Earthquake-induced landslide-susceptibility mapping using an artificial neural network

S. Lee, D. G. Evangelista

To cite this version:


HAL Id: hal-00299356
https://hal.archives-ouvertes.fr/hal-00299356

Submitted on 26 Jul 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Earthquake-induced landslide-susceptibility mapping using an artificial neural network

S. Lee¹ and D. G. Evangelista²

¹Geoscience Information Center, Korea Institute of Geoscience & Mineral Resources (KIGAM), 30, Gajeong-dong, Yuseong-gu, Daejeon, 305-350, Korea
²Mines and Geosciences Bureau, Department of Environment and Natural Resources, North Avenue, Diliman, Quezon City, Philippines

Received: 21 November 2005 – Revised: 29 May 2006 – Accepted: 10 June 2006 – Published: 26 July 2006

Abstract. The purpose of this study was to apply and verify landslide-susceptibility analysis techniques using an artificial neural network and a Geographic Information System (GIS) applied to Baguio City, Philippines. The 16 July 1990 earthquake-induced landslides were studied. Landslide locations were identified from interpretation of aerial photographs and field survey, and a spatial database was constructed from topographic maps, geology, land cover and terrain mapping units. Factors that influence landslide occurrence, such as slope, aspect, curvature and distance from drainage were calculated from the topographic database. Lithology and distance from faults were derived from the geology database. Land cover was identified from the topographic database. Terrain map units were interpreted from aerial photographs. These factors were used with an artificial neural network to analyze landslide susceptibility. Each factor weight was determined by a back-propagation exercise. Landslide-susceptibility indices were calculated using the back-propagation weights, and susceptibility maps were constructed from GIS data. The susceptibility map was compared with known landslide locations and verified. The demonstrated prediction accuracy was 93.20%.

1 Introduction

Landslides cause extensive damage to property and occasionally result in the loss of life. Recently, landslides have occurred in the Philippines. It is, therefore, necessary to assess and manage areas that are susceptible to land sliding in order to mitigate any associated damage. Among the many causes of landslides, those triggered by earthquake and heavy rainfall are the most common throughout the Philippines. The consequent requirement to predict such occurrences has led to the development of numerous stochastic and process-based models with an increasing emphasis on the use of a GIS.

To achieve a scientific assessment of an area susceptible to land sliding, an artificial neural network may be applied, and the objective of this study was to apply and verify models of landslide-susceptibility zonation in the Baguio city area of the Philippines using this approach.

Landslides may occur as a consequence of a number of determined trigger factors (Varnes, 1978). In order to assess the potential for landslide, it is a requirement to identify and analyze the influencing factors. In this study, the following parameters were used: slope, aspect, curvature, proximity to drainage, lithology, proximity to faults, land cover and geomorphologic/terrain units. The 16 July 1990 earthquake-induced landslides were used as a basis for landslide-susceptibility mapping (Arboleda and Regalado, 1990).

Using GIS as the basic analysis tool for landslide hazard mapping can be effective for spatial and data management and manipulation, together with some reasonable models for the analysis. In this regard, there have been many studies of landslide hazard mapping using GIS. There have been many studies carried out on landslide hazard evaluation using GIS; for example, Guzzetti et al. (1999) summarized many landslide hazard evaluation studies. Recently, there have been studies on landslide hazard evaluation using GIS, and many of these studies have applied probabilistic models (Jibson et al., 2000; Luzi et al., 2000; Parise and Jibson, 2000; Rautelal and Lakheraza, 2000; Baeza and Corominas, 2001; Lee and Min, 2001; Clerici et al., 2002; Donati and Turrini, 2002; Lee et al., 2002a, b, 2004a; Zhou et al., 2002; Lee and Choi, 2003). One of the statistical models available, the logistic regression models, has also been applied to landslide hazard mapping (Dai et al., 2001; Dai and Lee, 2002; Ohlmacher and Davis, 2003), as has the geotechnical model and the safety factor model (Gokceoglu et al., 2000; Romeo, 2000;
Refice and Capolongo, 2002; Carro et al., 2003; Shou and Wang, 2003; Zhou et al., 2003). As a new approach to landslide hazard evaluation using GIS, data mining using fuzzy logic, and artificial neural network models have been applied

(2) Study area

Terramont Foundation (1992) stated that within the study area of Baguio City, development presently continues without the benefit of a working extant land use plan. This has resulted in the creation of an uncontrolled urban sprawl, with a proliferation of squatter colonies on both private and public land, and a high population density reflected in an overcrowded city experiencing increasingly mixed urban land use. Most development involves massive movement of ground within and beyond feasible construction areas, giving rise to possible erosion, floods and landslide hazards.

The study area (Fig. 1) lies within 16°23′00″–16°29′00″ latitude and 120°34′00″–120°37′00″ longitude. It is located along the main trace and splays of the Philippine Fault which is a major seismic feature. Of particular interest are the northwest–southeast trending splays of this structure that are the western Tuba Fault and the Tebbo Fault to the east. Both of these faults are located less than 5 km away from the city center (Pinet and Stephan, 1990).

Geology plays an important role in landslide potential, and the composition of the study area (Fig. 2) was taken from the work of PINA (1994) and David (1997). Four formations occur, and these are from the base:

(2) **Kennon Formation**: principally massive biohermal limestone, calcarenite and calcirudite. The basal portion consists of wacke, conglomeratic calcarenite with volcanic diorite pebble and cobble clasts.

(3) **Klondyke Formation**: clastic sedimentary rock consisting mainly of polymictic conglomerate with interbedded sandstone, siltstone and shale and in places intercalated with flow breccia and pyroclastic rock. It rests unconformably upon the Kennon Limestone and underlies wide areas on the elevated western side of the Baguio City Quadrangle.

(4) **Baguio Formation**: tuff, volcanic conglomerate and breccia, glassy and porphyritic andesite, with minor sandstone layers.

3 Ground acceleration map

To verify and compare the landslide-susceptibility maps, the existing ground acceleration map (Fig. 3) was used. Although ground-motion hazard studies for the Philippines were made (Tenhaus et al., 1994) after the 1990 Baguio earthquake, the maps produced for peak horizontal ground acceleration are regional in scope and probabilistic in nature. Consequently, ground acceleration maps using a more deterministic approach have been employed. These engage the recent attenuation equation of Fukushima and Tanaka (1990). The relation is as follows:

$$\log_{10} A = 0.41 M - \log_{10}(R + 0.032 \times 100.41M) - 0.0034R + 1.30$$  (1)

where $A$ is the mean peak acceleration (in cm/s$^2$), $R$ is the shortest distance between the site and fault rupture (km) and $M$ is the magnitude of the earthquake.

The equivalent ground acceleration in terms of gravity, $g$, is computed using the equation:

$$g = A/(980 \text{ cm/s}^2).$$  (2)

In assessing the distance $R$ in Eq. (1), both the Tuba and Tebbo faults were regarded as seismic sources. $R$ in Eq. (1) is the shortest distance measured from a grid point to either the Tuba or Tebbo fault trace, whichever is the shorter measurement of the two. The final $g$ value in the map is further controlled by the geology in the area. The Baguio Formation is assigned here as the rock medium of slip. The Zigzag, Klondyke and Kennon formations are classified as hard rock, with a correction factor of 0.6 in considering the final $g$. The highest value computed for the grid points is 0.634092 g and this lies within the Baguio Formation. The lowest value derived is 0.3366503 from the Zigzag Formation.

4 Artificial neural network and weight determination

An artificial neural network is a “computational mechanism able to acquire, represent, and compute a mapping from one multivariate space of information to another, given a set of data representing that mapping” (Garrett, 1994). The back-propagation training algorithm is the most frequently used neural network method and is the method used in this study. The back-propagation training algorithm is trained using a set of examples of associated input and output values. The purpose of an artificial neural network is to build a model of the data-geneweight process, so that the network can generalize and predict outputs from inputs that it has not previously seen. This learning algorithm is a multi-layered neural network, which consists of an input layer, hidden layers, and an output layer. The hidden and output layer neurons process their inputs by multiplying each input by a corresponding weight, summing the product, and then processing the sum using a nonlinear transfer function to produce a result. An artificial neural network “learns” by adjusting the weights between the neurons in response to the errors between the actual output values and the target output values. At the end of this training phase, the neural network provides a model that should be able to predict a target value from a given input value.

There are two stages involved in using neural network for multi-source classification: the training stage, in which the internal weights are adjusted; and the classifying stage. Typically, the back-propagation algorithm trains the network until some targeted minimal error is achieved between the desired and actual output values of the network. Once the training is complete, the network is used as a feed-forward structure to produce a classification for the entire data (Paola and Schwengerdt, 1995).
A neural network consists of a number of interconnected nodes. Each node is a simple processing element that responds to the weighted inputs it receives from other nodes. The arrangement of the nodes is referred to as the network architecture (Fig. 4). The receiving node sums the weighted signals from all the nodes that it is connected to in the preceding layer. Formally, the input that a single node receives is weighted according to Eq. (3).

\[ \text{net}_j = \sum_i w_{ij} \cdot o_i \]  

where \( w_{ij} \) represents the weights between nodes \( i \) and \( j \), and \( o_i \) is the output from node \( j \), given by

\[ o_j = f(\text{net}_j). \]  

The transfer function \( f \) is usually a non-linear sigmoid function that is applied to the weighted sum of inputs before the signal propagates to the next layer. One advantage of a sigmoid function is that its derivative can be expressed in terms of the function itself:

\[ f'(\text{net}_j) = f(\text{net}_j)(1 - f(\text{net}_j)) \]  

The network used in this study consisted of three layers. The first layer is the input layer, where the nodes were the elements of a feature vector. The second layer is the internal or “hidden” layer. The third layer is the output layer that presents the output data. Each node in the hidden layer is interconnected to nodes in both the preceding and following layers by weighted connections (Atkinson and Tatnall, 1997).

The error, \( E \), for an input training pattern, \( t \), is a function of the desired output vector, \( d \), and the actual output vector, \( o \), given by:

\[ E = \frac{1}{2} \sum_k (d_k - o_k). \]  

The error is propagated back through the neural network and is minimized by adjusting the weights between layers. The weight adjustment is expressed as:

\[ w_{ij}(n + 1) = \eta(d_j \cdot o_i) + \alpha \Delta w_{ij} \]  

where \( \eta \) is the learning rate parameter (set to \( \eta=0.01 \) in this study), \( \delta_j \) is an index of the rate of change of the error, and \( \alpha \) is the momentum parameter (set to \( \alpha=0.01 \) in this study). The factor \( \delta_j \) is dependent on the layer type. For example, for hidden layers, \( \delta_j = (\sum_k \delta_k w_{jk}) f'(\text{net}_j) \)  

and for output layers, \( \delta_j = (d_j - o_j) f'(\text{net}_k) \)

This process of feeding forward signals and back-propagating the error is repeated iteratively until the error of the network as a whole is minimized or reaches an acceptable magnitude.

Using the back-propagation training algorithm, the weights of each factor can be determined and may be used for classification of data (input vectors) that the network has not seen before. Zhou (1999) described a method for determining the weights using back propagation. From Eq. (4), the effect of an output, \( o_j \), from a hidden layer node, \( j \), on the output, \( o_k \), from an output layer (node \( k \)) can be represented by the partial derivative of \( o_k \) with respect to \( o_j \) as

\[ \frac{\partial o_k}{\partial o_j} = f'(\text{net}_k) \cdot \frac{\partial (\text{net}_k)}{\partial o_j} = f'(\text{net}_k) \cdot w_{jk}. \]

Equation (10) produces both positive and negative values. If the effect’s magnitude is all that is of interest, then the importance (weight) of node \( j \) relative to another node \( j0 \) in the hidden layer may be calculated as the ratio of the absolute values derived from Eq. (10):

\[ \left| \frac{\partial o_k}{\partial o_j} \right| = \left| \frac{f'(\text{net}_k) \cdot w_{jk}}{f'(\text{net}_{k0}) \cdot w_{j0k}} \right| = \frac{|w_{jk}|}{|w_{j0k}|}. \]

We should mention that \( w_{j0k} \) is simply another weight in \( w_{jk} \) other than \( w_{ik} \).

For a given node in the output layer, the results of Eq. (11) show that the relative importance of a node in the hidden layer is proportional to the absolute value of the weight connecting the node to the output layer. When the network consists of output layers with more than one node, then Eq. (11) cannot be used to compare the importance of two nodes in the hidden layer.

\[ w_{j0k} = \frac{1}{J} \cdot \sum_{j=1}^{J} |w_{jk}| \]

\[ l_{jk} = \frac{|w_{jk}|}{J \cdot \sum_{j=1}^{J} |w_{jk}|} = \frac{J \cdot |w_{jk}|}{\sum_{j=1}^{J} |w_{jk}|} \]
Therefore, with respect to node $k$, each node in the hidden layer has a value that is greater or smaller than unity, depending on whether it is more or less important, respectively, than an average value. All the nodes in the hidden layer have a total importance with respect to the same node, given by

$$
\sum_{j=1}^{J} t_{jk} = J.
$$

Consequently, the overall importance of node $j$ with respect to all the nodes in the output layer can be calculated by

$$
t_{j} = \frac{1}{K} \cdot \sum_{j=1}^{J} t_{jk}.
$$

Similarly, with respect to node $j$ in the hidden layer, the normalized importance of node $j$ in the input layer can be defined by

$$
s_{ij} = \frac{|w_{ij}|}{\sum_{i=1}^{I} |w_{ij}|} = \frac{J}{I} \cdot \frac{|w_{ij}|}{\sum_{i=1}^{I} |w_{ij}|}.
$$

The overall importance of node $i$ with respect to the hidden layer is

$$
s_{i} = \frac{1}{J} \cdot \sum_{j=1}^{J} s_{ij}.
$$

Correspondingly, the overall importance of input node $i$ with respect to output node $k$ is given by

$$
S_{ti} = \frac{1}{J} \cdot \sum_{j=1}^{J} s_{ij} \cdot t_{j}.
$$

5 Data and methodology

Data preparation involved the digitization or creation of a GIS database which included the topographical, geomorphological, geological and land cover data. A digitized map of earthquake-induced landslide locations detected from satellite imagery and field survey was produced, and these digital data were included in the GIS. A vector-to-raster conversion was undertaken to provide raster data of landslides. The factors of slope, aspect, curvature, proximity to drainage, lithology, proximity to faults, land cover and geomorphologic/terrain units were used.

Contour and survey base points that had an elevation value read from the 1:10000 scale topographic map were extracted and a Digital Elevation Model (DEM) was constructed. Using the DEM, the slope gradient, slope aspect and curvature were calculated. The slope gradient of a surface refers to the maximum rate of change in $z$ values across a region of the surface and the slope aspect is the compass direction maximum rate of change in $z$ in a downward direction. The curvature represents the morphology of the topography. A positive curvature indicates that the surface is upwardly convex in that cell, and a negative shows that the surface is upwardly concave. A zero value represents a flat surface. The distance from drainage was calculated in 1 m intervals. The land use/land cover data were derived from the 1:10000 topographic map.

The lithology was taken from the 1:50000 scale geological map, and the distance from a lineament was measured to the nearest 1 m interval. Distance buffers on both sides of a fault were generated to note the occurrence of landslides with respect to fault lines. Most of the landslides that occurred following the 1990 earthquake were observed to be within the 500 m buffer.

Landslides occur in varying terrain units, which are floodplain, deep, wide or shallow valleys, basin, plateau, karst, or limestone hills. Floodplain is cultivated flat terrain. Pronounced meanders occur in narrow, deep valleys with wide drainage divides. Wide valley is a broad valley with a narrow floodplain dominated by active erosion processes. Shallow valley is characteristically narrow and shallow with steep slopes in generally rugged terrain. Basin is a shallow depression with rounded contours and poor development of drainage lines. Plateau is well-drained round ridges and peaks dominated by narrow plateaus rather than valleys. Karst is rugged terrain characterized by sinkholes and poorly defined drainage lines. Limestone hills are poorly drained, with rounded contours and an absence of sinkholes.

The study area was divided into a grid with 10 m×10 m cells, occupying 560 rows and 541 columns in total 295 637 grid cells and earthquake-induced landslides occurred in 61 of these.

Maps relevant to the landslide occurrences were used first to construct a vector-type spatial database using the ARC/INFO GIS software package. Secondly, landslide occurrence areas were detected during field survey of the study area. A map of the landslide locations was made as part of the GIS spatial database. Thirdly, for their weight determination, the landslide factors were entered into an ARC/INFO grid type, and then converted to ASCII data for use with an artificial neural network program.

The results of the analysis were converted to grid data using the GIS. For detected landslide location, the weight of each factor was determined by a neural network method, developed using MATLAB (Hines, 1997). For factor weight determination using this approach, the ground subsidence location was assigned as an experiment area. When the weights converged to a proper value, they were determined by back propagation between the neural network layers. Finally, ground subsidence hazard mapping was carried out considering the factor weight derived from our study, and the analytical results were verified in a comparison with ground subsidence locations. In this study, the GIS software ArcView 3.3 and ARC/INFO version 9.0 were used as the basic analysis tools for spatial management and data processing.
6 Landslide-susceptibility analysis using the artificial neural network

Landslide-prone occurrences from the 61 sites and locations without slide susceptibilities were selected as experiment sites using slope map and landslide location data. To assess the effect of the exercise site selection, they were identified as areas where differing slope values with a 5 degree interval were classified as “areas not prone to landslides”, and areas of landslides were assigned as “areas prone to landslides” sets. If the analysis selected more than 61 sites with the same value then the sites were chosen at random.

The back-propagation algorithm was applied to calculate the weights between the input and hidden layers, and between the hidden and output layers, by modifying the number of hidden nodes and the learning rate. A three-layered feed-forward network was established using the MATLAB software package based on the framework provided by Hines (1997) in which “feed-forward” denotes that the interconnections between the layers propagate forward to the next layer.

The number of hidden layers and the number of nodes in such a layer required for a particular classification problem are not easy to deduce. In this study, an $8 \times 16 \times 2$ structure was selected for the network with input data normalized in the range 0.1–0.9. The nominal and interval class group data were converted to continuous values ranging between 0.1 and 0.9. Continuous values were not, therefore, ordinal, but nominal data, and the numbers denote the classification of the input data. The learning rate was set to 0.01, and the initial weights were randomly selected to values between 0.1 and 0.3.

The back-propagation algorithm was used to minimize the error between the predicted and calculated output values. The algorithm propagated the error backwards, and iteratively adjusted the weights. The number of epochs was set to 2000, and the root mean square error (RMSE) value used for the stopping criterion was set to 0.1. The experimental data sets met the 0.1 RMSE goals in the case of 0 slope (Fig. 6). If the RMSE value was not achieved however, then the maximum number of iterations was terminated at 2000 epochs.

Figure 5 is the flowchart of the neural network exercise for weight determination. The weights between layers acquired by using the neural network were calculated reversely and the contribution or importance of each factor was derived for each set in the exercise. With slope at 0, the final weights of the 8 factors used to predict landslide susceptibility are shown in Table 1. For easy interpretation, the weight values were normalized. Distance from a fault has the minimum weight value of 1.00, and slope has the maximum value of 2.30. Finally, the weights were applied across the entire study area, and the landslide-susceptibility map was then developed (Fig. 7). Susceptibility was classified into four classes: highest 10%, second 10%, third 20% and remainder 60%, based on area for easy visual interpretation. The minimum value obtained was 0.0092 and 0.9971 was the maximum. The mean value is 0.3784 and the standard deviation 0.2534.

Table 1. Weights of each factor for each selection of training site.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Weight</th>
<th>Normalized Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slope (unit: degree)</td>
<td>0.2003</td>
<td>2.300</td>
</tr>
<tr>
<td>Aspect</td>
<td>0.1302</td>
<td>1.495</td>
</tr>
<tr>
<td>Curvature (Unitless)</td>
<td>0.1088</td>
<td>1.249</td>
</tr>
<tr>
<td>Distance from Drainage (unit: m)</td>
<td>0.1031</td>
<td>1.184</td>
</tr>
<tr>
<td>Geology</td>
<td>0.1278</td>
<td>1.467</td>
</tr>
<tr>
<td>Distance from Fault (unit: m)</td>
<td>0.0871</td>
<td>1.000</td>
</tr>
<tr>
<td>Land Cover</td>
<td>0.1033</td>
<td>1.186</td>
</tr>
<tr>
<td>Terrain mapping unit</td>
<td>0.1395</td>
<td>1.602</td>
</tr>
</tbody>
</table>

Fig. 6. Backpropagation training results.
Fig. 7. Landslide susceptibility mapping using artificial neural network.

7 Verification

Two basic assumptions are required in the verification of landslide-susceptibility calculation models. One is that landslides are related to spatial factors such as topography, geology and land cover, and the other is that future landslides will be triggered by a specific impact such as seismic shock or heavy rainfall (Brabb, 1984; Varnes, 1978). Both assumptions are satisfied in this study because the landslides were related to spatial factors and were triggered by earthquake.

The susceptibility analysis result was verified using known landslide locations compared with the landslide-susceptibility map. Rate curves were created and areas under the curve were calculated in each case. The rate illustrates how well the model and factor predict landslide and, the area under the curve enables an assessment of the prediction accuracy qualitatively. To obtain the relative ranking for each prediction pattern, the calculated index values of all cells in the study area were sorted in descending order. The ordered cell values were divided into 100 classes, with accumulated 1% intervals. In Fig. 8, for slope 0 the rate verification results appear as a line. However, the, 90 to 100% (10%) classification of the study area where the landslide-susceptibility index had a higher rank accounts for 75% of all the landslides. Furthermore, the 80 to 100% (20%) classification of the study area predicts 92% of the landslides.

To compare the result quantitatively the areas under the curve were recalculated with the total area as 1 which means perfect prediction accuracy. Consequently, the area under a curve can be used to assess the prediction accuracy qualitatively. In the case of slope 0, the area ratio was revealed as 0.9320 and we can say that the prediction accuracy is thus 93.20%.

To assess the effect of experiment site selection locations, sites were chosen as areas where slope values with a 5° interval were classified as prone or not prone to landslides. After considering the slope, the verification showed that with a slope increase, the accuracy decreased (Table 2). Above 36°, the prediction accuracy has a particularly low value. Because landslide is notably related to slope, the designation of such a test site as a “landslide not prone area” has an effect on the accuracy of a landslide-susceptibility map.

The landslide-susceptibility map was also visually compared with the ground acceleration map (Fig. 3). The landslide possibility and ground acceleration values are higher in the south than elsewhere in the study area.

8 Discussion and conclusions

Landslides are among the most hazardous of natural disasters. Government and research institutions worldwide have attempted for years to assess landslide hazard and risk and to show its spatial distribution. In this study, a data bank approach by means of a GIS shows considerable promise in the identification of areas susceptible to landslide caused
by earthquakes. In the verification of landslide-susceptibility maps, the artificial neural network showed a very high prediction accuracy of 93.20% in the case of 0 slope. The landslide-susceptibility map was spatially consistent when compared with the ground acceleration.

The back-propagation algorithm presented difficulties when trying to follow the internal procedures. The method involves a long execution time, has a heavy computing load, and there is a requirement to convert the database to another format. Landslide susceptibility can, however, be analyzed qualitatively. In addition to using a multifaceted approach to a solution, the extraction of reliable results for a complex problem is possible with continuous and discrete data processing.

These results can be used as basic data to assist slope management and land-use planning. The models used in the study are valid for generalized assessment purposes, although they may be less useful at a site-specific scale where local geological and geographic heterogeneities may prevail.

Edited by: F. Guzzetti
Reviewed by: C. Gokceoglu and D. Keefer

References


Donati, L. and Turrini, M. C.: An objective method to rank the importance of the factors predisposing to landslides with the GIS methodology: application to an area of the Apennines (Valnerina; Perugia, Italy), Eng. Geol., 63, 277–289, 2002.


Lee, S., Ryu, J. H., Min, K. D., and Won, J. S.: Landslide suscept-
S. Lee and D. G. Evangelista: Landslide susceptibility mapping using artificial neural network 695


www.nat-hazards-earth-syst-sci.net/6/687/2006/