

# Quantitative safety evaluation of drilling engineering by combining analytic hierarchy process with alternating condition expectation

Fan Kunkun<sup>1,2\*</sup>, Wang Fangxiang<sup>3</sup>, Wang Chunguang<sup>1,2</sup>, Chen Shaojie<sup>1,2</sup>, Lin Hai<sup>4</sup>, He Jia<sup>5</sup>, and Chen Zhongwei<sup>6</sup>

<sup>1</sup> State Key Laboratory of Mining Disaster Prevention and Control, Shandong University of Science and Technology, Qingdao 266555, Shandong, China

<sup>2</sup> College of Energy and Mining Engineering, Shandong University of Science and Technology, Qingdao 266555, Shandong, China

<sup>3</sup> Downhole Technology Service Company of Bohai Drilling Engineering Company, CNPC, Tianjin 300283, China

<sup>4</sup> Western drilling downhole company, CNPC, Karamay 834000, Xinjiang, China

<sup>5</sup> Tuha Oilfield, CNPC, Turpan 838202, Xinjiang, China

<sup>6</sup> School of Mechanical and Mining Engineering, The University of Queensland, QLD 4074, Australia

**Abstract:** Drilling engineering is important for development of underground resources. There are many potential risks in the process of drilling engineering. If the drilling safety is not effectively managed, it will lead to great loss. In order to predicate and prevent safety accidents, a safety evaluation index system is established and an AHP-ACE method is proposed. 23 groups of original drilling data are collected from a block of Xinjiang oil field, and the safety levels are quantitatively analyzed by characteristic values. First, the AHP method is applied to relate direct observation variables to potential variables by calculating weight and determining membership function of each index. Second, three potential variables are chosen as the input of the ACE model and a training set containing 17 groups of data is used to establish the mathematical relationship between the potential variables and drilling safety value. Finally, the validity of the proposed model is examined by matching it to the other 6 groups of data. This study provides a new and effective way to quantitatively evaluate the safety of drilling engineering.

## 1 Introduction

Drilling engineering is the foundation for exploration and development of underground resources, including oil, gas and geothermal energy. A large number of human financial resources have been invested in developing drilling equipment and technology for better accessing to the underground resources. Once drilling safety accident occurred, it will result

---

\* Corresponding author: skdrbc@sdust.edu.cn

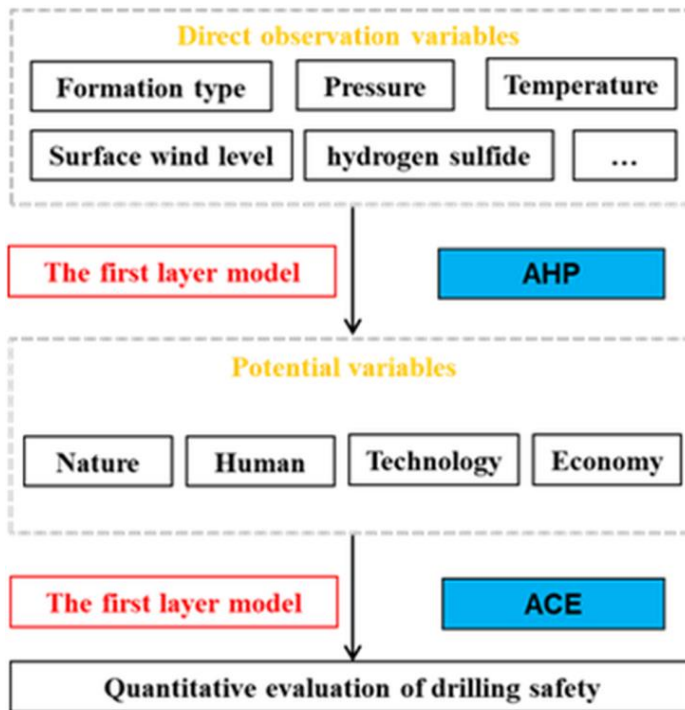
in economic loss, equipment damage, environmental pollution and even personal injury. Therefore, evaluation of drilling safety is of great importance on preventing the occurrence of accidents and maintaining sustainable development of underground resources<sup>[1]</sup>.

At present, several methods have been proposed to evaluate the safety of drilling engineering system, including analytic hierarchy process (AHP) method, fuzzy evaluation method, rough set theory and artificial neural network (ANN) method<sup>[2-5]</sup>. Among these methods, the AHP is the most popular method for both qualitative and quantitative analysis of drilling safety together with membership function. In the AHP method, a lot of influencing factors, such as temperature, pressure, as well as drilling fluids, well construction, wire-line and drilling services, are related to several intermediate variables, e.g. geology, economy and technology. However, the traditional AHP model requires a priori assumption of the function forms relating intermediate variables and drilling safety, which may yield erroneous results, because relationship between the independent variables and dependent is very complicated.

An approach by combining analytic hierarchy process (AHP) with alternating condition expectation (ACE) is proposed for evaluation of drilling safety in this paper. The ACE method is a nonparametric regression technique, which can relate the intermediate variables to drilling safety without a priori assumptions of functional form<sup>[6]</sup>. The mathematical form of the ACE is totally data driven. Therefore, the ACE-AHP method integrates expert experience and science calculation to improve the predication accuracy of drilling safety.

## **2 Background**

The AHP-ACE model is consist of three variable layers and two model layers, as shown in Fig. 1. The first variable layer is an input layer of original data, including 12 directly observed indexes. These indexes are used to reflect the safety of the drilling engineering and their details are introduced in the next section. The second variable layer contains four potential variables. Each variable is characterized by weights and membership functions corresponding to the direct observation variables. The weights are determined by the AHP method and the membership functions are represented by empirical function, such as trapezoidal function and triangular function. The third layer is the total target, which is the safety level of the drilling. Values of the drilling safety are associated with potential variables by using the ACE regression method.

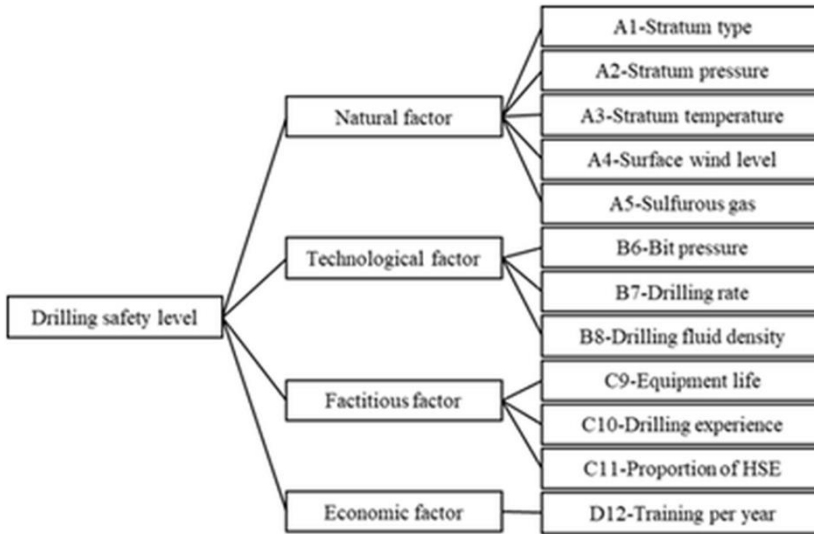


**Fig. 1.** Schematic diagram of the AHP-ACE model

## 2.1 AHP method

The AHP has been widely used in dealing with multi-criteria decision problems since it was proposed by Satty<sup>[7]</sup> in 1980. It can effectively quantify the complex and qualitative problems. The basic steps of the AHP are as follows:

- (1) Build hierarchy structure model.



**Fig. 2.** The AHP structure of drilling safety assessment

The first step of the AHP is to build the hierarchy structure for the index system. Construction approaches are usually based on published papers and experienced experts. In this paper, the hierarchical structure is established as shown in Fig. 2. Though this hierarchical model, the correlation structure between the four potential variables and the 12 direct observation variables is cleared.

(2) Establish judgement matrix.

In order to determine the weight of each factor, a judgement matrix as shown in Eq. (1) needs to be established. Dimension of the matrix depends on the number of influencing factors corresponding to each variable in the higher level. Elements of the matrix are determined by pairwise comparisons among the influencing factors in the lower level<sup>[2]</sup>.

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \quad (1)$$

where,  $A$  is the judgement matrix,  $n$  is the number of the influencing factors,  $a_{ij}$  is the value resulting from comparison between the  $i^{\text{th}}$  factor and the  $j^{\text{th}}$  factor. The evaluation of  $a_{ij}$  is based on principles in previous literature<sup>[8]</sup>.

(3) Calculate weight of each index.

The weight of each factor is determined by calculating the eigenvalue vector corresponding to the largest eigenvalue of the judgement matrix, as shown in Eq. (2). A Matlab algorithm is established to obtain these weights in this paper.

$$A\varpi = \lambda_{max}\varpi \quad (2)$$

where,  $\lambda_{max}$  is the largest eigenvalue,  $\varpi$  is the weight vector corresponding to  $\lambda_{max}$ .

(4) Consistency test

In order to determine the validity of the judgement matrix, it is necessary to conduct consistency test. If the judgment matrix does not meet the requirement of the consistency, the judgment matrix is invalid and need to reset. The consistency indicator  $CI$  is expressed as:

$$CI = \frac{\lambda_{max} - n}{n - 1} \quad (3)$$

The larger the value of  $CI$ , the worse the consistency. When  $CI/R.I. \leq 0.1$ , consistency meets the requirements and the judgement matrix is valid, where,  $R.I.$  is the average consistency indicator, its value is show in Table 1.

**Table 1.** Assignment table of  $R.I.$

n	1	2	3	4	5	6	7	8	9
$R.I.$	0	0	0.58	0.09	1.12	1.24	1.32	1.41	1.45

## 2.2 ACE method

Principle of the ACE method is to transform the dependent variable and a series of independent variables into a virtual space, where the maximum linear regression between the transformed dependent variable and a series of independent variables is accomplished<sup>[9]</sup>. The mathematical expression of the ACE model can be written as:

$$\theta(y) = \alpha + \sum_{i=1}^m \varphi_i(x_i) + \varepsilon \quad (4)$$

where,  $\varphi_i(\cdot)$  and  $\theta(\cdot)$  are transforming functions of independent variables and dependent variable, respectively.  $\alpha$  is constant and  $\varepsilon$  is an error.  $M$  is the number of independent variables.

In the ACE model, the original problem of correlating dependent variable with independent variables  $X(x_1, x_2, \dots, x_m)$  is replaced by evaluating  $m$  separate on-dimensional functions  $\varphi_i(\cdot)$  and  $\theta(\cdot)$ . The target of the ACE model is to establish optimal transformations can minimize the variance ( $\varepsilon^2$ ) of a linear correlation between the transformed dependent variable and the sum of transformed independent variables<sup>[10-11]</sup>. Expression of the  $\varepsilon^2$  is written as:

$$\varepsilon^2(\theta, \varphi_1, \dots, \varphi_m) = E\{[\theta(y) - \sum_{i=1}^m \varphi_i(x_i)]^2\} \quad (5)$$

With the constraints:  $E[\theta^2(y)] = 1$ ,  $E[\theta(y)] = E[\varphi_1(x_1)] = \dots = E[\varphi_n(x_n)] = 0$ .

In order to obtain the minimum  $\varepsilon^2$ , a series of single-function minimizations are accomplished, resulting in:

$$\varphi_i(x_i) = E[\theta(y) - \sum_{j \neq i}^m \varphi_j(x_j) | x_i] \quad (6)$$

$$\theta(y) = \frac{E[\sum_{i=1}^m \varphi_i(x_i) | y]}{\|E[\sum_{i=1}^m \varphi_i(x_i) | y]\|} \quad (7)$$

After finishing the minimization process, the optimal transformations  $\varphi_i^*(x_i)$  and  $\theta^*(y)$  can be determined on the basis of Eq. (6) and Eq. (7). In the transformed space, the dependent and independent variables are related as:

$$\theta^*(y) = \sum_{i=1}^m \varphi_i^*(x_i) + \varepsilon^* \quad (8)$$

where  $\varepsilon^*$  is the error in the ACE transformations and is assumed to have a normal distribution with zero mean.

## 3 Methodology of drilling safety evaluation

The proposed AHP-ACE system mainly consists of three Steps. First, analyze all direct observation factors by AHP and calculate weights of these factors for determination of potential variables. Then, choose the valid potential variables as the input of the ACE model and set up the ACE model. At last, carry out drilling safety evaluation after establishing the ACE model by historical data.

### 3.1 Preparation of data sets

Information of drilling safety collected from a block of Xinjiang Oilfield is used to establish the AHP-ACE model and to verify the validity of the model. 23 group data of 12 factors affecting drilling safety and their corresponding safety levels are obtained as shown in Appendix A. The safety is classified into four levels including safety, general safety, danger and very danger and each safety level is scored by experts in the range of [0.9, 1], [0.5, 0.9], [0.1, 0.5], and [0, 0.1], respectively.

### 3.2 Calculation of potential variables

**Table 2.** Discrimination matrix and weight of the natural factor

Matrix	A1	A2	A3	A4	A5	Weight
A1	1	3	3	2	1/8	0.1417
A2	1/3	1	1	2	1/9	0.069
A3	1/3	1	1	2	1/9	0.0690
A4	1/2	1/2	1/2	1	1/9	0.0507
A5	8	9	9	9	1	0.6696

**Table 3.** Discrimination matrix and weight of the technological factor

Matrix	B5	B6	B7	B8	Weight
B5	1	1	3	1/3	0.1998
B6	1	1	3	1/3	0.1998
B7	1/3	1/3	1	1/5	0.0781
B8	3	3	5	1	0.5222

**Table 4.** Discrimination matrix and weight of the factitious factor

Matrix	C9	C10	C11	Weight
C9	1	2	3	0.5152
C10	1/2	1	2	0.3304
C11	1/3	1/2	1	0.1544

On the basis of the AHP method, weights of the 12 direct observation factors can be determined as shown in Table 2, Table 3 and Table 4. Since economic factor as a potential variable only contains one influencing factor, it is not necessary to calculate the weight and default value is 1. Then, the original data is transformed into four potential variables by using the weights and membership functions. The membership function is established on the basis of related literature and knowledge of experienced experts as shown in Appendix B. The calculated values of the four potential variables for each group data using the AHP method is shown in Table 5.

### 3.3 Establishment of ACE model

Table 5 shows that values of the economic factor are constant to 1, which indicates that

the effect of the economic factor on the drilling safety level is negligible. Therefore, only three potential variables are selected as the input of the ACE model, including the natural factor, technological factor and human factor. In the process of building the ACE model, the data body is divided into two parts: a training set and a verification set. The samples in the training set are used to generate the regression model, and samples in the verification set are used to test and verify the accuracy of the model. In this paper, 13 group samples are selected as training set samples and the rest of 7 group samples are selected as verification set samples.

The GRACE software<sup>[12]</sup> is used to transform the independent variables (natural factor, technological factor and human factor) and the dependent variable (safety level) into virtual space. The optimal transformation is corresponding to a linear transformation of the dependent variable, which is plotted in Fig. 3.

**Table 5.** Calculated values of for potential variables

<b>Order</b>	<b>Natural factor</b>	<b>Technological factor</b>	<b>factitious factor</b>	<b>Economic factor</b>	<b>Safety</b>
1	0.867	0.717	0.906	1	0.48
2	0.919	0.950	0.953	1	0.91
3	0.875	0.619	0.485	1	0.06
4	0.751	0.578	1.000	1	0.38
5	0.932	0.906	1.000	1	0.95
6	0.995	0.053	0.906	1	0.04
7	0.226	0.500	1.000	1	0.03
8	0.869	0.672	0.858	1	0.33
9	0.875	0.834	0.811	1	0.53
10	0.917	0.231	1.000	1	0.15
11	0.847	0.694	1.000	1	0.62
12	0.922	0.698	1.000	1	0.72
13	0.856	0.783	1.000	1	0.82
14	0.857	0.672	1.000	1	0.55
15	0.941	0.806	1.000	1	0.87
16	0.860	0.414	1.000	1	0.21
17	0.912	0.815	0.906	1	0.75
18	0.868	0.868	0.811	1	0.63
19	0.900	0.668	0.858	1	0.35
20	0.955	0.773	0.917	1	0.85
21	0.237	0.756	1.000	1	0.09
22	0.901	0.784	0.811	1	0.49
23	0.850	0.763	0.906	1	0.53

Since the output of the transformed independent and dependent variables in ACE model is

in a discrete fashion, Polynomial functions are used to fit with the scattered points as shown in Fig. 3. The fitting results indicate that the transformed data for natural and human factor can be fitted by quadratic polynomial expression with reasonable accuracy,  $R^2 = 0.928$  for natural factor and  $R^2 = 0.998$  for human factor, respectively. As for technical factor, a linear function match with the transformed data well. Therefore, the optimal transformations of the independent variables can be expressed with a universal form as:

$$\varphi^*(x_i) = \sum_{k=1}^3 b_k x_i^{3-k} \quad (9)$$

where,  $b_k$  is coefficient of optimal transformation function, which is shown in Fig. 3.

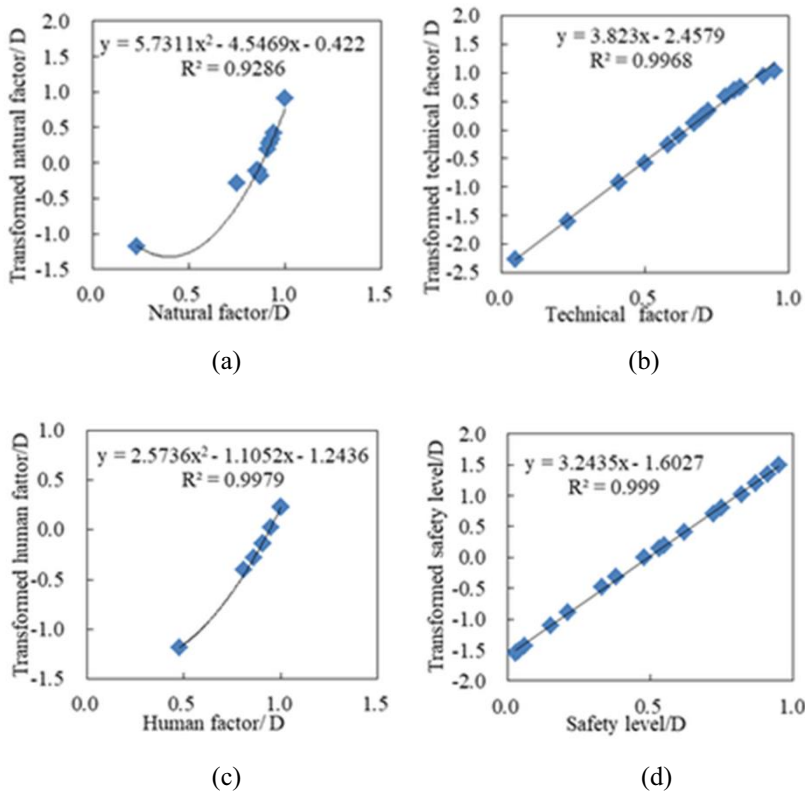
The linear transform of the dependent variable (Safety level, SL) is expressed as:

$$\theta^*(SL) = 3.235SL - 1.0627 \quad (10)$$

Substituting Eq. (4) and Eq. (9) into Eq. (10), we can obtain a regression prediction model of  $SL$  with an inverse transformation of Eq. (10), the safety level  $SL$  is written as:

$$SL = \sum_{i=1}^3 \sum_{k=1}^{M_k} c_{i,k} x_i^{M_k-k+1} + D \quad (11)$$

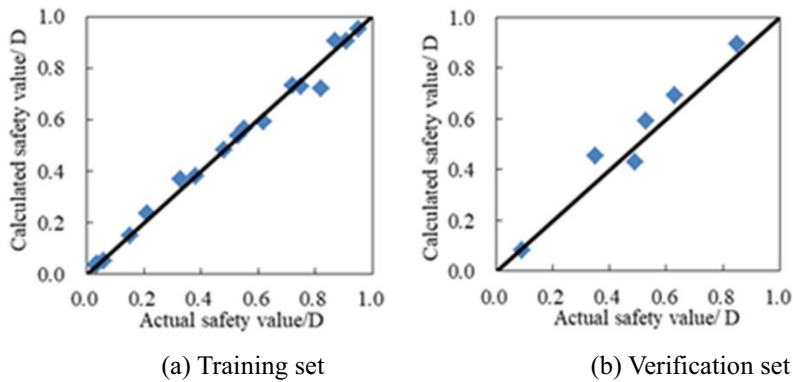
where,  $M_k$  is the order of the polynomial for the  $k^{\text{th}}$  independent variable,  $c_{i,k}$  is the coefficient of the  $i^{\text{th}}$  term of the  $k^{\text{th}}$  variable, which can be determined by the optimal transformation function shown in Fig. 3.  $D$  is the constant term, which covers the constant terms of all polynomials.



**Fig. 3.** Optimal transforms of the independent and dependent variables



## 4 Evaluation of the AHP-ACE model



**Fig. 4.** Relationship between actual and calculated safety value

Relationship between the actual and calculated  $SL$  values is shown in Fig. 4, which indicates that the cross points generally fall close to the 45 degree line. The average absolute error (AAE)<sup>[10]</sup> of the training and verification samples is 0.02, and the average absolute relative error (AARE)<sup>[10]</sup> of these two data sets is 5.29%, 4.71%, respectively. This shows that the proposed ACE model is reliable to fit training samples and to predicate verification samples.

## 5 Conclusions

The drilling safety level is evaluated by combining AHP and ACE method. Several important observations are summarized as follows:

- (1) The evaluation index system of drilling safety is established, including 12 direct observation variables in the first layer and four potential variables in the second layer.
- (2) Values of the potential variables are determined on the basis of the weights and membership functions. The weights are calculated on the basis of the AHP method.
- (3) The input of the ACE model is consist of three potential variables including natural factor, technical factor and human factor. The output of the ACE model is the safety value, which is corresponding to the safety Level of the drilling engineering.
- (4) The validity of the proposed AHP-ACE model for drilling safety evaluation is verified by both training and verification data set. This AHP-ACE model is rigorous in theory, which provides an effective approach for the safety evaluation of drilling engineering.

**Acknowledgement.** The authors would like to express their sincere appreciation to the support of the National Natural Science Foundation of China (Grant No. 41772154).

### Appendix A. Original data of safety index for drilling engineering

Order	A1	A2	A3	A4	A5	B6	B7	B8	C9	C10	C11	D12	Safety
1	0.4	40	56	2	0	6.2	7.3	1.21	3	12	10	20	0.48
2	0.9	32	65	2	0.2	5.1	5.2	1.13	5	25	11	20	0.91

3	0.4	46	58	3	0.3	6.3	1.2	1.28	2	3	15	20	0.06
4	0.4	42	63	4	0.5	6.5	3.5	1.23	6	26	16	24	0.38
5	0.9	36	52	2	0	4.5	5.3	1.16	4	13	12	24	0.95
6	0.9	65	72	2	0	7.9	2.3	1.46	8	26	10	20	0.04
7	0.7	30	42	2	1.8	5.5	3.2	1.09	7	25	12	20	0.03
8	0.9	42	53	3	0.2	4.3	1.8	1.36	2	23	9	24	0.33
9	0.9	46	59	1	0.3	5.6	5.3	1.32	4	36	8	24	0.53
10	0.9	25	32	2	0	3.8	6.3	1.01	8	18	15	24	0.15
11	0.4	26	33	4	0	4.2	8.9	1.05	4	25	14	24	0.62
12	0.9	29	35	3	0	4.5	3.6	1.08	6	13	18	24	0.72
13	0.4	32	42	2	0	5.6	2.3	1.11	3	24	13	24	0.82
14	0.4	33	53	4	0.3	6.9	3.6	1.39	4	13	14	24	0.55
15	0.9	45	62	2	0.2	6.2	5.2	1.25	2	16	12	24	0.87
16	0.4	35	39	3	0	6.5	1.3	1.31	6	23	15	20	0.21
17	0.7	42	45	4	0	5.8	5.2	1.33	4	42	10	24	0.75
18	0.4	41	48	2	0	5.2	6.2	1.22	3	36	8	24	0.63
19	0.7	33	39	3	0.3	4.5	6.5	1.12	6	12	9	24	0.35
20	0.9	86	52	2	0.4	3.6	6.2	1.14	2	8	11	24	0.85
21	0.7	38	46	1	1.3	4.1	2.3	1.12	3	13	12	24	0.09
22	0.7	34	39	3	0.2	3.5	5.3	1.05	4	41	8	24	0.49
23	0.4	28	33	4	0.2	3.2	4.6	0.95	3	23	10	24	0.53

Note: carbonate rock=0.4, sandy conglomerate=0.7, mud shale=0.9

**Appendix B. Membership functions for different observation factors**

Observation factors	Membership function	Observation factors	Membership function
A1	$a_1(x) = \begin{cases} 0.4 & \text{carbonate rock} \\ 0.7 & \text{sandy conglomerate} \\ 0.9 & \text{mud shale} \end{cases}$	B7	$b_7(x) = \begin{cases} 0 & x < 1.5 \\ \frac{x-1.5}{3} & 1.5 \leq x \leq 4.5 \\ \frac{8.5-x}{4} & 4.5 < x \leq 8.5 \\ 0 & 8.5 < x \end{cases}$
A2	$a_2(x) = \begin{cases} 0 & x < 25 \\ \frac{x-25}{50} & 25 \leq x \leq 75 \\ \frac{100-x}{25} & 75 \leq x \leq 100 \\ 0 & 100 < x \end{cases}$	B8	$b_8(x) = \begin{cases} 0 & x < 0.85 \\ \frac{x-0.85}{0.3} & 0.85 \leq x \leq 1.15 \\ \frac{1.35-x}{0.2} & 1.15 < x \leq 1.35 \\ 0 & 1.35 < x \end{cases}$

---

A3	$a_3(x) = \begin{cases} 1 & 0 \leq x \leq 60 \\ \frac{105-x}{45} & 60 < x < 105 \\ 0 & 105 \leq x \end{cases}$	C9	$c_9(x) = \begin{cases} 1 & 0 < x < 5 \\ \frac{8-x}{3} & 5 \leq x \leq 8 \\ 0 & 8 < x \end{cases}$
A4	$a_4(x) = \begin{cases} 1 & 0 < x < 4 \\ \frac{6-x}{2} & 4 \leq x \leq 6 \\ 0 & 6 < x \end{cases}$	C10	$c_{10}(x) = \begin{cases} 0 & 1 \leq x < 3 \\ \frac{x-3}{5} & 3 \leq x \leq 8 \\ 1 & 8 < x \end{cases}$
A5	$a_5(x) = \begin{cases} 1 & 0 < x < 0.4 \\ \frac{1-x}{0.6} & 0.4 \leq x \leq 1 \\ 0 & 1 < x \end{cases}$	C11	$c_{11}(x) = \begin{cases} 0 & 0 \leq x < 5 \\ \frac{x-5}{7} & 5 \leq x \leq 12 \\ 1 & 12 < x \end{cases}$
B6	$b_6(x) = \begin{cases} 0 & x < 3 \\ \frac{x-3}{2} & 3 \leq x \leq 5 \\ \frac{7-x}{2} & 5 < x \leq 7 \\ 0 & 7 < x \end{cases}$	D12	$d_{12}(x) = \begin{cases} 0 & 0 \leq x < 4 \\ \frac{x-4}{8} & 4 \leq x \leq 12 \\ 1 & 12 < x \end{cases}$

---

## References

1. Todd B., Reese M., Brown S., SPE 196521, (2019)
2. Zhao X. J., Zhan H. B., Zhao H. F. International Petroleum and Petrochemical Technology Conference, 167 (2019)
3. Li J., Li K., Wang B. [J] Journal of Southwes Petroleum University, 39:5, 120 (2017)
4. Zhang G. M., Qiu C. L., Li X., et al. Fourth International Conference on Natural Computation (2008)
5. Duan Y. M. [J] Journal of Safety Science and Technology, 13:11, 145 (2011)
6. Breiman, L.; Friedman, J. H. [J] Journal of the American Statistical Association, 80:391, 580 (1985)
7. Satty, T. L.. The Analytic Hierarchy Process, McGraw-Hill (New York, 1980)
8. Li F. W., Du X. L., Zhang M. J., et al. [J] Beijing Gongye Daxue Xuebao, 38:2, 167-172 (2012)
9. Wang D., Murphy M. [J] Journal of Data Science, 2:4, 329 (2004)
10. Feng Q., Zhang J., Zhang X., et al. [J] International Journal of Coal Geology, 121, 137 (2014)

11. Li Z., Wang S., Li S., et al. [J] *Energy & Fuels*, 1, 625 (2013)
12. Xue G., Datta-Gupta A., Valko P., et al. [J] *SPE* 35412, 85 (1997)