METHODOLOGIES FOR THE ASSESSMENT OF EARTHQUAKE- TRIGGERED LANDSLIDES HAZARD. A COMPARISON OF LOGISTIC REGRESSION AND ARTIFICIAL NEURAL NETWORK MODELS.

EGU Assembly 2009

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

In recent years, interest in landslide hazard assessment studies has increased substantially. They are appropriate for evaluation and mitigation plan development in landslide-prone areas. There are several techniques available for landslide hazard research at a regional scale. Generally, they can be classified in two groups: qualitative and quantitative methods. Most of qualitative methods tend to be subjective, since they depend on expert opinions and represent hazard levels in descriptive terms. On the other hand, quantitative methods are objective and they are commonly used due to the correlation between the instability factors and the location of the landslides. Within this group, statistical approaches and new heuristic techniques based on artificial intelligence (artificial neural network (ANN), fuzzy logic, etc.) provide rigorous analysis to assess landslide hazard over large regions. However, they depend on qualitative and quantitative data, scale, types of movements and characteristic factors used. We analysed and compared an approach for assessing earthquake-triggered landslides hazard using logistic regression (LR) and artificial neural networks (ANN) with a back-propagation learning algorithm. One application has been developed in El Salvador, a country of Central America where the earthquake-triggered landslides are usual phenomena. In a first phase, we analysed the susceptibility and hazard associated to the seismic scenario of the 2001 January 13th earthquake. We calibrated the models using data from the landslide inventory for this scenario. These analyses require input variables representing physical parameters to contribute to the initiation of slope instability, for example, slope gradient, elevation, aspect, mean annual precipitation, lithology, land use, and terrain roughness, while the occurrence or non-occurrence of landslides is considered as dependent variable. 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 with an adjustment of 95.1 % for ANN model and 89.4% for LR model. In addition, we make a comparative analysis of both techniques using the Receiver Operating Characteristic (ROC) curve, a graphical plot of the sensitivity vs. (1 - specificity) for a binary classifier system in function of its discrimination threshold, and calculating the Area Under the ROC (AUROC) value for each model. Finally, the previous models are used for the developing a new probabilistic landslide hazard map for future events. They are obtained combining the expected triggering factor (calculated earthquake ground motion) for a return period of 475 years with and the susceptibility map.

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 LR and ANN models 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 epicentre 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 epicentre 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.

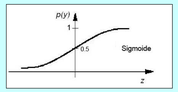
Figure. 2. Geographic localization of the El Salvador (A) and Centre-America Sismotectonic map (B)

LOGISTIC REGRESSION MODEL

The logistic regression is a statistical method appropriate for analysis of presence - absence of the dependent variable, which in our case is the evaluation of landslide susceptibility, such as the significance level of the factors that are involved in the model. this model describes the relationship of the presence/absence, which is the dependent variable and normally codes as 0 or 1 for its two possible categories, with a set of independent variables X1, X2,…, Xn, such as slope angle, aspect, lithology and land use. The quantitative relationship between the occurrence and its dependency on several variables can be expressed as a sigmoid function type:

$$p(y) = \frac{1}{1+e^{-z}} \longrightarrow Z = b_0 + b_1x_1 + b_2x_2 + ... + b_nx_n$$

where:
b_0 y b_i (i=0, 1,…, n) are the coefficients estimated from input data.
x_i (i=1,…, n) are the independent variables (landslide-related physical parameters)



If we substitute the logistic model formula for p(Y=1) (landslide) in the logit form. $logit = \ln\left(\frac{p(Y=1)}{1-p(Y=1)}\right)$

It follows that: $\eta = \ln\left(\frac{p(y=1)}{1-p(y=1)}\right) = b_0 + b_1x_1 + b_2x_2 + ... + b_nx_n$ ➡ The adjusted model results in the following expression:

$$\eta = -11.309 + 5.111 * R - 2.853 * S_1 + 3.073 * S_3 + 2.242 * S_4$$

where: R: roughness, S1: rock, S3: consolidated soil and S4: non consolidated soil.

The susceptibility map ***(Figure 7)*** was validated with the landslides inventory, then we obtained a concordance of 89.4% between the susceptible areas and the observed data.

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.

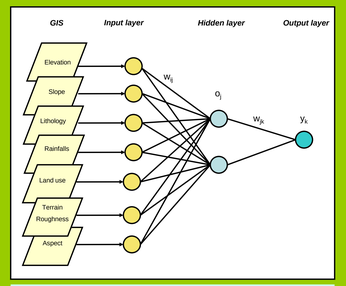


Figure 3. Artificial Neural Network Architecture

GIS METHODOLOGY FOR THE ASSESSMENT OF EARTHQUAKE-TRIGGERED LANDSLIDES HAZARD

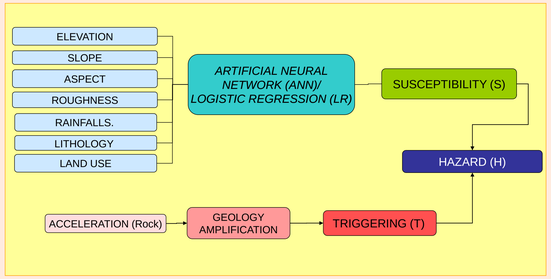


Figure 5. GIS Methodology for the assessment of earthquake-triggered landslides hazard

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 was estimated based on Logistic regression (LR) ***(Figure 7),*** and ANN model ***(Figure 8).*** Later, we calculate the Triggering map (T) in terms of PGA (Peak Ground Acceleration) for a return period of 475 years ***(Figure 9),*** with the local effects due to geology amplification ***(Table I)***. From the combination of both map, we obtained the earthquake-triggered landslides Hazard (H) using the GIS for both methods ***(Figure 10 and 11)*** .

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

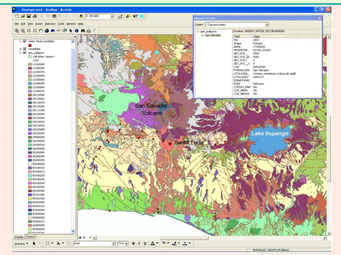


Figure 6. Window of the El Salvador GIS

SUSCEPTIBILITY

Figure 7. Susceptibility map based on Logistic Regression model

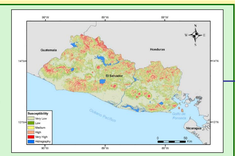
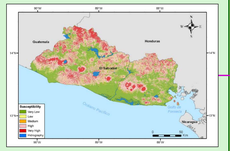
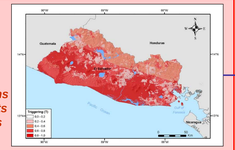


Figure 8. Susceptibility map Based on ANN model



TRIGGERING

Figure 9. Triggering Map in terms of PGA considering local effects for a return period of 475 years



EARTHQUAKE- TRIGGERED LANDSLIDES HAZARD MAP

HLR

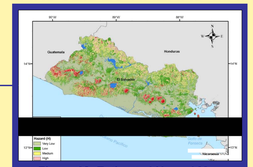


Figure 10. Estimated Hazard Map with LR model

HANN

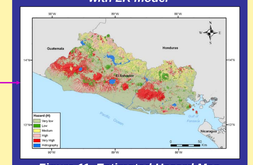


Figure 11. Estimated Hazard Map with ANN model

	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 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 LR model and ANN model.

ROC Curve (Receiver Operating Characteristic) visualize a classifier's performance in order to select the proper decision threshold. A probability of detection or sensibility versus probability of False Alarm curve. The results of both analysis, ROC Curve ***(Figure 12)*** and global percentage and AUCROC are shown in the ***Table II***.

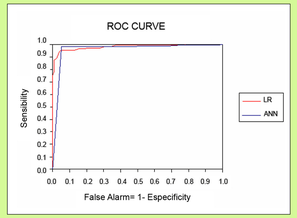


Figure 12. ROC curve of LR and ANN model

Table II. Global Percentage and AUCROC of LR and ANN model

	Global Percentage	*AUCROC*
LR Model	89.4%	98.0%
ANN Model	95.1%	96.3%

DISCUSSION

A landslide susceptibility map at a regional scale of El Salvador were derived by using LR and ANN models. 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 terrain roughness. The validation process was carried out using a ROC curve. Later, earthquake-triggered landslide hazard maps were calculated from the estimated Susceptibility (S) with LR and ANN models respectively, and Triggering (T) map (PGA for a return period of 475 years).

The *logistic regression technique* was applied. 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 approach, the probability was 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 lesser flexible and lesser capable of solving complex problems than ANN. The advantages of the *ANN method* and its possible application to the evaluation of landslide risk comes from the remarkable information processing characteristics of the artificial simulated biological system, including its 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 accuracy produced in our work with both methods was 95.1% for the ANN and 89.4% for the LR. For the ANN model, the black-box characteristic of the former did not allow investigating which variables were more influential in the response of variable. However, this is possible with LR because the weight of each factor is known. LR shows a more clearly defined area of susceptibility which corresponds to a steep ROC curve, while ANN gives a spread defined area of susceptibility which corresponds to a gradual rise of the ROC curve. In conclusion, there is no a preferred model for all cases: this depends on the number of training samples, the independent variables being considered, the scale, etc.

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.

Grupo de Ingeniería Sísmica (UPM)

CONTACT

ETSI Topografía, Geodesia y Cartografía,
Universidad Politécnica de Madrid, SPAIN
mjosegr@topografia.upm.es
+ 34 913366441