deviation	possible causes	consequences	protective measures
1	boundary	high	the crude oil tank farm has a short dehydration time and high crude oil water content;	the electric desalting has high boundary and high current, which will burn out the transformer; the primary distillation tower carries water, the tower is flushed, and in case of serious overpressure oil leakage;	track the water content of crude oil before dewatering and contact and report in time;
			large water injection FIC10102, FIC10101 control failure, the valve is fully opened	the electric desalination tank is overpressured, and the equipment, manholes, and flange gaskets leak;	electric desalination interface LI-10101/10102, LI-10201/10202 display, control, and alarm;
			LIC10101 control failure, the drain valve is fully closed.	crude oil leaks and fires occur in the case of open flames;	electric desalination current EI-10101A~C/EI-10201A~C display and alarm;
				the dirty oil enters the rainwater system and pollutes the environment.	electric desalination pressure PIC-10101/10201 display and alarm;
					a safety valve is set on the top of the electric desalination tank;
					the safety valve is checked regularly;
					there is a plan for the treatment of crude oil with water, and the plan drills regularly;
					the top pressure of the initial distillation column PI10401.
2	D-101A/B current	high	high electrical desalination level;	transformer trip;	D-101A/B current EI-10101A~C/EI-10201A~C display and alarm;
			the nature of crude oil becomes heavier;	the effect of electric desalting is poor.	electric desalination interface LI-10101/10102, LI-10201/10202 display, control, and alarm;
			water in crude oil;		regularly check the electrical desalination boundary.
			the D-101A/B current meter indicates low, and the electric desalination transformer gear is high;		
3	electric desalination pressure	High	the temperature of the crude oil entering the desalination tank is high;	the pressure of the electric desalination tank is high, the backpressure of the electric desalination water injection line and the demulsifier injection line is high, and the injection volume is reduced, which may cause backflow in severe cases;	electric desalination pressure PIC-10101/10201 display and alarm;
			the nature of crude oil becomes lighter;	the high pressure of the electric desalination tank causes the electrode plate and internal parts of the electric desalination tank to fall off, which affects the effect of the electric desalination;	inlet temperature of electric desalination TI-10102/TI-10201 display and alarm;
			serious water entrainment in crude oil;	the electric desalination tank is overpressured, and the equipment, manholes, and flange gaskets leak;	a safety valve is set on the top of the electric desalination tank;
			electric desalination pressure gauge shows low.	crude oil leakage encounters high temperatures or open flames and fires;	the safety valve is shall be calibrated regularly;
				the dirty oil enters the rainwater system and pollutes the environment.	install a check valve in front of the root valve of electric desalination water injection and demulsifier injection.
4	inlet temperature of electric desalination	high	the heat source temperature of the crude oil heat exchanger before removal is high or the flow rate is large;	large amount of crude oil gasification, high pressure of electric desalination tank;	inlet temperature of electric desalination TI-10102/TI-10201 display and alarm;
			the temperature of the crude oil entering the device is high;	increase device energy consumption;	the temperature of the crude oil entering the device is displayed on TI-11701;
			changes in crude oil properties;	cause damage to the internal components of the electric desalination tank;	a safety valve is set on the top of the electric desalination tank;
			sudden increase in crude oil feed;	electric desalination operation fluctuates and the effect of desalination and dehydration is poor; crude oil contains high salt and water content after removal, which intensifies equipment corrosion and affects product quality.	the safety valve shall be calibrated regularly.
			the inlet temperature meter of the electric desalination is low.		
5	C-101 top pressure of initial distillation column	high	the initial top naphtha return tower control valve FIC-10403/10404 fails to open;	the dry point of naphtha at the top of the primary distillation tower is low, which affects the quality of the product;	PI-10401 display and alarm for the pressure at the top of the C-101 column of the initial distillation column;
			the pressure of the liquid separation tank at the top of the initial distillation tower is high;	the liquid level at the bottom of the initial distillation tower is high, the load of P-101A/B is heavy, and the pump motor trips in serious occasions;	FIC-10403 display, control, and alarm for the flow rate of naphtha returning to the tower at the beginning;

Table 3. continued

no.	parameter Indicator	deviation	possible causes	consequences	protective measures
			the liquid level of the liquid separation tank at the top of the initial distillation tower is extremely high;	the overpressure of the initial distillation tower will cause the tower to flush in severe cases.	PIC-10901 displays, controls, and alarms the pressure of the liquid separation tank at the top of the initial distillation tower;
			after the top of the primary distillation tower is cooled, the temperature is high;		FIC-10404 display, control, and alarm for the flow rate of the initial top circulation tower;
			the feed volume of the initial distillation tower is increased;		after the top of the primary distillation tower is cooled, the temperature TI-10801 displays and alarms;
			the feed of the initial distillation tower has a lot of water;		regular laboratory testing of naphtha at the top;
			there is too much water in the reflux at the top of the initial distillation tower;		daily inspection on site;
			the circulating water temperature of the primary top cooler E-193A~D is high, the pressure is low, or the flow is interrupted.		set up fire-fighting facilities on site; on-site combustible gas alarm; set up cofferdams and have clean water and sewage diversion facilities.

a According to section 2.1.1.2, the HAZOP report in Table 3 is optimized and Table 4 is obtained.

$$\begin{cases} \Delta y_t = y_t - y_{t-1} = y_t - Ly_t = (1-L)y_t \\ \Delta^2 y_t = \Delta y_t - \Delta y_{t-1} \\ \quad = (1-L)y_t - (1-L)y_{t-1} \\ \quad = (1-L)^2 y_t \\ \vdots \\ \Delta^d y_t = (1-L)^d y_t \end{cases} \quad (4)$$

Among them, the relationship with the dth-order data sequence y_t is

$$X_t = \Delta^d y_t = (1-L)^d y_t \quad (5)$$

Then, X_t is a stable data sequence.

3. *Test the Granger causality.* According to the ARIMA model established by step 2, the process parameters of alarm are taken as the target sequence for Granger causality test. The process parameter sequence $\{x_{rt}\}$ is used here as an example to illustrate the Granger causality test process.

   (1) *Test data covariance stationarity and data preprocessing.* First of all, for $\{x_{rt}\},\{y_t\}$, the ADF test is performed to verify whether the covariance is stable; if the noncovariance of the time series is stable, the time series is processed by first-order difference. The first-order difference calculation is shown in eq 6:

   $$\nabla \omega_t = \omega_t - \omega_{t-1} \quad (6)$$

   where $\{\omega_t\}$ is the time series that needs difference operation, $\nabla \omega_t$ is the first-order difference of ω_t, and ω_{t-1} is a time series with one time unit difference.

   (2) *Construct the regression equation.* When studying whether the time series $\{x_{rt}\}$ is the Granger cause $\{y_t\}$, it is necessary to construct a regression equation containing the lag term of x_r and the lag term of y, as shown in eq 7:

   $$y_t = \sum_{i=1}^p \alpha_i x_{r(t-i)} + \sum_{i=1}^p \beta_i y_{t-i} + u_{1t} \quad (7)$$

   Next, the residual sum of squares of this regression equation (RSS_{UR}) is calculated. Then a regression equation of y for all lags y_{t-i} ($i = 1, ..., q$) and other variables in which the lag term $x_{r(t-i)}$ ($i = 1, ..., q$) of X is not included x_r is constructed, as shown in eq 8:

   $$y_t = \sum_{i=1}^p \beta_i y_{t-i} + u_{2t} \quad (8)$$

   Finally, the sum of squares of the residual errors of the regression equation (RSS_R) is calculated.

   (3) *Establish the null hypothesis and F test.* The null hypothesis $H_0: \alpha_i = 0$ ($i = 1, ..., q$) is established, even if $\{x_{rt}\}$ is not the Granger cause of $\{y_t\}$. The Granger causality of X and Y can be detected by the F test:

   $$F = \frac{\frac{RSS_R - RSS_{UR}}{q}}{\frac{RSS_{UR}}{n-k}} \sim F(q, n-k) \quad (9)$$

Table 4. HAZOP Report after Processing

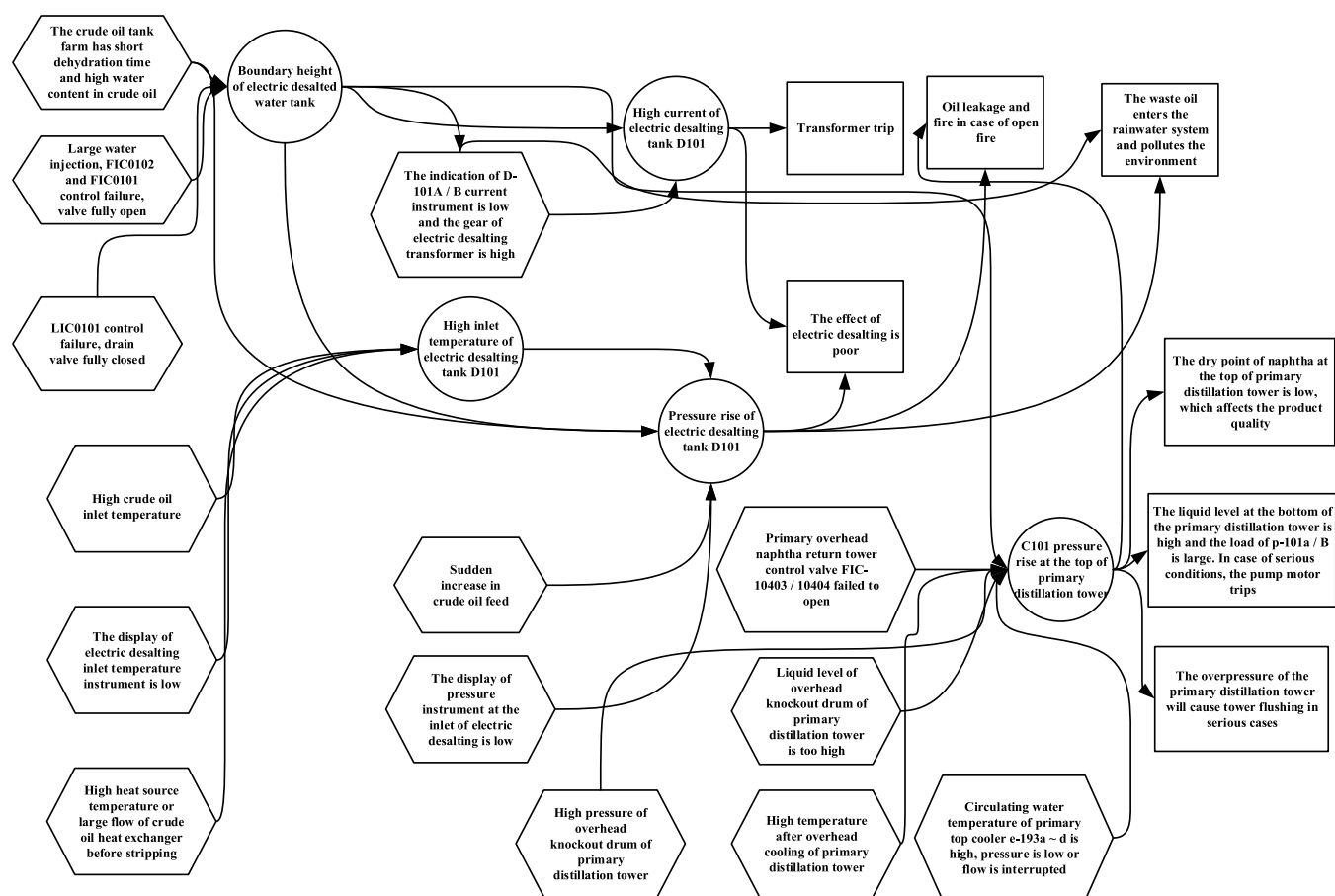| no.	deviation	alarm number	reason	alarm number	consequence	alarm number
1	high boundary of electric desalinated water tank	LI10101/10102, LI10201/10202	the crude oil tank farm has a short dehydration time and a high water content in crude oil	EI10101A/EI10201A	electric desalting has high boundary level and high current, which burns out the transformer	PDIC10102
			large water injection FIC10102, FIC10101 control failure, full valve		the initial distillation tower contains water, and the tower flushes, and the overpressure oil leaks in severe cases	PI10401
			LIC10101 control failure, the drain valve is fully closed		the electric desalination tank is overpressured, and the equipment, manholes, and flange gaskets leak; crude oil leaks and fires occur in case of open flames	PIC10101/PIC10201
					the dirty oil enters the rainwater system and pollutes the environment	
2	D-101A/B high current	EI10101	high electrical desalination level	LI10101	the electric desalination current is high and the transformer has tripped;	EI-10101A/EI10201A
			the nature of crude oil becomes heavier; crude oil with water	EI10101A/EI10201A	after electrical desalination, the water in crude oil is serious, causing fluctuations in the operation of the primary distillation tower and the stabilization tower	PI10401
			D-101A/B current meter indication is low, electric desalination transformer gear is high	EI-10101A		
3	high pressure in electric desalination tank	PIC10101	high temperature of crude oil into the electric desalting tank	TI-10102/TI-10201	the pressure of the electric desalination tank is high, the backpressure of the electric desalination water injection line and the demulsifier injection line is high, the injection volume is high, and the reverse flow is caused in severe cases.	PIC10101/PIC10201
			electric desalination pressure gauge shows low	PIC10101/PIC10201	the high pressure of the electric desalination tank causes the electrode plates and internal parts of the electric desalination tank to fall off, which affects the effect of the electric desalination.	PIC10101/PIC10201
			the nature of crude oil becomes lighter; Serious water entrainment in crude oil	EI10101A/EI10201A	electric desalination tank is overpressured; equipment, manholes, flange gaskets are leaking; crude oil leakage meets high temperature or open flames and fires; dirty oil enters the rainwater system, polluting the environment	PIC10101/PIC10201
			sudden increase in crude oil feed	FIC11703		
4	high inlet temperature for electrical desalination	TI10101	the heat source temperature of the crude oil heat exchanger before removal is high or the flow rate is large	TG11075	large amount of crude oil and gas, high pressure in the electric desalination tank	PIC10101/PIC10201
			the crude oil enters the device with high temperature	TI-11701	increase device energy consumption	
			electric desalination inlet temperature meter shows low	TII0101	damage to internal components of the electric desalting tank; the electric desalting operation fluctuates, the desalting and dehydration effect is poor, and the desalting crude oil has a high salt and water content, which intensifies the equipment corrosion and affects the product quality;	
5	high pressure at the top of the initial distillation column C-101	PI10404	the initial top naphtha return tower control valve FIC-10403/10404 failed to open	FIC-10403/10404	the dry point of naphtha at the top of the primary distillation tower is low, which affects the quality of the product	LI10102
			high pressure in the liquid separation tank at the top of the initial distillation tower	PIC-10901	the liquid level at the bottom of the initial distillation tower is high, the load of P-101A/B is heavy, and the pump motor trips at a serious time.	
			the liquid level in the top separation tank of the initial distillation tower is super high	LI10102	the initial distillation tower is overpressured, which will cause the flushing tower in severe cases	PI-10401
			after the top of the primary distillation tower is cooled, the temperature is high.	TI-10801		
			the circulating water temperature of primary top cooler E-193A~D is high, pressure is low or flow is interrupted	TI-10801, PI-10401		

Figure 11. Case relationship topology diagram.

This formula follows the F distribution of the degrees of freedom q and $(n - k)$, where n is the sample size and k is y_t that does not include the lag term $x_{r(t-i)}$ $(i = 1, ..., q)$ of x_r.

Next, the required significance level a is determined, and the F distribution table is checked to obtain the critical value F_a. If $F > F_a$, then the null hypothesis H_0 is rejected, indicating that the description $\{x_{rt}\}$ is the Granger cause of $\{y_t\}$, and the magnitude of the causality can be represented by the F value. $\{x_{rt}\}$ and $\{y_t\}$ for the Granger causality test process are shown in Figure 4.

Steps $(1)-(3)$ are repeated to test the Granger causality between the process parameter time series $\{x_{1t}\}$, $\{x_{1t}\}$, ..., $\{x_{rt}\}$, ..., $\{x_{mt}\}$ that may cause the alarm and the alarm process parameter time series $\{y_t\}$.

4. *Construct Granger causality diagrams between time series.* To show the Granger causality between time series more intuitively, researchers usually use Granger causality graphs to visualize the Granger causality between time series. A Granger causality graph is a kind of directed graph $G = \{V, E\}$, where each dot in the graph represents a time series, and the directed lines represent the causality, with the starting point denoting the cause variable and the ending point denoting the result variable, i.e., v_a points to v_b. The directed edge of v_a represents that the time series v_a Granger leads time series v_b. The magnitude of causality can also be marked next to the corresponding line. In

Figure 5, time series 1, 3, 4, and 5 are Granger results of time series 2, and time series 1, 2, 3, and 4 are Granger results of time series 5.

2.1.3. Revised Final Causal Network Model. As the result of HAZOP analysis has certain subjectivity and uncertainty, the topology map obtained may have redundancy and inconsistency problems, and it may also cause problems such as path inaccuracy and contradictions. Therefore, it is necessary to establish a preliminary causal relationship model (prototype) on the basis of the qualitative topological graph analyzed by HAZOP. And then, the causal relationship of error, redundancy, and omission in the prototype can be optimized and corrected by a sequential causal network so that the final diagnosis and tracing results can be more accurate and have a higher guiding significance for the actual situation. At the same time, when too many process parameters are selected, the established quantitative causal diagram will become more complicated and inconvenient for fault diagnosis. To improve the efficiency of reasoning, the topology diagram and the sequential causal network diagram are combined and simplified according to the following rules.

The specific steps are as follows:

1. Get the topology diagram through section 2.1.1.
2. Draw a sequential causal network diagram through section 2.1.2.
3. According to the results of Granger causality test, the time series causal network diagram obtained in step 2 is processed to a certain extent, including the following rules:
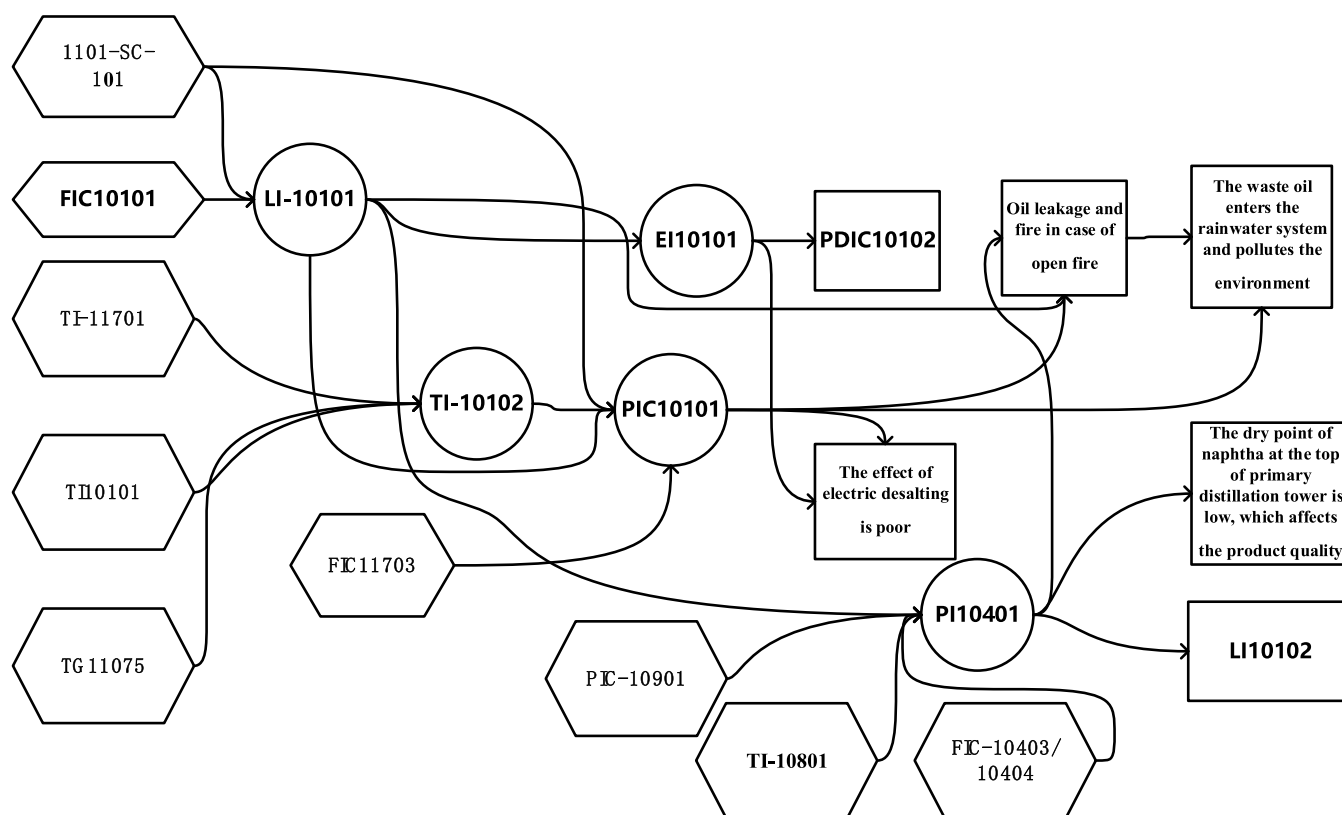
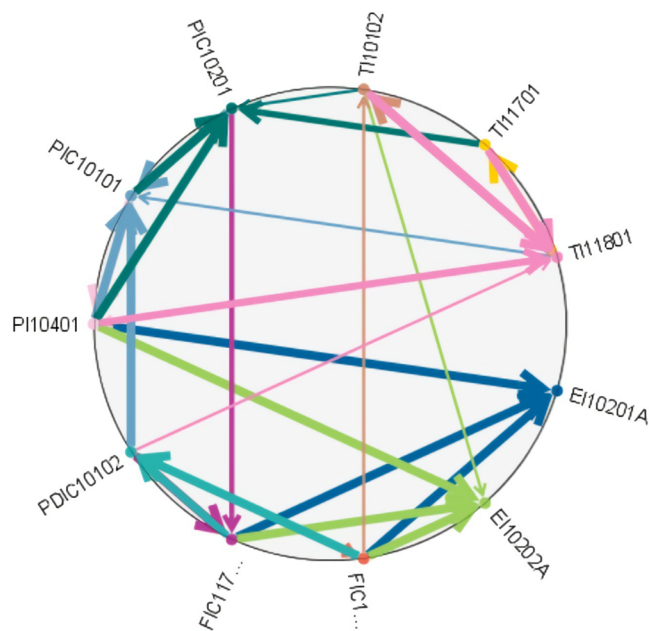**Figure 12.** Case location number topology diagram.

Figure 13. Time causal network diagram (spherical).

(1) On the basis of analyzing the interaction and influence relationship between process parameters, delete routes with less than the tangent point threshold in the quantitative causality diagram. Specifically, the cutoff point threshold first selects the maximum value in the alternative and then gradually decreases. When and only when the causality graph is programmed from the non-connected graph to the connected graph (i.e., any

two points in the graph are connected by paths), the cutoff point threshold stops decreasing.

(2) Judge whether there are cascade control loop points in the diagram. If yes, combine the points representing the same cascade control loop into one point, regardless of the causal relationship between them. The causal relationship between them and other points (including other merging points and ordinary control loop points) shall be integrated according to the following principles: the causal relationship value in a certain direction between merging points and ordinary control loop points shall be the maximum value of the causal relationship value in that direction between each point in merging points and this ordinary control loop point. For the causal relationship between merging point A and merging point B in a certain direction, take the maximum value of the causal relationship between each point in merging point A and each point in merging point B in this direction.

(3) The line can be marked according to the results of Granger causality test, and the size of the value can be expressed by the thickness of the line.

4. Match the time series causal network with the topology diagram, including variables and relationship lines. If the topological diagram is missing, add it according to the sequential causal network.

2.2. Traceability Mechanism. When a fault occurs in the chemical process, a large influx of alarm information forms the time series of events. These alarm messages have temporal constraints, that is, timing attributes. The timing attribute is an important attribute of the alarm information, which contains

rich fault-related information. In this paper, the time causal network is used as the rule and time constraint to construct the abductive reasoning network to comprehensively realize the diagnosis of alarm information.

2.2.1. Overview of Abductive Reasoning. The abduction reasoning can be traced back to Aristotle and was put forward by the modern American philosopher Peirce. Abductive reasoning is a kind of reasoning form of "from result to cause", which forms a hypothesis and then explains fact through explanatory reasoning. Its logical form is: (1) Observe a surprising phenomenon C; (2) If A is true, then C is a self-evident fact; (3) Therefore, there are reasons to believe that A may be true. The abduction reasoning is carried out mainly according to reasoning rules and related constraints. In abductive reasoning, rules are used to describe the causal relationship between events, and the constraints between events are described by formula constraints (FC). A complete abductive reasoning rule includes cause events, result events, and constraints between events.

2.2.2. Basis of Abductive Reasoning Based on Time.
2.2.2.1. Language Description Based on Temporal Abductive Reasoning. This paper adopts the concept of alarm time zone. Compared with the occurrence time of the original fault, the alarm time of other process parameters should be distributed in the corresponding time interval. Moreover, the time constraint relationship between the following events is defined: (1) $\sigma^- = (T_1, T_2, T_a, T_b)$ and (2) $\sigma^+ = (T_1, T_2, T_a, T_b)$, where σ^- indicates the time relationship of "earlier than", σ^+ indicates the time relationship of "later than", T_1 and T_2 represent the times when events 1 and 2 occur, respectively, and T_1 and T_2 satisfy the conditions $T_a \leq T_2 - T_1 \leq T_b$.

In the alarm information diagnosis of chemical process system, the relationship between events mainly includes two types, namely, event physical logic rules (inference rules) and time constraints. This article defines equipment failure as the source event, which is both the origin of the entire failure event and the end of the inference result. The abductive reasoning proposed in this paper is to select the appropriate subset from the inference rule set R according to the alarm information set A so that the source event and the alarm information set A meet the corresponding rule constraints.

In this abductive reasoning, "→" is used to describe the causal relationship between events. For event sets A and B, "$A \rightarrow B$" means that A explains B, or the occurrence of A directly leads to the occurrence of B. The causal relationship with constraints can be described as $(A_1, ..., A_n) \rightarrow (B_1, ..., B_m)$, $C(t_{A_1}, ..., t_{A_n}, t_{B_1} ... t_{B_m})$, where A_i and B_j represent events and $C(t_{A_1}, ..., t_{A_n}, t_{B_1}, ..., t_{B_m})$ represents the time constraints on the occurrence of events.

This article defines the basic event description language: $\Omega_{\text{start}}(S,T)$ indicates that the fault S fails at time T; $\Omega_{\text{between}}(B,T)$ indicates that an alarm occurs at the intermediate point A at time T; and $\Omega_{\text{end}}(E,T)$ indicates that the final point P gives an alarm at time T.

Assuming that the alarm time caused by fault transmission is $[1, 6]$ min, the figure shows the intermediate and final alarm retrospective reasoning rules in the case of a certain equipment failure. Supposing that the occurrence time of the fault S_1 is T_1, the alarm time of the intermediate point B_1 is T_2 and the alarm time of the final point E_1 is T_3. After the failure of S_1, the fault process that causes B_1 and E_1 alarms can be expressed as

$$\Omega_{\text{start}}(S_1, T_1) \rightarrow \Omega_{\text{between}}(B_1, T_2) \rightarrow \Omega_{\text{end}}(E_1, T_3),$$

$$\sigma^- = (T_1, T_2, 1 \text{ min}, 6 \text{ min}),$$

$$\sigma^- = (T_2, T_3, 1 \text{ min}, 6 \text{ min})$$

The rule description language can be expressed as

$$R_1: \quad \Omega_{\text{start}}(S_1, T_1) \rightarrow \Omega_{\text{between}}(B_1, T_2),$$

$$\sigma^- = (T_1, T_2, 1 \text{ min}, 6 \text{ min})$$

$$R_2: \quad \Omega_{\text{between}}(B_1, T_2) \rightarrow \Omega_{\text{end}}(E_1, T_3),$$

$$\sigma^- = (T_2, T_3, 1 \text{ min}, 6 \text{ min})$$

The reasoning rules in this article can be directly obtained from the revised topology diagram obtained in section 2.1, that is, the causal relationship between events.

2.2.2.2. Processing Method of Time Constraint. In fault diagnosis, the time constraint between events can be expressed by the temporal constraint network (TCN). TCN is a directed graph whose vertices are the moments when events occur, and the vertices are connected by directed edges to describe the uncertain time distance constraints between events.

Figure 7 is a time-constrained network of the rules shown in Figure 6. Supposing the time reference point X_0 is 0, the distance of T_1 relative to the reference point 0 is t_0.

In the time constraint problem, any time constraint can be represented by the interval $[a_{ij}, b_{ij}]$. That is to say, supposing that events i and j occur successively at t_i, t_j ($t_i \leq t_j$) and event i is a precursor to event j, the time points when the two events occur should meet their time constraints. That is, there is $a_{ij} \leq t_j - t_i \leq b_{ij}$. The time constraint problem at this time is called simple temporal problems (STP).

For STP, to check whether a certain alarm information meets the requirements of time consistency, the received alarm information needs to be checked pairwise with other alarm information in its corresponding minimum network M (in M, $(T_x, T_y) = [a_{ij}, b_{ij}]$ means $a_{ij} \leq T_y - T_x \leq b_{ij}$). Check whether it meets the specified time constraint. If the time constraint between the alarm information and all other alarm information is satisfied, the information is said to meet the time consistency in this alarm. If the alarm information does not meet the time constraint between all alarm information except itself constraint, it is said that the alarm information does not meet the time consistency in this alarm.

STP can be represented by a weighted directed graph $G_d = \{V, E_d\}$, where V represents a set of vertices and E_d is a set of directed edges with weights. G_d is also called a distance graph. The weight of the branch from i to j is b_{ij}, and the weight of the branch from j to i is $-a_{ij}$.

For STP, the sufficient and necessary condition for time consistency is that there is no negative ring in the distance graph G_d (i.e., the sum of the branch weights in the directed loop in the graph is less than 0). In the problem of abductive reasoning, because the corresponding event satisfies the causal relationship, there will be no event loop, and any simple loop on G_d satisfies $b_{ij} - a_{ij} \geq 0$, and there will be no negative loop, so it meets the requirements of time consistency.

The minimum network of STP is the minimum network matrix M, describing the distance graph G_d, $M_{ij} = \{[-d_{ji}, d_{ij}]\}$, $\forall i,j$. Among them, d_{ji} and d_{ij}, respectively, represent the shortest distance between any two nodes i and j on the distance graph G_a, which can be solved by the Floyd–Warshall algorithm. For the
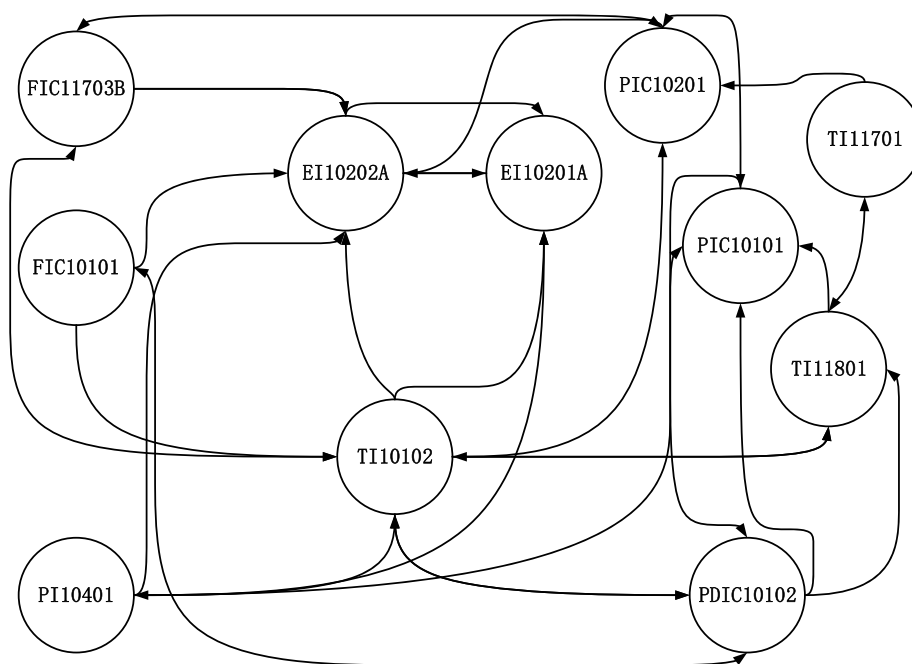
Figure 14. Time causal network model diagram (expanded topology).
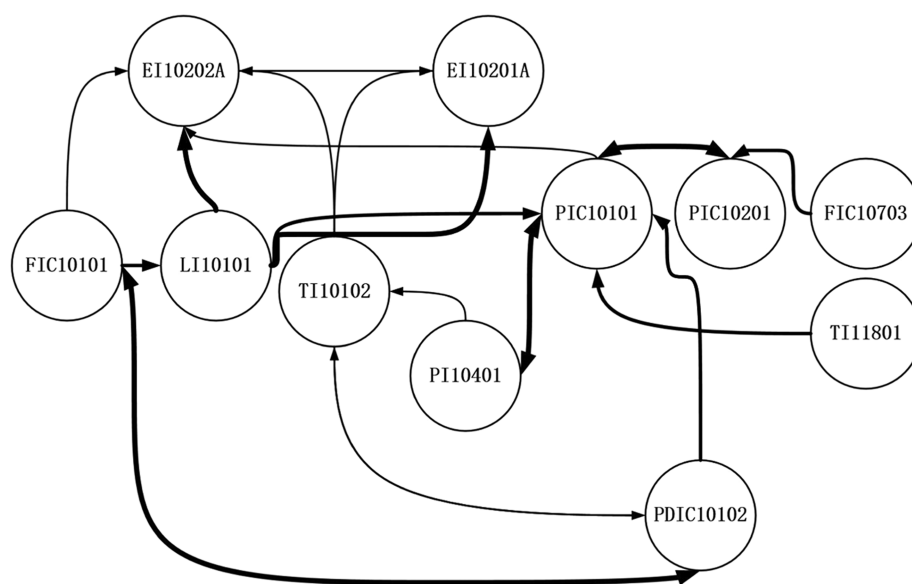

Figure 15. Granger causality correction time causal network diagram (expanded topology).

time constraint problem described in Figure 7, the distance graph is used to describe it, as shown in Figure 8.

If the observed values at the time of occurrence of several events in $T_0 - T_5$ are known, we can: (1) infer the size of t_0, that is, the time when the fault S_1 occurs; (2) infer the time range of the missing event; and (3) determine whether the observed values meet the time consistency.

2.2.3. Process of Abductive Reasoning Based on Time. The related definitions of the abduction reasoning algorithm in this article are as follows:

(1) **TCN**. According to the reasoning rules, the alarm information expressed in the form of the smallest network is a time-constrained network.

(2) *A*. The collection of received alarm information *m*.

(3) *S*. The candidate alarm information set is a set composed of alarm information S_i that may form a causal relationship with the information *m* to be diagnosed. Through rule matching, all of the rules whose source event is **m** are found, and the "result" information generated by these rules is R_i.

(4) *X*. The diagnostic information collection is composed of a series of alarm information to be diagnosed that meet the mutual time constraint relationship.

(5) *R*. The set of causal rules corresponding to the set *X* of diagnostic alarm information.

(6) Source event. Diagnose the cause of the event, which in this context refers to the original failure.

The algorithm uses a recursive method for diagnosis. The specific process description is shown in Figure 9.
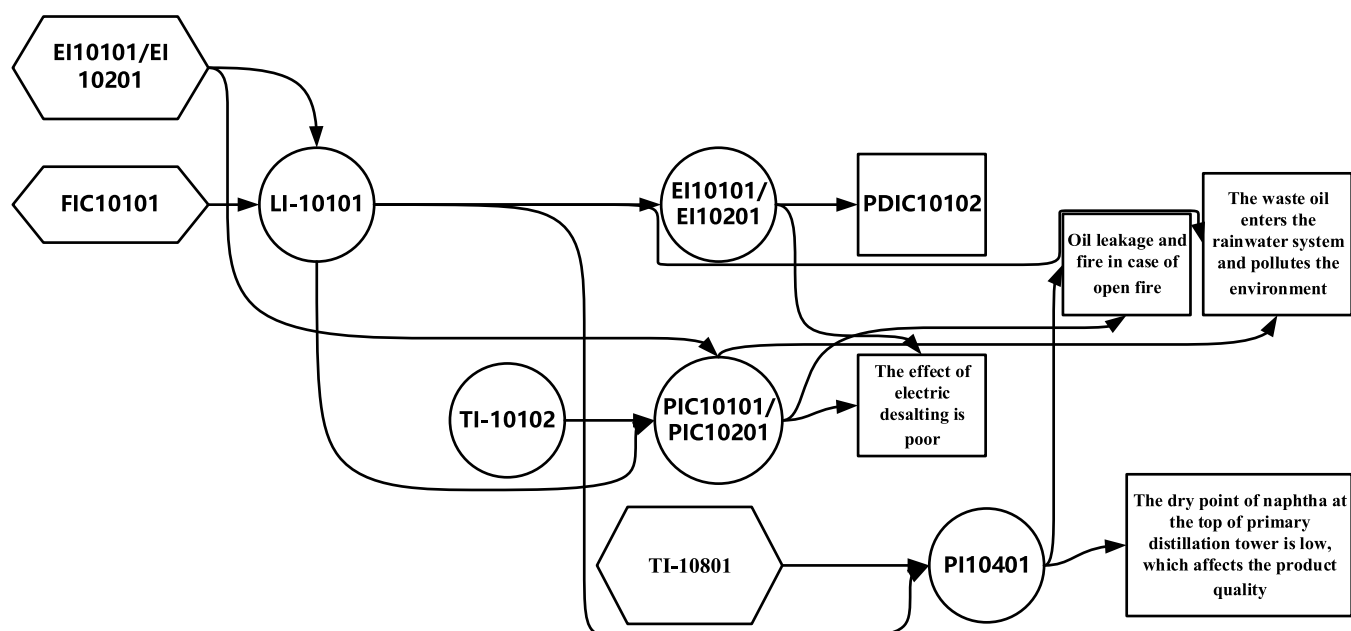
Figure 16. Final time causal network model.

Table 5. HAZOP Report after Processing

no.	alarm time	alarm tag number	alarm description	alarm device
T_0	02/09/20 06:43:35.386	EI10101	high high alarm	1101 atmospheric and vacuum distillation unit
T_1	02/09/20 06:44:37.392	LI10101	high alarm	1101 atmospheric and vacuum distillation unit
T_2	02/09/20 06:44:38.363	EI10101	high alarm	1101 atmospheric and vacuum distillation unit
T_3	02/09/20 06:48:52.370	PIC10101	high alarm	1101 atmospheric and vacuum distillation unit
T_4	02/09/20 06:49:12.375	PI10401	high alarm	1101 atmospheric and vacuum distillation unit

2.3. Implementation of Alarm Root Cause Diagnosis.

2.3.1. Accurate Location of Accident Chain and Root Cause. Starting from the final modified parameter causality topological diagram as the reasoning rule, and starting from the process parameters of alarm occurrence, we can find the path with the largest causal relationship in the graph and check whether the path meets the time consistency test in the traceability mechanism. If so, then the path is the propagation path of the fault in the system, and the process parameter at the end point is the root cause of the fault, as shown in Figure 9.

2.3.2. Design of Assistant Decision. After the root cause of the fault is found, the root cause is clearly displayed on the screen. At the same time, relevant keywords are automatically searched in the operation log, inspection record, emergency plan, and other databases, and the operation scheme will be displayed in the reminder box of the screen.

3. CASE STUDIES AND DISCUSSIONS

Based on the above research, the "Alarm Root Cause Diagnosis System Module" was developed and applied to a 10 million tons/year atmospheric and vacuum unit in a refinery. The system monitors the process operation status of the device in real time. When a device failure is found, the root cause of the alarm can be judged from multiple types of alarm information in time, and auxiliary decision-making prompts can be carried out. In this case, the electric desalination process part of the atmospheric and vacuum unit is intercepted, as shown in Figure 10. After the crude oil is sent to the unit from the raw oil pump outside the unit, it enters the crude oil-primary top oil-gas heat exchanger in four routes for heat exchange. Then, these four
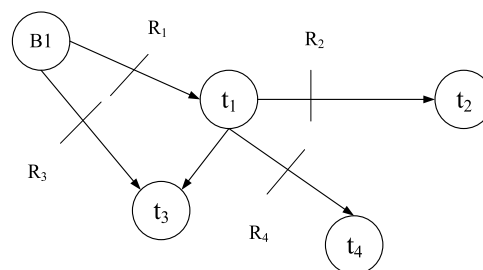

Figure 17. Example abductive reasoning rules.

routes are combined into two routes for heat exchange. After the two routes of crude oil are combined, the heat exchange temperature is 131 °C, which enters the electric desalting tank for desalination and dehydration. After that, the crude oil is divided into two routes to enter the post desalting crude oil heat exchange system. After the heat exchange, the two-way crude oils after removal are combined with a temperature of 233 °C and enter the initial distillation tower. The overhead gas of the primary distillation tower is sent to the crude oil-primary overhead oil and gas heat exchanger. After heat exchange, it is sent to the primary overhead air cooler and primary overhead water cooler to cool to 40 °C. Finally, it is sent to the primary overhead reflux and product tank. The specific process is shown in Figure 10.

3.1. Establishment of a Causal Network Model.

3.1.1. Establishment of Topology Based on HAZOP. First, select the relevant HAZOP report in this electric desalination process from the latest version of the HAZOP report of the

The Diagnostic Results

The Root Cause: The dehydration time of crude oil tank is too short and the water content of crude oil is high.
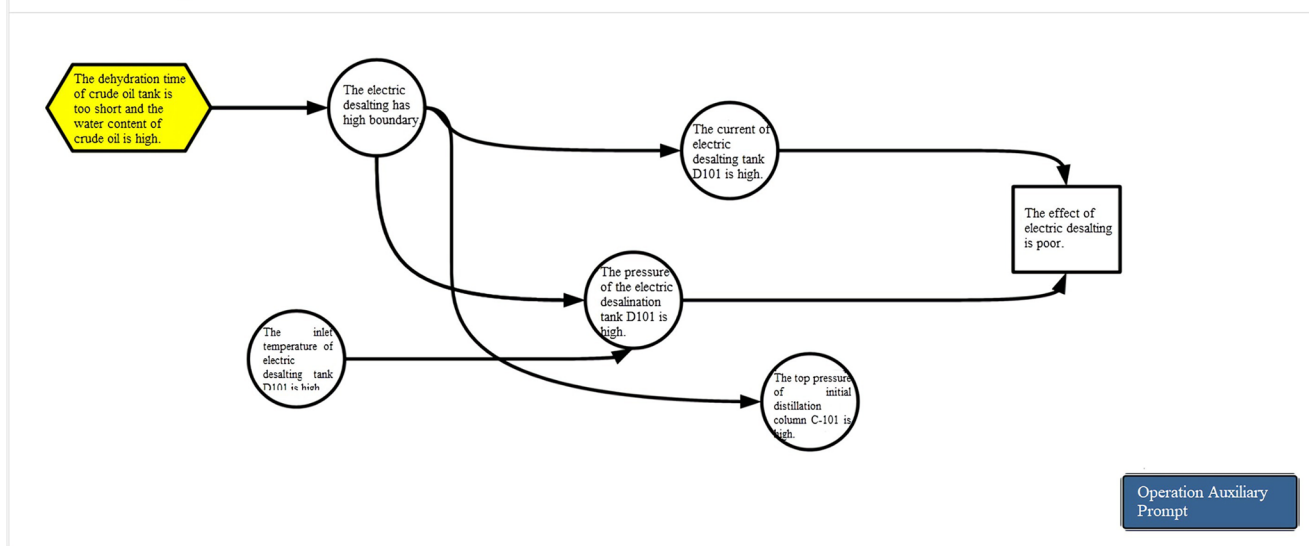
The Chain Diagram of Fault Path

Figure 18. Final diagnostic alarm results.

Figure 19. Final auxiliary decision.

refinery (due to the length of the report, only the relevant content of the analysis is retained), as shown in Table 3.

The correlation between the relevant process deviations in Table 4 is extracted, and the relationship topology diagram shown in Figure 11 and the related parameter topology diagram in Figure 12 are obtained.

3.1.2. Establish Time Cause and Effect Network Diagram. In Table 4, 18 relevant process parameters are extracted, and the relevant time series data are extracted from DCS respectively. The corresponding ARIMA model for each parameter is established. For the ARIMA model of process parameters that may cause alarms, the Granger causality test is performed
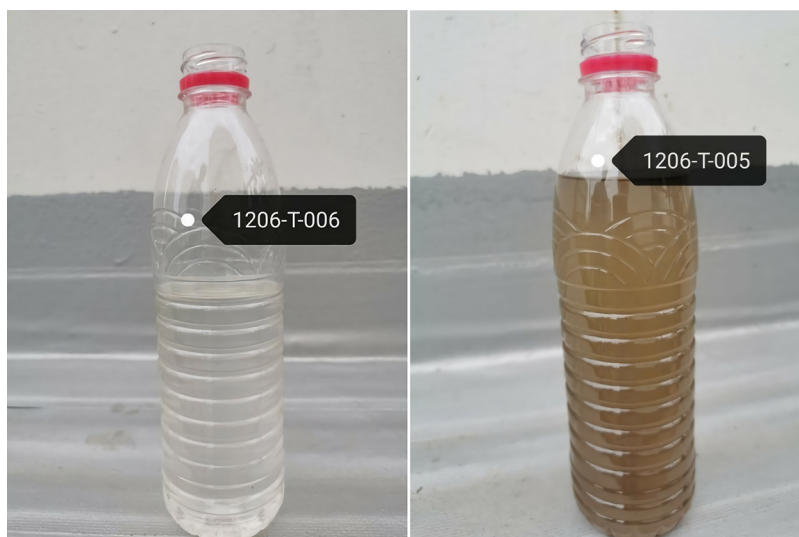
Figure 20. Comparison of the appearance of the first line before and after crude oil with water.

pairwise. The Granger causality between ARIMA models is shown in Figures 13 and 14.

3.1.3. Final Time Causal Network Model Diagram. According to the correction method of causal network model in section 2.1.3, the corrected time causal network diagram can be obtained as shown in Figures 15 and 16.

3.2. Results of Traceability Mechanism. The alarm records provided by the enterprise are as shown in Table 5. According to section 2.2.3 abductive reasoning algorithm, the reasoning rules shown in Figure 17 can be generated from Figure 16. Abductive reasoning is carried out according to the algorithm proposed in section 2.2.3, and the final diagnostic alarm results are obtained, as shown in section 3.3.

3.3. Final Result of Alarm Root Cause Diagnosis. The diagnosis result of the root cause of the alarm is that the dehydration time in the crude oil tank farm is short and the water content of crude oil is high. Accurate positioning of the accident chain and root cause is displayed in the system, as shown in Figure 18. The auxiliary decision-making interface is given in the system, as shown in Figure 19. The accident records of the enterprise show that the accident process is that around 6:00 on September 2, and the current of the electric desalination tank began to rise rapidly. The internal operation judged that the crude oil was carrying water seriously, immediately reduced the water injection volume, increased the water cutting of the electric desalination, required on-site monitoring of the total cut water with oil of the electric desalination, and timely informed the public works department to strengthen the monitoring. The reason is analyzed as follows: The incomplete dehydration of tank 2# of Basra crude oil is the root cause of this water-carrying crude oil incident.

The actual accident treatment records are consistent with the final results obtained by this method (Figure 20). It can be concluded that this method can accurately judge that the root cause of the alarm in a large number of alarms is crude oil with water and can provide auxiliary decision making to help operators make accurate judgments and quickly take action to shorten the time for root cause diagnosis. The method is scientific and accurate.

4. CONCLUSIONS AND PROSPECT

To suppress the phenomenon of alarm flooding and help operators accurately and quickly judge the root cause of the alarm when a large number of alarms pour in, this article provides a promising method. The conclusions are as follows:

(1) To obtain a more accurate and scientific causal model, a comprehensive use of the data-knowledge-driven method is proposed. In the application of the knowledge-driven method, we proposed a novel method using the HAZOP report as the basis to establish the industrial topology model of chemical process. And its transmission relationship is relatively clear and practical. On this basis, a data-driven method based on ARIMA and Granger causality test is added to make the results more reliable.

(2) The time abduction reasoning method is used as the traceability analysis mechanism. It can use the alarm timestamp record more scientifically so as to find the root cause of the alarm quickly and accurately.

(3) The alarm root cause diagnosis system is designed, which can display the alarm root cause and fault path, and provide auxiliary prompt for personnel action. The interface is very clear and friendly to operators.

However, some potential limitations should be noted and some problems should be intentionally considered. Therefore, future research should be undertaken to explore:

(1) One concern about the technology was that the knowledge-driven approach is based on high-quality, reliable, and complete HAZOP reports. If the quality of HAZOP report is not up to standard, it will have a certain impact on it. Therefore, the quality of HAZOP should be reviewed before it is used for alarm root diagnosis. In the future work, the rapid inspection of HAZOP report quality should be developed as a front-end module.

(2) The system can be used not only to provide auxiliary operation tips for DCS operators but also for safety engineers to train and test their fast and accurate characteristics of distinguishing accident causes. In view of the current phenomenon of alarm flooding in chemical enterprises, it is expected that the application range of the system is very wide. In the future, it is necessary to consider different systems of DCS and alarm management

software of different enterprises and make more scientific and reasonable adjustment and development.

■ AUTHOR INFORMATION

Corresponding Author

Dongfeng Zhao − *Center for Chemical Process Safety—China Section and College of Chemical Engineering, China University of Petroleum (East China), Qingdao, Shandong 266580, China;* ⊙ orcid.org/0000-0003-0394-4819; Email: zhaodf@vip.sina.com

Authors

Xiaomiao Song − *College of Mechanical and Electrical Engineering, China University of Petroleum (East China), Qingdao, Shandong 266580, China*

Qinglong Liu − *Qingdao OASIS Environmental & Safety Technology Co., Ltd., Qingdao, Shandong 266580, China*

Mingxin Dong − *College of Mechanical and Electrical Engineering, China University of Petroleum (East China), Qingdao, Shandong 266580, China*

Yifei Meng − *Center for Chemical Process Safety—China Section and College of Chemical Engineering, China University of Petroleum (East China), Qingdao, Shandong 266580, China*

Chuanrui Qin − *College of Mechanical and Electrical Engineering, China University of Petroleum (East China), Qingdao, Shandong 266580, China*

Fabo Yin − *Qingdao OASIS Environmental & Safety Technology Co., Ltd., Qingdao, Shandong 266580, China*

Jiangbo Jiu − *Qingdao OASIS Environmental & Safety Technology Co., Ltd., Qingdao, Shandong 266580, China*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acsomega.2c01529

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) *ISA-18.2: Management of Alarm Systems for the Process Industries*; International Society of Automation: Durham, NC, 2009.

(2) Goel, P.; Pistikopoulos, E.; Mannan, M.; Datta, A. A data-driven alarm and event management framework. *J. Loss Prev. Process Ind.* **2019**, *62*, No. 103959.

(3) Wang, J.; Zhao, Y.; Bi, Z. Criteria and algorithms for online and offline detections of industrial alarm floods. *IEEE Trans. Control Syst. Technol.* **2018**, *26*, 1722−1731.

(4) Pasman, H. J. Early warning signals noticed, but management doesn't act adequately or not at all: a brief analysis and direction of possible improvement. *J. Loss Prev. Process Ind.* **2021**, *70*, No. 104272.

(5) Wang, J.; Yang, F.; Chen, T.; Shah, S. L. Engineering. An overview of industrial alarm systems: Main causes for alarm overloading, research status, and open problems. *IEEE Trans. Control Syst. Technol.* **2015**, *13*, 1045−1061.

(6) Adeel, M.; Shang, C.; Zhu, K.; Lu, C. J. O. E. Nuisance alarm reduction: using a correlation based algorithm above differential signals in direct detected phase-OTDR systems. *Opt. Express* **2019**, *27*, 7685−7698.

(7) Kaced, R.; Kouadri, A.; Baiche, K.; Bensmail, A. Multivariate nuisance alarm management in chemical processes. *J. Loss Prev. Process Ind.* **2021**, *72*, No. 104548.

(8) Aslansefat, K.; Gogani, M. B.; Kabir, S.; Shoorehdeli, M. A.; Yari, M. Performance evaluation and design for variable threshold alarm systems through semi-Markov process. *ISA Trans.* **2020**, *97*, 282−295.

(9) Cheng, Y.; Izadi, I.; Chen, T. design. Pattern matching of alarm flood sequences by a modified Smith−Waterman algorithm. *Chem. Eng. Res. Des.* **2013**, *91*, 1085−1094.

(10) Hu, W.; Wang, J.; Chen, T. A local alignment approach to similarity analysis of industrial alarm flood sequences. *Control Eng. Pract.* **2016**, *55*, 13−25.

(11) Bouillard, A.; Buob, M.-O.; Raynal, M.; Salaün, A. Log Analysis via Space-Time Pattern Matching. In *2018 14th International Conference on Network and Service Management (CNSM)*; IEEE, 2018; pp 303−307.

(12) Xu, Y.; Fan, C.; Zhu, Q.-X.; Rajabifard, A.; Chen, N.; Chen, Y.; He, Y.-L. Novel pattern-matching integrated KCVA with adaptive rank-order morphological filter and its application to fault diagnosis. *Ind. Eng. Chem. Res.* **2020**, *59*, 1619−1630.

(13) Li, W.; Gu, S.; Zhang, X.; Chen, T. Pattern matching and active simulation method for process fault diagnosis. *Ind. Eng. Chem. Res.* **2020**, *59*, 12525−12535.

(14) Li, G.; Qin, S. J.; Yuan, T. Data-driven root cause diagnosis of faults in process industries. *Chemom. Intell. Lab. Syst.* **2016**, *159*, 1−11.

(15) Zhu, Y.; Patwardhan, R.; Wagner, S. B.; Zhao, J. Toward a low cost and high performance MPC: The role of system identification. *Comput. Chem. Eng.* **2013**, *51*, 124−135.

(16) Xie, L.; Kruger, U.; Lieftucht, D.; Littler, T.; Chen, Q.; Wang, S.-Q. Statistical monitoring of dynamic multivariate processes part 1. modeling autocorrelation and cross-correlation. *Ind. Eng. Chem. Res.* **2006**, *45*, 1659−1676.

(17) Yuan, T.; Qin, S. J. Root cause diagnosis of plant-wide oscillations using Granger causality. *J. Process Control* **2014**, *24*, 450−459.

(18) Xu, S.; Baldea, M.; Edgar, T. F.; Wojsznis, W.; Blevins, T.; Nixon, M. Root cause diagnosis of plant-wide oscillations based on information transfer in the frequency domain. *Ind. Eng. Chem. Res.* **2016**, *55*, 1623−1629.

(19) Luo, Y.; Gopaluni, B.; Xu, Y.; Cao, L.; Zhu, Q.-X. A novel approach to alarm causality analysis using active dynamic transfer entropy. *Ind. Eng. Chem. Res.* **2020**, *59*, 8661−8673.

(20) Amin, M. T.; Khan, F.; Ahmed, S.; Imtiaz, S. A data-driven Bayesian network learning method for process fault diagnosis. *Process Saf. Environ. Prot.* **2021**, *150*, 110−122.

(21) Amin, M. T.; Khan, F.; Ahmed, S.; Imtiaz, S. A novel data-driven methodology for fault detection and dynamic risk assessment. *Can. J. Chem. Eng.* **2020**, *98*, 2397−2416.

(22) Chen, H.-S.; Yan, Z.; Yao, Y.; Huang, T.-B.; Wong, Y.-S. Systematic procedure for Granger-causality-based root cause diagnosis of chemical process faults. *Ind. Eng. Chem. Res.* **2018**, *57*, 9500−9512.

(23) Lindner, B.; Auret, L.; Bauer, M.; Groenewald, J. W. Comparative analysis of Granger causality and transfer entropy to present a decision flow for the application of oscillation diagnosis. *J. Process Control* **2019**, *79*, 72−84.

(24) Liu, Y.; Chen, H.-S.; Wu, H.; Dai, Y.; Yao, Y.; Yan, Z. Simplified Granger causality map for data-driven root cause diagnosis of process disturbances. *J. Process Control* **2020**, *95*, 45−54.

(25) Ghosh, A.; Ahmed, S.; Khan, F. Modeling and testing of temporal dependency in the failure of a process system. *Ind. Eng. Chem. Res.* **2019**, *58*, 8162−8171.

(26) Zhang, K.; Kalander, M.; Zhou, M.; Zhang, X.; Ye, J. An Influence-Based Approach for Root Cause Alarm Discovery in Telecom Networks. In *International Conference on Service-Oriented Computing*; Springer, 2020; pp 124−136.

(27) Yang, F.; Xiao, D. Progress in root cause and fault propagation analysis of large-scale industrial processes. *J. Control Sci. Eng.* **2012**, *2012*, No. 478373.

(28) Parsa, K.; Hassall, M.; Naderpour, M. Process alarm modeling using graph theory: Alarm design review and rationalization. *IEEE Syst. J.* **2021**, *15*, 2257−2268.

(29) Xu, X.; Xu, H.; Wen, C.; Li, J.; Hou, P.; Zhang, J. A belief rule-based evidence updating method for industrial alarm system design. *Control Eng. Pract.* **2018**, *81*, 73−84.

(30) Milana, D.; Darena, M. S.; Bettio, N.; Cerruti, C.; Siliprandi, G.; Fidanzi, A.; Cerioli, P.; Silvestri, G.; Tarasconi, F.; Caserio, M. Natural Language Understanding for Safety and Risk Management in Oil and Gas Plants. Presented at the Abu Dhabi International Petroleum Exhibition & Conference, Abu Dhabi, UAE, November 2019. DOI: 10.2118/197668-MS.

(31) Wang, J.; Li, H.; Huang, J.; Su, C. Association rules mining based analysis of consequential alarm sequences in chemical processes. *J. Loss Prev. Process Ind.* **2016**, *41*, 178−185.

(32) Aziz, A.; Ahmed, S.; Khan, F. I. An ontology-based methodology for hazard identification and causation analysis. *Process Saf. Environ. Prot.* **2019**, *123*, 87−98.

(33) Zhu, Q.-X.; Meng, Q.-Q.; Wang, P.-J.; He, Y.-L. Novel causal network modeling method integrating process knowledge with modified transfer entropy: a case study of complex chemical processes. *Ind. Eng. Chem. Res.* **2017**, *56*, 14282−14289.

(34) Jeon, S. H.; Cho, S. J. Named Entity Normalization Model Using Edge Weight Updating Neural Network: Assimilation Between Knowledge-Driven Graph and Data-Driven Graph. *arXiv (Computer Science.Artificial Intelligence)*, June 14, 2021, 2106.07549, ver. 1. https://arxiv.org/abs/2106.07549 (accessed 2022-03-14).

(35) Ericson, C. A. *Hazard Analysis Techniques for System Safety*; John Wiley & Sons, 2015.

(36) Zhu, Q.-X.; Luo, Y.; He, Y.-L. Novel multiblock transfer entropy based Bayesian network and its application to root cause analysis. *Ind. Eng. Chem. Res.* **2019**, *58*, 4936−4945.

(37) Yuan, F.; Guo, J.; Xiao, Z.; Zeng, B.; Zhu, W.; Huang, S. A transformer fault diagnosis model based on chemical reaction optimization and twin support vector machine. *Energies* **2019**, *12*, 960.

(38) Li, Y.; Li, T.; Yi, X.; Liu, H. Expert System for Comprehensive Transmission Fault Diagnosis Based on Oil Sensor. In *2020 International Conference on Sensing, Diagnostics, Prognostics, and Control (SDPC)*; IEEE, 2020; pp 49−54.

(39) Shang, L.; Wang, L.; Niu, L.; Zhang, Y.; Li, H.; Ma, M.; Zhou, B. Finite State Machine Model of Fault Diagnosis for Distribution System under Time Sequence Constraints. In *2020 IEEE/IAS Industrial and Commercial Power System Asia (I&CPS Asia)*; IEEE, 2020; pp 982−986.

(40) Ye, L.; Ma, X.; Wen, C. Rotating Machinery Fault Diagnosis Method by Combining Time-Frequency Domain Features and CNN Knowledge Transfer. *Sensors* **2021**, *21*, 8168.