Risk assessment of equipment at gas storage depots based on clustering and principal component analysis

Jing Zhang 1, Jinlei Yuan 1, Renyuan Wang 1, Mingxing Luo 1, Yang Chen 1, Shoucheng Wang 1 and Yujie Peng 2, *

1 SINOPEC Northwest Oil Field Company, China
2 College of Safety and Ocean Engineering, China University of Petroleum, Beijing, 102249, China

*Corresponding author e-mail: pyj@student.cup.edu.cn

Abstract. In China, the number of gas storage depots under construction and planning has increased in recent years. The scale of construction is moving towards large-scale development, increasing the safety risk. This study analyzed the risk to ground equipment at gas storage depots by examining the site conditions of the Wen 96 gas storage depot. Furthermore, this paper proposed a calculation method to semi-quantitatively assess the risk implications to ground process equipment at gas storage depots and a quantitative model for the failure consequences using the API581 method. The API581 risk failure consequence analytical data was used to propose comprehensive risk assessment methods for process equipment at gas storage depots, including cluster and principal component analyses. Categories and equipment with principal component characteristics were suggested that considered the risk impact factors and rankings of vital equipment, improving the quantification of the API581 method. This work enhanced the quantitative risk evaluation accuracy when assessing equipment. Risk mitigation measures are provided according to the revised risk ranking, which is highly significant to the safe operation of gas storage depots.

1. Introduction

The Chinese energy development plan clearly states that by 2030, the effective working gas volume in underground Chinese gas storage depots must reach 35 billion cubic meters. To date, the effective working gas volume is only about 10 billion cubic meters, necessitating many additional large gas storage depots, the construction, and planning of which have been increasing lately. Gas storage depots generally have extended service lives and involve complex ground equipment for complicated, low-temperature processes, such as desulfurization, dehydration, and dehydrocarbonization, continuously increasing the operational safety risk. Therefore, research involving the ground risk assessment of underground gas storage depots is essential to improve operational safety and risk management.

Existing gas storage depots can be divided into four types according to different geological structures: depleted oil and gas storage depots, aquifer structural gas storage depots, salt cavern gas storage depots, and pit gas storage depots [1]. Of these, depleted gas storage depots account for the highest proportion due to their large reserves and low cost [2]. In this study, the ground equipment data of Wen 96, a depleted oil and gas storage depot, is collected, and a semi-quantitative evaluation method is proposed.
for underground gas storage based on API581. The type of ground equipment at the gas storage depot is assessed using cluster analysis, and the equipment was classified according to the risk consequence value. The correlation between the pipeline risk indicators and the subjectivity of weighting is reduced via numerical standardization and statistical analysis. The principal component factors of the pipeline risk indicators are extracted using SPSS software. Furthermore, the risk classification and ranking of different types of equipment are obtained while clarifying the key influencing factors. Compared with the API581 evaluation method, the proposed method makes decisions based on multiple risk values instead of only one, rendering the key factors more prominent. This method is more feasible in practice and can assess the risks of ground equipment at underground gas storage depots more accurately. Risk control measures are also proposed, guiding operational safety at underground gas storage depots.

2. Evaluation method model

Although the API581 risk assessment method displays universal characteristics, it lacks an in-depth analytical ability for assessing the risk to complex equipment and facilities. To render the risk assessment more targeted, the cluster analysis method is used to further refine the main influencing risk factors on the equipment and facilities at gas storage depots. Furthermore, principal component analysis is employed to identify the primary influencing risk factors, which are then used to classify the equipment.

2.1. Cluster analysis method

Cluster analysis is a standard method in mathematical statistics. It uses mathematical methods to calculate the vector modulus and dispersion of each sample matrix. The basis of clustering is the calculated sample distance [3]. Until now, various clustering methods have been established, of which systematic clustering and K-means clustering represent the two most frequently used methods. Hierarchical clustering is a fuzzy clustering method that analyzes the clustering of samples via fuzzy equivalence relations [4-5]. Generally, long-distance oil and gas pipelines are damaged by multiple factors (such as operating pressure, fluid temperature, and third-party damage). Therefore, it is necessary to adopt a correlation division for the elements in the research sample, namely fuzzy clustering. Samples obtained via the clustering method contain various influencing factors, reflecting the actual scenario more objectively [6].

The hierarchical cluster method is used for data processing. The basic idea of this technique is to first treat the elements in the sample as individual classes, determine the correlation between these elements by defining the distance between the classes, and aggregate two or more classes with the highest correlation into a new class. Then, clustering analysis is performed on the new and other classes according to the distance between these groups, aggregating two or more categories with the highest correlation into a new class. The final clustering results show that the correlation between similar elements is high, while the correlation between different types is low [7]. According to the different definitions of the distance between classes, the hierarchical cluster method can be divided into six types: the shortest distance method, the longest distance method, the class average method, the center of gravity method, the intermediate distance method, and the square deviation method [8]. Here, the shortest distance method was used [9].

The distance between classes was defined as the distance between the two closest samples, namely:

$$D_{KL} = \min_{i \in G_k, j \in G_l} d_{ij}$$  \hspace{1cm} (1)

Where $d_{ij}$ represents the distance between the i-th and the j-th sample, $G1, G2, ...$ represent different types, and $D_{KL}$ represents the distance between $GK$ and $GL$.

When the $GK$ and $GL$ classes were merged into $GM$, the distance between the new class $GM$ and other classes $GJ$ was calculated using the shortest distance method. The recurrence formula is:

$$D_{MJ} = \min_{i \in GM, j \in G_l} d_{ij} = \min_{j \in G_J} \left( \min_{i \in G_k} d_{ij}, \min_{i \in G_l} d_{ij} \right) = \min\left\{ D_{KL}, D_{LM} \right\}$$  \hspace{1cm} (2)
2.2. Principal component analysis

The principal component analysis is a widely used statistical method and is often employed in mathematics for dimensionality reduction to simplify complex problems. Its basic principle involves orthogonally transforming the variables via a specific method and converting the correlated variables in the sample into several sets of linearly independent comprehensive variables, also known as principal components. This method can accurately reflect the information initially required by the sample for expression while reducing the variables, in turn decreasing the burden of statistical analysis [10-11].

The implementation steps of principal component analysis are as follows:

1) Standardization of the raw data: The processing formula for standardizing the data is:

$$x'_i = \frac{x_i - \bar{x}}{\sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2}}$$

where $x'_i$ represents the standardized data value, $\bar{x}$ represents the average value of the original data of the group, $x_i$ ($i=1, 2, 3, \ldots, n$) represents the original data value, and $n$ is the original data number.

2) The correlation coefficient matrix of the variables is established, while its eigenvalues and corresponding eigenvectors are calculated. The eigenvalue is regarded as an indicator of the impact of the principal component. One of the extraction principles of the number of principal components asserts that the eigenvalue corresponding to the principal component exceeds 1 [12]. The correlation coefficient calculation formula is:

$$\gamma_{xy} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}}$$

where $\gamma_{xy}$ represents the correlation coefficient of the variable, $x$ and $y$ are two different variables, $\bar{y}$ represents the average value of the original data of the group, $x_i$ ($i=1, 2, 3, \ldots, n$) represents the value of each original data value, and $n$ is the number of original data in the group.

3) The number of principal components is determined, the coefficients of each principal component are calculated [13-14], and the principal component formula is established ($F_1, F_2, \ldots, F_k$).

4) The formula of the comprehensive index $F$ is established [15].

The total variance value remains the same during the mathematical transformation process. The first variable displays the most significant variance value, which is denoted as the first principal component, while the second variable is known as the second principal component. The principal components are linearly irrelevant. The number of variables represents the number of principal components [16]. The steep slope graph is drawn according to the degree of interpretation of the data variation of each principal component [13]. Each principal component in the figure represents a point, and the number of principal components to be extracted can be judged by the position where the "steep slope tends to be gentle."

3. Risk assessment of the ground equipment at a gas storage depot based on the clustering-principal component analysis method

A risk analysis was performed on the field data of the large-scale Wen 96 depleted oil and gas reservoir, which was an underground gas storage depot. The primary purpose of the gas storage depot was to inject and extract natural gas from the Yulin-Jinan gas pipeline. The overall planned storage capacity was as high as $5.88 \times 10^8$ m$^3$. The actual gas injection and production volume were $2.95 \times 10^8$ m$^3$, of which the storage capacity of the primary gas storage area accounted for 88.27%. A risk analysis of the gas storage depot was performed, the procedure of which is shown in Figure 1.
The risk consequence value of the equipment was calculated using the gas storage database and the standard calculation method.

The ground equipment was classified via cluster analysis.

Principal component analysis was performed of the equipment and pipeline with significant differences.

The key influencing factors were extracted.

The risk ranking of the equipment was obtained.

Risk control measures were proposed according to the analytical results.

Figure 1. Risk assessment process of gas storage depots

3.1. Analysis of risk factors affecting the primary equipment in the unit

According to the risk assessment process, it was first necessary to determine the risk factors of the evaluation object. Therefore, the risk factors of the primary equipment were identified, which included the ground separator at the gas storage depot, the gas production pipeline, the gas injection pipeline, the air compressor, the production separator, the fuel gas buffer tank, and the incoming air cooler. Since the types of station equipment were different, the influencing risk factors of each varied. The leading integrated influencing risk factors were fluid molecular weight, standard fluid boiling point (°C), fluid auto-ignition temperature (°C), ambient temperature (°C), inner pipe or equipment diameter (m), pipe length or equipment liquid level (m), operating pressure (MPa), the mitigation system category, the device condition category, winter temperature (°C), earthquake zone, number of equipment connections, number of pipe joints, number of pipeline branches, number of pipeline valves, type of construction specifications, design life (years), service starting date, design pressure (MPa), operating temperature (°C), production materials, number of planned shutdowns (/year), number of unplanned shutdowns (/year), the process stability category, the pressure relief valve maintenance status category, the pressure relief valve dirt working condition category, the pressure relief valve corrosion condition category, and the pressure relief valve cleaning condition category.

3.2. Calculation of risk consequences of station equipment based on API581 method

Combustion explosions are serious accidents caused by gas storage leaks. Therefore, the consequences of leakage were analyzed according to the area affected by combustion and explosions. The leakage of combustibles presented significant hidden risks that could cause many types of accidents. However, the RBI risk assessment determined the comprehensive result as the average of all possible results and was obtained based on the probability weighting theory.

In API581, the leakage consequence expression is (5) [17]:

$$A = ax^b$$

where \( A \) is the area affected by the leakage m², \( a \) and \( b \) are the correlation coefficients (constant), and \( x \) is the total leakage kg or leakage rate kg/s.

The equation of the affected area of natural gas leakage consequences is:
(1) The consequence equation of continuous leakage (x is the leakage rate kg/s)
Equipment damage area (m²): \[ A = 8.67x0.98 \]
Personnel lethal area (m²): \[ A = 21.83x0.96 \]
(2) The instantaneous leakage consequence equation (x is the total leakage kg)
Equipment damage area (m²): \[ A = 6.46x0.67 \]
Lethal area (m²): \[ A = 12.46x0.67 \]

The risk consequence value of the station equipment was determined based on the actual data of the station equipment and the ground process database of the gas storage depot (Table 1) using the calculation method for leakage risk consequences in API581.

**Table 1. Partial results of the equipment risk consequence value**

| Equipment                                                                 | Working temperature (°C) | Pressure (MPa) | 1/4 inch-- Damaged area of non-automatic ignition equipment (m²) | 1 inch-- Damaged area of non-automatic ignition equipment (m²) | 1/4 inch-- Lethal area of non-automatic ignition equipment (m²) | 1 inch-- Lethal area of non-automatic ignition equipment (m²) | Damaged area of automatic ignition equipment (m²) | Lethal area of automatic ignition equipment (m²) |
|--------------------------------------------------------------------------|--------------------------|----------------|---------------------------------------------------------------|---------------------------------------------------------------|---------------------------------------------------------------|---------------------------------------------------------------|---------------------------------------------|---------------------------------------------|
| 10m³ Residual liquid tank /G89501                                       | 25                       | 1              | 53.1165                                                      | 804.0213                                                      | 128.8904                                                      | 1845.7625                                                      | 319.5355                                    | 785.4473                                    |
| Inlet pipeline from unit A to manifold 2                                  | 25                       | 6              | 307.4809                                                    | 4654.3155                                                    | 719.8563                                                      | 10308.6320                                                    | 1752.9232                                    | 4083.3463                                   |
| Inlet pipeline of unit B                                                 | 25                       | 6              | 307.4810                                                    | 4654.3155                                                    | 719.8564                                                      | 10308.6321                                                    | 1752.9232                                    | 4083.3463                                   |
| Inlet pipeline of unit C                                                 | 25                       | 6              | 307.4810                                                    | 4654.3155                                                    | 719.8564                                                      | 10308.6321                                                    | 1752.9232                                    | 4083.3463                                   |
| The pipeline between FE-82103 and FL-82102                              | 25                       | 4              | 206.6564                                                    | 3128.1415                                                    | 487.7511                                                      | 6984.7912                                                      | 1192.5489                                    | 2811.9802                                   |
| The pipeline between FE-82403 and FL-82402                              | 25                       | 6              | 307.4810                                                    | 4654.3155                                                    | 719.8564                                                      | 10308.6321                                                    | 1752.9232                                    | 4083.3463                                   |
| The pipeline between FL-82102 and FL-82107                              | 25                       | 6              | 307.4810                                                    | 4654.3155                                                    | 719.8564                                                      | 10308.6321                                                    | 1752.9232                                    | 4083.3463                                   |
| Upstream pipeline of FL-82107                                            | 25                       | 6              | 307.4810                                                    | 4654.3155                                                    | 719.8564                                                      | 10308.6321                                                    | 1752.9232                                    | 4083.3463                                   |
| Upstream pipeline of FL-82207                                            | 25                       | 6              | 307.4810                                                    | 4654.3155                                                    | 719.8564                                                      | 10308.6321                                                    | 1752.9232                                    | 4083.3463                                   |

3.3. Cluster analysis
The station equipment data were sorted according to the risk consequence value and analyzed via clustering calculation using MATLAB software (Table 2 and Figure 2).
Table 2. Partial results of the risk consequence value-based cluster analysis of the station equipment

| Equipment category | Equipment number |
|--------------------|------------------|
| Type 1             |                  |
| 10m³ residual liquid tank /G89501 |                  |
| Inlet pipeline from unit A to manifold 2 |                  |
| Inlet pipeline of unit B |                  |
| Inlet pipeline of unit C |                  |
| Type 2             |                  |
| Various types of pipelines, pressure-regulating valves, electric ball valves, compressors, advanced orifice valves, separators, and check valves |                  |
| Type 3             |                  |
| Injection and production pipelines from each well site to the gathering and injection station |                  |

Figure 2. A hierarchical clustering diagram according to the risk consequence value

Several station equipment categories with different risk consequence values were obtained via clustering: Type 1: A 10m³ residual liquid tank/G89501, sewage tank, and slop oil tank/G-81001. Type 2: Various types of pipelines, pressure-regulating valves, electric ball valves, compressors, advanced orifice valves, separators, and check valves. Type 3: Injection and production pipelines from each well site to the gathering and injection station. Since there was no relative difference between the equipment in Type 1 and the injection and production pipelines in Type 3, and the equipment in the same category was placed in the same environment with the same operating parameters and did not display principal component characteristics, it was impossible to obtain accurate risk results using this method. Therefore, principal component analysis was only conducted for the process piping and equipment in Type 2.

3.4. Principal component extraction and analysis
(1) The principal component extraction and analysis for the process pipeline of the station. The main equipment risk factors in the unit were selected according to the analysis in Section 2.1 and included the X1 pipeline or inner equipment diameter, the X2 pipeline length or equipment liquid-level height, and X3 operation. The pressure and number of the X4 pipe joints were analyzed as factors influencing the process piping of the station. The correlation matrix of the influencing factors was obtained from the standardized data (Table 3). The data analysis showed that the correlation between the influencing factors was significant, allowing further principal component extraction. The eigenvalue was 2.345>1, while the cumulative contribution rate of variance was 58.626% (Table 4), containing most of the information of the original index. The steep slope chart (Figure 3) indicated that the data tended to
become flat after the second principal component. Therefore, the first principal component was extracted, reflecting the original index. The amount of information was sufficient.

**Table 3.** The correlation matrix of the factors influencing the process pipeline of the station

| Influencing factors (standardization) | Inner diameter of the pipeline or equipment (m) | Tube length or equipment liquid-level height (m) | Operating pressure (MPa) | Number of pipe joints (pieces) |
|---------------------------------------|-----------------------------------------------|-----------------------------------------------|--------------------------|-------------------------------|
| Inner diameter of the pipeline or equipment (m) | 1.000 | 0.484 | 0.812 | 0.122 |
| Tube length or equipment liquid-level height (m) | 0.484 | 1.000 | 0.543 | 0.260 |
| Operating pressure (MPa) | 0.812 | 0.543 | 1.000 | 0.279 |
| Number of pipe joints (pieces) | 0.122 | 0.260 | 0.279 | 1.000 |

**Table 4.** The characteristic values and contribution rates of the factors influencing the process pipeline of the station

| Component | Eigenvalues | Initial Eigenvalue | Variance contribution % | Cumulative contribution % |
|-----------|-------------|--------------------|--------------------------|----------------------------|
| 1         | 2.345       | 58.626             | 58.626                   | 58.626                     |
| 2         | 0.928       | 23.201             | 81.827                   |                             |
| 3         | 0.556       | 13.891             | 95.719                   |                             |
| 4         | 0.171       | 4.281              | 100.000                  |                             |

**Figure 3.** Steep slope map

The expression of principal component F1 was obtained by calculating the principal component coefficient and multiplying the obtained principal component coefficient with the standardized data:

\[ F_1 = 0.599ZX_1 + 0.566ZX_2 + 0.498ZX_3 + 0.272ZX_4 \]  

The principal component score was denoted by the corresponding factor score multiplied by the arithmetic square root of the corresponding variance [18]. The principal component score reflected the magnitude of the impact of this type of risk on different devices. The higher the ranking, the more the device was affected by this type of risk [19-20]. The principal component score and device ranking
(Table 5) indicated that the pipeline from manifold 1 to the compressor station via XV-80004 was most affected by operating pressure risk among similar pipelines.

**Table 5.** Principal component scores and rankings of the factors influencing the station process pipeline

| Equipment name | Equipment type | Principal component factor | Score   | Ranking |
|----------------|----------------|----------------------------|---------|---------|
| The pipeline from manifold 1 to the compressor station through XV-80004 | Pipeline | 1.4442 | 2.21107 | 1 |
| The pipeline between FE-82103 and FL-82102 | Pipeline | 1.30615 | 1.999716 | 2 |
| The pipeline between FE-82403 and FL-82402 | Pipeline | 1.30615 | 1.999716 | 3 |
| Upstream pipeline of FL-82407 | Pipeline | 1.17926 | 1.805447 | 4 |
| Upstream pipeline of FL-82507 | Pipeline | 1.17926 | 1.805447 | 5 |
| Upstream pipeline of FL-82607 | Pipeline | 1.17926 | 1.805447 | 6 |
| Inlet pipeline of XV-80001 | Pipeline | 0.97928 | 1.499278 | 7 |
| Inlet pipeline of XV-80002 | Pipeline | 0.97928 | 1.499278 | 8 |
| Inlet pipeline from unit A to manifold 2 | Pipeline | 0.47477 | 0.726873 | 9 |
| Inlet pipeline of unit B | Pipeline | 0.47477 | 0.726873 | 10 |

(2) Furthermore, the main equipment risk factors in the unit were identified according to the main component extraction, the analysis of the station process equipment, and the evaluation in Section 2.1, while X5 ambient temperature (°C), X6 operating pressure (MPa), and X7 equipment were selected. The number of pipes joints, the design pressure of X8 (MPa), and the operating temperature of X9 (°C) were analyzed as the factors influencing the process equipment of the station. The first principal component was extracted at (Table 6) an eigenvalue of 3.562>1 and a cumulative variance contribution rate of 71.240% (>70%), containing most of the information of the original indicator. The amount of information was considered sufficient.

**Table 6.** Characteristic values and contributions of the factors influencing the station process equipment

| Component | Initial Eigenvalue |
|-----------|--------------------|
| 1         | 3.562              |
| 2         | 0.730              |
| 3         | 0.579              |
| 4         | 0.122              |
| 5         | 0.007              |

The expression of principal component F2 was obtained by calculating the principal component coefficient and multiplying it with the standardized data:

$$F_2 = 0.359046X_5 + 0.515382X_6 + 0.506493X_7 + 0.49496X_8 + 0.322938X_9$$

(7)

It could be concluded from the principal component score and equipment ranking (Table 7) that the operating pressure and the risk of the number of joints among similar types of equipment exhibited the most significant effect on the electric drive reciprocating compressor/B engine.
Table 7. The principal component scores and rankings of the factors influencing the station process equipment

| Equipment name                        | Equipment type                  | Principal component factor | Score         | Ranking |
|---------------------------------------|---------------------------------|----------------------------|---------------|---------|
| Electric drive reciprocating compressor/B | Electric drive reciprocating compressor | 4.28629                   | 8.08822923   | 1       |
| Electric drive reciprocating compressor/A | Electric drive reciprocating compressor | 3.0536                    | 5.7621432    | 2       |
| Electric drive reciprocating compressor/C | Electric drive reciprocating compressor | 3.0536                    | 5.7621432    | 3       |
| Plug valve/PL-88303                    | Plug valve                      | 0.12142                   | 0.22911954   | 4       |
| Check valve/CH-80103                   | Check valve                     | 0.04428                   | 0.08355636   | 5       |
| Check valve/CH-80203                   | Check valve                     | 0.04428                   | 0.08355636   | 6       |
| Check valve/CH-80303                   | Check valve                     | 0.04428                   | 0.08355636   | 7       |
| Check valve/CH-80403                   | Check valve                     | 0.04428                   | 0.08355636   | 8       |
| Outbound safety valve/PSV88301         | Outbound safety valve           | -0.05817                  | -0.10976679  | 9       |
| Electric ball valve/XV-80203           | Electric ball valve             | -0.08703                  | -0.16422561  | 10      |

3.5. Risk control measures
(1) The risk assessment showed that the operating pressure posed the highest risk. Therefore, attention should be focused on the operating pressure during the daily management and maintenance of the ground process pipelines at the stations of underground gas storage depots. Consequently, the following recommendations are made: 1) The pipelines should be dynamically monitored. Regular inspections and maintenance should be performed while the operating, inlet, and outlet pressure of the pipeline should be recorded daily. 2) A pressure alarm should be installed that is activated when the pressure approaches the upper limit of safe operation, allowing the on-duty personnel to detect and address the issue in time; 3) The operators should be trained and requested to strictly follow the operating procedures, avoiding overly elevated operating pipeline pressure caused by maloperation. 4) A series of cut-off valves should be installed to ensure that they remain effective in the case of one or more operational failures, guaranteeing the reliability of the entire system.

(2) The risk assessment of the ground station process equipment in the underground gas storage depot indicated that the primary influencing factors are the operating pressure and the number of equipment joints. Therefore, the daily management and maintenance of this equipment should focus on the two points. The following measures are proposed: 1) Regular inspection and maintenance should be performed, while the operating status and pressure of each piece of equipment should be recorded daily, especially those with several joints. The causative factors of the problems identified during daily inspections should be determined and resolved timeously. A pressure alarm should be installed that is activated when the pressure approaches the upper limit of safe operation, allowing the staff on duty to identify and address the issue. Production should proceed in strict accordance with the operating procedures to avoid excessively high or low operating pressures in the valves and other equipment due to maloperation.

4. Conclusion
Focusing on the ground station technology equipment at the Wen 96 underground gas storage depot, this study investigates the risk assessment of the station equipment via data processing and calculations. The clustering-principal component analysis is used to improve the current quantitative API581 evaluation method. The gas storage equipment data is further calculated and analyzed to obtain the risk evaluation classification, which is more in line with the actual situation of equipment production at the
stations of gas storage depots. During daily production and maintenance, close attention should be paid to the operating pressure of the equipment and pipelines, while the operating status should be dynamically monitored.

Acknowledgments
The Science and Technology program for strategic cooperation of CNPC- China University of Petroleum (ZLZX2020-05).

References
[1] DING G S, LI W Y. Foreign and Domestic Underground Gas Storage Tanks: Current Situation and Development Trends [J]. International Petroleum Economics, 2002(06):23-26+63.
[2] DING G S, WANG S. Review on 20 years' UGS construction in China and the prospect[J]. Oil & Gas Storage and Transportation, 2020, 39(01):25-31.
[3] TIAN B. Hierarchical Clustering Method and Its Research about Application Hierarchical Clustering Method and Its Research about Application. Yinshan Academic Journal, 2014, 28(02):11-16.
[4] WANG Q L, ZHANG Y F. Research on the application of unified clustering method in the analysis of E-learning Behavior[J]. The Chinese Journal of Ict in Education, 2016(08):90-93.
[5] LI S, ZHANG Z, HUANG F C. Application of Fuzzy Clustering Analysis Technology in Risk Assessment of Oil and Gas Pipeline[J]. Journal of Shenyang University (Natural Science), 2019, 31(01):33-38.
[6] YANG Y L, SHAO C S, JING G. The application of systematic clustering method to the classification of chemical types of karst groundwater[J]. Geophysical and Geochemical Exploration, 2018, 42(04):738-744.
[7] LI D Q, ZHANG P, AI M Y, WANG Y L, CHEN L Q. Application of fuzzy clustering method to pipeline segmentation process of oil & gas pipeline risk assessment[J]. Natural Gas Industry, 2012, 32(07):63-67+107.
[8] A. Yajima, H. Wang, R.Y. Liang, H. Castaneda, A clustering based method to evaluate soil corrosivity for pipeline external integrity management, International Journal of Pressure Vessels and Piping, 126-127 (2015) 37-47.
[9] SUN W S. Admissibility of minimum distance clustering method[J]. Journal of Chuxiong Normal University, 2000, 03(01):1-6.
[10] ZHANG J. Welding defects recognition algorithm research based on Principal Component Analysis[D]. Xi’an Shiyou University, 2017.
[11] WANG X L, WEI Z B, PENG S T, GUAN W L. Comprehensive assessment model of liquid pipeline leakage consequences based on principle component analysis[J]. Journal of Safety Science and Technology, 2014, 10(05):84-89.
[12] LI Z, LI W G, ZHANG L, ZHENG L X. Feature frequency extraction algorithm based on principal component analysis and its application[J]. Journal of Vibration, Measurement & Diagnosis, 2018, 38(04):834-842+879.
[13] ZHAO H Q, ZHU J P. How to Use SPSS Software to Calculate Results of Factor Analysis[J]. Statistics and Decision, 2019, 35(20):72-77.
[14] LI L Q. Application of SPSS principal component analysis in economic benefit analysis of large medical equipment[J]. China Logistics & Purchasing, 2020(02):76-77.
[15] ZHANG J. Pipeline risk assessment method based on principle component-clustering analysis[J]. Oil & Gas Storage and Transportation, 2014, 33(02):139-143.
[16] LIU X B, ZHANG H, XIA M Y, LIANG L C, ZHENG W, LI M. Pipeline leakage recognition based on principal component analysis and neural network[J]. Oil & Gas Storage and Transportation, 2015, 34(07):737-740.
[17] XIAO F, JIANG J Q, LING A J, ZHAKO K. Research on analysis methods of combustion and explosion consequence based on API 581[J]. Journal of Safety Science and
[18] ZHANG Q L, WANG J, WANG X M. Failure Risk Assessment Model of Filling Pipeline Based on KPCA and PSO-SVM[J]. Gold Science and Technology, 2017, 25(03): 70-76.

[19] WANG S Y, WANG P, ZHANG Z P, DONG W Y, LEI X. Application of multi-attribute clustering method based on principal component analysis in fracture prediction[J]. Petroleum Geophysics, 2019, 17(01): 60-62+67.

[20] LI X Y, YAN B. The Formation and Comparison of Robust Principal Component Cluster Analysis Method[J]. Journal of Applied Statistics and Management, 2019, 38(05): 849-857.