Information Entropy Embedded Back Propagation Neural Network Approach for Debris Flows Hazard Assessment

Yan Yu¹ and Jun Wang²,*

¹Science and Technology Library of Guangdong Province/Guangdong Institute of Scientific & Technical Information and Development Strategy, Guangzhou 510070
²Guangdong Open Laboratory of Geospatial Information Technology and Application, Guangzhou Institute of Geography, Guangzhou 510070

*Corresponding author

Abstract. The purpose of the study is to propose a new method to assess debris flow hazard by embedding Shannon’s information entropy in back propagation neural network. The procedure was divided into two parts. One was to establish a mathematical model for debris-flow hazard assessment based on Shannon’s information entropy forming input and output data sets of back propagation neural network. The other was to establish a back propagation neural network technique for characterizing debris flow dynamic hazard. The proposed method was employed to assess debris flow hazard of Shenxi gully basin, Sichuan province, China. The result shows that the assessed result of proposed method is highly consistent with the result of field surveys. The proposed method can reflect the interaction, nonlinearity and dynamic process of background variable factors and can be used to debris flow hazard assessment, risk management and mitigation of debris flows.

1. Introduction
Debris flows are a common geological disaster in mountainous regions. Because of its sudden causes, it usually causes severe damage to people's lives and property. Therefore, debris flow hazard assessment is very significant for disaster reduction and mitigation.
In recent years, a lot of researches have been reported to the debris flow hazard assessment. A variety of related models have been established such as logistic regression model, information entropy model, geomorphic information entropy model and artificial neural network model. For example, a probabilistic prediction method was represented by a logistic regression model based on multiple rainfall and environmental factors [1][2]. The minimum entropy analysis method was used to characterize the selection of variables in the debris flow hazard evaluation, and provided quantitative criteria for selecting relevant variables [3]. Combining the specific geomorphic conditions, the geomorphic information entropy model was applied to carry out debris flow hazard zonings by using the mathematical software MATLAB [4]. Based on debris flow hazards and geomorphologic environment data, artificial neural network method was used to model the hazard of debris flow [5]. Among them, information entropy theory was the most commonly used tool for establishing natural disaster assessment models. In the family of information entropy methods, Shannon's information entropy plays a leading role in practice and has been successfully applied in natural disaster assessment. Meanwhile, Shannon entropy is an effective tool to describe the factors [6][7]. Because the debris flow is dynamic rather than static, Herein, back propagation neural network is introduced as a tool to characterize the dynamic debris flow hazard in this paper.
This study is to propose a new method to assess debris flow hazard by embedding Shannon’s information entropy in back propagation neural network. The general procedure is roughly divided into two parts. The first part introduces the numerical calculation model of the formulation and solution. The second part describes the method numerical simulation model embedded in a three-layer feedforward neural network model. Finally, the method was verified by a case study in Shenxi gully, Sichuan, China.

2. Methods

2.1. Debris-flow Hazard Assessment Mathematical Model

Debris flows is related with the numerous environment variables. To quantitative analysis of these factors, Shannon's entropy is a very useful tool for analysing their weights of these debris flow factors. In this paper, a numerical calculation model based on Shannon’s information entropy theory is described as follows.

2.1.1. Appraisal Matrix. A comprehensive appraisal matrix of debris-flow is expressed as follows:

\[
A = (a_{ij})_{mn} = \begin{pmatrix}
  a_{11} & a_{12} & \cdots & a_{1m} \\
  a_{21} & a_{22} & \cdots & a_{2m} \\
  \vdots & \vdots & \ddots & \vdots \\
  a_{n1} & a_{n2} & \cdots & a_{nm}
\end{pmatrix}
\]

where, \(a_{ij}\) is the \(j\)th \((j=1,\cdots,m)\) factor of the \(i\)th \((i=1,\cdots,n)\) study object. Above \(m\) factors are consisted of environment variables such as watershed area, average slope of the formation region, watershed cutting density, bed bending coefficient, loose solid material reserves, rainfall parameter, vegetation coverage, reclamation index and population density in the basin, etc. Then, the matrix \(A\) is quantified and normalized to the matrix \(B\).

\[
B = (b_{ij})_{mn} = \begin{pmatrix}
  b_{11} & b_{12} & \cdots & b_{1m} \\
  b_{21} & b_{22} & \cdots & b_{2m} \\
  \vdots & \vdots & \ddots & \vdots \\
  b_{n1} & b_{n2} & \cdots & b_{nm}
\end{pmatrix}
\]

2.1.2. Weights Calculation. The weight of the \(j\)th debris-flow factor of information entropy model is set up (Equation (3)). The formula (3) \(H_j\) indicates that the smaller information entropy value of debris-flow factor, the less contribution has.

\[
H_j = -\frac{\sum_{i=1}^{n} p(b_{ij}) \log p(b_{ij})}{\log n}
\]

where, \(p(b_{ij})\) is the \(j\)th factor’s frequency of the \(i\)th study object and satisfies \(0 \leq p(b_{ij}) \leq 1 (i=1,\cdots,n; j=1,\cdots,m)\) and \(0 \leq b_{ij} \leq 1 (i=1,\cdots,n; j=1,\cdots,m)\).

According to the definition of the debris-flow factor of information entropy (Equation (3)), the weight of the \(j\)th factor of debris-flow can be expressed as follows:
\[ w_j = \frac{1-H_j}{\sum_{j=1}^{n}(1-H_j)} \]  

where \( w_j \) is the weight of the \( j \)th factor.

At last, the quantitative hazard assessment model of debris flow is:

\[ D_i = \sum_{j=1}^{n} w_j b_{ij} \]  

where, \( D_i \) is the \( i \)th study object debris-flow hazard value and \( i = 1, \ldots, n \).

2.2. Debris-flow Hazard Assessment Mathematical Model Combined with Back Propagation Neural Network

In 1962, the famous artificial neural network theorem was given by Rosenblatt, who believed that an artificial neural network can learn everything it can express \(^8\). The artificial neural network learning process was a training procedure. In 1986, the parallel distributed processing group of Rumelhart and other researchers proposed independently a multi-layer network learning algorithm of back propagation and solved the problem of learning a multi-layer network. Therefore, the back propagation neural network (BPNN) was born.

The BPNN was a one-way transmission of the multi-layer feed forward neural networks. The BPNN used the error back propagation learning algorithm and it relied on adjusting the weights of each layer to make the network learn the laws expressed by training samples. The BPNN was composed of an input layer, an output layer and one or more hidden layers. Each layer was composed of several neurons. Each neuron had an input and an output. The central idea of back propagation neural network was that the weights were adjusted according to the networks and that it minimized the total. The learning process was divided into forward propagation and back propagation \(^9\).

Because of back propagation neural network has advantage in data prediction application. Simultaneously, neural networks have self-learning capabilities, they are constantly training new networks with new data, and they can continually modify prediction models based on environmental changes. In addition, when the network is trained, the calculation of the network is quite easy and fast, which is of great benefit for solving the predictions of real-time scheduling problem. Due to debris flow hazard results change with input data, therefore, this paper combines the proposed mathematical model with the neural network algorithm to train the data in a certain period of time and space, and then apply the trained network model to predict the results of data in other time or space. The specific steps are as follows:

2.2.1. Data Transformation and Processing. The data of debris-flow factors are from field investigations and literature. Meanwhile, the comprehensive appraisal matrix \( A \) is quantified and normalized to the matrix \( B \).

2.2.2. Input and Output Data-sets. By the Equation (2) and the Equation (6), we can obtain the input data-sets \( B \) and the output data-sets \( D \), respectively.

2.2.3. The Back Propagation Neural Network Learning Algorithm. The model is developed on the software MATLAB as follows:

- First, input and output data of the model is is determined input data-sets and output data-sets, respectively.
- Second, a network is created by program language fit net ().
- Third, the model is divided by the proportions of training, testing and verification data: net.divideParam.trainRatio; net.divideParam.valRatio; net.divideParam.testRatio.
Fourth, the network is trained by program language train().
Fifth, the network is saved by program language save('net').
Sixth, the saved network is used by program language load('net').
Finally, debris-flow hazard values in different periods are predicted by program language:
y_predict= sim (net, x_predict).
where, 'net' is the trained network, x_predict is the network input and y_predict is the network output by the 'net'.

2.2.4. Hazard Assessment of Debris Flow. The output value of this model is used to evaluate the degree of debris flow hazard according to the classification criteria of debris flow hazard (Table 1), where D is the normalized values.

| D        | Hazard degree         |
|----------|-----------------------|
| 0.8 ≤ D ≤ 1 | Extremely high       |
| 0.6 ≤ D < 0.8 | High               |
| 0.4 ≤ D < 0.6 | Moderate           |
| 0.2 ≤ D < 0.4 | Low                 |
| 0 ≤ D < 0.2  | Extremely low        |

The flow chart of the research method is shown in Figure 1.

3. Case Study

3.1. Study Site Description
The Shenxi gully, Dujiangyan City, Sichuan Province, was chosen as the study site. The study area covers a total area of 8.17 square kilometres and has a resident population of 600. The area is from 31°03′49″ N to 31°05′55″ N and from 103°36′27″ E to 103°37′38″ E. The elevation of Shenxi gully ranges from 870 m to 2141 m, the gradient of the main river is 137‰, and the length of the main river is about 4400 m. In terms of hydrometeorology, the region belongs to the humid subtropical climate, with an annual average temperature of 15.2 °C and an average annual rainfall of 1200 mm. Rainfall in
the region is seasonal, with about 70 percent of the total rainfall occurring between June and September during the monsoon season and the remaining 30 percent occurring in other seasons. From 2008 to 2014, there have been many debris flows in Shenxi gully watershed (Table2) [4].

![Shenxi gully location image]

**Figure 2.** The location of the Shenxi gully.

**Table 2.** Debris flow events recorded in the study area.

| Time       | Debris flow gullies                  |
|------------|-------------------------------------|
| Before 2008, 5, 12 | No record                  |
| 2008, 9, 24    | Guo Juanyan                    |
| 2009, 7, 17    | Guo Juanyan                    |
| 2010, 8, 13    | Guo Juanyan                    |
| 2010, 8, 17    | Guo Juanyan, Wu Xianmiao       |
| 2011, 7, 01    | Guo Juanyan                    |
| 2012, 8, 17    | Guo Juanyan, Wu Xianmiao       |
| 2013, 7, 09    | Guo Juanyan                    |
| 2013, 7, 26    | Guo Juanyan                    |
| 2014, 7, 18    | Guo Juanyan                    |

### 3.2. Application and Validation of the Model

According to the DEM (10m), the study area can be classified into 41 sub-watersheds by the ArcGIS 9.3 software (Figure 3). Field surveys in 2014 and literature reviews are the basic data sources for this regional debris flow hazard assessment. Through analysis, 5 variables were chosen as the main controlling indicators including: X1 - average gradient; X2 - basin area; X3 - main channel length; X4 - solid loose material reserves; X5 - the maximum 1 h rainfall. The source of data and the numerical value of five variables were obtained (Figure 4).

Based on the proposed approach, the five variables of 41 sub-basins were selected as the training input, and the numerical calculation model of 41 sub-basins were selected as the output. Therefore, the number of input layer nodes is 5, and the number of output layer nodes is 1. Then, the back propagation neural network used in the model is divided into three layers. Through analysis, when the number is 11, the calculation cycle and consumption time of the network are the least, so the model structure is set to 5-11-1. The model was built using the Levenberg-Marquardt algorithm on
MATLAB software. The error target is set to 0.001, and the maximum period of network training is set to 1000.

![Figure 3. The division of sub-watershed in Shenxi gully basin.](image)

Figure 3. The division of sub-watershed in Shenxi gully basin.

![Figure 4. The source of data of sub-watershed in Shenxi gully basin.](image)

Figure 4. The source of data of sub-watershed in Shenxi gully basin.

Through network training, a neural network model was obtained. To verify the feasibility of the model, considering other factors (average gradient, basin area and main channel length) have little change, the maximum 1 h rainfall data and solid loose material reserves have the characteristics of...
spatial-temporal dynamic change. Therefore, the dynamically varying data of the maximum 1 h rainfall data and solid loose material reserves from 2011 to 2014 of Guo Juanyan were input into the model, and the hazard of debris flow was predicted through the model and compared with the occurrence records of debris flow (Table 3 and Table 4), from which we can find that the result of the proposed model has a good agreement the field survey result of Guo Juanyan gully. This shows that the proposed model method is feasible. From Table 4 we can also find that the hazard degree of debris flow will change with the time.

**Table 3.** The statistics of maximum 1h rainfall of debris flow events in Guo Juanyan\(^{[10]}\).

| Events      | The maximum 1 h rainfall/mm | Triggered debris flow |
|-------------|-----------------------------|-----------------------|
| 2011, 7, 01 | 41.5                        | Yes                   |
| 2012, 8, 17 | 42.4                        | Yes                   |
| 2013, 7, 09 | 32                          | Yes                   |
| 2013, 7, 26 | 18.9                        | Yes                   |
| 2014, 7, 18 | 32.5                        | Yes                   |
| 2011, 8, 20 | 26.8                        | No                    |
| 2011, 9, 05 | 16.2                        | No                    |
| 2012, 6, 16 | 27.0                        | No                    |
| 2012, 8, 03 | 26.7                        | No                    |
| 2013, 8, 06 | 17.1                        | No                    |

**Table 4.** The results of proposed model in Guojuyan gully.

| Time          | Hazard degree | Triggered debris flow |
|---------------|---------------|-----------------------|
| 2011, 7, 01   | Extremely high| Yes                   |
| 2012, 8, 17   | Extremely high| Yes                   |
| 2013, 7, 09   | Extremely high| Yes                   |
| 2013, 7, 26   | Extremely high| Yes                   |
| 2014, 7, 18   | Extremely high| Yes                   |
| 2011, 8, 20   | High          | No                    |
| 2011, 9, 05   | Moderate      | No                    |
| 2012, 6, 16   | High          | No                    |
| 2012, 8, 03   | High          | No                    |
| 2013, 8, 06   | Moderate      | No                    |

### 3.3. Results and Verification

According to the classification criteria of debris flow hazard (Table1), the results of the proposed model in 2014 were shown in Table 5 and Figure 5. From Table 2 and Table 4, it is found that the predicted result of proposed model is agree will with the actual situation. For example, watersheds of 0 (Wuxianmiao gully) and 2 (Guanjuanyan gully) have extremely high hazard degree when using the proposed model. Actually, these two gullies occurred debris flow events after the earthquake (Table 2) and thus should be classified into extremely high hazard degree. Therefore, the model proposed in this paper is reasonable for debris flow hazard assessment in Shenxi gully. Although Huang Nigang gully (No. 4) and Chen Jiaping gully (No. 7) have not experienced debris flow so far, the gullies has a wealth of loose solid materials. These potential debris flow gullies need high vigilance and the hazard degree should be extremely high.
Above analysis indicate that No. 0, No. 2, No. 4 and No. 7 sub-watersheds ought to be extremely high hazard degree in the study area which require high vigilance. The assessment results of other debris flow gullies are shown in Table 5.

![Debris flow hazard zoning map in the Shenxi gully.](image)

**Figure 5.** Debris flow hazard zoning map in the Shenxi gully.

**Table 5.** The results of proposed model in Shenxi gully.

| NO. | Hazard degree | NO. | Hazard degree |
|-----|---------------|-----|---------------|
| 0   | Extremely high| 21  | Extremely low |
| 1   | High          | 22  | Low           |
| 2   | Extremely high| 23  | Low           |
| 3   | Low           | 24  | Low           |
| 4   | Extremely high| 25  | Moderate      |
| 5   | Moderate      | 26  | Moderate      |
| 6   | Extremely low | 27  | Low           |
| 7   | Extremely high| 28  | Low           |
| 8   | Extremely low | 29  | Low           |
| 9   | Low           | 30  | Extremely low |
| 10  | Low           | 31  | Extremely low |
| 11  | Extremely low | 32  | Low           |
| 12  | High          | 33  | Low           |
| 13  | Moderate      | 34  | Extremely low |
| 14  | Low           | 35  | Low           |
| 15  | Low           | 36  | Extremely low |
| 16  | Extremely low | 37  | Extremely low |
| 17  | Moderate      | 38  | Low           |
| 18  | Extremely low | 39  | Low           |
| 19  | Low           | 40  | Low           |
| 20  | Low           | -   | -             |

**4. Conclusions**

This study incorporated the Shannon’s information entropy theory with BPNN approach to propose a new model for hazard evaluation of debris flow. Through the case study of Shenxi gully, Sichuan Province, China, it is found that this proposed model is suitable for hazard assessment of debris flow in this area.
In the specific application of the model, to calculate the area debris flow information entropy reasonable and effective, the selected variable factors of three conditions (the topographic, solid loose mass reserves and hydrodynamic) should be as much as possible and rigorous, adaptive, and the self-learning back propagation neural network approach should be adopted to reflect the interaction, nonlinearity and dynamic process of these variable factors. This model can comprehensively reflect the interaction mechanism of variable factors of three conditions (the topographic, solid loose mass reserves and hydrodynamic) in the debris flow formation process and has a broad application prospect in the regional debris flow hazard assessment.

Acknowledgments
This research was supported by Natural Science Foundation of Guangdong Province (2018A030310469), the Science and Technology Project of Guangzhou City (201804010126; 201803030025), GDAS’ Special Project of Science and Technology Development (2017GDASCX-0807; 2018GDASCX-0101; 2018GDASCX-0403; 2019GDASYL-010308) and the Water Resource Science and Technology Innovation Program of Guangdong Province (2016-15).

References
[1] Xu W B, Yu W J, Zhang G P 2012 Prediction method of debris flow by logistic model with two types of rainfall: a case study in Sichuan, China (Natural Hazards vol 62 no 2) pp733-744
[2] Xu WB, Jing SC, Yu WJ, et al 2013 A comparison between Bayes discriminant analysis and logistic regression for prediction of debris flow in southwest Sichuan, China (Geomorphology vol 201) pp45-51
[3] Chen CC, Tseng CY, Dong JJ2007 New entropy-based method for variables selection and its application to the debris-flow hazard assessment (Engineering Geology vol94 no1-2) pp19-26
[4] Wang J, Yu Y, Yang S, et al. 2014A modified certainty coefficient method (M-CF) for debris flow susceptibility assessment: A case study for the Wenchuan earthquake meizoseismal areas (Journal of Mountain Science vol 11(5)) pp1286-1297
[5] Liang WJ, Zhuang DF, Jiang D, et al.2012 Assessment of debris flow hazards using a Bayesian network (Geomorphology vol 171) pp94–100
[6] Sharma LP, Patel N, Ghose MK, et al.2012 Influence of Shannon’s entropy on landslide-causing parameters for vulnerability study and zonation—a case study in Sikkim, India (Arabian Journal of Geosciences vol5 no3)pp 421-431
[7] Rakoczy R, Kordas M, Story G, et al. 2014 The characterization of the residence time distribution in a magnetic mixer by means of the information entropy (Chemical Engineering Science vol105) pp191-197
[8] Rumelhart DE, and McClelland JL, PDP Research Group1986 Parallel distributed processing: explorations in the microstructure of cognition (Cambridge: MIT Press) pp1-110.
[9] Samarasinghe S 2006 Neural networks for applied sciences and engineering: from fundamentals to complex pattern recognition (New York: The Chemical Rubber Company Press) pp 69-190
[10] Pan HL, Jiang YJ, Wang J, et al. 2018 Rainfall threshold calculation for debris flow early warning in areas with scarcity of data (Natural Hazards and Earth System Sciences vol 18 no 5)pp1395-1409