Integration of dynamic rainfall data with environmental factors to forecast debris flow using an improved GMDH model

Hui Zhang a, b, Xiangnan Liu a, *, Erli Cai a, Gang Huang c, Chao Ding a

a School of Information Engineering, China University of Geosciences, Beijing 100083, China
b School of Computer and Information Engineering, Tianjin Institute of Urban Construction, Tianjin 300384, China
c Information center, Unit 91917, Beijing 100841, China

A R T I C L E   I N F O

Article history:
Received 6 December 2012
Received in revised form
30 January 2013
Accepted 5 February 2013
Available online 17 February 2013

Keywords:
GMDH model
Dynamic rainfall data
Environmental factors
Simulated annealing algorithm
Genetic algorithm

A B S T R A C T

The objective of this study was to apply an improved Group Method of Data Handling (GMDH) network model for prediction of debris flow by integrating dynamic rainfall data and environmental factors. The rainfall data were collected from weather information, and the environmental data were extracted from RS, GIS, drilling data, and geophysical data. The input variables used in the SAGA-GMDH model were derived from six variables acquired by Kernel Linear Discriminant Analysis (KLDA). The results showed that the GMDH for prediction of debris flow performed well using the training, validation, and testing sets ($R^2$ above 0.80 and ARE below 3.54%). The SAGA-GMDH model was subsequently compared with a back-propagation (BP) neural network model and adaptive network fuzzy interference system (ANFIS). The accuracies of the SAGA-GMDH model prediction were slightly better than those of other two models, which demonstrated that the SAGA-GMDH model was more suitable for prediction of debris flow.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Rainfall-triggered debris flow is a serious type of natural hazard that occurs in certain places in the world. In the past few decades, rainfall-triggered debris flow has caused serious loss of human lives and extensive damage to property (Larsen and Simon, 1993; Crosta, 2004; DeGraff et al., 2011). A number of methods for forecasting of rainfall-triggered debris flow have been demonstrated (Tsaparas et al. 2002; Santoso et al. 2011). The general approaches can be classified into two types: the field-based qualitative approach or the data-driven quantitative approach (Nandi and Shakoor, 2010; Xu et al., 2012; Kung et al., 2012). Traditionally, the assessment of a debris flow disaster has been carried out using physical model experiments (Cui, 1991) and long-term outdoor observation. With the development of RS and GIS, the environmental data are easy to obtain. Because remote sensing has many advantages over physical model experiments or long-term outdoor observations, i.e., lower cost, faster data acquisition, better spatial and temporal continuity, great progress has been facilitated by remote sensing (Ou et al., 2006; Vahidnia et al., 2010; Oh and Pradhan, 2011). However, the existing work on prediction of debris flow only considers such rainfall factors as accumulated rainfall capacity, rainfall intensity, and rainfall capacity. In addition to the precipitation, the geological environmental factors are also important conditions in the process of debris flow forecasting. If the influence of the geological environmental factors is included in the debris flow prediction model, the forecast result will result in higher objectivity and credibility. Through a combination of geological environmental factors and rainfall factors, an accurate and valid forecast for debris flow in single gully can be achieved. Therefore, it is necessary to develop a significant method that couples the environmental factors with rainfall factors to predict the debris flow in a single gully.

Due to the uncertainties and complexities of the factors involved in the causes of debris flow, it is generally difficult to quantitatively analyze these influences and to predict the occurrence of debris flow. System identification techniques are applied in many fields in order to model and predict the behaviors of unknown and/or very complex systems based on given input-output data. Soft computing methods, which concern computation in an imprecise environment, have gained significant attention. Among these methodologies, the Group Method of Data Handling (GMDH) is an effective method for establishing a mathematical model of a complex system using a heuristic self-organization approach (Ivakhnenko, 1976). One characteristic of the GMDH is that it provides an automatic modeling mechanism (Zhu et al., 2012). This automatic modeling mechanism has been
successfully applied to build Bayesian networks (Xiao et al., 2009) and Mamdani-type fuzzy models (Mueller and Lemke, 2000), among others. Another desirable characteristic of the GMDH is its immunity to noise. It is well known that when the data contain noise, the most dangerous effect is over-fitting (Schittenkopf et al., 1997), which produces that models tend to be excessively complex and characterized by poor generalization. A detailed discussion of noise immunity in the GMDH can be found in (Madala and Ivakhnenko, 1994). The objective of this research is to apply the GMDH model to predict the debris flow in a single gully using a combination of environmental factors and rainfall factors. This is to say, a kind of system identification technique is applied to predict debris flow.

2. Materials and methods

2.1. Study areas

This research focuses on the areas of Xiu Yan county, the town of An Shan, and the Liao Ning province in China (Fig. 1) due to the frequent occurrence of debris flow in this region, which is approximately located between latitudes 40°00’N to 40°39’N and longitudes 122°52’E to 123°46’E. The Xiu Yan county covers an area of 4502 km², and the temperature of the study area ranges from −30.9° to 37.3°. The terrain consists of hills, low mountains, and middle-sized mountains.

The average annual precipitation is 896 mm, with a minimum precipitation of 172.2 mm and a maximum of 1451.3 mm. The rainy season occurs primarily from June to September, with precipitation in this period accounting for 73.5–80.2% of the yearly amount. The precipitation distribution decreases from the southeast to the northwest.

The debris flows in the Xiu Yan country represent natural degradation processes that are largely triggered by heavy rain due to either a single heavy thunderstorm/rain or successive days of moderate rain (especially during the rainy season), which causes flash flooding that leads to failure of the rock surfaces along fractures, joints, and cleavage planes (Jin, 2011).

2.2. Data collection

The formation of debris flow is a complicated process that requires favorable terrain conditions, source conditions and hydrodynamic conditions. Precipitation is the most important triggering factor for inducement of a debris flow disaster. The debris flow is the joint result of antecedent rainfall and short-term intense rainfall (Tang et al., 1994; Chen, 1985). Several factors were selected as the environmental factors for debris flow: the drainage area (DA), the basin relative height difference (BRHD), supply length ratio (SLR), brook longitudinal slope (BLS), vegetation coverage (VC), ditch shore hillside slope (DSHS), and loose materials reserves along the ditch (LMRD). The 5-day accumulative rainfall (FAR), the maximum hours of rainfall intensity (MHRI), and the daily rainfall (DR) were selected as components of the rainfall influence index of the debris flow warning model. The rainfall intensity is an important initiation factor for debris flow, and the daily rainfall data can be provided by the meteorological observatory 24-h rainfall forecast; the greater the rainfall, the greater the likelihood of debris flow occurrence. The long-term continuous rainfall significantly increases the void water pressure in deep-slope soil, and the soil produces sliding fracture behavior, which is an important contribution to the occurrence of debris flow. Therefore, the accumulative rainfall amount responds to the saturated liquefaction degree of the soil, to a certain extent, according to the degree of influence of the antecedent precipitation. When debris flow occurs in the Xiu Yan area, the 5-day accumulative rainfall acts as the effect index of the anteecedent precipitation.

The largest quantity of debris flow that mobilizes at one time in a single gully (BQDR) directly determines the destructiveness of the debris flow, and thus this quantity was viewed as the dominant assessment index of the debris flow risk. The abbreviations and units of the predictors of debris flow are shown in Table 1.

2.3. Methodology

The most important tasks in this method are the establishment of a data retrieval model and selection of the input variables for
the model. The data process and a conceptual diagram of the SAGA-GMDH are illustrated in Fig. 2. The main processes include data acquisition, data processing, modeling with the SAGA-GMDH, and model validation. A SAGA-GMDH model was developed to predict the BQDR. This model consisted of an input layer, an output layer, and several hidden layers, with the hidden layers applying evolutionary algorithms (combining simulated annealing and genetic algorithms) to identify the coefficients of the polynomial. The environmental factors and rainfall data were taken into consideration as input variables for the model. The model was subsequently used to predict the BQDR. In this research, the environmental factors were selected to represent the terrain conditions whereas the rainfall factors were important as induced factors for triggering of debris flow.

3. The model of forecasting debris flow based on the SAGA-GMDH

3.1. Data normalization

To avoid the impact of certain small or large data on the network, the input variables in this model were normalized based on their possible ranges using the equation

$$X_{\text{norm}} = \frac{(X_i - X_{\text{min}})}{(X_{\text{max}} - X_{\text{min}})}$$

where $X_i$, $X_{\text{min}}$, $X_{\text{max}}$, and $X_{\text{norm}}$ are the primitive input variable, the minimum input variable, the maximum input variable and its normalized value, respectively. The output from the SAGA-GMDH model is an indexed value that corresponds to the input variable.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Abbreviations</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>The largest quantity of debris flow that mobilizes at one time in a single gully</td>
<td>BQDR</td>
<td>$10^6$ m$^3$</td>
</tr>
<tr>
<td>Drainage area</td>
<td>DA</td>
<td>Km$^2$</td>
</tr>
<tr>
<td>Basin relative height difference</td>
<td>BRHD</td>
<td>m</td>
</tr>
<tr>
<td>Supply length ratio</td>
<td>SLR</td>
<td>%</td>
</tr>
<tr>
<td>Brook longitudinal slope</td>
<td>BLS</td>
<td>$%$</td>
</tr>
<tr>
<td>Vegetation coverage</td>
<td>VC</td>
<td>$%$</td>
</tr>
<tr>
<td>Ditch shore hillside slope</td>
<td>DSHS</td>
<td>$\circ$</td>
</tr>
<tr>
<td>Loose materials reserves along the ditch</td>
<td>LMRD</td>
<td>m$^2$</td>
</tr>
<tr>
<td>The 5-days accumulative rainfall</td>
<td>FAR</td>
<td>mm</td>
</tr>
<tr>
<td>The maximum hours of rainfall intensity</td>
<td>MHRI</td>
<td>mm</td>
</tr>
<tr>
<td>The day rainfall</td>
<td>DR</td>
<td>mm</td>
</tr>
</tbody>
</table>
To obtain the real predicted value, the indexed output value must be de-normalized according to the following equation:

$$Y_i = Y_{\text{min}} + Y_{\text{norm}}(Y_{\text{max}} - Y_{\text{min}})$$  \hspace{1cm} (2)

where $Y_{\text{min}}, Y_{\text{max}}, Y_{\text{norm}}, Y_i$ are the minimum and the maximum output value, the indexed output value, and the real predicted value, respectively.

### 3.2. Reduction of the dimensions of the input variables

Dimensionality reduction is a crucial step for pattern recognition tasks, and identification of a suitable low-dimensional subspace has an important effect on the recognition performance. One of the most important steps involves selection of an appropriate set of inputs prior to the SAGA-GMDH model development process. Currently, the PCA, kernel principal component analysis (KPCA), linear discriminant analysis (LDA) and kernel linear discriminant analysis (KLDA) are the most popular feature extraction and dimensionality reduction methods and have aroused considerable interest in the fields of pattern recognition and machine learning (Muller et al., 2001; Schokopf et al., 1998). The purpose of such PCA-based methods as KPCA is to retain as much information as possible in terms of variance and find the directions that have the minimal reconstruction error (Duda et al., 2001). The purpose of such LDA-based methods as KLDA is to optimize the low-dimensional representation of the data and focus on the most discriminating feature extraction (Fukunaga, 1990). Due to the ability to acquire the most discriminating information contained in the data set, the LDA-based methods, and especially KLDA and its variants, have been widely used in many real-world applications (Baudat and Anouar, 2000; Yang et al., 2004; Howland et al., 2006). Moreover, the KLDA is a type of nonlinear dimensionality reduction algorithm. In this study, the selection of the input variables is based on the KLDA.

### 3.3. Model building

#### 3.3.1. Model structure

In this paper, an improved GMDH network (SAGA-GMDH) model was constructed to assess the degree of danger of debris flow in a gully. The characteristics of the SAGA-GMDH can be summarized as follows: (1) automatic modeling mechanism, (2) noise-immunity, and (3) introduction of the advantages of the simulated annealing and genetic algorithm into the GMDH network. In this paper, the simulated annealing algorithm and the genetic algorithm are combined to identify and optimize the binary quadratic polynomial coefficients. The basic structure of the SAGA-GMDH is also shown in Fig. 3.

The layerd operation of the GMDH is as follows:

The input layer – Each neuron in this layer represents an input variable, $x_{i,j}=1, 2, \ldots, r$, which are used to provide the values of the independent variables from the learning set to the subsequent layers of the network. The hidden layer – In the construction of a hidden layer, an initial population of units is generated, and each unit corresponds to Ivakhnenko’s polynomial form:

$$X_{n+1,k} = a + bX_{n,j} + cX_{n,j} + dX_{n,j}^2 + eX_{n,j}^2 + fX_{n,j}X_{n,j}$$  \hspace{1cm} (3)

In Eq. (3), $a, b, c, d, e, f$ are the coefficients of partial descriptions (PD) in the fully binary quadratic polynomial. In the structure of the GMDH network, the inputs of the processing unit are taken from two outputs of the different units. The output variable of the processing unit is the quadratic polynomial of the input variables, and the input variables for the entire GMDH network going forward are progressive layers; thus, the number of the polynomial increases to second-order, and eventually, the entire network is able to form a $2k$-order polynomial $(k$ is the number of layers of the GMDH). This process shows that the GMDH network is able to express the structural model of the discrete polynomial series $(k-G)$ polynomial.

The output layer – Each neuron in this layer represents an output variable as a summation of the incoming signals. The SAGA-GMDH model considers a unique output variable, the BQDR. The output of a neuron in this layer can be described as the optimal Ivakhnenko polynomial:

$$y = a_0 + \sum_{i} a_i x_i + \sum_{i} \sum_{j} a_{ij} x_i x_j + \sum_{i} \sum_{j} \sum_{k} a_{ijk} x_i x_j x_k + \cdots$$  \hspace{1cm} (4)

where $x_i$ is the input variable, $y$ is the estimation equation and $a_0, a_i, a_{ij}, a_{ijk}, \ldots (i, j, k = 1, 2, \ldots) m$ are the coefficients.

#### 3.3.2. Simulated annealing and genetic algorithm for identification of the coefficients of the polynomial

In Eq. (3), the least squares method is commonly applied to identify the coefficients $a, b, c, d, e, f$. However, this approach is prone to becoming trapped in local minima, performs poorly in identifying nonlinear systems, and is not suitable for severe interference (Chen, 2005). Therefore, the effect of the traditional GMDH network in the application is not ideal. Many scholars have attempted to improve the identification method of the traditional GMDH network (Oh et al., 2003; Hernandez and Herrera, 2012; Khalkhali and Safikhani, 2012). The genetic algorithm is a global optimum algorithm that was successfully used in system recognition and parameter optimization (Grefenstette, 1986). However, early-maturation problems occur in the genetic algorithm for global searches, and it cannot guarantee convergence to the global optima (Liu et al., 2003). To avoid convergence to local optima, the simulated annealing algorithm uses the Metropolis accepted standards and eventually converges asymptotically to the global optimal solution (Wang and Wang, 1997). In this paper, the simulated annealing algorithm and the genetic algorithm were combined to identify and optimize the parameters. This strategy not only ensures global optimization but also prevents premature convergence and further enhances the global and local optimization capabilities.

We use the simulated annealing genetic algorithm (SAGA algorithm) to identify the coefficients $a, b, c, d, e, f$ of the complete
quadratic multinomial (3); the specific procedure is described follows:

(1) Code
Adopt binary 0/1 character coding, convert decimal into binary; the length of each coefficient to be estimated is 15-binary-character string such that the $a, b, c, d, e, f$ binary strings are ordered into a 90-character chromosome, and each chromosome expresses a complete binary quadratic multinomial.

(2) Initialize Genetic Algorithm
Set population scale $M = 50$, largest evolution algebra $N = 50$, exchange probability $PC = 0.75$, and mutation probability $PM = 0.05$, and use a random number generator to produce an initial group as the parent generation.

(3) Calculate the fitness of every chromosome
The fitness is a quality index that reflects the solution represented by the chromosome, and usually, the better the performance, the greater the fitness. In this article, we adopt the reciprocal of the square error as the fitness function.

$$f(x) = \frac{1}{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2} \quad (5)$$

where $y_i$ is the actual observed value, $\hat{y}_i$ is the model estimation, $n$ is number of training samples, and $f_i(i = 1, 2, ..., N)$ is the best individual fitness of every generation.

(4) Produce a new group according to the following operation

Operation 1: Selection operation. The $f_i$ is the fitness of the $i$th chromosome. Constructing the discrete probability distribution $p_i = f_i / \sum_{i=1}^{M} f_i(i = 1, 2, ..., M)$, we randomly choose a group of chromosomes according to the size of the probability and add to the new groups after reproduction.

Operation 2: Cross operation. This operation exchanges the two randomly selected chromosomes with high-status variation properties and operates on the result of the crossover operation with the simulated annealing operation. According to the Metropolis criteria, this step decides whether to accept the new chromosomes after the crossover operation. If not, operation two is repeated; otherwise, we continue to operation three, and the new individual produced enters the new group.

Operation 3: Variation operation. Based on the mutation rate, the step adopts a low-status large-probability variation method (Wang and Wang, 1997). After changing a selected character of a certain chromosome, the simulated annealing operation is carried out on the result of the variation operation. According to the Metropolis standards, a judgment is made whether to accept the new chromosomes from the variation operation. If not, operation three is repeated; otherwise, we continue to operation four, and the new individual is added to the new group.

Operation 4: Temperature operation. Temperature operation using $T_{i+1} = \alpha T_i$.

(5) Judgment on whether the result meets the termination conditions. If it does not meet the conditions, the genetic operation continues repeatedly. If it meets the termination conditions, it outputs the current best individual, and the algorithm ends.

The specific process of the SAGA algorithm is shown as Fig. 4.
3.3.3. Modeling of the GMDH network based on the simulated annealing and genetic algorithm

To carry out the forecasting with the GMDH network, the method must set up a network model in the form of network self-organization. This process requires the following several steps to build a GMDH model.

(1) Extract the basic unit. The input data of the GMDH network are \( \{Y, x_1, x_2, \ldots, x_m\} \), which consists of multiple input variables and a single output variable structure. After selecting \( n_1 \) samples from \( n \) data samples as the modeling samples (\( n_1 < n \), generally \( n_1 = 1/2 \cdot n \)) and randomly picking two variables \( x_{ij} (j = 1,2, \ldots ,m; i \neq j) \) from \( m \) variables of these samples, \( Y \) is the output, with the simulated annealing genetic algorithm identifying the parameters \( a, b, c, d, e, f \). Therefore, \( m \cdot (m-1)/2 \) basic units are used to produce the first level, forming the initial network.

(2) Establish the input layer. To set a threshold value \( E_g \), first, the relevant variables of the remaining \( n_2 (n_2 = n - n_1) \) samples are fed into the processing unit above. This unit will calculate the variance \( E \) of the processing unit output and actual output and compare \( E \) with \( E_g \), retaining the unit whose variance is less than the threshold (assuming there are \( u_1 \) units) and recording the minimum variance \( E_{m1} \) in these units. This step will define the first layer of the unit.

(3) Build the median unit. The relevant variables from all data are fed into the first layer unit for calculation and identification of the outputs of the first layer \( Y^1(u) (u = 1,2, \ldots ,u_1) \). Taking \( Y^1(u) \) as the input of the second layer, this step obtains the processing units whose number is \( u_2 \), the output \( Y^2(u) (u = 1,2, \ldots ,u_2) \), and the minimum variance \( E_{m2} \) of the second layer. If \( E_{m2} < E_{m1} \), then the second layer is built successfully, and the process continues to construct the next layer.

(4) Establish the output layer. Prior to the \( (k+1) \)th layer, if the process finds the minimum variance of this layer \( E_{m(k+1)} > E_{m(k)} \), the modeling is terminated. The unit whose variance is the least in the \( k \)th layer is taken as the output unit.

Finally, this process ensures that the upper units are related to the output unit for a layered connection. Therefore, the other unrelated units are not included in the network structure. Thus far, this method builds the GMDH network based on the samples. By inputting the corresponding samples into the network model, it can subsequently be used to make predictions.

3.4. Model evaluation

To quantify the performance of the SAGA-GMDH model for forecasting debris flow, three evaluation parameters for the gap between the measured values and predicted values were calculated: the correlation coefficient \( R^2 \), the relative error \( RE \), and the average relative error \( ARE \). The equations for these calculations are listed as follows:

\[
R^2 = \left[ \frac{\sum_{i=1}^{N} (y_i - \bar{y}) \cdot \sum_{i=1}^{N} (y_m - \bar{y})}{\sum_{i=1}^{N} (y_i - \bar{y})^2 \cdot \sum_{i=1}^{N} (y_m - \bar{y})^2} \right]^2
\]

\[
RE = \left| y_i - \bar{y} \right| / y_i
\]

\[
ARE = \frac{1}{n} \sum_{i=1}^{n} RE_i
\]

where \( y \) is the measured value, \( \bar{y} \) is the predicted value, and \( n \) is the predicted node number. In the above three parameters, \( R^2 \) was calculated to analyze the goodness of the predicted values versus the measured values. The values of \( R^2, RE \) and \( ARE \) range from 0 to 1. The higher the value of \( R^2 \), the stronger the linear relationship between the measured and predicted values. Additionally, \( RE \) and \( ARE \) indicate estimation errors. Lower values of \( RE \) and \( ARE \) indicate fewer prediction errors.

4. Results and discussion

4.1. Training of the GMDH model

First, the input variables were selected according to the KLDA (Table 2). The higher similarity degree indicated that the independent variables were significantly correlated to the BQDR. As shown in Table 2, six parameters, i.e., DA, BRHD, SLR, VC, MHRI and DR, showed strong correlations with the QDFR and similarity degrees greater than 0.85; these were each taken as input variables.

Second, the SAGA-GMDH was established based on the selected input variables and the output variables. Approximately 80% of the data is normally sufficient to train the network, and the remainder is often used to test the final architecture of the model (Swingle, 1996). In this study, among the 39 cases of debris flow in gullies, 31 cases (80%) were selected for calibration of the SAGA-GMDH, and the remaining eight cases (20%) were used for validation testing. In the SAGA-GMDH, the QDFR served as the output variables, and DA, BRHD, SLR, VC, MHRI and DR were taken as input variables. The results are shown in Fig. 5. The RE increased sharply when the 27th sample entered the model, and the \( ARE \) and \( R^2 \) between the measured value and the predicted value were 0.9125 and 3.54%, respectively.

4.2. Validation and testing of the GMDH model

To examine the credibility and the stability of the GMDH, the model was verified using the validation and testing sets based on the network structure in the training stage. It shows the results for the measured and predicted values of the BQDR in the validation and testing sets in Fig. 6. The \( R^2 \) values for the validation and test sets were 0.8528 and 0.8216, respectively. The \( ARE \) values were 2.19% and 5.95%, respectively. The SAGA-GMDH provided a more suitable model for both data sets.

4.3. Comparison of different models

In this study, to examine the performance of the SAGA-GMDH model more thoroughly, a BP neural network model and ANFIS were also applied to estimate the BQDR. The BP neural networks are popular neural network architectures in ANN (artificial neural network) models (Ou et al., 2006; Yi et al., 2007). The ANFIS is the most well known FNN (fuzzy neural networks),

| Table 2 |
|-----------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
|                | BQDR    | DA      | BRHD    | SLR     | BLS     | VC      | DSHS    | LMRD    | FAR     | MHRI    | DR      |
| Ordering       | 1       | 0.87    | 0.931   | 0.856   | 0.509   | 0.899   | 0.746   | 0.756   | 0.750   | 0.866   | 0.920   |

The degree of similarity between the principal factors and the relevancy factors.
which has been widely used in many different applications over the last few decades (Jang, 1993; Kurtulus and Flipo, 2012; Naderloo et al., 2012).

For clarity, only the training set was verified using three different models. The result is shown in Figs. 7–9. A comparison of the three predicted results is also shown in Fig. 10. The SAGA-GMDH model performed well in terms of the correlation coefficient ($R^2$).

As shown in Figs. 7–10, the optimal model for estimating the BQDR was the SAGA-GMDH, with $R^2=0.8528$ and $ARE=3.54\%$. Additionally, for all three models, the accuracy of the SAGA-GMDH model predictions was always better than those of the other models. The high accuracy of the SAGA-GMDH in predicting the BQDR can be explained by several factors: first, the SAGA-GMDH model possesses the advantages of both the simulated annealing and the genetic algorithm in identifying the coefficients of the polynomial. Secondly, the SAGA-GMDH model provides an automatic modeling mechanism and is noise immune. Furthermore, the GMDH can be used to model...
complex systems without having specific knowledge of the system. Additionally, this model integrates the environmental factors with rainfall factors to predict the debris flow in a single gully. This study reveals that due to its extended prediction capability compared with that of the other models, the SAGA-GMDH model is well suited for prediction of debris flow.

5. Conclusions

In the current study, we focused on proposing a new methodology and developing novel ideas for predicting debris flow using rainfall factors and environmental factors as the input variables.

Simulated annealing and genetic algorithm had been used in GMDH type neural network for each neuron searching its optimal set of connection with the preceding layer. This strategy is good at identifying and optimizing the parameters.

The preliminary input variables for the model included rainfall factors and environmental factors. The input variables used in the SAGA-GMDH model were derived from the six input variables acquired by the KLDA. The results confirm that the SAGA-GMDH model for prediction of the BQDR demonstrated both a high level of accuracy and a compact network structure. In general, the SAGA-GMDH performed better than the BP network model and the ANFIS model for predicting the BQDR according to the above evaluation parameters.

This research also verified that the SAGA-GMDH model developed in this work could provide an effective and accurate prediction of the debris flow. Furthermore, an accurate and fast prediction of debris flow can be achieved by establishing a proper data retrieval model and selecting suitable input variables for the model.

Competing interests

The authors declare that they have no competing interests.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (40771155) and the National High-tech R&D Program of China (863 Program) (2007AA12Z174).

References
