Risk Assessment of Debris Flow Disaster Based on Principal Component Analysis

Dejie Yu* and Lansong Lv
Shandong Geological Environmental Monitoring Station, Jinan 250014, Shandong, China

*Corresponding author: lvlins@163.com

Abstract. Debris flow disasters have the characteristics of large energy, strong explosiveness, and great harm. They cause irreversible damage to the environment and at the same time, seriously threaten the safety of human life and property. Therefore, it is essential to evaluate the risk of debris flow. In this paper, nine factors such as the basin area, the length of the main ditch, and the cutting density of the basin along the Jinsha River are used to calculate the cumulative contribution rate to screen the principal components using the principal component analysis method, and then comprehensively score the linear combination of independent principal components. Therefore, the risk assessment of each debris flow channel is carried out, and the results obtained are compared with the methods in different documents, which shows that this method has better applicability and feasibility. This method of determining weights overcomes the shortcomings of the subjective method of determining weights and is more reasonable and objective. The research results can provide decision-making references for relevant departments and industries.

Keywords: Geological hazard, debris flow, principal component analysis, risk assessment.

1. Introduction
Debris flow is an impressive torrent that occurs in mountainous areas. It is a viscous fluid mixed with a large number of solid materials such as sand and gravel and water. It is a natural geological phenomenon that is concentrated in rainstorms or rapid flow in valleys under gravity [1]. Debris flow disasters have the characteristics of enormous energy, muscular explosiveness and significant harm, which cause irreversible damage to the environment, and at the same time seriously threaten the safety of human life and property [2]. The formation of debris flow must meet the three conditions of a large amount of loose solid matter, sufficient water source and steep terrain. Our country has a complex topography and numerous mountains. Most mountainous areas have abundant precipitation and large elevation differences [3-4], prone to debris flow disasters and even trigger the chain of debris flow disaster effects [5-6]. Therefore, it is of great significance to conduct prior risk assessments in areas where debris flow disasters are likely to occur, reduce the probability of debris flow disasters, timely prevent and control debris flow disasters, and reduce personal and property casualties.
Many experts and scholars have conducted a lot of research on the risk prediction of debris flow. Liu Xilin [7] was the first to study the risk of debris flow, Duan Liyao [8] established a debris flow risk evaluation model based on the analytic hierarchy process to predict the risk of debris flow, Liu Hongjiang [9] used GIS to carry out debris flow risk. According to the regional division principle, the hazard level of debris flow in each area is calculated. Xiang Liangjun [10] conducted a three-dimensional flow field numerical simulation of Xiaojiagou debris flow and conducted a debris-flow risk assessment.

The principal component analysis method is to linearly combine many related indicators and recombine them into several sets of new independent comprehensive indicators [11]. By calculating the total indicator values, objective weighting can avoid the influence of subjective factors. Objectively reflects the actual relationship between the samples [12]. This paper takes the debris flow along the Jinsha River as an example to study, apply the principal component theory to the risk assessment of debris flow, determine the principal component and weight through the calculated comprehensive contribution rate, and reflect the comprehensive.

2. Principal Component Analysis

The principal component analysis is a multivariate analysis technique [13]. It uses the idea of dimensionality reduction to transform high-dimensional variables into new replacement indicators based on linear transformations. It becomes a principal component as a comprehensive indicator to objectively determine the weight of each factor to achieve simplify dimensionality reduction and comprehensive evaluation methods. The necessary steps of principal component analysis are as follows:

2.1. Raw data standardization

If m principal component analysis index variables \((x_1, x_2, \ldots, x_m)\) have n evaluation objects, the value of the j-th index of the i-th evaluation object is \(x_{ij}\), and the value of \(x_{ij}\) each index is converted into a standardized index \(\tilde{x}_{ij}\):

\[
\tilde{x}_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}, \quad (i = 1, 2, \ldots, n; j = 1, 2, \ldots, m)
\]

\[
\bar{x}_j = \frac{1}{n} \sum_{i=1}^{n} x_{ij}
\]

\[
s_j = \frac{1}{n-1} \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2, \quad (j = 1, 2, \ldots, m)
\]

Where \(\bar{x}_j, s_j\) are the sample mean and sample difference of the j-th indicator, corresponding to \(\tilde{x}_i = \frac{x_i - \bar{x}_j}{s_j}, \quad (i = 1, 2, \ldots, m)\) is a standardized indicator variable.

2.2. Build a correlation matrix between variables \(R\)

Correlation matrix \(R = (r_{ij})_{n \times m}, \quad r_{ij} = \frac{\sum_{k=1}^{n} \tilde{x}_{ik} \cdot \tilde{x}_{kj}}{n-1} (i, j = 1, 2, \ldots, m)\)

Where \(r_{ii} = 1\), \(r_{ij} = r_{ji}\), \(r_{ij}\) is the correlation coefficient between the i-th index and the j-th index.
2.3. Calculate the eigenvalues and eigenvectors of the correlation coefficient matrix

Calculate the eigenvalues \( \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_m \geq 0 \) of the correlation coefficient matrix \( R \) and the corresponding eigenvectors \( u_1, u_2, \ldots, u_m \), where \( u_j = (u_{1j}, u_{2j}, \ldots, u_{nj})^T \), eigenvectors form \( m \) new index vectors

\[
\begin{align*}
y_1 &= u_{11}\tilde{x}_1 + u_{12}\tilde{x}_2 + \ldots + u_{1n}\tilde{x}_n \\
y_2 &= u_{21}\tilde{x}_1 + u_{22}\tilde{x}_2 + \ldots + u_{2n}\tilde{x}_n \\
&\vdots \\
y_m &= u_{m1}\tilde{x}_1 + u_{m2}\tilde{x}_2 + \ldots + u_{mn}\tilde{x}_n
\end{align*}
\]

(4)

Where \( y_1 \) is the first principal component, \( y_m \) is the first principal component

2.4. Calculate the principal components

Calculate the information contribution rate of eigenvalues \( \lambda_j \) \((j = 1, 2, \ldots, m)\):

\[
b_j = \frac{\lambda_j}{\sum_{k=1}^{m} \lambda_k} \quad (j = 1, 2, \ldots, m)
\]

(5)

\[
\alpha_p = \frac{\sum_{k=1}^{p} \lambda_k}{\sum_{k=1}^{m} \lambda_k}
\]

(6)

\( b_j \) is the information contribution degree of the main component \( y_j \), and \( \alpha_p \) is the cumulative contribution rate of the main component \( y_1, y_2, \ldots, y_p \). When \( \alpha_p = 0.85 \), the first \( p \) index variables are selected as the \( p \) principal components, instead of the original \( m \) index variables, so that the \( p \) principal comprehensive analysis of ingredients.

3. Debris flow risk assessment

According to the current research results and engineering status, five typical debris flow ditches along the Jinsha River were selected for risk assessment, and the Aibagou, Zhuluhe, Fujiagou, Heizhe, and Daganggou were selected for risk assessment. Drainage area \( S_1 \), Length of the main ditch \( S_2 \), Maximum relative height difference of the watershed \( S_3 \), Watershed cutting density \( S_4 \), Mud-sand replenishment section length ratio \( S_5 \), Maximum daily rainfall \( S_6 \), the maximum washout of debris flow \( S_7 \), Frequency of debris flow \( S_8 \), Basin population density \( S_9 \), nine indicators are used as the influencing factors of debris-flow risk. The value of each factor is shown in Table 1.
Table 1. Index values of factors [12].

| Debris flow name                      | Aibagou | Zhuluhe | Fujiagou | Heizhe | Daganggou |
|---------------------------------------|---------|---------|----------|--------|-----------|
| Drainage area (km²)                   | 5.84    | 152.60  | 8.62     | 51.7   | 18.90     |
| Length of the main ditch (km)         | 5.08    | 26.3    | 5.16     | 13.9   | 5.10      |
| Maximum relative height difference of | 148     | 1.3     | 1.53     | 1.31   | 0.59      |
| watershed (km)                        |         |         |          |        |           |
| Watershed cutting density             | 8.79    | 4.32    | 6.34     | 5.13   | 10.95     |
| Mud-sand replenishment section length ratio | 1.19   | 1.7     | 1.26     | 1.15   | 1.11      |
| Maximum daily rainfall (mm)           | 0.62    | 0.08    | 0.44     | 0.12   | 0.74      |
| Maximum rush out of a debris flow (10^4 m³) | 8.37  | 31.2    | 12.9     | 15.03  | 13.50     |
| Frequency of debris flow (times/100a) | 72.00   | 10.5    | 66.3     | 2.3    | 74.18     |
| Basin population density (number of people /km²) | 68.00 | 4.00    | 10.00    | 9.00   | 15.00     |

3.1. Data standardization

Positive index:

\[ x_{ij} = \frac{x_{ij} - x_{j\min}}{x_{j\max} - x_{j\min}}, i = 1, 2, \ldots, m; j = 1, 2, \ldots, n; \]  

Negative indicators:

\[ x_{ij} = \frac{x_{ij} - x_{j\min}}{x_{j\max} - x_{j\min}}, i = 1, 2, \ldots, m; j = 1, 2, \ldots, n; \]  

From Drainage area, Length of the main ditch, Maximum relative height difference of watershed, Watershed cutting density, Mud-sand replenishment section length ratio, Maximum daily rainfall, Maximum rush out of a debris flow, Frequency of debris flow, Basin population density, the nine indicators through indicator standardization, removes the restriction of indicator units and ensures the accuracy of the calculation results.

3.2. Calculating principal components and expressions

First, establish the correlation coefficient matrix between the variables, and calculate the principal components by calculating the eigenvalues and eigenvectors. The calculated principal components can replace the original nine indicator variables to calculate the principal components and their expressions. In most information, a few comprehensive indicators are used to replace the original indicators for analysis.

3.3. Judging the risk level of debris flow

Substitute the principal components obtained in step (2) into the standardized data in step (1), and conduct comprehensive analysis through the final comprehensive score. According to the standard of the principal component analysis method, the risks of debris flow disasters at five locations are combined with the site conditions.

4. Risk Assessment

4.1. Data standardization

Standardize the data in Table 2, according to formula 7-8, and the results obtained are shown in the following table:
Table 2. Standardized index value.

| Debris flow name | Aibagou | Zhuluhe | Fujiaogou | Heizhe | Daganggou |
|------------------|---------|---------|-----------|--------|-----------|
| Drainage area    | 0.0000  | 1.0000  | 0.0189    | 0.3125 | 0.0890    |
| Length of the main ditch | 0.0000  | 1.0000  | 0.0038    | 0.4156 | 0.0009    |
| Maximum relative height difference of watershed | 0.9468  | 0.7553  | 1.0000    | 0.7660 | 0.0000    |
| Watershed cutting density | 0.6742  | 0.0000  | 0.3047    | 0.1222 | 1.0000    |
| Mud-sand replenishment section length ratio | 0.1356  | 1.0000  | 0.2542    | 0.0678 | 0.0000    |
| Maximum daily rainfall | 0.8182  | 0.0000  | 0.5455    | 0.0606 | 1.0000    |
| Maximum rush out of a debris flow | 0.0000  | 1.0000  | 0.1984    | 0.2917 | 0.2247    |
| Frequency of debris flow | 0.9697  | 0.1141  | 0.8904    | 0.0000 | 1.0000    |
| Basin population density | 1.0000  | 0.0000  | 0.0938    | 0.0781 | 0.1719    |

4.2. Calculating principal components and their expressions

4.2.1. Build a correlation matrix between variables. By calculating Drainage area, Length of the main ditch, Maximum relative height difference of watershed, Watershed cutting density, Mud-sand replenishment section length ratio, Maximum daily rainfall, Maximum rush out of a debris flow, Frequency of debris flow, Basin population density. The correlation coefficient and characteristic value of the indicator are analyzed. The number of principal components is determined according to the cumulative contribution rate. Correlation coefficient matrix:

\[
R = \begin{pmatrix}
1.000 & 0.988 & 0.042 & -0.672 & 0.897 & -0.763 & 0.978 & -0.754 & -0.491 \\
0.988 & 1.000 & 0.131 & -0.752 & 0.861 & -0.844 & 0.945 & -0.838 & -0.481 \\
0.042 & 0.131 & 1.000 & -0.634 & 0.273 & -0.439 & -0.020 & -0.211 & 0.253 \\
-0.672 & -0.752 & -0.634 & 1.000 & -0.649 & 0.961 & -0.633 & 0.829 & 0.469 \\
0.897 & 0.861 & 0.273 & -0.649 & 1.000 & -0.624 & 0.915 & -0.491 & -0.337 \\
-0.763 & -0.844 & -0.439 & 0.961 & 0.624 & 1.000 & -0.704 & 0.950 & 0.535 \\
0.978 & 0.945 & -0.020 & -0.633 & 0.915 & -0.704 & 1.000 & -0.666 & -0.601 \\
-0.754 & -0.838 & -0.211 & 0.829 & -0.491 & 0.950 & -0.666 & 1.000 & 0.521 \\
-0.491 & -0.481 & 0.253 & 0.469 & -0.337 & 0.535 & -0.601 & 0.521 & 1.000
\end{pmatrix}
\]

According to the principle of selecting the number of principal components in the principal component analysis method, it can be seen from the table that the feature values of component 1 and component 2 are both greater than 1, and the cumulative contribution rate reaches 85.486%, so a total of two principal components are extracted. These two principal components summarize the nine index factors' vital information, and the two principal components are independent and irreplaceable.

Table 3. Total explained variance.

| Ingredient | Total | Initial eigenvalue variance % | Grand total% | Extract the sum of squares and load variance % | Grand total% |
|------------|-------|------------------------------|--------------|---------------------------------------------|-------------|
| 1          | 6.179 | 68.660                       | 68.660       | 6.179                                       | 68.660      |
| 2          | 1.514 | 16.826                       | 85.486       | 1.514                                       | 85.486      |
| 3          | 0.888 | 9.862                        | 95.348       |                                             |             |
| 4          | 0.419 | 4.652                        | 100.00       |                                             |             |
| 5          | 1.00E-013 | 1.027E-013                    | 100.00       |                                             |             |
| 6          | 1.00E-013 | 1.023E-013                    | 100.00       |                                             |             |
| 7          | -1.00E-013 | -1.000E-013                   | 100.00       |                                             |             |
| 8          | -1.001E-013 | -1.010E-013                   | 100.00       |                                             |             |
| 9          | -1.002E-013 | -1.022E-013                   | 100.00       |                                             |             |
Therefore, two principal components \((Y_1, Y_2)\) can replace the original nine indicators. Let \(Y_1\) and \(Y_2\) represent the first and second principal components, respectively. Since the component score coefficient matrix needs to be processed before it can be used in the expression, the processed results are shown in Table 4:

| Ingredient | \(S_1\) | \(S_2\) | \(S_3\) | \(S_4\) | \(S_5\) | \(S_6\) | \(S_7\) | \(S_8\) | \(S_9\) |
|------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| 1          | 0.0611  | 0.0632  | 0.0173  | -0.0571 | 0.0547  | -0.0603 | 0.0595  | -0.0567 | -0.0386 |
| 2          | 0.1211  | 0.0634  | -0.5039 | 0.2284  | 0.0244  | 0.1349  | 0.1731  | 0.0455  | -0.2463 |

Therefore, the linear combinations of principal component expressions are

\[
\begin{align*}
\text{Ingredient 1:} & \quad Y = 0.0611S_1 + 0.0632S_2 + 0.0173S_3 - 0.0571S_4 + 0.0547S_5 - 0.0603S_6 + 0.0595S_7 - 0.0567S_8 - 0.0386S_9 \\
\text{Ingredient 2:} & \quad Y = 0.1211S_1 + 0.0634S_2 - 0.5039S_3 + 0.2284S_4 + 0.0244S_5 + 0.1349S_6 + 0.1731S_7 - 0.0455S_8 - 0.2463S_9
\end{align*}
\]

After calculating the principal component score, each principal component's contribution rate is defined as the weight. The formula for calculating the total score is:

\[
F = 0.6866Y_1 + 0.16826Y_2
\]

It reflects the risk level of debris flow, and the specific principal component analysis values are shown in Table 5:

| Debris flow name | Aibagou | Zhuluhe | Fujiagou | Heizhe | Daganggou |
|------------------|---------|---------|----------|--------|-----------|
| Ingredient 1      | -0.1577 | 0.2452  | -0.0600  | 0.0600 | -0.1620   |
| Ingredient 2      | -0.4115 | 0.0066  | -0.3002  | -0.2528| 0.4162    |
| Risk assessment   | -0.1775 | 0.1694  | -0.0917  | 0.0028 | -0.0412   |

It can be seen from the content in Table 6 that there are specific differences between the method in this paper and literature 12 and 13, but it is generally consistent with the comprehensive evaluation result of literature 13.

| Debris flow name | Aibagou | Zhuluhe | Fujiagou | Heizhe | Daganggou |
|------------------|---------|---------|----------|--------|-----------|
| Method of this article | Moderate | Moderate | Mild | Mild | Mild |
| Literature [12]  | Severe  | Mild    | Mild     | Mild   | Moderate  |
| Literature [13]  | Moderate | Moderate | Mild     | Mild   | Moderate  |

5. Conclusion

In this paper, the principal component analysis method is applied to the risk assessment of debris flow along the Jinsha River. By comparing with methods in different documents, it is concluded that this method has better applicability and feasibility. In the comprehensive evaluation, each principal component's weight is determined by the contribution rate, reflecting the proportion of the amount of information contained in the original data of the principal component to the total amount of information. This method of determining weight overcomes the shortcomings of the subjective weight determination method, and has more Reasonability and objectivity provide a reference for the risk assessment of debris flow.

References

[1] Tang Bangxing, Mudslides in China, Chengdu Institute of Mountain Hazards and Environment [J]. 2000.
[2] Wei Fangqiang, Hu Kaiheng, Cui Peng, Zhong Dunlun, Xie Hong, He Yiping. Decision support system of debris flow mitigation for mountain towns [J]. Journal of Natural Disasters, 2002 (02): 31-36.

[3] Xue, Y., Kong, F., Qiu, D. et al. The classifications of water and mud/rock inrush hazard: a review and update. Bull Eng Geol Environ (2020).

[4] Zhou, B.H., Y.G. Xue, S.C. Li, D.H. Qiu, Y.F. Tao, K. Zhang, X.L. Zhang and T. Xia. 2020. Probabilistic analysis of tunnel collapse: Bayesian method for detecting change points. Geomechanics and Engineering 22 (4): 291-303.

[5] Wu Qian, Qiao Haitao, Deng Gangyang, Ji Lei, Wang Xiaowei, Liu Dunlong. Construction of a GIS-based early warning system for debris flow disasters in Southwest China [J]. Technology Innovation and Application, 2021 (04): 99-102.

[6] Xue Yiguo, Kong Fanmeng, Yang Weimin, Qiu Daohong, Su Maoxin, Fu Kang, Ma Xinmin. Main unfavorable geological conditions and engineering geological problems along Sichuan—Tibet railway [J]. Chinese Journal of Rock Mechanics and Engineering, 2020, 39 (03): 445-468.

[7] Liu Xilin, Tang Chuan. Evaluation of the Risk of Debris Flow [M]. Beijing: Science Press, 1995.

[8] Duan Liyao, Xiong Mingming, Li Daming, Liu Xunping, Wang Jing, Li Peiyan. Assessment of local debris flow disaster in Tianjin based on AHP method [J]. Journal of Meteorology and Environment, 2019, 35 (06): 109-116.

[9] Liu Hongjiang, Tang Chuan, Cui Peng. Debris Flow Risk Division in Dongchuan District Supported by GIS [J]. Arid Land Geography, 2005 (04): 29-33.

[10] Xiang Liangjun. Three-dimensional flow field numerical simulation and risk assessment of the Xiaojiagou debris flow in the dam area of Jinsha Hydropower Station [D]. Jilin University, 2014.

[11] Zhang Wenlin. The application of principal component analysis in SPSS [J]. Marketing Research, 2005 (12): 31-34.

[12] Li Yanshuang, Zeng Zhenxiang, Zhang Min, Yu Shujie. Application of Principal Component Analysis in Multi-index Comprehensive Evaluation Method [J]. Journal of Hebei University of Technology, 1999 (01): 96-99.

[13] ABDI H, WILLIAMS L J W I R C S 2010. Principal component analysis. 2: 433-459.

[14] Zhang Chen, Wang Qing, Chen Jianping, Gu Fuguang, Zhang Wen. Evaluation of debris flow risk in Jinsha River based on combined weight process [J]. Rock and Soil Mechanics, 2011, 32 (03): 831-836.

[15] Zhao Minghua, Liao Binbin, Liu Sisi. Calculation of anti-slide piles spacing based on soil arching effect [J]. Rock and Soil Mechanics, 2010, 31 (04): 1211-1216.