Estimation and Prediction of Debris Flow Potential Using Discrimination Analysis

Estimation et prédiction du potentiel d'écoulement de boue utilisant une analyse discriminante

Lin M.L., Lin Y.S.

Department of Civil Engineering, National Taiwan University, Taiwan

ABSTRACT: Taiwan is situated at the juncture of tectonic plates, which caused complex and fragile geological condition with steep mountain terrain. Being frequently struck by typhoons and earthquakes, the landslide and debris flow hazard occurs frequently. In this research, the estimation model of regional debris flow potential was constructed based on the geo-morphological and hydrological conditions of the research area. For constructing the estimation model of debris flow potential, the Fisher's discrimination analysis was used. A study area of Nantou County in Central Taiwan was selected. Influence factors were identified and a database for both debris flow torrents and non-debris flow torrents were constructed. Estimation model was constructed using the Fisher's analysis by random sampling of the debris flow and non-debris flow torrents. The estimation model is validated and then used for prediction of debris flow potential. The final model can be determined by evaluating the estimation stability and prediction rate with each additional influence factor. The resulting potential estimation of the study area appears to be satisfactory. The influence factor stability of the Fisher's discriminant model and the prediction rate associated with the differences in influence factors were discussed.

RÉSUMÉ : Taiwan est situé à la jonction de plaques tectoniques, ce qui engendre un terrain escarpé avec des situations complexes et des conditions géologiques fragiles. L'île étant régulièrement frappée par des typhons et des tremblements de terre, des glissements de terrain et des écoulements de boue se produisent fréquemment. Pour cette étude, le modèle d'estimation des coulées de boue régionale potentiel a été construit sur la base des conditions géomorphologiques et hydrologiques de la zone de recherche. Pour la construction de celui-ci, l'analyse discriminante de Fisher a été utilisée. Une zone d'étude du comté de Nantou qui se situe au centre de Taïwan a été choisie. Les facteurs d'influence ont été identifiés et des bases de données pour les torrents d'écoulement de boue et torrents d'écoulement de non-boue ont été construites. Le modèle d'estimation a été construit en utilisant une analyse de Fisher par échantillonnage aléatoire des torrents d'écoulement de boue et des torrents d'écoulement de non-boue. Le modèle d'estimation est validé, puis utilisé pour la prédiction de potentiels écoulements de boue. Le modèle final peut être déterminé en évaluant l'estimation de la stabilité et la fréquence prédite avec chaque facteur d'influence additionnel. L'estimation résultante potentielle de la zone d'étude semble être satisfaisante. Le facteur d'influence de stabilité du modèle discriminant de Fisher et la fréquence prédite associée aux différences des facteurs d'influence ont été discutés.

KEYWORDS: debris flow, potential estimation, Fisher's discriminaton analysis, influence factor, prediction model.

1 INTRODUCTION

More than 70 percent of areas in Taiwan are in mountain region and with steep and fragile slopes. The earthquakes and heavy rainfall introduced by typhoons often induced significant landslide and debris flow hazards in Taiwan, which lead to significant loss of properties and lives. For effective mitigation of the debris flow hazards, it is important to evaluate the potential of debris flow torrents, which supports decision on mitigation measures and priority.

This research is based on the data of the 1,420 debris flow torrents published by Soil and Water Conservation Bureau in 2003. The fundamental data of the debris flow torrents for the study area of Nantou County in Central Taiwan were collected, and the basic database along with the related influence factors were established utilizing the geographic information system software, Arcview. The influence factors database included watershed area, stream length, hypsometric integral, stream mean slope, form factor, slope distribution, slope aspect and geology category, were extracted from the fundamental data. The statistic analysis was performed on all influence factors to discuss their significances. The multi-variant discrimination analysis was used to discriminate debris flows and non-debris flows. The analysis model was verified, and accordingly the potential of debris flow torrents in Nantou County was evaluated.

2 STUDY AREA AND FUNDAMENTAL DATABASE

The Nantou County located in central Taiwan is selected as the study area in this research, which has a large area with high mountains and rugged topography. The debris flow hazard in Nantou County came into great concern since 1996, when Typhoon Herb caused severe losses of properties and lives. In addition, severe debris flow hazard struck this area frequently after the Chi-Chi earthquake, 1999. According to the data published by Soil and Water Conservation Bureau (2003), the number of debris flow torrents reaches 199 in Nantou County, which are for the potential analysis. The distribution of the debris flow torrents and the study area are as shown in Figure 1. The fundamental database used for construction of related database for analysis include: the digital elevation model published in 1989 with a resolution of 40m x 40m, 1/500,000 Taiwan geology map produced by the Central Geological Survey in 1986, aerial photographs, topographical map.

The primary geologic formations of the study area include slate, phyllite, sandstone, and shale. According to the engineering geology zonation proposed by Hung in 1997, the research area contents C zone (metamorphic rocks, metamorphic sandstone, shale, slate, phyllite), D zone (sedimentary rocks), E zone (lateritic tableland), and G zone (basin and plain). As shown in

Figure 1. Study area and distribution of Debris Flow Torrents and Non-Debris Flow Torrents

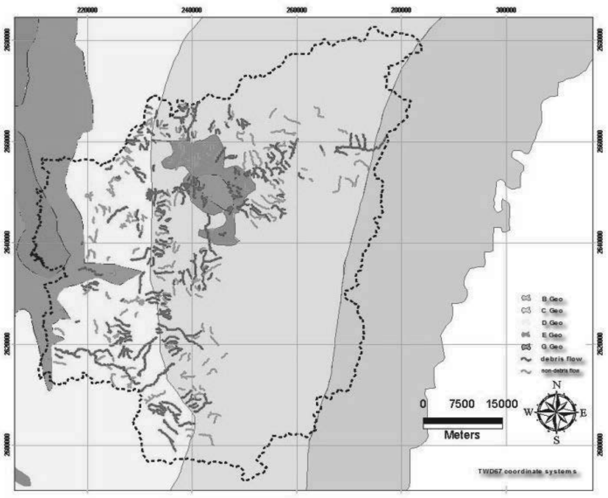

Figure 1, most of the debris flow torrents in Nantou County distribute in C zone and D zone.

In order to perform the discriminant analysis, database for both debris flow torrents and non-debris flow torrents are constructed as discussed in the followings:

a. Debris flow torrents: the debris flow torrents in research area are based on the debris flow torrents published by Soil and Water Conservation Bureau in 2003 in 1/5,000-scale. The total number of debris flow torrents in the research area is 199. The watershed of each debris flow torrent was generated using DEM accordingly.

b. Non-debris flow torrents: The procedures for establishing the non-debris flow torrents are to establish watersheds and streams based on the 1/25,000 topographic map and DEM, and are not debris flow torrents by SWCB (Lin, et al. 2003). The non-debris flow should be in the neighboring area of debris flow torrents and with similar geologic traits and watershed size. Accordingly, a total number of 175 non-debris flow torrents was generated in the research area. Cross examination of these torrents were performed by comparing the torrent distribution to the aerial photograph. The results appeared to be consistent, and the distribution of the non-debris flow torrents is as shown in Figure1.

3 INFLUENCE FACTORS AND THEIR SIGNIFICANCES

The necessary conditions for triggering debris flow include affluent debris materials, sufficient water supply, and appropriate geomorphological conditions. For evaluation of debris flow potential, the influence factors were selected based on the three conditions, and the database were generated. Based on previous researches (Lin, et al. 2003, Lin and Chen, 2005, and Lin and Wen, 2006), eight influence factors are selected as: watershed area, streams length, hypsometric integral, form factor of basin, stream mean slope, slope angle, slope aspect, and geology formation, respectively.

1. Watershed area: BA, in hectare. The watershed with larger area usually contributes more water, which is regarded as a factor related to the water supply.
2. Stream length: SL, in meter. Stream length means the largest length of the stream in the related watershed. The longer the stream length is, the larger the watershed area is. Stream length is also regarded as another factor of sufficient water supply.
3. Hypsometric Integral: HI, dimensionless. The hypsometric integral is determined from the integral of the basin's height versus area percentage curve, which is a characteristic of geomorphic evolution of the watershed. It also stands for terrain ruggedness and is regarded as a factor related to the abundance of debris material.
4. Form factor: Form factor, FF (dimensionless), is proposed by Horton in 1920, and is defined as the basin area divided by the square of stream length. A higher form factor suggests broader basin shape, and different basin shapes affect flow hydrograph of the stream (Robet and Raymond, 1978).
5. Stream mean slope: SMS, in degree. Stream slope is an important factor related to flow velocity; the steeper the slope is, the faster the stream flow. The stream mean slope serves as an indicator of appropriate slope conditions.
6. Slope aspect: N, NE, E, ES, S, SW, W, WN, in percentage of area ratio. Most debris flow disasters were triggered by typhoon or heavy rainfall. According to the landing path of typhoons, rain falls concentrated more on windward side. Strong wind also influences the weathering process. Slope aspect serves as an indicator of geomorphlogical condition.
7. Slope distribution: SD10(0°~10°), SD15(10°~15°), SD20(15°~20°), SD30(20°~30°), SD45(30°~45°), SD90(45°~90°), in percentage of area ratio. Steeper slope has a higher tendency of slope failure and leads to more debris material.
8. Geological formation: E1, E2, EO, Mj, Ml, MS, My, O1, O2, O3, P1, P2, Q2, Q3, Q0, as listed in Table 1, in area ratio percentage. Weak rock quality, complicated geologic structure, highly developed fault and fold, and intensive tectonic activity zone tend to provide abundant debris material. Different geological formation contributes to different material strength, degree of fracture, and soil type.

The influence factors were derived from fundamental data through performing spatial and hydrological analysis using GIS and the related database were established. The distributions and characteristics of the influence factors of debris flow torrents and non-debris flow torrents are compared and discussed.

Table 1. Geological formations of the study area

Category	Symbol	Rock Type
Metamorphic Rock (C Zone)	E1	Slate, interlaminations of slate and sandstone
	E2	Indurate sandstone with carbonaceous slate interbeds
	EO	Slate, phyllite, with sandstone interbeds
	Mj	Sandstone, shale
	Ml	Argillite slate phyllite sandstone interbeds
	MS	Sandstone, shale
	O1	Quartzitic sandstone, slate, graphitic shale
	O2	Argillite, indurate sandstone, slate
	Q2	Gravel, laterite, clay, sand
	Q3	Clay, sand, gravel
Sedimentary Rock (D Zone)	My	Sandstone, shale
	O3	Sandstone, shale
	P1	Shale, sandy shale, mudstone
	P2	Sandstone, mudstone, shale
	Q0	Sandstone, mudstone, shale, conglomerate (limestone)

During the analysis, it was found that the distribution of factors for different engineering geological region had different characteristics with significant effects on triggering of the debris flow; the potential analysis of the study area was conducted on whole area with combined geological zones, Zone C, and Zone D, respectively. The number of debris flow and non-debris flow torrents in Zone C are 99 and 95, and 54 and 54 in Zone D as shown in Figure 1. In order to conduct the statical analysis, the independcy and significance level of each influence factor was checked using analysis of covariances and Pearson Test. A significance level of 0.1 with 90% of confidence was chosen, and it appeared that the influence factors selected were independent of each other. The resulting significant influence factors for the whole area, Zone C, and Zone D are listed in order of level of significance in Table 2. Due to the complexity of the geological characteristics when the whole area was analyzed, the principal component analysis was performed on slope distribution and aspect distribution. In Table 2, the PS2 factor indicated a resulting principal component of slope for the whole area. As shown in Table 2, the significant factors are quite different for the whole area, Zone C, and Zone D except with the Hypsometric Integral, HI. This justified that the debris flow potential would be better understood if separate analysis were conducted for regions of different geological properties.

Table 2 The significant influence factors for different geological zones

Order of significance	Whole area	Zone C	Zone D
1	HI	HI	HI
2	PS2	E1	Mj
3	Q0	O1	FF
4	NE	WN	Q0
5	ES	Ml	ES
6	My	SD10	O3
7	Ml	-	SD10

4 ESTIMATION MODEL OF DEBRIS FLOW POTENTIAL

In this research the multi-variant variables discrimination analysis is used to establish the differential function for debris flow torrents and non-debris flow torrents. The discrimination analysis is to form a linear combination of variables for each associated group to provide estimation values, where the coefficient of each individual variable represents its contribution to the associated group. The differential function of discrimination analysis defines the line which differentiates two groups, and its coefficients help to discriminate properties of each group. This research uses the commercial statistic software, SPSS, with Fisher's discrimination analysis, and analyses are performed for the whole area, Zone C and Zone D. Random sampling of the debris flow and non-debris flow torrents were used assuming normal distribution of each factor. For each analysis, the contributing influence factor was added following the order of significance, and the improvement of the rate of accuracy was checked with each additional factor. The definition of accuracy rate is expressed as the sum of accurately estimated debris flow torrents and non-debris flow torrents divided by the total number of torrents.

1. Whole area with combined geological zones: The analysis was performed over the whole area using 87 sets randomly sampled out of 199 and 175 debris and non-debris flow torrents. It was found that the HI appeared to be the most significant factor; the additional factors were added following the significant sequence of PS2, Q0, NE, ES, My, and Ml, with accuracy rate of 78.9%, 81%, 82.2%, 82.2%, 83.3%, 84.5%, and 85.1%. The resulting discrimination function, y, is:

$$y=5.108(HI)+0.090(PS2)-0.020(Q0)-0.027(NE)-0.065(ES)-0.018(My)+0.003(Ml)-1.911 \quad (1)$$

The accuracy rate increases more or less steadily with the additional parameters, but the trend is not significant with addition of NE, and the amount of increase in accuracy was not steady, suggesting different contribution of the parameters compared to their significance level.

2. Zone C: The analysis was performed for the Zone C using 40 sets randomly sampled out of 80 sets debris and non-debris flow torrents. It was found that the HI appeared to be the most significant factor; the additional factors were added following the significant sequence of E1, O1, WN, Ml, SD10, with accuracy rate of 82.5%, 83.8%, 85.0%, 85.0%, 83.8%, and 83.8%. The resulting discrimination function, y, is:

$$y=19.050(HI)+0.018(E1)-0.016(O1)+0.009(WN)-0.025(M1)+0.082(SD10)-11.388 \quad (2)$$

The accuracy rate increases with the additional parameters up till O1, and then remains the same and decreases. Thus, the amount of increase in accuracy does not increase beyond parameter O1. Although the rest of the parameters appear to be significant, they do not contribute to the estimation model

3. Zone D: The analysis was performed for the Zone D using 40 sets randomly sampled out of 54 sets debris and non-debris flow torrents. It was found that the HI appeared to be the most significant factor; the additional factors were added following the significant sequence of Mj, FF, Q0, ES, O3, and SD10, with accuracy rate of 61.3%, 70%, 68.8%, 75%, 80%, 80%, and 80%. The resulting discrimination function, y, is:

$$y=-5.070(HI)+0.036(Mj)-1.516(FF)-0.015(Q0)-0.083(ES)+0.024(O3)+0.009(SD10)+4.731 \quad (3)$$

The accuracy rate increases with the additional parameters till ES and then remains the same. It suggests that the addition of O3 and SD10 parameters does not improve the accuracy rate, although both parameters are significant.

Observing the estimation models for the three regional analyses, the accuracy rate has a tendency to increase with the additional factors, and the HI factor appears to be the most effective factor in all three models. For all three models, the coefficient of each parameter indicates the contribution of the parameter, and is consistent with the variation in accuracy rate. However, the effectiveness of the influence factors is not fully in accord with the order of factor significance shown in Table 2. Therefore, the level of significance of the parameter could not be correlated to the contribution of the parameter to the estimation model.

5 VALIDATION AND PREDICTION

In order to verify the feasibilities of the potential estimation model discussed previously, the data sets of debris flow torrents and non-debris flow torrents not used in developing the estimation models were used for validation and prediction. A total of 112 debris flow torrents and 87 non-debris flow torrents were used for the prediction of the whole area using Eq.1. A total of 40 sets of debris flow and non-debris flow torrents were used for the Zone C, and a total of 14 sets of debris flow and non-debris flow torrents were used for prediction using Eqs. 2, and 3, respectively. The prediction accuracy rates were compared to the estimation accuracy rates for whole area with combined geological zones, Zone C, and Zone D, as shown in Figure 2, Figure 3, and Figure 4, respectively.

From Figure 2, it was found that the accuracy rate for prediction increased steadily up to ES but then decreased with additional factor for whole area with combined geological zones compared to the estimation model. Therefore, the factors used for the model are only up to ES, and the model is rectified as:

$$y=4.955(HI)+0.090(PS2)-0.0205(Qo)-0.027(NE)-0.065(ES)-1.741 \quad (4)$$

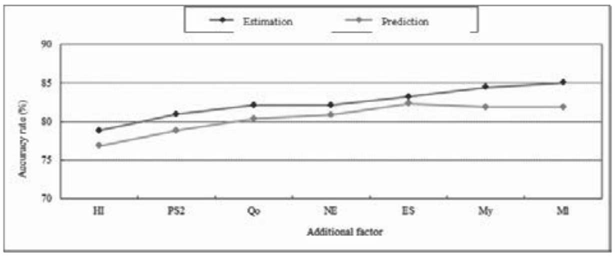

Figure 2. Accuracy rates of estimation and prediction for whole area

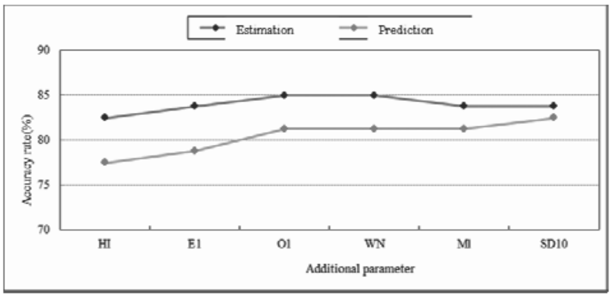

Figure 3. Accuracy rates of estimation and prediction for Zone C

The estimation accuracy rate is 83.3%, and the prediction accuracy rate is 82.4% with a difference of 0.9%.

From Figure 3, it was found that the accuracy rates for both estimation and prediction increased steadily with additional factor up to O1 but then varied for Zone C. It appears that the accuracy rate would become stable only with factors up to O1, and the model is rectified as:

$$y=15.531(HI)+0.019(E1) - 0.007(O1) - 9.255 \tag{5}$$

The estimation accuracy rate is 85.0%, and prediction accuracy rate is 81.3% with a difference of 3.7%.

From Figure 4, the accuracy rates varied for both estimation and prediction, but the accuracy rates increased steadily with additional factor up to ES for Zone D. It appears that the accuracy rate would become stable with the factors up to ES only, and the model is rectified as:

$$y= - 5.685(HI)+0.032(Mj) - 1.504(FF) -0.015(Qo) -0.085(ES) +5.380 \tag{6}$$

The estimation accuracy rate is 80.0%, and the prediction accuracy rate is 78.6% with a difference of 1.4%.

For all three analyses, the prediction curves appear to be quite stable and approach the estimation curves with the similar trends. The resulting accuracy rates are satisfactory and the differences between prediction accuracy rate and estimation accuracy rate of all three models are all within 4%. Thus the discrimination functions used is good for differentiating debris flow torrents from non-debris flow torrents with satisfactory results. Comparing the three modified function equations, not all factors contributed effectively to the model for all three models. The prediction accuracy rate actually reduced if all the factors were included, which suggested over-fitting of the estimation models. Observing the number of factors used for the whole area prediction model is higher than the Zone C and Zone D, it may doe to the complexity of multi-geological zones included. The accuracy rate of the Zone D analysis appeared to be lower than the other two analyses, which might due to less samples were available for estimation and prediction compared to number of samples in whole area and Zone C.

Based on the analysis results, the potential of debris flow torrent can be evaluated using distribution of the calculated discrimination function. The accumulated distribution density of 30% and 70% values are selected as the boundary indicators for low-moderate and moderate-high potential. For the whole area analysis, the discrimination function value for 30% is 0.506 and 3.092 for 70%. Thus the debris flow torrents with discrimination values greater than 3.092 are classified as high potential and moderate potential when discrimination values range within 0.506 to 3.092. The debris flow torrents are considered as low potential when the values are less than 0.506. Similar procedures were applied to Zones C and D. For Zone C the discrimination value for 30% accumulation is 0.765, and 3.143 for 70% accumulation. The discrimination value for 30% accumulation is 0.325 and for 70% is –1.720 for Zone D. Thus the potential of debris flow torrents can be evaluated in all three analysis models.

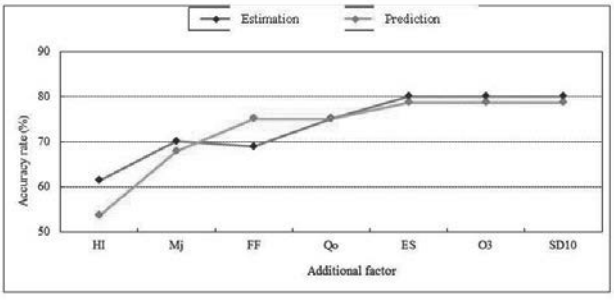

Figure 4. Accuracy rates of estimation and prediction for Zone D

6 CONCLUSIONS

In this research, the potential analysis of the debris flow torrents in Nantou County was performed using discrimination analysis. The influence factors included: watershed area, stream length, form factor, hypsometric integral, stream mean slope, slope distribution, slope aspect, and geological formation. Base on the previous discussions, most influence factors selected are within significance level. Among the factors used, the hypsometric integral appears to be the most important factor, which is a characteristic of geomorphic evolution of the watershed. The second most important factor is the geological properties of the analysis area. For all three models, the accuracy rates are about 80%, which suggests that the discrimination analysis provide satisfactory results. The accuracy rate for Zone D is slightly smaller than the accuracy rates for both the whole area and Zone C, which might due to smaller sample sizes in both estimation and prediction compared to the sample sizes of the whole area and Zone C. The final model could be determined based on the stability of accuracy rates in both estimation and prediction. However, not all the significant factors are contributing to the model effectively. The final potential of the debris flow torrents can be evaluated based on the values of discrimination function in all three models properly.

7 REFERENCES

Hung J.J. 1997. Engineering Geology Zonation of Taiwan

Lin M.L. and Chen T.C. 2005. Frame in renovation and feedback mechanism of database of potential debris flow torrents. SWBC-94-048 (Chinese)

Lin M.L. and Lien H.P. and Hsieh C.L. 2003. Follow-up investigation and observation in developed tendency of potential debris flow torrents. SWCB-92-107. (Chinese)

Lin M.L. and Wen H.Y. 2006. Field investigation and trend analysis of potential debris-flow rivers. *Sino-geotechnic*. 110, 45-54 (Chinese)

Robet L.S. and Raymond J.K. 1978. *Landslide analysis and control*. Special Report 176, National Academy of Science, Washington, D.C. 17-27

Taiwan Geology Map (1/500,000), 1986, Central Geological Survey