Evaluating Susceptibility of Debris Flow Hazard using Multivariate Statistical Analysis in Hualien County

Che-Wei Shen[1,2*], Wen-Chun Lo[3], Chen-Yu Chen[4]

[1] Disaster Prevention Technology Research Center, Sinotech Engineering Consultants INC., B1F, No.7, Lane 26, Yat-Sen Rd, Taipei City 11071, Taiwan

[2] Department of Civil Engineering, National Taiwan University, No. 1, Sec. 4, Roosevelt Rd, Taipei City 10617, Taiwan

[3] Department of Civil Engineering, National Chiao Tung University, 1001 University Road, Hsinchu 300, Taiwan

[4] Ujigawa Open Laboratory, Department of Civil and Earth Resource Engineering, Kyoto University , Kyoto daigaku-katsura, Nishikyo-ku, Kyoto 615-8530, Japan,

[*] cwshen@sinotech.org.tw

Abstract

Taiwan was recently ravaged by frequent natural disasters. The typhoons and heavy rainfall triggered landslides, often causing incalculable damage to human's life or properties. In this study, we evaluated the susceptibility of debris flow hazard of 160 potential debris flow creeks in Hualien County by using multivariate statistical analysis. Five categories of topographic factors, including watershed form, slope, height, aspect, and landslide, were used as independent variables and debris flow inventory as dependent variable. By using principal component and correlation analysis, 10 topographic factors are correlated with dependent variable at 85% significance level. Then, we evaluated susceptibility of debris flow form the 10 topographic factors by using logistic regressionand verified by the classification error matrix, ROC curve, and debris flow inventory. The study has established debris flow susceptibility maps and method of debris flow hazard assessment for debris flow prevention and risk management.

Keywords: debris flow, multivariate statistical analysis, logistic regression, topographic factors, susceptibility map.

Introduction

Taiwan is characterized by frequent landslide and debris flow hazard in the mountainous area. The Soil and Water Conservation Bureau, Taiwan (SWCB) has announced 1,660 potential debris flow creeks in Taiwan in 2010 The risk of debris flow are categorized into three levels with high, medium, and low potential based on their watershed topography, geological setting, and risk to protection target [15,16]. In 2001~2004, 685 of the 1,660 debris flow were triggered by heavy rainfall. The typhoon-induced intensive rainfalls in recent years have increased the frequency of massive debris flows in Taiwanese mountains [3].

Topographic form controls the location of debris flow sources (Montgomery and Dietrich, 1994a; Vandaele et al., 1996), and a threshold relation exists between slope angle and the contributing area (Dietrich et al., 1992; Montgomery and Dietrich, 1994b). Numerous studies used topography of source area and channel gradient related to initiation of debris flows. Wichmann et al. (2007) modeled debris-flow initiation locations in relation to channel gradient, discharge and sediment contributing area using GIS. Godt and Coe (2007) show that slope angles >32° and upslope contributing areas < 3000 m^2 critical threshold for debris flow initiation in the central Front Range, Colorado. Other studies have also related to the slope of source areas, with typical values between 27-38° [8,22,29]. A channel gradient greater than 25° is also necessary for debris flow initiation, and it decreases with an increasing catchment area. Millard (1999) indicated that debris flows from channel sidewalls tend to be larger and occur on steeper slopes than those from headwalls. Although this inference agrees with the concept of sediment transport limit [17], it was based on data for a coastal environment and may have only limited applicability to mountains.

This study analysis the susceptibility of 160 potential debris flow creeks in Hualien County by using 20 topographic factors, e.g. watershed form, slope, height, aspect, and landslide calculating the form factors, slope aspect, height, slope grade, and landslide factors in catchment. Statistical significance for the 20 variables were tested by using principle component analysis (PCA) and correction analysis. Then, the empirical formula for susceptibility of debris flows were determined by using logistic regression model.

Study Area

Terrain and Geology: Hualien County is a mountainous region, with plains making up only a small part. In addition to the plains distributed around the Meilun River alluvial fan, majority is distributed along the two sides of the East Rift Valley in strips. Mountains account for majority of Hualien County's total area (approximately 87%), with 40 mountains exceeding 3,000 m. Of those belonging to the Central Range, Hsiukuluan Peak is the tallest (3,833m). Of those belonging to the Coastal Range, Xingang Peak is the tallest (1,628m). The topographic map is shown in Figure 1.

The county's terrain can be divided into Central Range area, Coastal Range area, and Rift Valley plain area. Hualien's geology (as shown in Figure 2) can be roughly divided into three eras and igneous rock: (1) the First Tertiary and Paleogene Periods metamorphic rock region, with Tananao Schist, Xicun Formation (Eocene slate and phyllite in Snow Mountain Range), Xingao Formation (Eocene slate and phyllite in Backbone Range) distributed along the east side of the Central Range; (2) Miocene Lushan Formation is distributed in the mountains to the west of Dawu and Chihpen. Tuluanshan Formation is distributed along the Coastal Range. (3) Dagangkou Formation and Chimei Formation from Pliocene and Pleistocene Periods, Lichi Melange, Puyuma Hill Conglomerate, accumulation of red earth platform, uplifted coral reef, and alluvial layer were distributed along Coastal Range, Taitung Longitudinal Valley, and the east coast (There is no Puyuma Hill Conglomerate in the Hualien area.). (4) Igneous rock: Serpentinite and mafic igneous rock from the First Tertiary Period are distributed around the Central Range. Andesite from the Miocene Period as well as gabbro, peridotite, basalt, serpentinite, and agglomerate from an unknown period are distributed in the Coastal Range.

River System: The drainage system within Hualien County is composed of primarily three river basins: Heping River, Hualien River, and Hsiukuluan River. Heping River is located in the northeastern part of Taiwan, on the border of Ilan and Hualien. To its east is the Pacific Ocean, and to the west is Tachia River. To its south lies the Liwu River, and to the north, Nanau river and Lanyang River. The River distribution map is shown in Figure 3.

Fault Distribution: The faults within Hualien County include the Chimei Fault, Yuli Fault,Yuemei Fault, Meilun Fault, and Chihshang Fault (as shown in Figure 4).

Topographic Factors

This debris flow susceptibility assessment model in this study considers both topographic and landslide characteristic factors. Using 5m-resolution digital elevation model (DEM) and extraction from the Typhoon-Aere landslide inventory, 20 factors under 5 categories are obtained. The topographic factors definition and formula of the 160 potential debris flow creeks are listed in Table 1. The following summarizes the topographic factor categories considered [26].

Form Factor: The catchment factors of form include catchment area, catchment length, catchment circumference, form factor, elongation ratio, and circularity ratio. River basin form factor was raised by Horton in 1932 and is defined as $F=A/L^2$, where L is the total river length, and A is the catchment area [10]. Form factor represents river basin area per main river length, primarily reflecting the length of time of concentration in catchment in order to accurately reflect the effect on catchment by river width.

Slope Aspect Factor: The catchment slope aspect factors include the average slope aspect X vector, average slope aspect Y vector, slope aspect standard deviation, and average slope aspect. Catchment slope aspect factors mainly reflect the strike of slope.

Height Factor: The catchment height factors include Hypsometric Integral (or residual soil rate), average elevation, elevation standard deviation, and elevation coefficient of variation. Height factors reflect the degree of ups and downs in the terrain of the study area. Greater ups and downs signify greater erosion by rain or wind.

Slope Grade Factor: The catchment slope grade factors include average slope grade, first 5% average slope grade, first 10% average slope grade, and first 15% average slope grade. Slope grade factors reflect the fact the steeper the slope, the more susceptible it is to failure.

Landslide Characteristic Factor: Based on the satellite images after Typhoon Aere (2004), the total landslide area and the rate of collapse within 50 m of the two sides are obtained through digitization. The distribution of the landslides in catchment is considered, mainly reflecting the possibility of existing landslides and exposed areas to collapse again.

Establishing a Flow Chart Through susceptibility Model

Based on the landslide inventory established by the "Hazard Factors Investigation in Areas of Potential Debris flow" conducted between 2006 and 2008, potential debris flow creek catchments are divided into debris flow creeks and non-debris flow creeks. Debris flow creeks are those that had landslides and had actual damages and losses, and the opposite is true for non-debris flow creeks. The following points illustrate the steps in establishing susceptibility model [26]. Analysis is shown in Figure 5.

Establishing Overflow Points and Calculating the Topographic Factors: First, establishing overflow point is needed. This study used the definition by SWCB: " First, the affected range of 105° sector area as calculated by Hiroshi Ikeya Formula is used as base map. Possible overflow points (such as valley entrance, obstructions, or where terrain suddenly becomes smooth) are repositioned through site inspection. Afterward, they are corrected based on the current terrain, and the areas where debris flow would be impossible are eliminated." Based on this rule, the distribution of the 160 potential debris flow creeks within Hualien and the overflow points are completed, as shown in Figure 6. Through 5 m digital elevation model (DEM) and the landslide inventory made after Typhoon Aere (2004), extraction of the topographic factors in catchments above overflow points is made.

Principle Component Analysis: Principal component analysis (PCA) is a mathematical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of

uncorrelated variables called principal components. The number of principal components is less than or equal to the number of original variables. This transformation is defined in such a way that the first principal component has as high a variance as possible (that is, accounts for as much of the variability in the data as possible), and each succeeding component in turn has the highest variance possible under the constraint that it be orthogonal to (uncorrelated with) the preceding components. Principal components are guaranteed to be independent only if the data set is jointly normally distributed. PCA is sensitive to the relative scaling of the original variables.

PCA is the simplest of the true eigenvector-based multivariate analyses. Often, its operation can be thought of as revealing the internal structure of the data in a way which best explains the variance in the data. This is done by using only the first few principal components so that the dimensionality of the transformed data is reduced. PCA is closely related to factor analysis; indeed, some statistical packages deliberately conflate the techniques. True factor analysis makes different assumptions about the underlying structure and solves eigenvectors of a slightly different matrix. PCA can be considered as a type of low-rank approximation. This study chooses a representativeness (also known as cumulative proportion) of 85% and 5 principle components (Stone and Brooks, 1990). Through eigenvalues and cumulative variance percentages, representativeness for the principle components is obtained (as shown in Table 2), and the first five principle components are selected.

Correlation Analysis: Formally, dependence refers to any situation in which random variables do not satisfy a mathematical condition of probabilistic independence. In loose usage, correlation can refer to any departure of two or more random variables from independence, but technically it refers to any of several more specialized types of relationship between mean values. There are several correlation coefficients, often denoted ρ or r, measuring the degree of correlation. The most common of these is the Pearson correlation coefficient, which is sensitive only to a linear relationship between two variables. The most familiar measure of dependence between two quantities is the Pearson product-moment correlation coefficient, or "Pearson's correlation." It is obtained by dividing the covariance of the two variables by the product of their standard deviations. The population correlation coefficient $\rho_{X,Y}$ between two random variables X and Y with expected values μ_X and μ_Y and standard deviations σ_X and σ_Y is defined as:

$$\rho_{X,Y} = corr(X,Y) = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y]}{\sigma_X \sigma_Y} \qquad (1)$$

where E is the expected value operator, cov means covariance, and, corr, a widely used alternative notation for Pearson's correlation. The Pearson correlation is defined only if both of the standard deviations are finite, and both of them are nonzero. It is a corollary of the Cauchy–Schwarz inequality that the correlation cannot exceed 1 in absolute value. The correlation coefficient is symmetric: *corr(X,Y) = corr(Y,X)*.

The Pearson correlation is +1 in the case of a perfect positive (increasing) linear relationship (correlation), −1 in the case of a perfect decreasing (negative) linear relationship (anticorrelation), and some value between −1 and 1 in all other cases, indicating the degree of linear dependence between the variables. As it approaches zero, there is less of a relationship (closer to uncorrelated). The closer the coefficient is to either −1 or 1, the stronger the correlation between the variables.

If the variables are independent, Pearson's correlation coefficient is 0, but the converse is not true because the correlation coefficient detects only linear dependencies between two variables. The sample correlation coefficient is written as:

$$\gamma_{XY} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{(n-1)s_x s_y} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2 \sum_{i=1}^{n}(y_i-\bar{y})^2}} \qquad (2)$$

where $\bar{x}$ and $\bar{y}$ are the sample means of X and Y, and s_x and s_y are the sample standard deviations of X and Y. This can also be written as:

$$\gamma_{XY} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{(n-1)s_x s_y} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x^2{}_i - (\sum x_i)^2}\sqrt{n\sum y^2{}_i - (\sum y_i)^2}} \qquad (3)$$

If x and y are measurements that contain measurement error, as commonly happens in biological systems, the realistic limits on the correlation coefficient are not -1 to +1 but a smaller range.

This study uses Pearson product-moment correlation coefficient to examine the independence of the various topographic factors. The range is between –1 and 1. When the value approaches 1, the two variables are positively correlated. When the value approaches –1, the two variables are negatively correlated. When the value equals 0, the two variables are completely independent [5]. A correlation matrix is obtained through the analysis, as shown in Figure 7. Based on the outcome of the analysis, this study eliminates the dependent factors selected in Principle Component Analysis, and the combination of significant topographic factors in the studied area is obtained. So 10 significant topographic factors are chosen, as shown in Table 3.

Sample Selection: The ratio between debris flow creek samples and non-debris flow samples is often disproportional. Considering the importance of unbiased estimate of the samples, analysis should have a sample ratio of 1:1 between debris flow creeks and non-debris flow creeks. This study uses Simple Random Sampling to select

the training sample for analysis [12]. The procedure includes all of the debris flow creek samples and uses Random Sampling to select non-Debris flow creek samples so that there are equal number of debris flow creek samples and non-debris flow-creek samples. The two types of samples are combined to form the training sample used in the analysis.

Susceptibility Analysis: In statistics, logistic regression (sometimes called the logistic model or logit model) is used for prediction of the probability of occurrence of an event by fitting data to a logistic function. It is a generalized linear model used for binomial regression. An explanation of logistic regression begins with an explanation of the logistic function, which, like probabilities, always takes on values between zero and one:

$$f(z) = \frac{e^z}{e^z + 1} = \frac{1}{1 + e^{-z}} \tag{4}$$

The input is z, and the output is $f(z)$. The logistic function is useful because it can take as an input any value from negative infinity to positive infinity, whereas the output is confined to values between 0 and 1. The variable z represents the exposure to some set of independent variables, while $f(z)$ represents the probability of a particular outcome, given that set of explanatory variables. The variable z is usually defined as:

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + ... + \beta_k x_k \tag{5}$$

where β0 is called the "intercept", and β_1, β_2, β_3, and so on, are called the "regression coefficients" of x_1, x_2, x_3 respectively. The intercept is the value of z when the values of all independent variables are zero (e.g. the value of z in someone with no risk factors). Each of the regression coefficients describes the size of the contribution of that risk factor.

Susceptibility Analysis uses Logistic Regression in Data Mining Prediction Algorithm. It was executed through data mining software Polyanalyst6.0. Through multivariate statistics, a set of linear combination functions composed of topographic factors and regression coefficients that represent the degree of contribution of each factor are obtained. Through induction, we obtained a discerning debris flow susceptibility analytical empirical formula. The following summarizes the basis for Logistic Regression. Logistic Regression is a special form of Log-Linear Model. It is a function that can be applied to a binary variable as dependent variable (ie. debris flow and non-debris flow) and defines a series of independent variables (the debris flow topographic factors in this study). Logistic Regression can be expressed as following (Hosmer and Lemeshow, 1989):

$$P(y_i = 1 \mid x_i) = \frac{1}{1 + e^{-Z}} \tag{6}$$

$$Z = \alpha + \beta x_i \tag{7}$$

Z is a linear polynomial of probability factors that affect the occurrence of an event. Its range lies between negative infinity and positive infinity. Substituting the value of Z into equation (2), we can obtain the value for P, which lies between 0 and 1. This represents the possibility for an event to occur. x_i is an independent variable, and α and β are regression intercept (constant) and regression coefficient respectively. In the analysis of the probability for debris flow to occur, x_i is the value for each debris flow impact factor. β is the weight of each factor. Substituting Z into (1), we can obtain the value for P, the debris flow susceptibility value. Next, we follow the suggestion by Lillesand and Kiefer (2000)[13] to use Classification Error Matrix to indicate the accuracy of pattern analysis. This study recommends classification accuracy greater than 70% as the outcome of the pattern analysis is better.

Statistical Tests: Individual parameter testing may follow recommendation from Hair (1998), using Cox & Snell R square value in model summary testing. The primary goal is to determine the significance of model variables [7]. In other words, the higher the Cox & Snell R square value, the more the independent variables used in model can discern debris flow set from non-debris flow set, and the more accurate the model is.

In terms of overall model testing, indices are used to determine the Goodness of Fit of the overall model, and they are, respectively, Chi-square test (χ2 value) and Hosmer-Lemeshow test [9]. When χ2 is significant, it indicates that there needs at least one independent variable that can predict samples' probability value in the dependent variables and that value for P is not significant (P>0.05) for the Goodness of Fit of the overall model to be good.

Let us consider a two-class prediction problem (binary classification), in which the outcomes are labeled either as positive (p) or negative (n) class. There are four possible outcomes from a binary classifier. If the outcome from a prediction is p and the actual value is also p, then it is called a true positive (TP); however, if the actual value is n, then it is said to be a false positive (FP). Conversely, a true negative (TN) has occurred when both the prediction outcome and the actual value are n, and false negative (FN) is when the prediction outcome is n while the actual value is p. Let us define an experiment from P positive instances and N negative instances. The four outcomes can be formulated in a 2×2 contingency table or confusion matrix, as shown in Figure 8.

Receiver Operator Characteristic curve (ROC curve) was raised by Swets in 1988 [23]. ROC curve primarily shows the accuracy of analysis model. The proportion of incorrect interpretations from Classification Error Matrix makes up the x-axis, and the proportion of correct interpretations makes up the y-axis. A curve is thus drawn. The higher the proportion of correct interpretations, the greater the area is under the curve (AUC). This indicates the analysis model

has better classification accuracy, and AUC value falls between 0 and 1. This study recommends setting AUC to 0.7 as benchmark for judging analysis model. If AUC value is greater than 0.7, the requirement for model accuracy is met.

Susceptibility Analysis Results

Logistic Regression Equation: The outcome of debris flow susceptibility analysis needs to be verified by Classification Error Matrix and ROC curve for classification accuracy. After the model significance is verified by statistical tests, a debris flow susceptibility map can be drawn based on susceptibility classification criteria.

Through Logistic Regression, debris flow susceptibility model in Hualien County is obtained (the Logistic Regression Equation), as shown in equation (3). Substituting equation (3) in Logistic Regression Equation (2), each of the debris flow susceptibility value can be obtained. The values are between 0 and 1. Greater the susceptibility values means greater probability of debris flow occurring. They are relative values.

$$\lambda = 0.002X_1 - 5.829X_2 + 8.941X_3 + 12.286X_4 - 0.011X_5 - 0.357X_6 - 0.016X_7 + 0.114X_8 - 0.020X_9 + 0.026X_{10} - 0.406 \quad (8)$$

X_1 is the catchment area; X_2 is the form factor; X_3 is the circularity ratio; X_4 is Hypsometric Integral; X_5 is height coefficient of variation; X_6 is the average slope gradient; X_7 is the average slope aspect; X_8 is the average slope aspect for the first 5 percent; X_9 is the standard deviation for slope aspect; X_{10} is the landslide area within 50 m of the two sides.

Susceptibility Classification Method: Using 0.5 as Logistic Regression index P to divide the two groups, this study recommends classifying those susceptibility values greater than 1.5 times that of Logistic index (in other words, susceptibility value$\geqq$ 0.75) as high susceptibility class, classifying susceptibility values between Logistic index and 1.5 times that of Logistic index as intermediate-high susceptibility class (in other words, 0.5$\leqq$susceptibility value<0.75), classifying susceptibility values between Logistic index and half of Logistic index as intermediate susceptibility class (in other words, 0.25$\leqq$susceptibility value<0.5), and classifying susceptibility values less than half of Logistic index as low susceptibility class (susceptibility value<0.25).

Pattern Analysis Accuracy: Classification Error Matrix is made from the result of susceptibility analysis, indicating the accuracy of pattern analysis. The result is shown in Table 4, and the Receiver Operator Characteristic curve is shown in Figure 9. After comparison with disaster inventory (debris flow disaster inventory from SWCB(2006)" The investigation of vulnerability factors in debris flow areas and it's counter measurements report" Chapter3 and SWCB(2008)"The investigation of vulnerability factors and risk analysis, risk management of debris flows report" Chapter3) [24, 26], the result indicates that the overall accuracy of the analysis result for debris flow susceptibility in Hualien County is close to 80%. Receiver Operator Characteristic curve also meets the significance requirement (AUC>0.7).

Results of Statistical Tests: The result from the above analysis then undergoes Homer and Lemeshow test as well as Cox&Snell R square test. Hosmer and Lemeshow test indicates that p value is greater than the significance level of 0.05. This test assumes if p value is greater than the significance level 0.05, then the analysis model is significant. The greater the Cox&Snell R square value is, the better the "Goodness of fit" will be. In this model, the correlation is the medium relationship of Cox&Snell R square test. The statistical testing results are compiled in Table 5. The results indicate that the statistical tests passed the requirement, and a susceptibility map can be drawn based on the susceptibility classification, as shown in Figure 10.

Conclusion

1. This study raises a susceptibility analysis flow chart that is based on statistical theories and is available for future debris flow related studies.
2. The overall accuracy for debris flow susceptibility analysis in Hualien County exceeds 80%. Moreover, the statistical tests meet the requirement for a significant pattern. These indicate that the analysis model is fairly reliable.
3. From the debris flow susceptibility map, the probability of landslides happening in the potential debris flow creek catchments is obtained. The result is available for classification management and disaster prevention..

Acknowledgement

We would like to extend our gratitude to the Soil and Water Conservation Bureau in Executive Yuan's Council of Agriculture for their funding and technical support. It allowed this project to be smoothly completed. We would also like to thank everyone who had participated or assisted in this project as a part of the Sinotech team.

References

1. Chen, S.H., Su, H.B., Tian, J., Zhang, R.H., Estimating soil erosion using MODIS and TM images based on support vector machine and a trous wavelet, International Journal of Applied Earth Observation and Geoinformation, 13(4), 626-635.(2011)
2. Chen, C.Y., Chen, T.C., Yu, F.C., and Hung, F.Y., "A Landslide Dam Breach Induced Debris Flow – A Case Study on Downstream Hazard Areas Delineation", Environmental Geology, 47, 91-101.(2004)
3. Chen, C.Y., Chen, L.K., Yu, F.C., Lin, S.C., Lin, Y.C., Lee, C.L., Wang, Y.T., Cheung, K.W., "Characteristics analysis for the flash flood-induced debris flows", Natural Hazards, 47, 245–261. (2008)
4. Dietrich, W.E., Wilson, C.J., Montgomery, D.R., McKean, J., Bauer, R., "Erosion thresholds and land surface morphology", Geology, 20, 675–679.(1992)

5. Fisher, R. A., “The general sampling distribution of the multiple correlation coefficient”, Proceedings of the Royal Society of London, Ser. A, 121, 654-673.(1928)

6. Godt, J.W., Coe, J.A., “Alpine debris flows triggered by a 28 July 1999 thunderstorm in the central Front Range, Colorado”, Geomorphology, Vol. 84, pp. 80–97, 2007.

7. Hair, J.F. et al., Multivariate Data Analysis, 5th edition, Prentice Hall.(1998)

8. Hungr, O., Morgan, G.C., Kellerhals, P., “Quantitative analysis of debris hazards for design of remedial measures”, Canadian Geotechnical Journal, 21, 663–677.(1984)

9. Hosmer, D. W and Lemeshow, S., Applied Logistic Regression, John Wiley & Sons, Inc.(1989)

10. Horton, R.E., “Drainage Basin Characteristics”, Trans. Amer. Geophys. Union, 13, 350-361.(1932)

11. Johnson, A.M., Rodine, J.R., Debris flow, In: Brunsden, D., Prior, D.B. (Eds.), Slope Instability, John Wiley and Sons Ltd., Chichester, UK, 257–361.(1984)

12. Lee, C. T., Huang, C. C., Lee, J. F., Pan, K. L., Lin, M. L., Dong, J. J., “Statistical approach to earthquake-induced landslide susceptibility”, Engineering Geology, 100(1-2), 43-58. (2008)

13. Lillesand, T. M. and Kiefer R. W., Remote Sensing and Image Interpretation, Fourth Edition, John Wiley & Sons, Inc.(2000)

14. Lin, M.L., Wang, K.L., and Huang, J.J., “Debris Flow Run Off Simulation and Verification – Case Study of Chen-You-Lan Watershed”, Taiwan, Natural Hazards and Earth System Sciences, 5, 439-445.(2005)

15. Lin, P.S., Lin, J.Y., Hung, J.C., Yang, M.D., “Assessing debris-flow hazard in a watershed in Taiwan”, Engineering Geology, 66, 295–313.(2002)

16. Lin, P.S., Lin, J.Y., Lin, S.Y., Lai, J., “Hazard assessment of debris flows by statistical analysis and GIS in Central Taiwan”, International Journal of Applied Science and Engineering, 4, 165–187.(2006)

17. Marshall, T.J., Holmes, J.W., Rose, C.W., Soil Physics, 3rd Edition. Cambridge University Press.(1996)

18. Millard, T., “Debris flow initiation in coastal British Columbia gullies”, Forest Research Technical Report, TR-002 Geomorphology, 23-24.(1999)

19. Montgomery, D.R., Dietrich, W.E., Landscape dissection and drainage area–slope thresholds, In: Kirkby, M.J. (Ed.), Process Models and Theoretical Geomorphology. John Wiley, New York, 221-246.(1994a)

20. Montgomery, D.R., Dietrich, W.E., “A physically-based model for the topographic control of shallow landsliding”, Water Resources Research, 30, 1153-1171.(1994b)

21. Pike, R. J. and Wilson, S. E., “Elevation-relief ratio, hypsometric integral, and geomorphic area-altitude analysis”, Geological Society of America Bulletin, 82, 1079-1084.(1971)

22. Rickenmann, D., Zimmermann, M., “The 1987 debris flows in Switzerland: documentation and analysis”, Geomorphology, 8, 175-189.(1993)

23. Swets, J.A., “Measuring the Accuracy of Diagonstic Systems”, Science, 204(4857), 1285-1293.(1988)

24. Sinotech Engineering Consultants, INC., “The investigation of vulnerability factors in debris flow areas and it’s counter measurements“, Report to the Soil and Water Conservation Bureau, Council of Agriculture, Executive Yuan.(2006) (in Chinese)

25. Sinotech Engineering Consultants, INC., “The investigation of vulnerability factors of debris flows torrents and it’s counter measurements“, Report to the Soil and Water Conservation Bureau, Council of Agriculture, Executive Yuan.(2007) (in Chinese)

26. Sinotech Engineering Consultants, INC., “The Investigation of Vulnerability Factors and Risk Analysis, Risk Management of Debris Flows”, Report to the Soil and Water Conservation Bureau, Council of Agriculture, Executive Yuan.(2008) (in Chinese)

27. Sinotech Engineering Consultants, INC., “The Investigation of Vulnerability Factors and Risk Analysis”, Risk Management of Debris Flows, Report to the Soil and Water Conservation Bureau, Council of Agriculture, Executive Yuan.(2009) (in Chinese)

28. Stone, M. and Brooks, R. J., “Continuum Regression: Cross-validated Sequentially Constructed Prediction Embracing Ordinary Least Squares”, Partial Least Squares, and Principal Components Regression, Journal of Royal Statistical Society, 52(2), 237-269.(1990)

29. Takahashi, T., “Estimation of potential debris flows and their hazardous zones; soft countermeasures for a disaster”, Journal of Natural Disaster Science, 3, 57–89.(1981)

30. Vandaele, K., Poesen, J., Govers, G., van Wesemael, B., “Geomorphic threshold conditions for ephemeral gully incision”, Geomorphology, 16, 161–173.(1996)

31. Wichmann, V., Heckmann, T., Haas, F., Becht, M., “Modelling alpine sediment cascades: process interaction and landscape connectivity”, Geophysical Research Abstracts 9 SRef-ID: 1607-7962/gra/EGU2007-A-06140.(2007)

Table.1 Results of principal component analysis

Principle Component	Eigenvalue	Variance Percentage, %	Cumulative Variance,%
First Principle Component	10.079	50.393	50.393
Second Principle Component	2.863	14.313	64.706
Third Principle Component	1.648	8.241	72.947
Fourth Principle Component	1.337	6.687	79.634
Fifth Principle Component	1.164	5.821	**85.455%>85%**

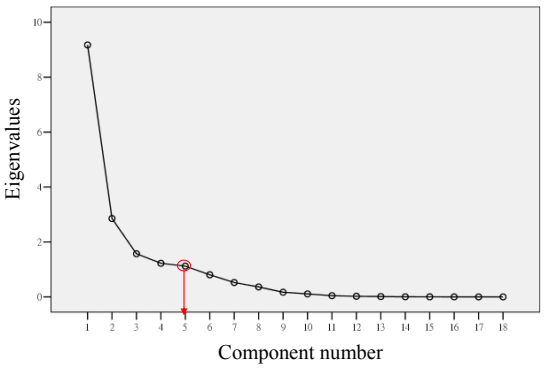


Table.2 Combination of topographic factors significantly List

No.	Debris Flow Susceptibility Factor Categories	Susceptibility Factor Types	Principle Component Type	Unit
1	Catchment Factors of form	Catchment Area	Second Principle Component	km^2
		Form Factors	Second Principle Component	(Dimensionless)
		Circularity Ratio	Fifth Principle Component	(Dimensionless)
2	Catchment Slope Aspect Factors	Slope Aspect Standard Deviation	Third Principle Component	(Dimensionless)
		Average Slope Aspect	Fourth Principle Component	Degree
3	Catchment Height Factors	Hypsometric Integral	Fourth Principle Component	(Dimensionless)
		Height Coefficient of Variation	Fifth Principle Component	(Dimensionless)
4	Catchment Slope Grade Factors	Average Slope	Third Principle Component	Degree
		Average Slope for the First 5%	First Principle Component	Degree
5	Catchment Landslide characteristic factors	Total landslide Area Within 50 m of Each Side	First Principle Component	m^2

Table.3 Classification error matrix results

Accuracy Types	Hualien County
Non-Debris flow Set Accuracy (%)	81.5
Debris flow Set Accuracy (%)	77.8
Overall Accuracy (%)	79.6

Table.4 Logistic regression model statistical test results

Test Types / Area	Hosmer and Lemeshow Test			Model Summary
	Chi-square test	Degree of Freedom	p Value	Cox & Snell R square
Hualien County	9.22	8	0.102＞0.05:significant	0.359:medium relationship

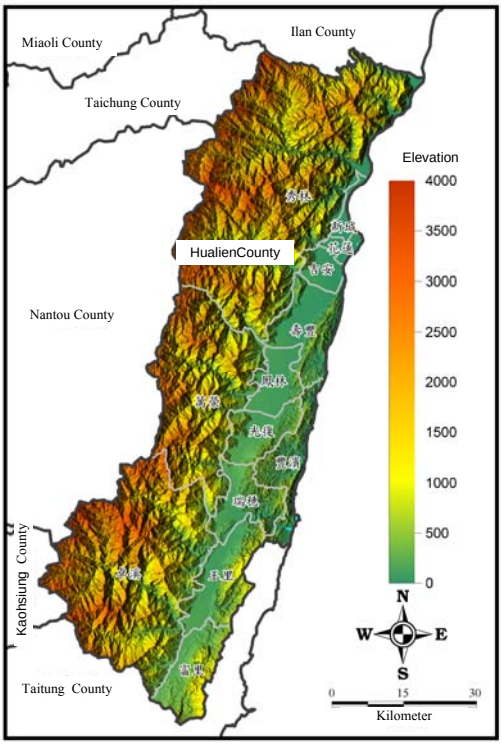


Fig.1 topographic map (redraw from Sinotech,2008)

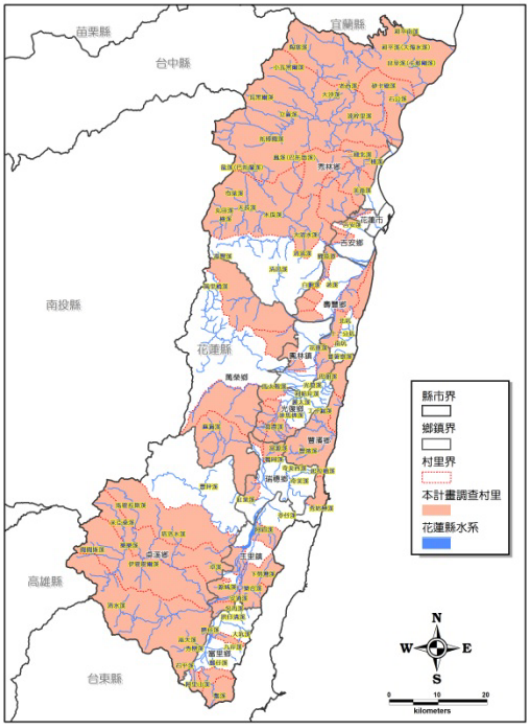


Fig.3 River Distribution map(redraw from SWCB,2008)

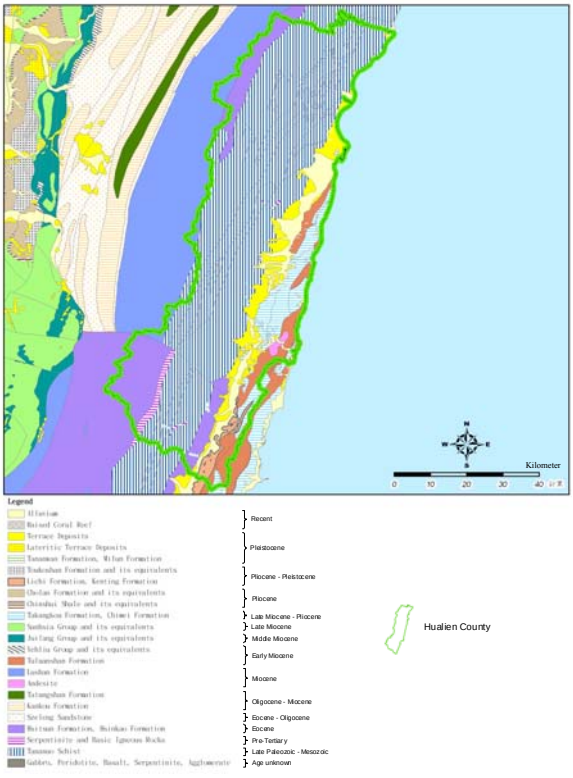


Fig.2 1/250,000 geologic map(redraw from CGS-MOEA geologic map)

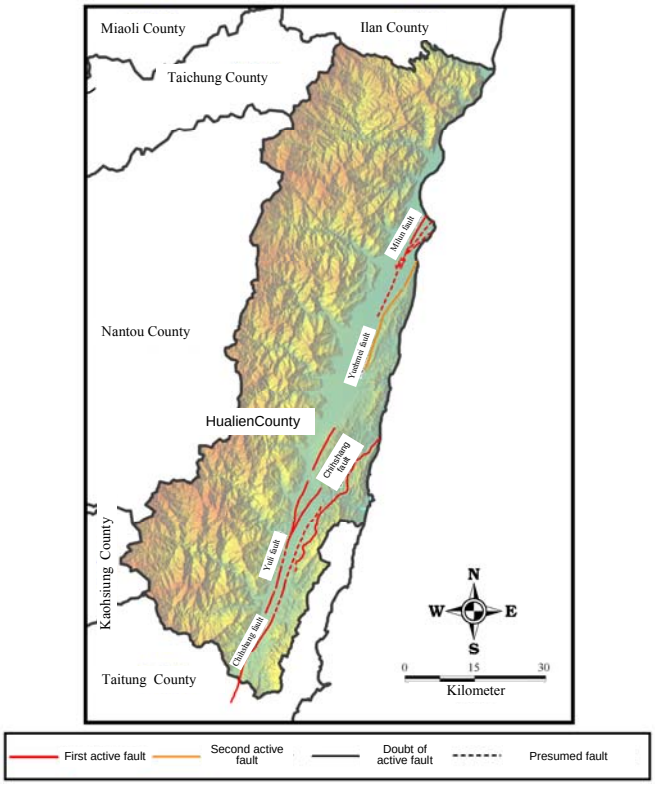


Fig.4 Fault Distribution map(redraw from CGS-MOEA activity fault map)

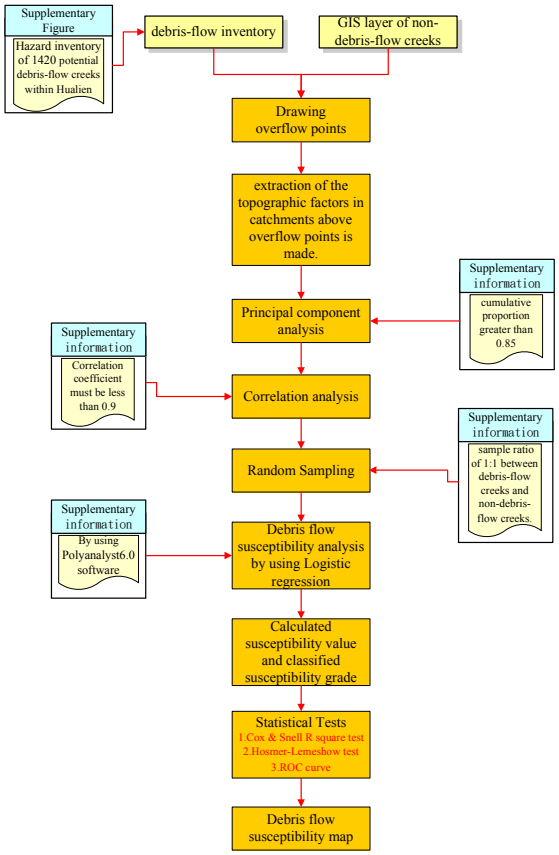
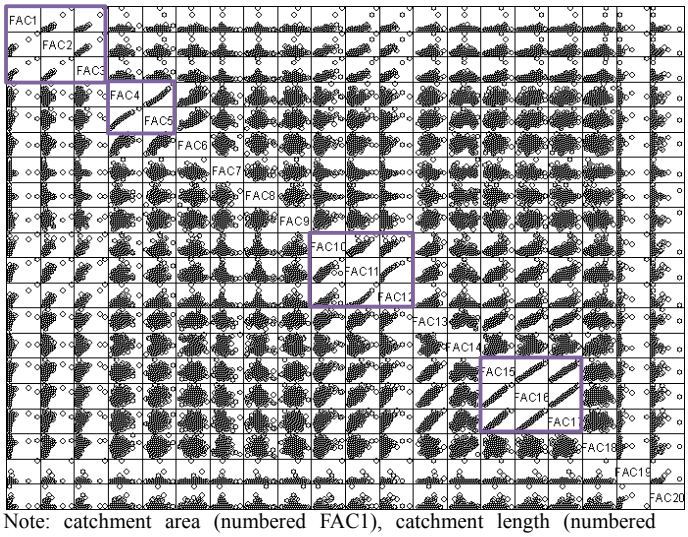


Fig.5 Flow chart of debris flow susceptibility analysis



Note: catchment area (numbered FAC1), catchment length (numbered FAC2), catchment circumference (numbered FAC3), form factor (numbered FAC4), elongation ratio (numbered FAC5), circularity ratio (numbered FAC6), the average slope aspect X vector (numbered FAC7), average slope aspect Y vector (numbered FAC8), slope aspect standard deviation (numbered FAC9), average slope aspect (numbered FAC10), Hypsometric Integral (numbered FAC11), average elevation (numbered FAC12), elevation standard deviation (numbered FAC13), elevation coefficient of variation (numbered FAC14), average slope grade (numbered FAC15), first 5% average slope grade (numbered FAC16), first 10% average slope grade (numbered FAC 17), first 15% average slope grade (numbered FAC18), the total landslide area (numbered FAC19) and the rate of landslide (numbered FAC20) within 50 m of the two sides.

Fig.7 Correlation matrix of topographic factors in Hualien County

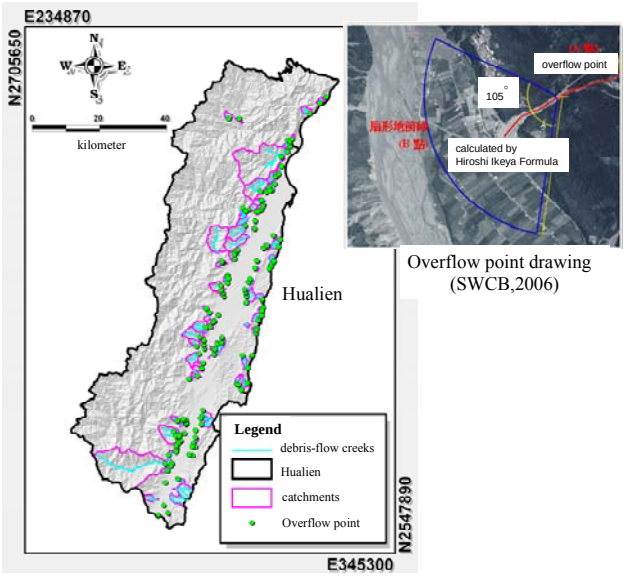


Fig.6 Debris flow catchment area and the overflow point distribution in Hualien County. (Sinotech，2008)

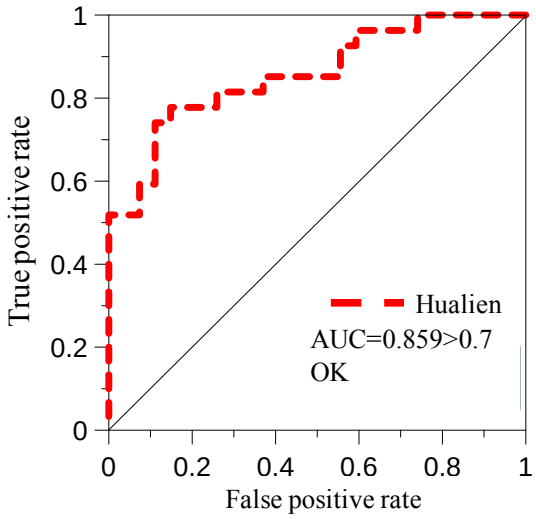


Fig.8 Receiver Operator Characteristic curve of debris flow susceptibility in Hualien County

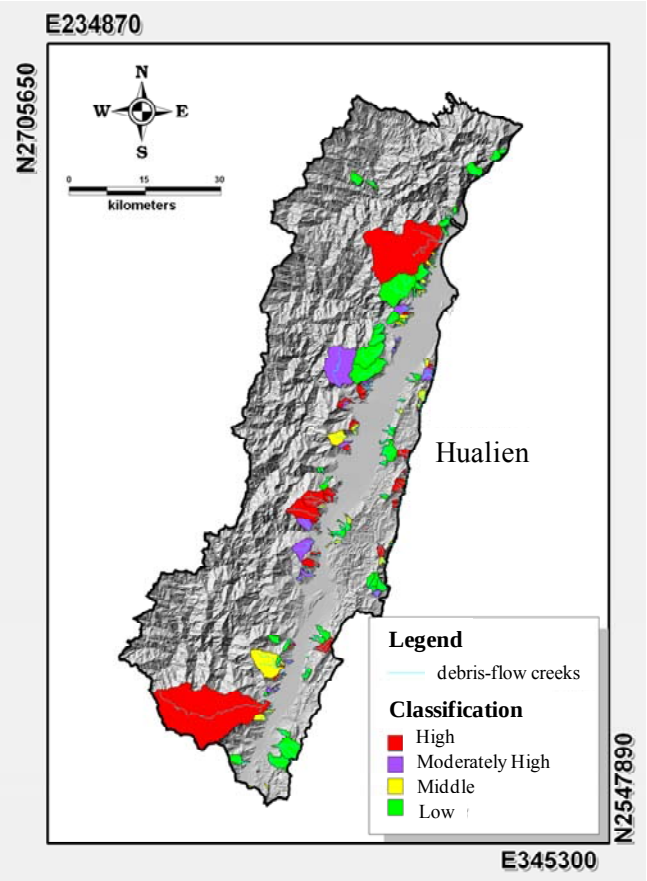


Fig.9 Debris flow susceptibility map of Hualien County