Comparison of two occurrence risk assessment methods for collapse gully erosion
——A case study in Guangdong province

K Sun¹,², D B Cheng²*, J J He¹, Y L Zhao²
¹Huazhong Agricultural University resource and environmental college, Wuhan 430070
²Changjiang River Scientific Research Institute, Wuhan 430010
Corresponding author E-mail: xiao2005zhu@163.com

Abstract. Collapse gully erosion is a specific type of soil erosion in the red soil region of southern China, and early warning and prevention of the occurrence of collapse gully erosion is very important. Based on the idea of risk assessment, this research, taking Guangdong province as an example, adopt the information acquisition analysis and the logistic regression analysis, to discuss the feasibility for collapse gully erosion risk assessment in regional scale, and compare the applicability of the different risk assessment methods. The results show that in the Guangdong province, the risk degree of collapse gully erosion occurrence is high in northeastern and western area, and relatively low in southwestern and central part. The comparing analysis of the different risk assessment methods on collapse gully also indicated that the risk distribution patterns from the different methods were basically consistent. However, the accuracy of risk map from the information acquisition analysis method was slightly better than that from the logistic regression analysis method.

1. Introduction
Risk is a general term of the probability distribute function of the adverse events occurs in the future and the possibility of damage caused by it. Risk assessment refers to the process of recognizing the risk factors of the adverse events, calculating the risk of the studying area and determining the risk level[1-4]. The concept of risk assessment in the field of soil erosion is still in the exploratory research statement. Collapsing gully erosion is a special form of soil erosion in the red soil area of southern China, and it is also a significant sign of the seriousness of soil erosion[5-7]. At present, the study of collapse is mainly in-depth gradually in the meso and micro level, further expansion are required in the macro, risk assessment relevant researches are rare. Cheng Dongbing leads the research team, applying the concept of risk assessment to the collapse gully erosion. In geological and ecological view, he referenced the process of risk assessment, gradually defined the connotation of collapse gully erosion assessment, prepared the methods of it, and constructed the index system of risk assessment, put forward the procedure of collapse gully erosion. After nearly three years of researching, the risk evaluate collapse gully erosion in south China was completed preliminarily.

Based on the project preliminary work, Guangdong Province was selected as a case study, this paper focuses on the comparison of two risk assessments methods. It aims to optimize the method of collapse gully erosion risk assessment scientifically, and provides a reference decision-making to control the collapse gully erosion in Guangdong province.
2. Method

2.1 Information acquisition analysis method

The information acquisition analysis method characterizes the probability of events through the reduction of information entropy [8]. In the risk assessment of collapse gully erosion, the reduction of information entropy means the occurrence of collapsing. Information entropy is the expectation of the amount of information, so the amount of information is used to measure the possibility of collapse. The amount of information is determined by the probability of occurrence of collapse, as shown in equation (1)

\[ I(f, x_1, x_2, \ldots, x_n) = \log_2 \frac{p(f|x_1, x_2, \ldots, x_n)}{p_f} \]  

Where: \( f, x_1, x_2, \ldots, x_n \) is the information that significant factor combination \( x_1, x_2, \ldots, x_n \) has provided to collapse gully erosion; \( P(f|x_1, x_2, \ldots, x_n) \) is the probability of occurrence of collapse gully erosion under combined conditions \( x_1, x_2, \ldots, x_n \).

2.2 Logistic regression analysis method

Logistic regression analysis belong to the discriminant analysis, the value of dependent variable are 0 or 1 [9]. The logistic model can be used to analyze the collapse occurred as a dependent variable. The model can simulate the dependent variable and the independent variable of the binary response. According to the establishment mode, the probability of collapse occurrence in each unknown grid is predicted, and then the risk is evaluated based on the probability of its occurrence. The expression is:

\[ \ln \left( \frac{p}{1-p} \right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n \]  

Where the dependent variable \( P \) represents the probability of event occurrence; the independent variable is \( x_1, x_2, \ldots, x_n \); \( \beta \) is the explanatory variable coefficient of logistic equation, and the influence degree of the explanatory variable on the dependent variable can be used as a measure of \( e^\beta \). \( e^\beta \) is a ratio of the events occurrence frequency and the non-occurrence frequency, indicating the change of the variable due to the each addition of one unit.

The collapse density of each factor is used as the independent variable of logistic regression model, which not only shows the contribution of each factor to the collapse of every pile collapse, but also shows its contribution. The formula used is shown in equation (3):

\[ S = \sum_{i=1}^{n} D(i) \times V_i \]  

Where: \( S \) is the contribution value of each sampling grid to pile collapse; \( D(i) \) is the density of the collapse gully erosion of factor \( Xi \); \( V_i \) contributes to the collapse of the factor in the factor.

3. Results and analysis

3.1 Evaluation of indicators

The data of collapse points in Guangdong Province and the data of each factor were analyzed by using the Arcgis and other softwares. Each number of collapsing points of different factors was calculated, and the density of collapse points of every factor was also calculated.

According to the analysis result, it shows regularly that the distribution of every collapse point is related with the temperature change, but the change annual average temperature is not obviously on the regional scale of Guangdong Province. And it is not suitable to regarded it as a evaluation index. In addition, some factors that are only have tiny influence in collapse density are also directly removed, such as the residents and the construction sites in the use of land. Therefore, the collapse gully erosion of Guangdong Province is mainly distributed in the following factors: the annual rainfall of 1500mm to 1800mm, the lithology type of sedimentary rocks, granite, and metamorphic rocks, and the soil is Latosol, Latosolic red soil, and yellow soil. sloping in 0° to 30°, elevation at 0m to 400m, fluctuation degree is 0m to 150m, and the vegetation coverage is less than 20%, or at 80% to 90%. The land type
3.2 assessment of collapse risk

3.2.1 information acquisition analysis method. The density of the collapse of the corresponding risk assessment index in 3.1 is calculated by the formula (1), the amount of information available in each factor, as shown in Table 1:

| Factor classes | Factors | Density (%) | Factor classes | Factors | Density (%) |
|----------------|---------|-------------|----------------|---------|-------------|
| Rainfall (mm)  | 1500–1600 | 4.05        | 0-100          | 3.58    |
|                | 1600–1700 | 5.14        | 100-200        | 4.42    |
|                | 1700–1800 | 5.16        | 200-300        | 4.65    |
|                | 1800–1900 | 3.45        | 300-400        | 4.52    |
|                | Granite   | 2.28        | 400-500        | 3.77    |
| Geology        | Igneous rock | 5.44        | 500-600        | 1.95    |
|                | Metamorphic rock | 3.28        | 0-50          | 4.84    |
|                | Sedimentary rock | 5.37        | 50-100        | 5.82    |
|                | Latosol    | 1.57        | 100-150        | 3.75    |
| Soil type      | Latosolic red soil | 50.85      | 150-200        | 0.22    |
|                | Red soil   | 45.02       | <20%          | 5.48    |
|                | Yellow soil | 2.56        | 20%-40%       | 3.93    |
|                | 0-10       | 4.72        | 40%-60%       | 2.63    |
|                | 10-20      | 5.37        | 60%-80%       | 0.08    |
|                | 20-30      | 4.65        | 80%-90%       | 4.91    |
|                | 30-40      | 2.71        | >90%          | 1.49    |
| Slope (°)      | 5.96       | 0           | 0             | 0       |
| Land use       | Farmland and bare land | 5.13        |                |         |

According to the amount of information on the factors in the above table, we can establish the amount of information prediction equation:

\[ P_f = 4.05X_1 + 5.14X_2 + \cdots + 5.13X_{34} \]  \hspace{1cm} (4)

In Arcgis, according to equation (4), all the factor layers are superimposed by the grid calculator. The risk distribution diagram of collapse in Guangdong Province was obtained after normalization.

3.2.2 logistic regression analysis method. In Arcgis, the factor data of each grid are obtained by spatial superposition analysis of each factor class data of the screening and 10,000 sampling grids (5,000 pieces of collapse points and no collapse points) obtained by random sampling. Find all the factors in which the factor range of data, each factor class range has a corresponding collapse density. The data of all the factors in 10000 sampling grid are replaced with the corresponding collapse density. The binary logistic regression in SPSS is used to analyze the data. The variable is the presence or absence of the collapse point in the sampling grid. Collapse point for 1, no collapse point for 0. The results of the analysis are shown in Table 2:

| Factor classes | Regression Coefficient | Standard Error | Wald  | df  | Significant Level | Exp   |
|----------------|-----------------------|----------------|-------|-----|-------------------|-------|
| Rainfall (X_1) | 0.033                 | 0.002          | 286.089 | 1.00 | 0.000             | 1.034 |
| Geology (X_2)  | 0.022                 | 0.002          | 206.897 | 1.00 | 0.000             | 1.022 |
Land use ($X_1$)  0.010  0.001  60.567  1.000  0.000  1.011  
Vegetation coverage ($X_2$)  0.013  0.002  71.371  1.000  0.000  1.013  
Hypsography ($X_3$)  0.010  0.001  50.482  1.000  0.000  1.010  
DEM ($X_6$)  0.092  0.003  763.335  1.000  0.000  1.097  
Slope ($X_7$)  -0.003  0.002  1.648  1.000  0.199  0.997  
Soil type ($X_8$)  0.013  0.001  101.388  1.000  0.000  1.013  
Constant term  -4.793  0.134  1276.810  1.000  0.000  0.008  

From the regression coefficients $\beta$ in Table 2, the regression model is available:

$$P_f = \frac{e^{-4.793+0.033X_1+0.022X_2+0.011X_3+0.013X_4+0.01X_5+0.092X_6-0.003X_7+0.013X_8}}{1+e^{-4.793+0.033X_1+0.022X_2+0.011X_3+0.013X_4+0.01X_5+0.092X_6-0.003X_7+0.013X_8}}$$  (5)

In Arcgis, according to Eq. (5), all the factor layers are superimposed by the grid calculator. The risk distribution diagram of collapse in Guangdong Province was obtained after normalization.

3.3 Comparison of risk characteristics based on two evaluation methods

The distribution of the occurrence risk of collapse gully erosion in Guangdong Province is as follows:

![Comparison chart of Guangdong province collapse gully erosion occurrence risk based on two methods](image)

The overall trend of risk distribution of collapse gully erosion of information acquisition analysis method is basically the same as logistic regression analysis method. The distribution of risk regions is more decentralized in information acquisition analysis method, and logistic regression analysis method is more close. The former high risk area is widely distributed, the middling risk area is slightly smaller, the latter vice versa.

In summary, The deviation of the two risk assessment methods in the spatial distribution of occurrence risk may be caused by the following two points: First, random sampling in a large number of sampling grid, different sampling results will lead to the sampling grid factor data changes, the regression results have a certain impact. Second, the selected sampling grid is 1000m × 1000m, the satellite image resolution is 30m × 30m, a single sampling grid may not collapse or exist multiple collapse points, but whether it exists or not, logistic regression analysis when the variable is uniformly calculated by 0 or 1, which may have an impact on the regression results.

3.4 Precision comparison

The prediction accuracy is evaluated by comparing the number of sampling grids and the number of sampling grids with collapse risk greater than 0.8. The accuracy of the evaluation results is expressed in empirical probabilities:
\[ E = \frac{N_{0.8}}{N_{pG}} \]  

(6)

Where \( N_{pG} \) is the total number of sampling grids with collapse gully erosion; \( N_{0.8} \) is the number of sampling grids in the region with collapse risk of greater than 0.8.

The accuracy of the risk assessment of information acquisition analysis method 83.64%, and the logistic regression analysis method is 82.95%.

Compared with the accuracy of two methods, the information acquisition analysis method is slightly higher than the logistic regression analysis method. The reason of the difference is that the information acquisition analysis method considers the different effects of each factor on the collapse gully erosion and retains the the impact value of different factors to collapse gully erosion. Such as vegetation coverage in 80% to 90% of the area accounted for more than 40% of the study area. Pinus massoniana is one of the main tree types in Guangdong Province. Excessive tree cover will consume limited soil moisture and nutrients, and the volatile substances of pine needles inhibit the survival of other plants. And the raindrops will re-gather in the canopy, the size of the raindrip may be greater than the natural state, indirectly increasing the gravity potential of the raindrops, so that the bare land is splashed. Therefore, excessive coverage area of trees has a higher collapse gully erosion. Logistic regression analysis method directly multiplies the data of factor classes with the sampling grid, leave the value of influence different factors of same factor class out of consideration. the accuracy is slightly lower than the information acquisition analysis method.

4. Conclusions
The occurrence risk of collapse gully erosion in Guangdong Province is mainly distributed from the east to the north, but the occurrence risk is also high in the west. The risk is lower in the Pearl River Delta and the southwest comparing with others. Neither there are no significant differences in the risk distribution between two methods, nor the accuracy of calculating the risk of collapse gully erosion. Both method’s accuracy is above 80%. And the two methods are feasible and reliable for the risk assessment of collapse gully erosion in Guangdong Province. Comparing with the actual collapse area in study area, the evaluation accuracy is slightly higher.

In contrast, the accuracy of information acquisition analysis method mainly depends on the number of factor classes: the more number of factor classes, the more quantity of information, the more credible results would receive. Logistic regression analysis method requires that the study area must be divided into several sample cells. The higher accuracy of calculation will be gained if the sample unit is divided small enough. Therefore, the method is mainly suitable for making assessment in a small scale. In this paper, a wide range of Guangdong province is selected as the study area, and more influencing factors category were adopted, as the result, the former one has a relatively better accuracy than the latter. From the risk distribution diagram of collapse gully erosion, we can see that the risk distribution of information acquisition analysis method has better a uniformity in the space, and less small pattern. So, we can be sure the latter one has a more value in use than the former one.

References
[1] Luk S H, Yao Q Y, Gao J Q, et al 1997 J. Environmental analysis of soil erosion in Guangdong Province: A Deqing case study. Catena. 29(2):97-113.
[2] Angima S D, Stott D E, O’Neill M K, et al 2003 J. Soil erosion prediction using RUSLE for central Kenyan highland conditions. Agr Ecosyst Environ. 97(1/3):295-308.
[3] Zhong B, Peng S, Zhang Q, et al 2013 J. Using an ecological economics approach to support the restoration of collapsing gullies in southern China. Land Use Policy. 32:119-124.
[4] Luk S H, Dicenzo P D, Liu X Z 1997 J. Water and sediment yield from a small catchment in the hilly granitic region, South China. Catena. 29(2):177-189.
[5] Munro D S, Huang L J. Rainfall 1997 J. evaporation and runoff responses to hillslope aspect in the Shenchong Basin. Catena. 29(2):131-144.
[6] Vrieling A, Sterk Gbeaulieu N 2002 J. Erosion risk mapping:A methodological case study in the
Colombian Eastern Plains. *J Soil Water Conserv.* 57(3):158-163.

[7] Woo M K, Huang L, Zhang S, et al 1997 *J.* Rainfall in Guangdong province, South China. *Catena.* 29(2):115-129.

[8] Woo M, Guoxiang F, Dicenzo P D 1997 *J.* The role of vegetation in the retardation of rill erosion. *Catena.* 29(2):145-159.

[9] Cohen M J, Shepherd K D, Walsh M G 2005 *J.* Empirical reformulation of the universal soil loss equation for erosion risk assessment in a tropical watershed. *Geoderma.* 124(3/4):235-252.