Designing intelligent disaster prediction models and systems for debris-flow disasters in Taiwan

Hsu-Yang Kung\textsuperscript{a}, Chi-Hua Chen\textsuperscript{b,*}, Hao-Hsiang Ku\textsuperscript{c}

\textsuperscript{a}Department of Management Information Systems, National Pingtung University of Science and Technology, 1 Shuefu Road, Neipu, Pingtung 912, Taiwan, ROC
\textsuperscript{b}Institute of Information Management, National Chiao Tung University, 1001 University Road, Hsinchu 300, Taiwan, ROC
\textsuperscript{c}Department of Computer Science and Information Engineering, Hwa Hsia Institute of Technology, 111 Gong Jhuan Road, Chung Ho, Taipei 235, Taiwan, ROC

\textbf{1. Introduction}

Due to Taiwan’s distinctive location on the seismic belt, debris-flow disaster prediction and notification has become an urgent task for Taiwan authorities. To build disaster-prevention mechanisms, the government is now actively promoting national disaster prevention programs. Damage to slopeland caused by earthquake has lowered the threshold of debris-flow occurrence year by year. If the area with a slope greater than 15° exceeds 3 hectares, then serious debris-flow disasters may easily occur and some areas near mountains were particularly severely hit by mass debris flows (Shieh, Chen, Tsai, & Wu, 2009). The authorities concerned therefore spent significant amounts of money setting up a debris-flow monitoring and detecting system. Nevertheless, the established monitoring and detecting stations often break down, and the accuracy rate is below 30% (Tamotsu, 2002). In this way, the detection results are only used as references for studies, rather than for real-time disaster prevention. Moreover, debris flows often destroy communication lines, paralyzing communication systems. As a result, setting up an effective real-time debris-flow disaster prevention and notification system without the limit of time and space has become an urgent task for the government.

Taiwan’s dense population accelerates the development of slopeland. Therefore, debris-flow disasters extend to the over-developed slopeland areas. For instance, in 2003, Typhoon Xangsane brought floods and mudflows to the residential areas on Xizhi’s mountaintops. Some residential regions, such as the slopeland areas in Nantou County, have been equipped with communicative infrastructures. In these areas, easy-to-carry information and communication equipments, such as PDAs combined with cellphones, can be used for real-time multimedia disaster information transmission, including data, sounds and photos, as well as disaster decision and prevention.

In this study, a \textit{Real-time Mobile Debris-Flow Disaster Forecast System} (RM(DF)\textsuperscript{2}) was designed to achieve the prediction and guard against of debris-flow disasters. In RM(DF)\textsuperscript{2}, users can apply a PDA or other handheld devices to input the information of the related disaster area and send the information back to the decision support system of the rescue center via \textit{Global System for Mobile Communications} (GSM), \textit{General Packet Radio Service} (GPRS), \textit{Universal Mobile Telecommunications System} (UMTS), or \textit{Long Term Evolution} (LTE) networks for disaster analysis. This study proposes three strict debris-flow mathematical analysis models based on (1) multiple linear regression, (2) multivariate analysis and (3) back-propagation network. Selecting appropriate disaster factors is the most important part of setting up a debris-flow prediction model. The correct selection of mudflow disaster factors can improve the accuracy of prediction model and then predict precisely the debris-flow...
precise and objective debris-flow disaster prediction.

(i) Non-real-time factors result from long-term environmental changes, including effective watershed, effective channel length, effective channel slope, and the rocks in the effective watershed.

(ii) Real-time factors can lead to immediate hazards. Such factors include the collapsed area within the effective watershed, effective accumulated precipitation, effective rainfall intensity, and vegetation index.

These disaster factors were extracted by using GIS and RS technology. Additionally, the accuracy of each debris-flow prediction model was examined herein by analyzing the historical indices of the 181 potential debris-flow hazardous torrents in Nantou County, Taiwan. Then, the best prediction model was chosen for precise and objective debris-flow disaster prediction.

This study has seven sections. Section 2 describes the relevant references and their advantages and disadvantages. Section 3 describes the configuration of the debris-flow disaster factor database. The selection and extraction of the occurring factors of debris-flow disasters are the main discussion in this section. Section 4 presents the architecture of the proposed debris-flow prediction model. Section 5 presents the analysis of three prediction models. Section 6 presents the implementation of the RM(DF)² system. The final section draws a conclusion and suggests subjects for future research.

2. Related work

Debris-flow disaster notification and warning systems are typically divided into contact and non-contact alarm models (Wei, Gao, & Cui, 2006): (1) Contact alarm models use debris-flow transmission and sensing equipment to generate the alarm signals. (2) Non-contact alarm models are classified into three types. (i) Image monitoring alarms use cameras to identify debris flows. However, this alarm is restricted to the weather and light conditions. (ii) Supersonic bit alarms use supersonic monitoring instruments to detect debris flows. The alarm produces a warning when the test value reaches a predetermined warning value. (iii) Debris-flow geoaoustic alarms produce alarm signals by sensing the special geoaoustic wave formed by debris-flow movement. Wei et al. proposed the disaster mitigation decision support system, which uses mass instruments to conduct debris-flow disaster prevention and testing. However, too many potential debris-flow hazardous torrents exist, meaning that the mechanical equipment requires adjustment and maintenance at any moment. In other words, performing the disaster prevention tasks would cost a fortune.

Cheng proposed the following classes of debris-flow model: (i) periodicity, (ii) randomness, and (iii) near periodicity (Cheng, 2002). Cheng’s debris-flow occurring models mainly relate to the random events caused by rainstorms. The scale of occurrence is relevant to the rainfall and the quantity of scree accumulated in the valley. Cheng also found that the near-periodicity of debris-flow occurrence is closely linked to the local climate. However, the accurate occurring timing is always hard to predict precisely. A sudden heavy rain may produce a severe debris-flow disaster. Therefore, the debris-flow occurring model parameter should be strictly and carefully defined to enhance the prediction accuracy.

In the study of Chang et al., mechanical vision was used to assess the occurrence of mudflows and give warning signals (Chang, Huang, & Lee, 2005). Chang et al.’s study integrates machinery, electronics, optics and computers to assess the following four image processes: (i) moving objects, (ii) the wave front of debris flows, (iii) objects in the stream on the scene and (iv) the grain of debris-flows and floods. First, the image bits are transformed and de-speckled. The distinctive objects are separated from the background image to be compared with the previous one and checked. Then, the signals are transformed into readable messages or data to measure the disaster condition and achieve an early warning of debris-flow disasters. However, the image processing method requires leased lines and high bandwidth for photo transmission. Chang et al.’s system has a high setup cost, and most images cannot be easily distinguished among the noise factors.

Both Yu’s study of slump-mud-flow forecasts (Yu, 2002) and Tan’s study of “the distribution of mudflow channel critical rainfall line” (Tan, Luo, & Wang, 2000) use rainfall parameters, including rainfall and rainfall intensity, to assess the possibility of debris-flow occurrence. Professors Yu and Tan also emphasize that rainfall parameters are important debris-flow occurrence factors. However, debris-flow occurrence is not only determined by a single rainfall factor. Other relevant factors should also be considered to improve the accuracy of the inference.

Chen took the 528 potential debris-flow hazardous torrents in northern Taiwan as sample spaces. Chen analyzed the hazard levels of debris-flow occurrences (Chen, 2002) based on rainfall factors, including hourly rainfall and effective accumulated rainfall, and used simple fuzzy theory to assess the possibility of debris-flow disasters resulting from potential debris-flow hazardous torrents. Chen undertook fieldwork, and then used fuzzy theory and previous experiences to create a hazard level analysis of the potential debris-flow hazard torrents. The result of the analysis was charted and taken as early warning suggestions. Chen’s study focused on the complete collection of information, making the analytical steps complicated and time-consuming. The proposed fuzzy theory sets up the assessing model by only using the case-inference method, and gives early warnings in spread sheets. Chen’s study does not apply information technology effectively enough.

Liu used hydrological and physiographical factors to evaluate the potential energy of debris-flow occurring from the sample torrents, which can be considered as the rainfall threshold for a debris-flow early warning (Liu, 2000). Liu’s study consists of three steps. (i) Choose a sample stream, list its characteristics, and determine the related potential debris-flow factors by statistical inspection. (ii) Combine the results of precipitation analysis to formulate the rainfall threshold for debris-flow occurrence. (iii) Compare the threshold rainwater line of debris-flow occurrence with the frequency cycle of rainstorm recurrence, and evaluate the hazard level. Although this study infers the prediction formula of the debris-flow line threshold and debris-flow occurrence, the maximum accuracy of the prediction formulae is only 80%. Additionally, the prediction formulae are only suitable for some streams. The reliability of the prediction formulae requires further discussion.

Considering the flaws found in the above studies, the design principles and functions of RM(DF)² are as follows.

(1) The proposed system requires only cheap portable devices, e.g., a PDA and a cell phone, and not a high-priced hardware system.

(2) The occurrence of debris flows belongs to a random event. Therefore, the debris flows should be evaluated by real-time functions and immediate on-line analysis. If users are disconnected, then they can use the regulation and case inference engine input on a PDA to aid assessment (George & William, 1999; Huang, 2000).

(3) To predict and assess the occurrence of debris flows, three numerical analysis models are considered, (i) multiple linear regression, (ii) multivariate analysis and (iii) back-propagation network. Additionally, a precise numerical inference, which is quicker and more objective than pattern recognition, was conducted on the selected factors.
The choice of debris-flow danger factors is the most important part in the foundation of the debris-flow prediction model. The appropriate choice of disaster factors can improve the accuracy of the prediction model and help predict the recurrence of debris-flow disasters. The disaster factors chosen in this study are divided into non-real-time factors and real-time factors. (i) Non-real-time factors are derived from long-term environmental changes. For instance, the rocks in the effective watershed is a non-real-time factor. (ii) Real-time factors can lead to immediate hazards. Effective rainfall intensity is an example of a real-time factor. Decision strategies and assessment of related danger factors are discussed in Sections 4 and 5.

3. The setup of the database of debris-flow disaster factors

This study adopted multiple linear regression, multivariate analysis and back-propagation network for regression analysis, and examined these three approaches to obtain the optimal analysis model. The accuracy of the prediction model is dependent on the selection of debris-flow danger factors and the establishment of the database. Related research methods and foundation strategies are described below.

3.1. The selection of debris-flow danger factors

In 2000, 181 streams located in the effective watersheds with slope greater than 15° in Nantou County were defined as potential debris-flow hazardous torrents. In 2001, Typhoon Toraji brought debris flows to all 181 potential debris-flow hazardous torrents. This study utilized the data and information of the debris flows in the 181 hazardous torrents as research samples. The choice of debris-flow danger factors is the most significant part in the foundation of a real-time debris-flow prediction model. The correct choice can make the prediction model achieve expected effects and precisely predict the occurrence of debris-flow disasters. A debris-flow is defined by three major conditions, (i) large water supply, (ii) large channel deposits and material sources, and (iii) steep slope conditions (Shieh et al., 2009).

The chosen danger factors in this study are divided into non-real-time factors and real-time factors. (i) Non-real-time factors are derived from long-term environmental changes, and therefore we require a long-range monitor to determine the changes. Non-real-time factors include (1) the effective watershed, (2) the length of the effective channel, (3) the slope of the effective channel and (4) the rocks in the effective watershed. (ii) Real-time factors result from immediate changes in disaster environment, and include (1) the collapsed area within the effective watershed, (2) effective accumulated precipitation, (3) effective rainfall intensity and (4) the vegetation index. Except the uncertainty of the factor, the collapsed area within the effective watershed, the real-time information of effective accumulated precipitation and effective rainfall intensity can be achieved from the environmental cognition agent designed in the system. The change potential of the disaster environment can also be monitored in real time by using the messages from the real-time factors.

This study applied satellite Remote Sensing (RS) technology to calculate the effective area. In RS technology, the radiation values of Infrared (IR) wave band and NDVI reflected from the earth surface denote the collapsed areas within effective watersheds. If
the average IR/NDVI value is low, then the average collapsed area within the effective watershed is large. Fig. 1 illustrates the boundary line, which is transformed and classified from the SPOT satellite images of Nantou disaster area in 2001, between the vegetation area and the collapsed area.

3.2. The extraction of debris-flow danger factors

The accuracy of the extraction of debris-flow danger factors is crucial to the success of the regression model. If a bad factor is extracted, then the variable becomes a noise among many data in the regression model, or even lowers the reliability of the regression model (Shieh et al., 2009). Herein, GIS/RS information technology was employed to extract the debris-flow danger factors and set up the database of debris-flow danger factors in Nantou County. The eight extraction methods adopted in the research are described below.

(i) Extraction of non-real-time factors: Non-real-time factors include effective watershed, effective channel length, effective channel slope, and the variety of rocks in the effective watershed. The extraction of these slow-changing environmental factors requires a long-term monitoring and complicated geomorphic calculation (Antenucci, 1991; Tamotsu, 2002). This study used GIS and RS technology, and defined as a slope steeper than 15° as the effective watershed of the drainage area (Shieh et al., 2009). Fig. 2 shows the extracted effective catchment area. Fig. 3 illustrates the calculation of the effective watershed, stream length, stream slope and lithological properties based on the overlying of disaster coverages. The danger factor information can be used for numerical regression analysis.

(ii) Extraction of real-time factors: The real-time factors include effective accumulated precipitation, effective rainfall intensity, IR mean value in the effective watershed and NDVI mean value in the effective watershed. The real-time information

Fig. 2. Delimitation of effective watershed of debris flows.

Fig. 3. Extraction of danger factors via GIS/RS information.
of effective accumulated precipitation (mm) and effective rainfall intensity (1 h/mm) can be obtained from the environmental cognition agent designed in the RM(DF)² system (Kung & Ku, 2003). As shown in Fig. 4, the IR mean value and NDVI mean value in the effective watershed are extracted from the high-resolution satellite images filmed in different times (Gilabert & González-Piqueras, 2002; Yang et al., 2004).

The extracted non-real-time and real-time factors are transmitted to the rear-end professional system to calculate the real-time regression and estimate the precise on-site potential debris-flow hazard level. Fig. 5 displays the database of debris-flow danger factors based on the danger-factor extracted values of the 181 streams in Nantou County. Such values are actually generated by using object-oriented programming using the connective database.

4. The theories and architecture designs of debris-flow prediction model

This research utilized multiple linear regression, multivariate analysis, and back-propagation network to generate the debris-flow disaster prediction scheme. The related research theories and system designs are described below.

4.1. Multiple linear regression method

Suppose that $Y$ is a variable and that $X_i$ ($i = 1, 2, \ldots, k$) is an independent variable. Let the expected value of variable $Y$ be the linear function of independent variable $X_i$, and modify the error accuracy according to the independent random error variable $e_i$ (Iliadis, Papastavrou, & Lefakis, 2002; Shieh et al., 2009). The hypothesis of the multiple linear model is:

$$
Y_1 = B_0 + B_1 x_{11} + B_2 x_{12} + \cdots + B_k x_{1k} + e_1 \\
Y_2 = B_0 + B_1 x_{21} + B_2 x_{22} + \cdots + B_k x_{2k} + e_2 \\
\vdots \\
Y_n = B_0 + B_1 x_{n1} + B_2 x_{n2} + \cdots + B_k x_{nk} + e_n
$$

The matrix is as follows:

$$
\begin{bmatrix}
Y_1 \\
Y_2 \\
\vdots \\
Y_n
\end{bmatrix} =
\begin{bmatrix}
1 & x_{11} & x_{12} & \cdots & x_{1k} \\
1 & x_{21} & x_{22} & \cdots & x_{2k} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & x_{n1} & x_{n2} & \cdots & x_{nk}
\end{bmatrix}
\begin{bmatrix}
B_0 \\
B_1 \\
\vdots \\
B_n
\end{bmatrix} +
\begin{bmatrix}
e_1 \\
e_2 \\
\vdots \\
e_n
\end{bmatrix}
$$

The regular equation sets are transformed by least binary multiplication and partial differential:
The coefficient matrix in the above regular equation sets is a symmetric matrix. Let $A$ be the coefficient matrix, and let $B$ be the right-end constant term matrix. The matrix equations are given below:

\[ Ab = B \]

where $b$ represents the unknown number in the regular equation, and $b = [b_0, b_1, \ldots, b_k]$. The reversed matrix of $(x' \cdot x)$ exists when the coefficient matrix has full rank. Thus, coefficient $b$ can be given by Formula (2):

\[ b = (x' \cdot x)^{-1}x' \cdot Y \]

Then, the multiple linear model, Formula (3), can be obtained (by OR with OR using) Formula (2):

\[ Y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + \cdots + b_k x_k \]
4.2. Multivariate analysis method

The multivariate analysis method focuses on analyzing the disaster factor variability and estimating the variance of potential debris-flow torrents. The calculated variance values can be used to evaluate the influential ratio, and each danger factor is given an evaluation value. Then, the disequilibrium index, $D_t$, of each factor can also be calculated by applying a high-adaptive statistic assessment model (Lin, 1994). The disequilibrium index $D_t$ represents the relative level, which takes a value between 1 and 10, where a value of 10 implies a high probability of debris-flow occurrence. The definition of $D_t$ is described below:

$$D_t = d_1^{w_1} \times d_2^{w_2} \times d_3^{w_3} \times \ldots \times d_n^{w_n}$$

The disequilibrium index of potential danger factors; $d_1, \ldots, d_n$ evaluated value of variable influential factors; $w_1, \ldots, w_n$ weighted value of each variance factor.

The classified progression evaluated value of each variance factor is calculated as follows:

$$d_n = a\left(\frac{X_i - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}}\right) + 1$$

$a$ specific constant value; $d_n$ evaluated variance factor; $X_i$ classified percentage of debris-flow occurrence for each factor; $X_{\text{max}}$ and $X_{\text{min}}$ maximum and minimum of $X_i$, respectively.

The coefficient of variation for each factor is (calculated OR computed) as follows:

$$v = \frac{\sigma}{\bar{X}} \times 100\%$$

$v$ and $\sigma$ variation coefficient and standard deviation, respectively; $\bar{X}$ mean value of classified destruction percentage for each factor.

Finally, the variation coefficient of each factor is divided by the sum total of the coefficient of variation of all factors, (yielding OR

---

**Fig. 6.** The flow chart of back-propagation network algorithm.
resulting in) the effective weighted value of the specific factor. The definition of \( W_i \) is given below:

\[
W_i = \frac{v_i}{v_1 + v_2 + \ldots + v_n}
\]  

(7)

\( W_i \) effective weighted value of the specific factor; \( v_i \) coefficient of variation of each factor.

4.3. Back-propagation network algorithm

Neural networking is an information management technology derived from research on the human brain and neural system. By means of information output and input a system model will be completed to make inferences, predictions, decisions and diagnoses. A neural network is a non-linear statistical technique (Skapura, 1995). This research applied the back-propagation network algorithm to analyze the potential debris-flow hazard level. This algorithm is a typical supervised learning network (Lin & Chang, 2003), which learns the internal reflection regulations between the input and output. The regulations are expressed with the connected weighted value of each network processing unit. To analyze any new cases, users simply need to key in input values or independent variables to obtain the inferential related output values quickly. Therefore, this study used back-propagation network algorithm to estimate the hazardous levels of potential debris-flow hazard torrents and the probability of debris-flow occurrence.

4.3.1. Data pre-processing analysis

Before the data from each influential factor are input into the back-propagation network system, they need to be pre-processed to map the source input variables to the same interval. Although the input processing units of the supervised back-propagation network accept any variable values, the significance of the small-range variables can not be expressed, if different processing units accept too large variable ranges, causing large-ranged variables to take control of the entire network learning process and affect the network learning process. Therefore, this study adopts probability mapping to transform the input variable range, as described below:

1. Define the mean value of a statistical variable as \( \mu \), and the standard deviation as \( \sigma \).
2. Define the demanded minimum as \( D_{\text{min}} \) and the demanded maximum as \( D_{\text{max}} \).
3. Define the data format Formula (8) as follows:

\[
X_{\text{new}} = \left( \frac{1}{2} \right) \times \left( \frac{X_{\text{old}}(\mu - k\sigma)}{k\sigma} \right) \times (D_{\text{max}} - D_{\text{min}}) + D_{\text{min}}
\]  

Here \( X \) is the independent variable; \( \mu \) refers to the mean value; \( \sigma \) means the standard deviation; \( D_{\text{min}} \) and \( D_{\text{max}} \) stand for the minimum and maximum, respectively, and \( k \) is the transition coefficient.

If \( k = 3.29 \), then 99.9% data are mapped into \([D_{\text{min}}, D_{\text{max}}]\) interval, while \( k = 2.81, 2.58, 1.96, 1.65 \) and 1.28 map the variable

![Image](image_url)

Fig. 7. Error of the operation of multiple linear regression model.
values 99.5%, 99%, 95%, 90% and 80%, respectively, onto the 
\[ D_{\text{min}}, D_{\text{max}} \] interval (Skapura, 1995).

Data pre-processing can be used to analyze every factor, and to measure the mean value and the standard deviation of each example variable. Then, the probability mapping formula can be applied to convert the variable ranges for each sample.

4.3.2. The algorithmic flow of a back-propagation network

The operation and algorithmic process of a back-propagation network has the following steps.

1. Set up a parameter.
2. Set up the weighted matrices, \( W_{xh} \) and \( W_{hy} \), and the initial values of partial weighted vector, \( \theta_h \) and \( \theta_y \), as uniformly random numbers.
3. Calculate the output quantity of the hidden layer.
4. Determine the tolerant difference quantity between the output layer and the hidden layer.
5. Calculate the difference quantity, \( \delta \), between the output layer and the hidden layer.
6. Determine whether the difference quantity between the output layer and the hidden layer is larger than the tolerant difference. If the difference quantity is smaller than the tolerant difference quantity, then the regression model is optimal.
7. If the difference quantity is larger than the tolerant difference quantity, then calculate the weighted matrices and the corrections of partial weighted values in the output and the hidden layers.
8. Revise the weighted matrices and the partial weighted values in the output layer and the hidden layer, and repeat Steps 3–8 until the difference quantity lies within the range of the tolerant difference quantity. Then, compare the correlation of sensitivity correction to determine the optimal regression model. The flow chart of the above algorithm is illustrated in Fig. 6.

4.3.3. The sensitivity analysis of the network model

In a back-propagation network, if the relation of input and output units is homogeneous, then users can input the mapped output unit sensitivity of each unit respectively onto the linked weighted analysis network. The positive or negative value of sensitivity, increases in proportion to the positive or negative correlation between the input and output units. This network learning process uses the gradient steepest descent method to minimize the energy function; that is, when a training example is input, the network needs to adjust the weighted value within a small range. The sensitivity of the adjusted range and error function needs to be directly proportional to the weighted value, which means the error function and the partial differential value are in direct proportion (Skapura, 1995). The formula is described below:

\[
\Delta W_{ij} = -\eta \frac{\partial E}{\partial W_{ij}}
\]

In the formula \( W_{ij} \) represents the joint weighted value between the \( i \)th and \( j \)th units, and \( \eta \) denotes the error adjusted range used by the learning velocity to control the gradient steepest descent method.

5. Verification and analysis of the debris-flow prediction (scheme OR system OR model)

This section verifies and compares the three analytical models, multiple linear regression, multivariate analysis and back-propagation network, and then presents the optimal debris-flow prediction (scheme OR system OR model).
5.1. Verification and analysis of hazard prediction (scheme OR system OR model)

5.1.1. Verification and analysis of multiple linear regression

This analytical method is intended to find the weighted coefficient of each independent variable \( X \) based on the expected value of variable \( Y \). In the expected value matrix, \( Y = 1 \) signifies debris-flow occurrence, while \( Y = 0 \) signifies non-occurrence. Each factor is treated as an independent variable matrix for the sum of the expected value matrixes. Then, regression analysis and verification were performed through a MATLAB computer program. The analysis is conducted in a 95% confidence interval, and the analytical results are graded by using statistical techniques based on upper and lower threshold intervals. The regression model is defined in the following Formula (10):

\[
Y = 1.641 - 0.1571 IR + 0.299 \text{NDVI} + 0.0335 R_e + 0.0262 E_r \\
+ 0.0153 E_s + 0.0289 E_e + 0.0109 E_t + 0.0029 E_g
\]  

\( Y \) = the hazardous degree of the debris-flow; \( IR \) = effective IR mean; \( R_e \) = effective accumulated precipitation; \( E_r \) = effective rainfall intensity; \( E_s \) = effective channel slope; \( E_e \) = rocks in the effective watershed; \( \text{NDVI} \) = effective NDVI mean (NDVI is the vegetation index).

5.1.2. Verification and analysis of multivariate analysis

The primary steps of this analysis method are taken to analyze the variability of the debris-flow danger factors and calculate the potential debris-flow variance values. The variation values are sequenced to indicate the influential weighted ratio of each danger factor, which is given a weighted evaluation value. Then, the slope disequilibrium index \( D_t \) is adopted to calculate the high-adaptive statistic assessment scheme. The formula is defined as below:

\[
D_t = \frac{0.15 \times \text{NDVI} + 0.16 \times E_t + 0.17 \times E_s + 0.06 \times E_g}{0.16 \times E_t + 0.06 \times E_s + 0.09 \times E_g}
\]  

\( D_t \) = disequilibrium index; \( IR \) = effective IR mean; \( R_e \) = effective accumulated precipitation; \( E_r \) = effective rainfall intensity; \( E_s \) = effective channel slope; \( E_e \) = rocks in the effective watershed; \( \text{NDVI} \) = effective NDVI mean.

The graded evaluation values are taken as the independent variable matrix, which is used in Formula (11). Therefore, the error

---

**Fig. 7** shows the distribution of the factors from the 181 potential debris-flow hazardous torrents operating in the multiple linear regression model. As shown in the distribution, the multiple linear regression model had an error probability of about 25–35%.

**Fig. 9.** Optimal model training processed by PCNeuron.
The probability of this model is tested and verified by means of a graded observation based on upper and lower interval statistical analysis. As illustrated in Fig. 8, the error probability is about 3.39–6.07%, which is lower than that in the multiple linear regression model.

5.1.3. Verification and analysis of back-propagation network

A back-propagation network system is used to output and input data and to construct the debris-flow prediction model. In the expected-value matrix, a value of 1 signifies occurrence of debris-flow, and a value of 0 signifies non-occurrence. The resulting of danger factor classification scheme is taken as the independent variable matrix. Then, the PCNeuron software package is applied to train the optimal numbers of the network hidden layers and the overall network simulation architecture. In this method, the learning initial velocity is 0.5, the reduction coefficient is 0.95, and the lower threshold value is 0.05. The result of the training reveals that the “error oscillation (over-learning) phenomenon” is relatively low, and that when the circulating learning numbers reach 30,000 times, the training can achieve almost complete convergence. Fig. 9 illustrates the optimal training system. The training of the 20 developed network hidden layers reveals that the error oscillation in hidden layer 1 is the highest, that in hidden layer 3 is reduced, and that in hidden layer 5 is the lowest. Hence, the optimal number of hidden layers is five. Fig. 10 shows the operated training results and the overall network architecture. This infrastructure adopts the optimal network model in which eight input layers produce five hidden layers and one output layer. Fig. 11 displays the distribution of verified results and the error probability of the network system. The back-propagation network
is verified via a MATLAB program, and analyzed through a 95% confidence interval. The back-propagation network model is verified means of a disorder matrix and the classified observation of the upper and lower thresholds. The analytical result indicates that the error probability of back-propagation network prediction model is about 1.16–1.87%, which is the best among the three prediction models.

5.2. Analyses of regression results and the comparison of model accuracy

5.2.1. Analyses of regression results

The danger factors are ranked in Table 1.

(i) When these three regression models are analyzed with factor correlation, the potential debris-flow occurrence rate has a negative correlation. Smaller NDVI and IR mean values of the effective watershed imply a larger collapsed area. A negative correlation may also mean the great volume of soil deposited upstream and a high probability of potential debris-flow occurrence. This result is corresponding to the fact. The vegetative conditions and the IR radiation value can also be applied to assess the collapse and debris deposition. The weighted rankings of NDVI and IR factors are placed first and second in the multiple linear regression and back-propagation network. In other word, the collapsed area within effective watershed significantly affects the debris-flow occurrence. The other danger factors reveal a positive correlation with the debris-flow occurrence. The selected danger factors are ranked 2nd, 3rd, 4th and 5th, respectively, which suggests that longer rainfall duration or greater intensity leads to higher probability of debris-flow occurrence.

(ii) The factor of the slope of the effective channel (Es) ranks 1st in Multivariate Analysis, while the models in the other two analyses rank 4th and 5th, respectively. The varieties of rocks denoted by Eg are the basic components of the slope. In this study, the lithological properties are divided into sedimentary and metamorphic rocks known as the values 1 and 2, respectively. The sorted potential debris-flow hazardous torrents in Nantou County consist of 88 torrents of sedimentary rocks and 93 torrents of metamorphic rocks. After

Table 1
Weighted ranking comparison of the three regression model factors.

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Linear regression</th>
<th>Multivariate analysis</th>
<th>Back-propagation network</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Factor</td>
<td>Weighted coefficient</td>
<td>Factor</td>
</tr>
<tr>
<td>1</td>
<td>IR</td>
<td>0.2991</td>
<td>Es</td>
</tr>
<tr>
<td>2</td>
<td>NDVI</td>
<td>0.1571</td>
<td>Re</td>
</tr>
<tr>
<td>3</td>
<td>Re</td>
<td>0.0335</td>
<td>Er</td>
</tr>
<tr>
<td>4</td>
<td>Es</td>
<td>0.0289</td>
<td>NDVI</td>
</tr>
<tr>
<td>5</td>
<td>Er</td>
<td>0.0262</td>
<td>IR</td>
</tr>
<tr>
<td>6</td>
<td>El</td>
<td>0.0153</td>
<td>Eg</td>
</tr>
<tr>
<td>7</td>
<td>Ea</td>
<td>0.0109</td>
<td>El</td>
</tr>
<tr>
<td>8</td>
<td>Eg</td>
<td>0.0029</td>
<td>Ea</td>
</tr>
<tr>
<td>Error probability</td>
<td>28.18–35%</td>
<td>3.39–6.07%</td>
<td>1.16–1.87%</td>
</tr>
</tbody>
</table>

Table 2
Accuracy comparison of the regression models.

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Regression analysis model</th>
<th>Overall error probability (181 data)</th>
<th>Verified error probability (100 random data)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Minimal (%)</td>
<td>Average (%)</td>
</tr>
<tr>
<td>1</td>
<td>Back-propagation network</td>
<td>1.16</td>
<td>1.43</td>
</tr>
<tr>
<td>2</td>
<td>Multivariate analysis</td>
<td>3.39</td>
<td>5.18</td>
</tr>
<tr>
<td>3</td>
<td>Linear regression</td>
<td>25.18</td>
<td>32.77</td>
</tr>
</tbody>
</table>

Fig. 12. Optimal numerical model flowchart of back-propagation network.
Table 3
Classification of network output values and debris-flow hazardous degrees.

<table>
<thead>
<tr>
<th>Network output value</th>
<th>Total</th>
<th>Occurring numbers</th>
<th>Occurring rate (%)</th>
<th>Potential hazardous degree</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Output value &lt; 0.1</td>
<td>64</td>
<td>0</td>
<td>0</td>
<td>Low risk degree</td>
<td>Ordinary (green light)</td>
</tr>
<tr>
<td>0.1 &lt; Output value &gt; 0.8</td>
<td>7</td>
<td>4</td>
<td>57</td>
<td>Middle risk degree</td>
<td>Near warning (yellow light)</td>
</tr>
<tr>
<td>Output value &gt; 0.8</td>
<td>110</td>
<td>109</td>
<td>99</td>
<td>High risk degree</td>
<td>Warning (red light)</td>
</tr>
</tbody>
</table>

Fig. 13. Distribution of Back-Propagation analysis output values.

Fig. 14. Classification of network output values and debris-flow hazard levels.
Typhoon Toraji’s invasion, debris flows occurred in 40 torrents of sedimentary rocks and 72 torrents of metamorphic rocks because metamorphic rocks are weaker than sedimentary ones. The occurrence rate of debris flows in metamorphic rocks is 77.42%, which is obviously higher than the rate of 44.94% in sedimentary rocks. Additionally, the positive correlation result in the regression analysis shows a high occurrence rate in metamorphic rocks, which corresponds to observations. Additionally, the factors of effective watershed and stream length correlate positively with the debris-flow occurrence. A larger effective watershed implies the stream contains more water, increasing the probability of debris-flow occurrence. A longer effective channel length increases the distance for debris-flow movements. These two inferences both correspond to empirical observations.

5.2.2. The accuracy comparison of the regression models

Table 2 lists the overall error probability data of the three regression models, and the verified error probability of 100 randomly selected data. The data reveal that back-propagation analysis is more accurate than the other two methods. Hence, this backpropagation analysis was adopted as the optimal numerical system for the debris-flow real-time prediction model.

5.3. The optimal numerical debris-flow (systems OR schemes OR models) and the classification of potential hazard levels

For the optimal analysis result and the minimal error probability, this study takes the output value of back-propagation model as the potential hazard levels of the potential debris-flow hazardous torrents in Nantou. Furthermore, the potential hazard levels are classified to distinguish among levels of debris-flow risk.

Fig. 15. Real-time Mobile Debris Flow Disaster Forecast System (RM(DF)²) architecture.
Fig. 16. PDA function display in RM(DF)^2.

Fig. 17. PC function display in RM(DF)^2.
the eight classified values into the optimal model flow, transfers the classified values via the linear system, and then calculates the output volume of the hidden layer by incorporating the weighted matrix and partial weighted vector into the hidden layer non-linear system. Besides, the expected output volume is inferred from the output layer non-linear system, and then the expected output volume is finally transferred into the optimal network estimate value.
Because the back-propagation analysis model does not predict the stream hazard level, this study re-classifies all output values, leading to the result shown in Fig. 13. The potential hazardous degrees were divided into low, intermediate and high levels, because having more than three levels may lower the efficiency of hazard classification. That is, when too many levels are applied,
The debris-flow occurrence rates of the middle levels are too close to their neighboring levels to categorize levels efficiently. Fig. 13 and Table 3 indicate that back-propagation analysis has the best effect and the lowest error probability among the tested models. Fig. 14 shows the relationship between the network output values and the classification of debris-flow hazard levels. The output values are classified into corresponding intervals. The expected interval division values are set to 0.8 and 0.1. An output value lower than 0.1 indicates a low probability of debris-flow occurrence. An output value in the range 0.1–0.8 is classified as a middling probability of debris-flow occurrence, while the output value above 0.8 is classified as a high probability.

6. System implementation

Fig. 15 shows the Real-time Mobile Debris Flow Disaster Forecast System (RM(DF)$^2$) architecture, which consists of the mobile user, front-end application server and rear-end decision support system. RM(DF)$^2$ users can utilize a Tablet PC, PDA, Notebook or desktop computer to obtain access to remote disaster prevention data (Meggers, Park, Fasbeder, & Kreller, 1998; Mohan, Smith, & Li, 1999). The data is simplified by the environmental cognition agent in the application server and is displayed in a virtual reality format. The user performs debris-flow field work by using a mobile device to input relevant information for decision inference and analysis (Hare, 2002; Pang & Poon, 2001). The required software and hardware devices are shown in Tables 4 and 5.

Figs. 16 and 17 illustrate the functions of information input from the disaster area and the debris-flow multimedia information reception. If the mobile communication system is destroyed, then users can make inferences from the regulation inference mechanism installed in the portable devices. In this case, a PDA is easy to carry and suitable for manipulating data.

The front-end application server focuses on providing related multimedia application and environmental cognition agent services (Huhns & Singh, 1997; Kung & Ku, 2003; Lo, Chen, Cheng, & Kung, 2011; Nwana & Ndumu, 1997). The application server performs information classification and cleanup, and cooperates with WEB/WAP functions to proceed with information transmission and enquiry as well as VR topography simulation. In Fig. 18 the rainfall agent is shown by using the WEB interface. Fig. 19 illustrates the rainfall data received by the rainfall agent on the PDA. Fig. 20 displays the on-the-spot simulations of the VR digital topography or contour map on the PDA. Fig. 21 shows the application of GPS technology on roads in Nantou County. The GPS files are transmitted to GIS layers for analysis, and converted to high accuracy satellite images such as the Quickbird satellite images with resolution 67 cm or aerial photos to improve the positioning function. Finally, the satellite images and aerial photos are analyzed by means of the RS system.

The rear-end decision support system (DSS) comprises a modeling base, database and dialogue unit (Efrain & Jay, 1998). The modeling base includes multiple linear regression, multivariate analysis and back-propagation network. These models are analyzed and estimated by using eight numerical factors: effective watershed, the effective channel length, effective channel slope, rocks in the effective watershed, collapsed area in the effective watershed, effective accumulated precipitation, effective rainfall intensity and vegetation index. When the inferred value is evaluated as dangerous, the DSS sends the danger signal back to the mobile user. Finally, the system images are displayed in Fig. 22.

7. Conclusions

This study proposed multiple linear regression, multivariate analysis and back-propagation network to form debris-flow disaster...
prediction models. The effective watershed, effective channel length, effective channel slope, the rocks in the effective watershed, collapsed area in the effective watershed, effective accumulated precipitation, effective rainfall intensity and vegetation index were defined as the influential factors to estimate the debris-flow occurrence. In addition, a Real-time Mobile Debris Flow Disaster Forecast System (RM(DF)²) was implemented to verify the feasibility and effectiveness of the designed model.

The historical data of the 181 potential debris-flow dangerous torrents in Nantou County were taken as example cases. When analyzed with multiple linear regression, the average threshold was set to 5.0927, and the upper and lower thresholds were set to 5.4463 and 4.7392, respectively. The analytical result reveals that the maximal error rate reached 35%, meaning that multiple linear regression model cannot effectively distinguish among the data of all the potential torrents in most situations, and therefore cannot precisely predict the occurrence of debris-flow. When analyzed with multivariate analysis, the average threshold was set to 5.256, and the upper and lower thresholds were set to 5.5757 and 4.9583, respectively. The analytical result shows that the maximal error rate dropped to 6.07%. Therefore, the assessment accuracy of multivariate analysis is much higher than that of multiple linear regression. Obviously, multivariate analysis has better distinguishability than multiple linear regression. When analyzed by 8–5–1 back-propagation network, the average threshold was set to 0.6174, and the upper and lower thresholds were set 0.6757 and 0.559, respectively. The maximal error rate was only 1.87%, indicating that this analysis model can evaluate the 181 cases almost distinctly. Among these three analytical models, Back-Propagation Network has the best prediction accuracy. However, multiple linear regression has the lowest Time Complexity, the slightest burden on the system operation.

RM(DF)² adopts a three-tier architecture. The front-end application server takes charge of the transmission. The rear-end decision support system infers the possibility of debris-flow disasters. The mobile user can obtain the required information promptly and precisely, and thus obtain the inferred result. However, if the water volume of the potential debris-flow dangerous torrent increases, the rear-end decision support system might fail to infer the danger of the streams. Therefore, a future strategy for improvement is to convert the decision support system into a distributed system. Additionally, future research will further emphasize gathering new potential torrent data and used 3G mobile communicative networks to improve self-examination of danger areas by cell phone users. Hopefully, by means of such a system, early disaster warnings can be in time to the victims so that they can evacuate from the disaster area safely.

Acknowledgment

The research is supported by the National Science Council of Taiwan under the Grant Nos. NSC 100-2218-E-020-001 and NSC 100-2218-E-020-002.

References


Chen, M. T. (2002). A study on risky grade and warning mode for potential debris flow in north region of Taiwan – Application of neural network and fuzzy theory. Taiwan: Department of Civil Engineering, National Taiwan University.


