An Optimal Risk Matrix Based on Factor Analysis for Oilfield Operation Management

Zhao Chunlan, Yuan Xihang and Xie Xiangjun

July 20, 2019

An Optimal Risk Matrix Based on Factor Analysis for Oilfield Operation Management

1st Chunlan Zhao*
College of Science
Southwest petroleum university
Chengdu, China
email address:308303451@qq.com

2nd Xihang Yuan*
basic subjects teaching
Southwest petroleum university
Nanchong, China
email address:33479538@qq.com

3rd Xiangjun Xie
School of Science
Southwest petroleum university
Chengdu, China
email address:1391005814@qq.com

***Abstract*—Traditional Risk Matrices are usually considered as applicable method for risk assessment. However, such techniques, mainly analyzing in the likelihood and consequence of danger, are often limited. Thus, this method can not be simply used with facing multidimensional data set. Therefor, this paper proposes an approach to solve the problem by combining Risk Matrix with Factor Analysis. It will be based on data of oilfield operation risk, on Factor analysis and on Risk Matrix providing an Optimal Risk Matrix. In order to prove the effect of the approach, both graphical analysis of traditional and optimal risk matrix are clearly depicted. Finally, some views and perspectives of this approach are proposed.**

***Index Terms*—optimal risk matrix; factor analysis; data analysis; risk assessment; oilfield management.**

I. INTRODUCTION

Risk matrices, an important decision-making tool, have unique advantages in risk analysis. Its main purpose is to ensure that managers can find potential risk factors and avoid accidents as much as possible. Currently, this method is commonly used because of its uncomplicated theory and graphic depictions. The earliest definition and usage of a risk matrix was in the late 1990s by the US Air force and the MITRE Corporation. The risk matrix what we are talking about currently, is also known as a probability consequence diagram (Cox, 2008; Ale, 2015) [1], [2]. Generally, it consists of two broad categories, $probability$ or $likelihood$ and $consequence$ or $severity$. The $likelihood$ indicates the likelihood of occurrence of accident in different conditions. And $consequence$ indicates the influence of accident in different situations. Therefore, the risk matrix is an evaluation tool concerning the two aspects of one accident.

At present, there are 3 kinds of articles on the risk matrix. The first type of article is mainly about the use of the risk matrix. Applications, besides injury control [3]–[6], range over landslide and loss assessment [7], [8],agriculture planning [9], [10], tunneling [11], air traffic [12]–[14],and much besides.

Besides the impact on diverse field, the second one mainly concerned with the improvement and research of the risk matrix itself. Based on theoretical research, these articles are good at putting forward some possible problems in using risk matrix. There is inevitably tangible and intangible factor in constructing the risk matrix. The reliability and utility of risk matrices had been amply discussed [15]. Industry-generated risk matrix data, revealed evidence of human cognitive biases in the judgment of likelihood and consequence, could improve risk matrix based risk analysis prevalent in industry [16].

Considering the calculation process of risk matrix, Ni (2010) proposed diverse risk index in arithmetic pattern for risk matrices [17]. Furthermore, in order to enhance the arithmetic applicability of risk matrix, a exponential continuous risk index was proposed to manage the resolution of conventional risk matrices [18], [19].

Combining risk matrices with other theory, the last type of article also promoted effective and creative combination. A bow-tie model, comprised of a fault tree and event tree, is widely applied in risk analysis, including probability calculation [20], [21]. In Lu's paper, combination of the bow-tie model and risk matrix creates an effective method for the comprehensive risk evaluation including pipeline management and risk factors reduction [22]. Analytic hierarchy Process, a structured technique for organizing and analyzing complex decisions, can also be combined with risk matrix [13]. Hsu proposed this revised risk matrix with continuous scale was also useful for assessing risk factors.

In recent years, the remarkable and effective fuzzy theory can also be applied to risk matrix. The concept of the fuzzy risk matrix, proposed by Markowski and Mannan(2008) [23], was used for analysis of distillation column unit. It was also referenced in Khaleghi et al. (2013), Shapiro (2013), Ataallahi and Shadizadeh (2015), Liu et al. (2016),and Skorupski (2016) [14], [24]–[27].

In summary , the application of a risk matrix requires a valid and effective system and applicable data sets; namely, the 2-dimensional data sets. In fact, data sets may have some different characteristics, one of which is multidimensional in practical problem. A large quantity of data is valuable for risk analysis, so how to deal with the various dimensions of data sets is a considerable question. Flage and Røed (2012) [28] mentioned multidimensional data sets that concerning manageability, uncertainty, and criticality three aspects. So the application of traditional risk matrix is influenced by multidimensional limitation problem.

This research is supported by the National Science & Technology Major Project.

To solve the problem, improve the approach, and implement risk management, this paper proposes an optimal risk matrix. The optimal risk matrix is based on risk matrix and factor analysis. We haven't changed much in the framework of risk matrix; namely, it is similar to the traditional risk matrix. But the factor analysis plays a crucial role in this paper. The usage of factor analysis makes it possible to use the multidimensional data set in risk matrix. And the usage of coordination transformation breaks the limitation of original interval after factor analysis. And the results of optimal risk matrix is more significant than traditional risk matrix.

The subsequent parts of this paper cover the following section. Section II introduces the basic theory such as risk matrix, factor analysis and optimal risk matrix. Section III illustrates the approach of optimal risk matrix and its limitation. Section IV presents the experimental procedure and discusses the results. Section V concludes this works.

II. METHODS

In this section, we will respectively introduce the basic methods of risk matrix, factor analysis. In the subsection of $Risk\ Matrix$, the structure and meaning of different components will be introduced. And in the subsection of $Factor\ Analysis$, the meaning of parameter and operational procedures of the factor analysis will be introduced.

A. Risk Matrix

The risk matrix is a classic analytical method that researches the probability and consequence of risk events. The basic idea of this semi-quantitative analysis method is to evaluate both the likelihood level and the consequence level of diverse risk accidents by locating them in different colored matrix cells. Generally, the vertical and horizontal axes of one risk matrix are consequence level and likelihood level respectively. The scales of probability and severity usually range from very low to very high. The different combinations decide the different colors, ranging from green to red (Cox, 2008) [1]. Fig. 1 shows a risk matrix diagram, and C represents the consequence level and L represents the likelihood level. In that risk matrix, the risk levels are divided into three categories, and the different colors represent different risk levels. Details of the risk matrix of case study design will be illustrated in Section IV.

L / C	Very low	Low	Medium	High	Very high
Critical	M	M	H	H	H
Serious	M	M	M	H	H
Moderate	L	M	M	M	H
Minor	L	L	M	M	M
Negligible	L	L	L	M	M

Fig. 1: 5×5 risk matrix

B. Factor Analysis

Factor analysis is a multivariate statistical technique used for extracting the main information of a large number of variables and evaluating their contribution to the total variation by common factors. This method enables the number of dimensions of the data sets to be reduced. The main information of the data sets changes into extracted common factors that represents most of the information. At that point, the extracted common factors can be used for analysis instead of using complete data sets directly. The advantage of this approach is that managers can retain the main effective information while reducing the volume of data. The mathematic model of factor analysis can be expressed as:

$$x_i = a_{i1}f_1 + a_{i2}f_2 + \cdots + a_{im}f_m + \varepsilon_i, i = 1, 2, \ldots, p, \quad (1)$$

where x_i is primitive variable, f_j is the common factor, and ε_i is the specific factor representing the specific factor of corresponding x_i, a_{ij} is the factor loading that is the covariance of x_i and f_j. It represents the dependent degree (or relative importance) of x_i to f_j. The significance of factor loadings is that they reflect the relationship between x_i and f_j, and they structure the component matrix, determining which common factors should be integrated. Meanwhile, the calculated component matrix represents the results; namely, the results of reducing dimension.

Note that Eq. (1) satisfies following 4 conditions:

1. $m \leq p$ it means the number of common factors is less than the number of primitive variables; namely, the number of dimensions may be reduced.

2. $cov(f_j, \varepsilon_i) = 0$; it means the common factors and special factors are uncorrelated; namely, they are independent.

3. Every common factor are uncorrelated, and each variance is 1. As shown in Eq. (2)

$$D([f_1, f_2, \ldots, f_m]') = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix}. \quad (2)$$

4. Every special factors are uncorrelated, and the variances may not be equal. As shown in Eq. (3)

$$D([\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_p]') = \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_p^2 \end{bmatrix}. \quad (3)$$

Generally speaking, the process of factor analysis can be summarized as following points:

Step1: Choose the appropriate data set. the data set must be tested for suitability; namely, we should prove that it can be used for factor analysis. Usually, we accept the data set when value of Kaiser-Meyer-Olkin test(KMO) is more than 0.7. Otherwise, we don't think this set of data can be used for factor analysis.

Step2: calculate correlation matrix. The original data set can form an observation matrix, and the calculation of correlation matrix is based on the observation matrix. For example,

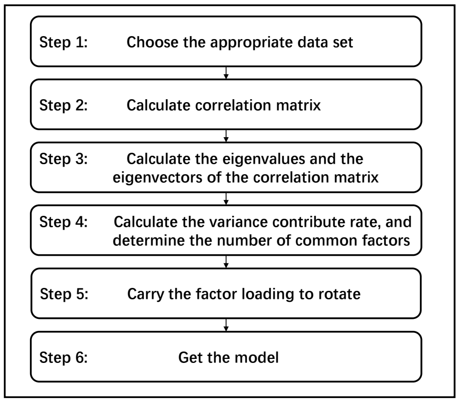


Fig. 2: The risk matrix with traditional method

the Eq. (4) is the observation matrix, and it represents there are p samples evaluated in n aspects.

$$(X_1, X_2, ..., X_n) = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ x_{p1} & x_{p2} & \dots & x_{pn} \end{bmatrix}. \tag{4}$$

Therefore, the correlation coefficient can be calculated by Eq. (5)

$$r = \frac{Cov(X_i, X_j)}{\sqrt{D(X_i)}\sqrt{D(X_j)}}, \tag{5}$$

where the $Cov(X_i, X_j)$ represents the covariance between X_i and X_j, and the $D(X_i)$ and $D(X_j)$ represent the variance of X_i and X_j respectively.

And the correlation coefficient matrix can be formed as Eq. (6) shown.

$$R = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ r_{21} & r_{22} & \dots & r_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ r_{n1} & r_{n2} & \dots & r_{nn} \end{bmatrix}, \tag{6}$$

Step3: Calculate the eigenvalues and the eigenvectors of the correlation matrix. In order to get the factor loadings, the eigenvalues and eigenvectors of the correlation matrix will be used for calculating. Suppose there are n eigenvalues, they ranged from big to small in order of Eq. (7), and their corresponding eigenvectors are represented as μ_n. Therefore, the factor loading matrix can be calculated as Eq. (8).

$$\lambda_1 > \lambda_2 > \cdots > \lambda_n, \tag{7}$$

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} = \begin{bmatrix} \mu_{11}\sqrt{\lambda_1} & \dots & \mu_{n1}\sqrt{\lambda_n} \\ \mu_{12}\sqrt{\lambda_1} & \dots & \mu_{n2}\sqrt{\lambda_n} \\ \dots & \dots & \dots \\ \mu_{1n}\sqrt{\lambda_1} & \dots & \mu_{nn}\sqrt{\lambda_n} \end{bmatrix} \tag{8}$$

Step4: Calculate the variance contribute rate, and determine the number of common factors. Commonly, there were two way to confirm the number of factors, depending on the number of eigenvalue (λ_i needs more than 1) and cumulative contribution rate (cumulative contribution rate need more than 80%). Here, we adopted observation method of cumulative contribution rate, calculating as Eq. (9):

$$c_k = \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{n} \lambda_i}, \tag{9}$$

where c_k is the cumulative contribution rate, k is the number of factors, λ_i is eigenvalue.

Step5: Carry the factor loading to rotate. Sometimes, there is no significant difference of each factor on the different variables in the non-rotating factor loading matrix. In order to distinguish each factor, we carry the factor to rotate.

Step6: Get the model. The factor analysis model can be built by the rotated factor loading matrix.

Note that the author of this paper suggests using software for calculation due to the complex calculation of factor analysis.

III. Optimal Risk Matrix

In this section, the optimal risk matrix will be proposed, according to the method of risk matrix and factor analysis.

A. The Limitations of Optimal Risk Matrix

The optimal risk matrix is similar to traditional risk matrix in terms of frames. Although optimal risk matrix is based on the traditional risk matrix, the usage of factor analysis makes it possible that the multidimensional data can be used in risk matrix. Please NOTE that the usage of multidimensional data set does not mean that there is no correlation between the data set. On the contrary, the data used here are related to the likelihood factors and the consequence factors indeed. Because the risk matrix aims at likelihood and consequence; and the factor analysis is a approach which integrates the similar data set. Therefore, the usage of optimal risk matrix still has its own limitations.

B. The Build Steps of Optimal Risk Matrix

In order to illustrate how to implement optimal risk matrix, the build steps of it will be introduced by a flow chart and explained in detail. The flow chart of optimal risk matrix is shown as follow:

Step1 : Test and process data. In this step, what we need to focus on involves the reliability of data set and the usage of factor analysis. On the once hand, the reliability means the trustworthiness of the data source. The data we get is not always reliable such as questionnaire. Therefore, it is necessary to ensure data validity before experiment. Generally, the Cronbachs alpha coefficient is used for confirming the validity [29], [30]. One the other hand, the data processing in factor analysis follows the steps and methods in Fig. 2, Section II-B. And the processed data will contain two factors, which represents the likelihood factor and the consequence factor.

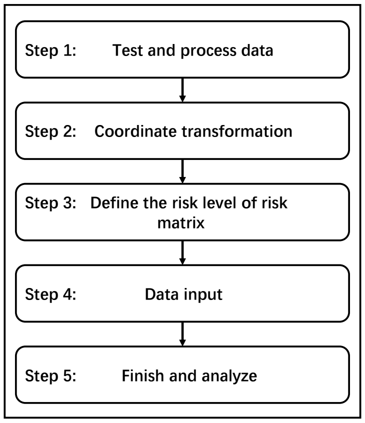


Fig. 3: Flow chart of build steps

Step2 : Coordinate transformation. The two processed factors should have been used in risk matrix but there are some other problems. Generally, the processed factors contain negative values which could not be used in the interval [0,5] of risk matrix. But the use of coordinate transformation is a method of converting the range of factors into interval [0,5]; the same as the interval on the risk matrix scale. The specific method of coordinate transformation is as Eq. (10) shown:

$$f_i^{'} = \frac{d-c}{b-a}(f_i - a), i = 1, 2, \tag{10}$$

where f_1 and f_2 represent likelihood factor or consequence factor respectively after processed. For example, if we consider f_1 first, and a is the minimum value in the components of f_1, b is the maximum value of the original components of f_1; c is the lower limit of the interval to which it will be transformed; namely, 0; and d is the upper limit of the interval to which it will be transformed; namely, 5. So Eq. (10) can also be depicted as:

$$f_i^{'} = \frac{5-0}{f_{imax} - f_{imin}}(f_i - f_{imin}), i = 1, 2, \tag{11}$$

Therefore, $f_1^{'}$ or $f_2^{'}$ will respectively be the X coordinates and Y coordinates of the risk matrix.

Step3 : Define the risk level of risk matrix. In the third step, what we should do is a partition, namely, the risk matrix will be divided into several regions, each region representing different risk situations. Generally, the tri-colored risk matrix is the most common diagram. The green represents low risk, the yellow represents moderate risk, and the red represents the high risk.

Step4 : Data input. The different $f_i^{'}$ will respectively represent X or Y coordinates.

Step5 : Finish and analyze. Complete the construction of optimal risk matrix. The information and significance of finished optimal risk matrix about the data should be analyzed and discussed on subsequent work.

IV. Case Study

In this section, we will take the data of an oil field as the sample, and make a comparative analysis by using traditional risk matrix and optimal risk matrix.

A. Questionnaire And The Data Sets

The data sets which used in our case study derived from a questionnaire of an oilfield at the Tarim Basin in Xinjiang Province, China. In the questionnaire, there are 123 risk questions, and each question would be evaluated from 4 aspects: The likelihood of an accident(x_1), The influence of an accident(x_2), The public opinion of an accident(x_3) and The economic losses of an accident(x_4). Each respondent will fill in his own opinion from the $'degrees'$ according himself. The $'degrees'$ which were divided into 5 showed in Table I. Finally, the average score of the answers were made into the original data sets.

In this case, the result of Cronbachs alpha coefficient was 0.652, so the data sets were comparatively trusted. And the KMO value is 0.734 that means the data can be used for factor analysis. So We can use the data for analysis.

B. The Building Process of Traditional Risk Matrix

In this subsection, we will use the original data sets in a traditional risk matrix. The traditional risk matrix is established here to design a contrast for the subsequent optimal risk matrix.

The four-aspects data sets which mentioned in Section IV-A is too many to be used in a traditional risk matrix. Considering x_2, x_3, x_4 are related, we gave a weight (generally, average value was used) to each risk aspect; namely, let

$$\bar{x} = \frac{x_2 + x_3 + x_4}{3}. \tag{12}$$

So $\bar{x}$ represents the consequence of accident.

Therefore, x_1 represents the likelihood level, which determines the abscissa value of the point; and $\bar{x}$ represents the consequence level, which determines the ordinate value of the point; That is how the points are determined in the traditional risk matrix.

Before inputting the data sets, the risk levels of the risk matrix scales were defined as follows:

In Table II, L_j is the likelihood level, categorized as 5 levels $(L_1, L_2, L_3, L_4, L_5)$ ordered from low to high: L_1 to L_5. C_i is the consequence level, also categorized 5 levels $(C_1, C_2, C_3, C_4, C_5)$ ordered from low to high: C_1 to C_5. $R_{ij}$is the risk level, and the i and j of R_{ij} are the location of i at C_i and the location of j at L_j. In this way, the risk table was divided into a 5×5 matrix of risk cells. Thus, each combination of C_i and L_j mapped to a risk cell. Here, the common operation rules such as addition rule and multiplication rule can be used to calculate the risk level. So, we decide to adopt multiplication rule for the mapping; namely, $R_{ij} = C_i \times L_j$. In this paper, the risk levels are divided into three categories (green, yellow, and red) according to their different values. The different colors represent different risk levels: the green,

TABLE I: Evaluative aspects of each sample.

Aspects	Description	Scale
	Once per 5 years	Very low
	Once per 3 $\sim$ 5 years	Low
x_1	Once per years	Medium
	Once per 6 months	High
	Once per months	Very high
	No injuries	Negligible
	At most 3 people suffered minor injuries	Minor
x_2	3$\sim$10 people suffered minor injuries	Moderate
	More than 10 people suffered minor injuries or 2$\sim$4 people suffered serious injuries	Serious
	Having casualty or more than 5 people suffered serious injuries	Critical
	Criticism by sub-committee	Negligible
	Reporting by local county government and media	Minor
x_3	Reporting by municipal government and media	Moderate
	Reporting by province government and media	Serious
	Reporting by national government and media	Critical
	Loss of 0$\sim$5 thousand RMB	Negligible
	Loss of 5$\sim$20 thousand RMB	Minor
x_4	Loss of 20$\sim$200 thousand RMB	Moderate
	Loss of 200$\sim$2000 thousand RMB	Serious
	Loss of more than 2000 thousand RMB	Critical

TABLE II: The risk level categories.

		The likelihood level				
		L_1	L_2	L_3	L_4	L_5
The consequence level	C_5	$R_{51} = 5$	$R_{52} = 10$	$R_{53} = 15$	$R_{54} = 20$	$R_{55} = 25$
	C_4	$R_{41} = 4$	$R_{42} = 8$	$R_{43} = 12$	$R_{44} = 16$	$R_{45} = 20$
	C_3	$R_{31} = 3$	$R_{32} = 6$	$R_{33} = 9$	$R_{34} = 12$	$R_{35} = 15$
	C_2	$R_{21} = 2$	$R_{22} = 4$	$R_{23} = 6$	$R_{24} = 8$	$R_{25} = 10$
	C_1	$R_{11} = 1$	$R_{12} = 2$	$R_{13} = 3$	$R_{14} = 4$	$R_{15} = 5$

yellow, and red colors are low, medium, and high-risk levels respectively. The points in a red cell have a quantitative value of at least 15, the points in a green cell have a value of at most 4, and the remaining points are in yellow cells.

The multiplication method has a small problem concerning the symmetry of risk cells. Baybutt (2015) [31] explained that the same consequence, likelihood, and risk level may be regarded as different results by different analysts, even during the same study. Generally, the points with a risk value of R_{41} (i.e., 4) should be in a green cell. But in this case, the color of the risk level that was green at R_{41} was not compliant with reality, so we changed it from green to yellow.

The risk matrix diagram with traditional method is shown in Fig. 4. Note that in the traditional risk matrix diagram the samples are excessively crowded except for some special points. Those crowded points obviously have a low degree of differentiation and cannot be analyzed clearly by managers. And we know that each value of the data sets from 0 to 5 was determined when we designed the questionnaire. Therefore the method of coordinate transformation can not be directly used here.

In this case, the traditional risk matrix is not a feasible approach to deal with multidimensional data sets. The reasons have 2 points.

First, multidimensional data sets cannot be directly used on traditional risk matrix because of its limited dimensions. Also, the method of using weighting is not effective in this case. Second, the essential theory of this phenomenon is that the inner relation of each data set is strong, so it leads to this

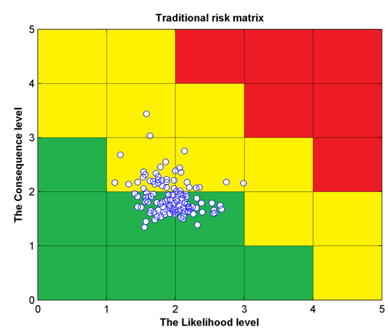


Fig. 4: The risk matrix with traditional method

unsuccessful risk matrix. The inner relation can be explained by correlation coefficient matrix.

The 123 questions and four aspects formed a 123×4 observation matrix as Eq. (13) shown.

$$X = \begin{bmatrix} x_{1,1} & x_{1,2} & x_{1,3} & x_{1,4} \\ x_{2,1} & x_{2,2} & x_{2,3} & x_{2,4} \\ \vdots & \vdots & \vdots & \vdots \\ x_{123,1} & x_{123,2} & x_{123,3} & x_{123,4} \end{bmatrix}. \quad (13)$$

And the correlation coefficient matrix can be calculated by Eq. (5) and shown in Eq. (14)

$$R = \begin{bmatrix} 1 & -0.129 & -0.325 & -0.256 \\ -0.129 & 1 & 0.777 & 0.844 \\ -0.325 & 0.777 & 1 & 0.886 \\ -0.256 & 0.844 & 0.886 & 1 \end{bmatrix}, \tag{14}$$

On the one hand, it is obvious that the values of r_{23}, r_{24}, r_{33} are comparatively high, which means they have a strong correlation. On the other hand, considering the problem about percentage of occupancy, the weight of each of the risk aspect may be different. Thus, the identical weights are inappropriate to be used for calculating, so the traditional risk matrix is not an effective and accurate approach to handle this problem.

C. The Building Process of Optimal Risk Matrices

To avoid the problem mentioned in the last paragraph, Section IV-B, we will design a optimal risk matrix to experiment. In the optimal risk matrix, factor analysis is used to reduce dimension of data sets. Namely, we wish reduce the dimensions and let it shown in Fig. 5. And the 2 dimensions can represent the likelihood factor and consequence level respectively.

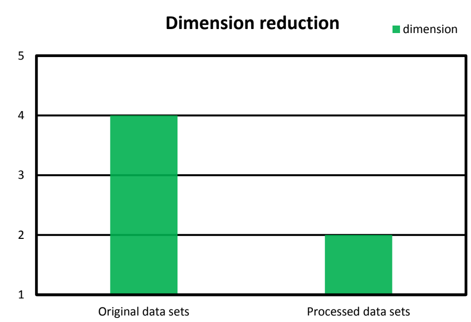


Fig. 5: Factor analysis for reduction dimension

The use of factor analysis makes it possible to achieve a quantitative dimensionality reduction. The 4-dimension data sets will be reduced to 2-dimension, as shown in Fig. 6.

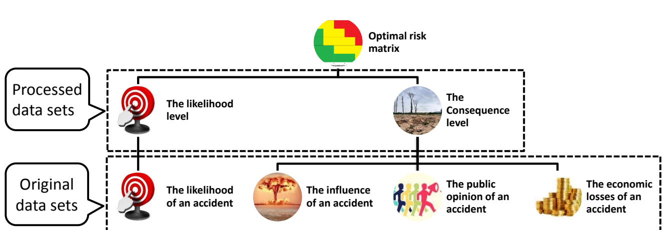


Fig. 6: Dimension reduction and integration

In Fig. 6, we let original 4-dimensional data sets change into 2-dimensional data by factor analysis. After factor analysis, one of the dimensions is the likelihood level, and it can be represented by the data set of $the\ likelihood\ of\ an\ accident$ in our case study; the other is the consequence level, and it actually integrated by factor analysis from $the\ influence\ of\ an\ accident$, $the\ public\ opinion\ of\ an\ accident$ and $the\ economic\ losses\ of\ an\ accident$.

Meanwhile, the $Component\ Matrix$ and the $Component\ Score\ Matrix$ can be calculated by SPSS.

TABLE III: Component Matrix And Component Score Matrix

(a) Component Matrix

	Component 1	Component 2
x_1	-0.379	0.922
x_2	0.901	0.259
x_3	0.946	0.011
x_4	0.958	0.110

(b) Component Score Matrix

	Component 1	Component 2
x_1	-0.137	0.992
x_2	0.325	0.279
x_3	0.342	0.012
x_4	0.346	0.118

In Table III(a), the $Component\ Matrix$, it represents the 2 dimensions after factor analysis. In $Component$ 1, we can observe the x_2 , x_3 , x_4 have a high and similar correlation coefficient values. Considering the meaning of x_2, x_3, x_4, it means that the $component$ 1 is referred to the consequence, and we name it the consequence factor. Similarly, in $Component$ 2, the x_1 has a high value, so it means the $Component$ 2 refers to the likelihood. Therefore, we name it the likelihood factor. Thus, the factor model can be expressed as Eq. (15)

$$\begin{cases} x_1 = & -0.379 f_1 + 0.922 f_2 + \varepsilon_1 \\ x_2 = & 0.901 f_1 + 0.259 f_2 + \varepsilon_2 \\ x_3 = & 0.946 f_1 + 0.011 f_4 + \varepsilon_3 \\ x_4 = & 0.958 f_1 + 0.110 f_4 + \varepsilon_4. \end{cases} \tag{15}$$

And the elements in $Component\ Score\ Matrix$ (TableIII(b)) are similar to assign weights to different x_i. Thus, we assign different weights to x_2, x_3 and x_4, which take full account of the different factors with different occupancy rates rather than traditional average method(Eq. (12)). Therefore, the factor score formula can be expressed in Eq. (17).

$$\begin{cases} f_1 = & -0.137 x_1 + 0.325 x_2 + 0.342 x_3 + 0.346 x_4 \\ f_2 = & 0.992 x_1 + 0.279 x_2 + 0.012 x_3 + 0.118 x_4, \end{cases} \tag{16}$$

where f_1 represents the consequence level, f_2 represents the likelihood level, x_i is standardized original data sets.

The coordinate transformation will be used like it shown in Eq. (11), to change the inconsistent interval to consistent interval [0,5].

$$\begin{cases} f_1 \xrightarrow{Eq.(11)} f_1^{'} \\ f_2 \xrightarrow{Eq.(11)} f_2^{'} \end{cases}, \tag{17}$$

where $f_1^{'}$ represents the values of the vertical coordinate; $f_2^{'}$ represents the values of the abscissa coordinate; and the coordinate $(f_2^{'}, f_1^{'})$ will be input into the risk matrix diagram.

In this study, 123 points are scattered in the optimal risk matrix by reducing the number of dimensions in the aspects

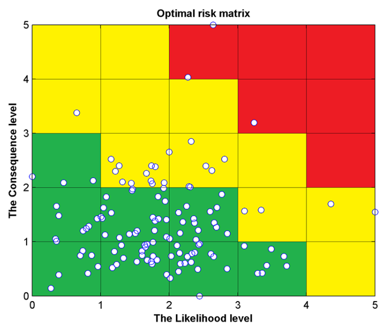


Fig. 7: The optimal risk matrix

of data sets in Fig. 7. And some points will be selected as examples to illustrate.

Points represent a high potential risk level when they are in the red risk cells. For example, in Fig. 8, the 6^{th} sample the risk of fire explosion, its coordinates are (2.638, 5); the 60^{th} sample the risk of fire and explosion and their subsequent influence, its coordinates are (2.271, 4.030); the 61^{st} sample the risk of poison gas or harmful gas, its coordinates are (3.238, 3.196).

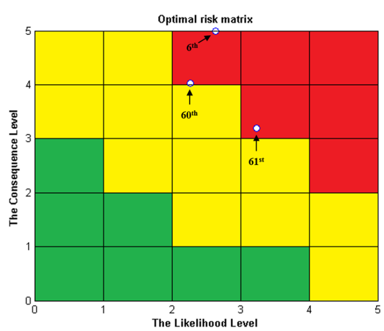


Fig. 8: Samples in high-risk level

The three points that are in high-risk level should be paid more attention to. Points represent a moderate risk level when they are in the yellow cells. For example, in Fig. 9, the 5^{th} sample the risk of oil tank collapse, its coordinates are (0.650, 3.379); the 26^{th} sample the risk of inside and outside of pipeline corrosion, its coordinates are (1.437, 5); and the 119^{th} sample the risk of oil pipeline break, its coordinates are (2.807, 2.524), they and other 35 points are in a moderate state.

And the remaining points represent a relatively low risk level when they are in the green cells. For example, in Fig. 10., the 2^{nd} sample the risk of oil tank wreck, its coordinates are (0, 2.199); the 103^{rd} sample the risk of staff occupational moral, its coordinates are (2.441, 0); and the 106^{th} sample the risk of staff health conditions, its coordinates are (1.749, 0.593), they and the rest of the 79 points are in a relatively safe state. Considering the main products of oilfield, they are inflammable and explosive. Therefore, the manager should pay due attention to the fire prevention and control of oilfield management problem.

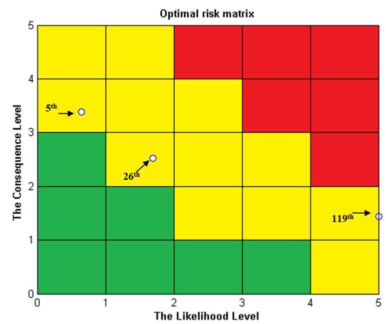


Fig. 9: Samples in moderate risk level

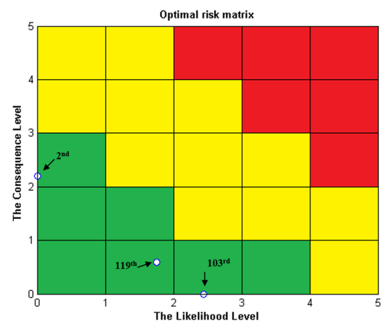


Fig. 10: Samples in low risk level

D. Results And Analysis

In a sharp contrast with the traditional risk matrix and the optimal risk matrix shown in Fig. 11. The points in Fig. 11(b) are more discrete and distinct than the points in Fig.11(a). It means that the method of optimal risk matrix is superior to the traditional risk matrix. Moreover, we also have discussed some shortcomings of the optimal risk matrix in the Section III-A. First, the multidimensional data needs to be able to refer to the likelihood and consequence of data. Second, there should be a strong internal relationship between the data. Third, in the process of using factor analysis, the dimension problem should be paid more attention to.

V. Conclusion

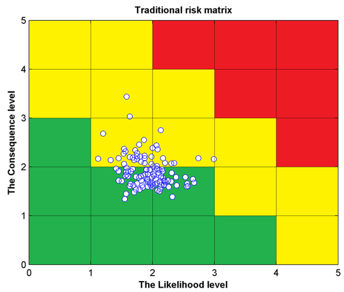
This study proposed a new approach in regard to a optimal risk matrix. That covers methods, computational model, problem, usage of limitation and comparative analysis. The case study results and comparative analysis indicate the efficiency of optimal risk matrix. The optimal risk matrix is an acceptable choice due to three reasons:

1. The usage of factor analysis makes multidimensional related data to be simplified.

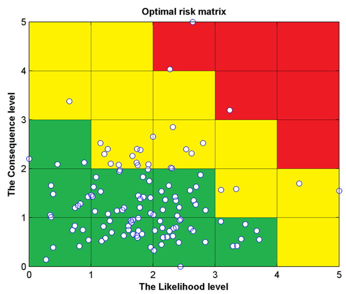
2. It has a significant degree of dispersion.

3. As a decision-making tool, it can also be used in many field.

In summary, the major contribution of this work is to propose a approach by combining risk matrix and factor analysis. As we known, the multidimensional data sets is usually more common in practical problems. We hope that



(a) Traditional Risk Matrix



(b) Optimal Risk Matrix

Fig. 11: Result comparison

this study can contribute to the both Management Decision Science and Statistics.

ACKNOWLEDGEMENTS

This research is supported by the National Science & Technology Major Project, Southwest Petroleum University, China ($No.2016ZX05020$) and funded by the project of $'DrillingEngineeringIntegrationSoftware'$ from Science & Technology Development Project of China National Petroleum Corporation. And author is grateful for their patience and support.

REFERENCES

[1] A. T. Cox, Louis, What's wrong with risk matrices?, Risk analysis 28 (2) (2008) 497–512.

[2] B. Ale, P. Burnap, D. Slater, On the origin of PCDS C (Probability consequence diagrams), Safety Science 72 (2015) 229–239.

[3] H. Per, S. R. Bolle, H. Eva, Video calls from lay bystanders to dispatch centers - risk assessment of information security, Bmc Health Services Research 11 (1) (2011) 1–7.

[4] B. Fautrel, B. Granger, B. Combe, A. Saraux, F. Guillemin, X. L. Loet, Matrix to predict rapid radiographic progression of early rheumatoid arthritis patients from the community treated with methotrexate or leflunomide: results from the ESPOIR cohort, Arthritis Research & Therapy 14 (6) (2012) R249.

[5] E. Henriksen, T. M. Burkow, E. Johnsen, L. K. Vognild, Privacy and information security risks in a technology platform for home-based chronic disease rehabilitation and education., Bmc Medical Informatics & Decision Making 13 (1) (2013) 85.

[6] J. E. Arnetz, L. Hamblin, J. Ager, D. Aranyos, M. J. Upfal, M. Luborsky, J. Russell, L. Essenmacher, Application and implementation of the hazard risk matrix to identify hospital workplaces at risk for violence., American Journal of Industrial Medicine 57 (11) (2014) 1276–84.

[7] J. Blahut, T. Glade, S. Sterlacchini, Debris flows risk analysis and direct loss estimation: the case study of Valtellina di Tirano, Italy, Journal of Mountain Science 11 (2) (2014) 288.

[8] J. Klimeš, J. Blahut, Landslide risk analysis and its application in regional planning: an example from the highlands of the Outer Western Carpathians, Czech Republic, Natural Hazards 64 (2) (2012) 1779–1803.

[9] R. Ullah, G. P. Shivakoti, A. Kamran, F. Zulfiqar, Farmers versus nature: managing disaster risks at farm level, Natural Hazards 82 (3) (2016) 1931–1945.

[10] F. Zulfiqar, R. Ullah, M. Abid, A. Hussain, Cotton production under risk: a simultaneous adoption of risk coping tools, Natural Hazards 84 (2) (2016) 959–974.

[11] M. R. Moradi, M. A. E. Farsangi, Application of the Risk Matrix Method for Geotechnical Risk Analysis and Prediction of the Advance Rate in Rock TBM Tunneling, Rock Mechanics & Rock Engineering 47 (5) (2014) 1951–1960.

[12] P. C. Cacciabue, M. Cassani, V. Licata, I. Oddone, A. Ottomaniello, A practical approach to assess risk in aviation domains for safety management systems, Cognition Technology & Work 17 (2) (2015) 249–267.

[13] W. K. K. Hsu, S. H. S. Huang, W. J. Tseng, Evaluating the risk of operational safety for dangerous goods in airfreights C A revised risk matrix based on fuzzy AHP, Transportation Research Part D 48 (2016) 235–247.

[14] J. Skorupski, The simulation-fuzzy method of assessing the risk of air traffic accidents using the fuzzy risk matrix, Safety Science 88 (2016) 76–87.

[15] S. Khaleghi, S. Givehchi, S. Karimi, Consensus Statement on Urethral Trauma - The Journal of Urology, World Applied Programming 1 (6) (2013) 1007–1014.

[16] H. C. Liu, Y. Z. Chen, J. X. You, H. Li, Risk evaluation in failure mode and effects analysis using fuzzy digraph and matrix approach, Journal of Intelligent Manufacturing 27 (4) (2016) 805–816.

[17] E. Ataallahi, S. R. Shadizadeh, Fuzzy consequence modeling of blowouts in Iranian drilling operations; HSE consideration, Safety Science 77 (2015) 152–159.

[18] E. D. Smith, W. T. Siefert, D. David, Risk matrix input data biases, Systems Engineering 12 (4) (2010) 344–360.

[19] D. J. Ball, J. Watt, Further thoughts on the utility of risk matrices, Risk Analysis 33 (11) (2013) 2068C–2078.

[20] N. J. Duijm, Recommendations on the use and design of risk matrices, Safety Science 76 (2015) 21–31.

[21] A. S. Markowski, M. S. Mannan, Fuzzy risk matrix, Journal of Hazardous Materials 159 (1) (2008) 152–157.

[22] A. F. Shapiro, Fuzzy post-retirement financial concepts: an exploratory study, Metron 71 (3) (2013) 261–278.

[23] P. C. Wierenga, L. Lie-A-Huen, S. E. de Rooij, N. S. Klazinga, H. J. Guchelaar, S. M. Smorenburg, Application of the Bow-Tie model in medication safety risk analysis: consecutive experience in two hospitals in the Netherlands, Drug Saf 32 (8) (2009) 663–73.

[24] L. Lu, W. Liang, L. Zhang, H. Zhang, Z. Lu, J. Shan, A comprehensive risk evaluation method for natural gas pipelines by combining a risk matrix with a bow-tie model, Journal of Natural Gas Science & Engineering 25 (2015) 124–133.

[25] H. Ni, A. Chen, N. Chen, Some extensions on risk matrix approach, Safety Science 48 (10) (2010) 1269–1278.

[26] Y. Duan, J. Zhao, J. Chen, G. Bai, A risk matrix analysis method based on potential risk influence: A case study on cryogenic liquid hydrogen filling system, Process Safety & Environmental Protection 102 (2016) 277–287.

[27] N. Khakzad, F. Khan, P. Amyotte, Quantitative risk analysis of offshore drilling operations: A Bayesian approach, Safety Science 57 (57) (2013) 108–117.

[28] R. Flage, W. Røed, A Reflection on Some Practices in the Use of Risk Matrices, in: European Safety and Reliability Conference, 2012.

[29] C. M. Fiehler, B. L. Cypher, L. R. Saslaw, Effects of oil and gas development on vertebrate community composition in the southern San Joaquin Valley, California, Global Ecology & Conservation 9 (C) (2017) 131–9141.

[30] C.L. Zhao, X.J. Cao, X.J. Xie, X.H. Yuan and B.Wang, "Environmental Aspects of Drilling Operation Risks: A Probability Prediction based on Fuzzy Time Series," Ekoloji, vol. 27, no.106, pp. 1493–1502, 2018.

[31] A. Leontitsis, J. Pagge, A simulation approach on Cronbach's alpha statistical significance, Mathematics & Computers in Simulation 73 (5) (2007) 336–340.