ASSESSMENT OF EARTHQUAKE- TRIGGERED LANDSLIDES HAZARD IN EL SALVADOR BASED ON AN ARTIFICIAL NEURAL NETWORKS (ANN) MODEL

EGU Assembly 2008

M.J. García-Rodríguez[1], J.A. Malpica[2], B. Benito[1]

1. Universidad Politécnica de Madrid. Escuela Técnica Superior de Ingenieros en Topografía, Geodesia y Cartografía.(ETSITGC).Departamento de Ingeniería Topográfica y Cartografía. Madrid 28031, Spain. Corresponding author: mjosegr@topografia.upm.es
2. Universidad de Alcalá. Escuela Politécnica. Departamento de Matemáticas. Madrid, Spain.

ABSTRACT

This study is an approach for assessing earthquake-triggered landslides hazard using artificial neural networks (ANN). The computational method used for the training process is a back-propagation learning algorithm. It is applied to El Salvador, one of the most seismically active regions in Central America, where the last severe destructive earthquakes occurred in January 13th (Mw 7.7) and February 13th, 2001, (Mw 6.6). The first of these earthquakes triggered more the 600 landslides, included the most tragic in Las Colinas landslide, and killed at least 844 people.
The ANN is designed and trained for developing landslide susceptibility analysis techniques at a regional scale. This approach uses an inventory of landslides and different parameters of slope instability: slope gradient, elevation, aspect, mean annual precipitation, lithology, land use, and terrain roughness. The obtained ANN is applied in combination with a Geographic Information System (GIS) for landslide susceptibility mapping.
A previous logistic regression analysis was done, taking as independent variables the same parameters considered in the neural network, while the variable occurrence or non-occurrence of landslides is considered as dependent variable. The logistic approach determined the importance of terrain roughness and soil type as key factors within the model.
The results of the landslide susceptibility analysis are checked using landslide location data. These results show a high concordance between the landslide inventory and the high susceptibility estimated zone. Finally, we make a comparative analysis of the statistics techniques, logistic regression and artificial neural network.

INTRODUCTION

Figure. 1. Earthquake-triggered landslide in Las Colinas, Santa Tecla (El Salvador, 13 January 2001), where 585 people died. The epicentre was located off the western coast of El Salvador, in the subduction zone between the Cocos and Caribbean plates, with a magnitude of Mw 7.7 and a focal depth of 40km (Benito et al., 2004).

Landslide hazard studies are appropriate for evaluation and mitigation plan development in potential areas for the occurrence of landslides. There are several techniques available for landslide hazard research, however heuristics approaches provide computational models to assess landslide hazard over large regions. In January and February of 2001, El Salvador experienced several destructive earthquakes, which caused hundreds of landslides of various sizes. In this study, we have used an ANN model to assess the susceptibility of earthquake-induced landslides for the whole country of El Salvador.
The first earthquake occurred on January 13, 2001, with the epicenter located offshore the western coast of El Salvador, in the subduction zone between the Cocos and Caribbean plates *(Figure 2)*. The earthquake had a magnitude Mw 7.7 and it produced more than 600 landslides, the most important of them taking place in Santa Tecla, Las Colinas *(Figure 1)*, where 585 people died.
The second destructive earthquake occurred one month later, on February 13, with a magnitude Mw 6.6, and with the epicenter located to the west of San Miguel, in the mainland. This earthquake was associated to local faults crossing the country from east to west.
The available information is composed by numerous strong-motion records, topographical and geologic maps. They provide a complete documental base for the application of this methodology and a data source to contrast with observations of landslides. The calibrated models will allow the post-evaluation of the landslide risk associated to future earthquakes in El Salvador.

Figure. 2. Geographic localization of the El Salvador (A) and Center-America Sismotectonic map (B)

ARTIFICIAL NEURAL NETWORKS MODEL

Artificial neural networks (ANN) are generic non-linear function approximators extensively used for pattern recognition and classification, they were developed by McCulloch y Pitts (1943). ANN are networks of highly interconnected neural computing elements that have the ability to respond to input stimuli and to learn to adapt to the environment. ANN establishes rules during the learning phase and uses the rule to predict the output. As input into the networks, topographic, climatic, geological, geotechnical and seismological data from different sources are collected, and stored in a Geographic information System (GIS).
ANN have the ability to handle imprecise and fuzzy data, so they can work with continuous, categorical and binary data without violating any assumptions. These techniques have been successfully applied to many problems, including forecasting and prediction problems (Bishop, 1996). In order to cope with non-linearly separable problems, such as the landslide phenomena, ANN have probed their effectiveness. The neural network model is employed to analyse specific elements presenting the study area that contributed to landsliding in the past. The resulting information can then be used in the prediction of areas that may face landsliding in the future. The ANN are trained with error correction learning, which means that some desired response for the system must be known (Werbos (1974) and Parker (1985)).
The behaviour of an artificial neural network depends on the architecture of the network and on both the weights assigned to the connections and the transfer function *(Figure 3)*. In our case we choose 7 input layer, 1 hidden layer and 1 output layer. First, we present the system with the input data and obtain the output. Second, we adjust the system, to make the output closer to what it should be. The first stage is refereed as feedforward, the second as back-propagation.

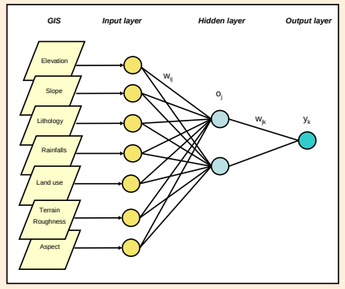

Figure 3. Artificial Neural Network Architecture

Back-propagation Algorithm

The back-propagation algorithm was applied to calculate the weights between the input layer and the hidden layer, and between the hidden layer and the output layer, by modifying the number of hidden node and the learning rate. The nodes perform nonlinear input-output transformations by means of sigmoid activation functions. This structure of nodes and connections, known as the network topology, together with the weights of the connections, determines the final behaviour of the network. It is characterised by the flexibility and adaptability for the modelling of a great number of problems.

The sequence of the back-propagation algorithm is the following *(Figure 4)*. The receiving node sums the weighted signals from all nodes to which it is connected in the preceding layer.
The expression is explicated by the equation **(eq.1)** $h_j = \sum_i w_{ij} x_i$

where w_{ij} is the weight between the node *i* and the node *j*, and *x* is the output value from node *i*. The value produced by the hidden node *j*, is the activation function, *f*, evaluated at the sum produced within the node *j* **(eq.2)**:

$$o_j = f(h_j) = f\left(\sum_{i=1}^{n} w_{ij} x_i\right)$$

In turn, the output value is a function of weight between the hidden and output layer, and the outputs of the input nodes **(eq. 3)**: $y_k = f\left(\sum_j^k w_{jk} o_j\right)$

The activation function *f* is normally a non-lineal sigmoid function, which is applied to the sum *f* the weight of the inputs before proceeding to the next layer.
The main advantage of the sigmoid function **(eq. 4)**: $o_j = f(h_j) = \frac{1}{1+e^{-h_j}}$ and its derivative is $f'(h_j) = f(h_j)(1 - f'h_j))$ **(eq. 5)**.

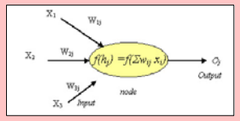

Figure 4. A back-propagation neural network and sigmoid function was used as activation function of ANN

In the developing a neural network for a particular problem, the data set is usually separated into two subsets of training and validation data. In the training phase, the selection of samples indicates the most relevant aspects for solving the problem, while validation data are useful for the evaluation of the neural network. The network architecture can be modified (changing the number nodes in the hidden layer, learning rate, momentum parameter, etc.), consequently, the process is repeated until the tests are satisfactory.

GIS METHODOLOGY FOR THE ASSESSMENT OF EARTHQUAKE-TRIGGERED LANDSLIDES HAZARD

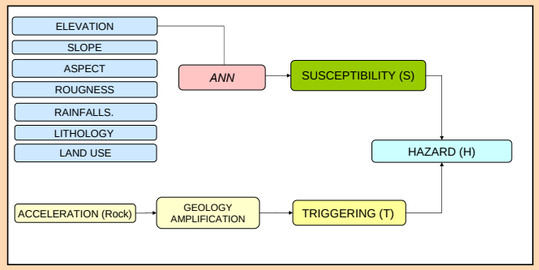

Figure 5. GIS Methodology for the assessment of earthquake-triggered landslides hazard

Table I. Geology Amplification Function

Lithology	Amplification Function (F_G)
Class I	F_G =1.00
Class II	F_G =0.66*Acceleration +1.33
Class III	F_G =0.24* Acceleration +1.62
Class IV	F_G =2.44

The methodology for the assessment of earthquake-triggered landslides hazard using a GIS is shown in the *Figure 5.* The evaluation of susceptibility requires data input of variables representing physical parameters known to contribute to the initiation of landslides. El Salvador GIS *(Figure 6)* is composed of diverse information: 1:100.000 geological map, 1:25.000 Digital Cartography, precipitation data, ground strong motions records, epicentres catalogue, inventory of landslides,…. All these layers from different sources and formats have been integrated into GIS. And there are other layers generated maps and requirements (Digital Terrain Model (DTM), slope map, aspect, roughness, precipitation, etc.) for the assessment hazard.

The susceptibility map *(Figure 7)* was estimated based on ANN model. Later, we calculate the Triggering map (T) *(Figure 8)*, with the local effect due to geology amplification *(Table I)*. From the combination of both map, we obtained the earthquake-triggered landslides Hazard (H) using the GIS *(Figure 9)* .

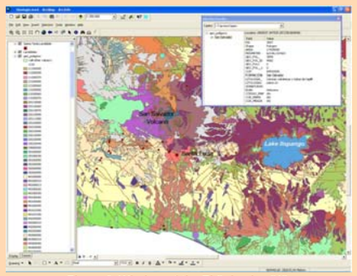

Figure 6. Window of the El Salvador GIS

Susceptibility Map

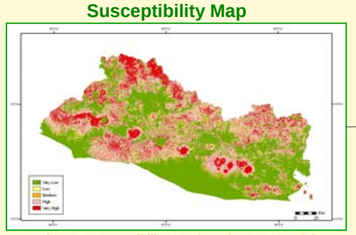

Figure 7. Susceptibility Map based on ANN model

Triggering Map

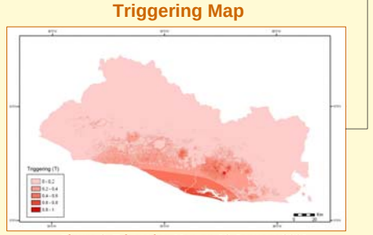

Figure 8. Triggering Map on 13 January 2001

Earthquake- Triggered Landslides Hazard Map

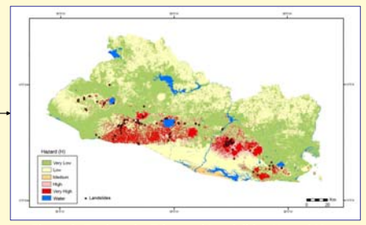

Figure 9. Earthquake-Triggered Landslides Hazard Map classified in five levels: very low, low, medium, high and very high.

	Hazard Level				
	Very Low	Low	Medium	High	Very High
Nº pixels	658218	1153517	21134	76030	162303
Percent (%)	31.8%	55.7%	1.02%	3.7%	7.8%

Validation Process

The ***back-propagation algorithm*** calculates the error, E for an input layer using the expression:

$$E = \frac{1}{2}\sum_k (d_k - y_k) \qquad \textbf{(eq. 6)}$$

where d_k is the desired value of the output vector, and y_k , and the actual output vector of the back-propagation algorithm for the input layer *t*. The main aim of the back-propagation algorithm is to minimize the error quadratic E. The ideal scenario is to achieve a zero error for all samples training, however this is virtually impossible, so it looks for the weights that minimize the error.

The total error is defined by the formula: $E_T = \sum_k E$ **(eq. 7)**

The error back propagated through the neural network which is minimized by changing the weight between layers. The adjusted weight can be expressed by equation of the generalized delta

$$w_{ij}' = w_{ij} + \Delta w_{ij} \qquad \textbf{(ec.8)}$$

where Δw_{ij} is the incremental difference of weight and η is the learning rate parameter, positive and less than one

$$\Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}} \qquad \textbf{(ec.9)}$$

The validation of the analysis was carried out by comparing the training sites and the estimated landslide map obtained by applying the weights derived from the ANN model.
A **confusion matrix** (contingency matrix) is calculated to know the accuracy of a classification result, by comparing the location and class of each ground truth pixel with the corresponding location and class in the classification image. An overall accuracy (95.1%), kappa coefficient (0.9013) confusion matrix, and errors of commission and omission are reported.

ROC Curve (Receiver Operating Characteristic) visualize a classifier's performance in order to select the proper decision threshold. A probability of detection (Pd) versus probability of False Alarm (Pfa) curve.

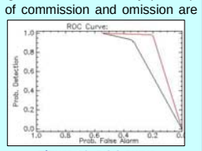

Figure 10. ROC curve

DISCUSSION

A landslide susceptibility map at a regional scale of El Salvador was derived by using an ANN, employing a back-propagation-learning algorithm. This ANN was designed and trained with the landslide inventory and information included in a Geographic Information System (GIS). The factors considered are: slope gradient, slope aspect, elevation, land-use, lithology, mean annual precipitation and roughness terrain. The validation process was carried out using a ROC curve and confusion matrix, with an overall accuracy of 95.1%. Later, an earthquake-triggered landslide hazard map was calculated from the estimated Susceptibility (S) and Triggering (T) map.
Recently, we have worked with the logistic regression techniques using the same input data, the results illustrate the importance of terrain roughness and soil type as key factors within the model-using only these two variables the analysis returned a significance level of 89.4% *(García-Rodríguez et al., 2008)*. In the Logistic regression (LR), the probability is estimated according to the logistic formula, which provides a deterministic model for the data and yields weighting factors for each contributing factor. It also allows the calculation of the odds ratio, which represents the degree of risk associated with each factor. However, LR fits the data to a fixed function, so it is less flexible and less capable of solving complex problems than ANN.
The advantages of the ANN and its possible application to the evaluation of landslide risk comes from the remarkable information processing characteristics of the artificial simulated biological system, including ability to handle imprecise and fuzzy information, fault and failure tolerance, high parallelism, non-linearity, robustness, capability to generalise and tolerance to noise data (Basheer and Hajmeer, 2000). On the other hand, they are known as black-box methods since it is not known exactly how ANN learn particular problems and apply the extracted rules to new cases, or how conclusions can be drawn from the trained networks.
The ANN has many advantages compared with statistical methods. Firstly, the ANN method is independent of the statistical distribution of the data and there is no need of specific statistical variables. Compared with the statistical methods, neural networks allow the target classes to be defined with much consideration to their distribution in the corresponding domain of each data source (Zhou, 1999).

REFERENCES

- Basheer, I.A., Hajmeer, M. (2000). Artificial neural networks: fundamentals, computing, design and application. Journal of Microbiological Methods 43, 3– 31.
- Benito, B., Cepeda, J.M., Martínez Díaz, J.J. (2004). Analysis of the spatial and temporal distribution of the 2001 earthquakes in El Salvador. In: Rose, W.I., Bommer, J.J., López, D.L., Carr, M.J., Major, J.J. (Eds.), Natural Hazards in El Salvador. Geological Society of America Special Paper, vol. 375. Boulder, pp. 339–356.
- Bishop, C.M. (1996). Neural Networks for Pattern Recognition. Oxford University Press, USA; 1 edition
- McCulloch, W. S. and Pitts, W. H. (1943). A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics, 5,115-133.
- García-Rodríguez, MJ., Malpica, J.A., Benito, B., Díaz, M. (2008). Susceptibility assessment of earthquake-triggered landslides in El Salvador using logistic regression. Geomorphology 95,172-191
- Parker, D.B. Learning logic. (1985).Technical report TR-47, Center for Computational Research in Economics and management Science, MIT, Cambridge, MA.
- Zhou, W. (1999). Verification of the nonparametric characteristics of backpropagation neural networks for image classification. IEEE Transactions on Geoscience and Remote Sensing 37, 771– 779.
- Werbos, P. Beyond. (1974). Regression: New Tools for Prediction and Analysis in the Behavioral Sciences. Tesis doctoral, Harvard Cambridge, MA..

ACKNOWLEDGEMENTS

This research has been developed within the framework of a ANDROS Project financed by the Spanish Ministry of Science and Education (CGL2005-07456-C03-03/BTE) and Polithenic University of Madrid - Latinoamerica Relationships (AL08-PID-038). Cartography data have been provided by *Ministerio de Ambiente y Recursos Naturales* (MARN) of the El Salvador. The authors thank these contributions.