Revista de Matemática: Teoría y Aplicaciones 1–10
cimpa – ucr   issn: 1409-2433

Combining neural networks and geostatistics for landslide hazard assessment of San Salvador Department, El Salvador

Combinando redes neuronales y geoestadística para evaluación de deslizamientos de tierra en San Salvador,Departamento de El Salvador

Ricardo Ríos *[1] and Carlos Gámez **[2]

[1]Department of Mathematics, Science and Mathematics Faculty, University of El Salvador, El Salvador.
[2]Department of Mathematics, Science and Mathematics Faculty, University of El Salvador, El Salvador.

24 de mayo de 2015

Resumen

Esta contribución describe como se obtuvo un modelo de evaluación de deslizamiento de tierra para San Salvador, departamento de El Salvador. El análisis inicio con la obtención de una foto área del SNET-MARN con un total de 939407 puntos georeferenciados con el

*ricardo.rios.sv@gmail.com
**gamez9e@yahoo.com

fin de producir un inventario de deslizamiento. En esta evaluación de los deslizamientos se uso 4792 eventos previamente foto-interpretados y 7 factores condicionantes incluyendo: geomorfología, geología, precipitaciones máximas, aceleraciones sísmicas, pendiente del terreno, distancia a carretera y falla geológica. Redes Neuronales Artificiales(RNA) fueron usadas para la evaluación de la susceptibilidad a deslizamiento de tierra, logrando que más del 80 % de deslizamientos fueran apropiadamente clasificados usando un criterio dentro y fuera de la muestra con la que se estimaron los parámetros del modelo. Regresión Logística fue usada como base de comparación, obteniendo este modelo un rendimiento inferior que el de RNA con un porcentaje de correcta clasificación abajo del 70 %. Para completar el análisis se realizo la interpolación de puntos usando el método krigging proveniente del enfoque geoestadístico. Finalmente, los resultados muestran que es posible obtener un mapa de riesgo a deslizamiento de tierra, haciendo uso de una combinación de RNA y técnicas geoestadísticas con lo cual la presente investigación puede ayudar a la mitigación de deslizamientos de tierra en El Salvador.

Palabras clave: deslizamiento de tierra, evaluación de riesgo, San Salvador, El Salvador, RNA, geoestadística.

Abstract

This contribution describes how we obtained a landslide hazard assessment model for San Salvador, department in El Salvador. The analysis started with an aerial photointerpretation from SNET-MARN ((Servicio Nacional de Estudios Territoriales - Ministerio de Medio Ambiente y Recursos Naturales) with a total amount of 939407 georeferenced points to produce a landslide inventory. In this landslide assessment we have used 4792 events previously photo-interpretaded and 7 conditioning factors including: geomorphology, geology, rainfall intensity, peak ground accelaration, slope angle, road and fault distance. Artificial Neural Networks (ANNs) were applied for the assessment of susceptibility to landslides, achieving more than 80 % of landslide were properly classified using in-sample and out of sample criteria. Logistic regression was used as base of comparison, obtaining this model a performance lower than ANNs with a percentage of correct classification under 70 %. To complete the analysis we have performed interpolation of the points using kriging method from geostatistical approach. Finally, the results show that is possible to derive a landslide hazard map, making use of a combination of ANNs and geostatistical techniques wherewith the present study can help landslide mitigation in El Salvador.

Keywords: landslide, hazard assessment, San Salvador, El Salvador, ANN, geostatistics.

Mathematics Subject Classification: 62P12.

1. Introduction

El Salvador, one of the smallest and most crowded nations in Central America, extends about 240 kilometers westward from the Gulf of Fonseca to the border with Guatemala. This country is very vulnerable to landslide due to social, political and economic factors. In the year of 2001, El Salvador was struck by two devastating earthquakes within a month. One of the most spectacular aspects featuring heavily in the first earthquake was the damage inflicted by landslides. Among them, Las Colinas landslide was the most tragic. A huge amount of soil mass (about 200,000 m^3) was thrown off the rim of a mountain ridge rising south behind Las Colinas area of Nueva San Salvador (Santa Tecla), and flushed many houses and therefore more than 500 lives to death.

Due to the above it is necessary to implement mechanisms that allow us to quantify the hazard of a given geographic area to landslides, usually this is done with development of susceptibility maps which present in a graphical way the zones more susceptible to landslides and represent a practical tool for urban planning.

In this paper we propose an ANN model which enable to estimate the susceptibility to landslide in the georeferenced points belong to the Metropolitan Area of San Salvador (MASS) and later a map of susceptibility was generated by means of kriging method from geostatistical.

2. Brief State of Art

Since the pioneering work ([1]), several mathematics and statistics models have been proposed to model landslide susceptibility: deterministic models ([2], [3] and [4]), probabilistic models ([5], [6] and [7]) and neural networks ([8],[9] and [10]).

There are two important applications of the neural network models to landslide susceptibility. The first ([11])

Despite of the recent use of maps in those published works have not given enough attention on the use of geostatistical models to interpolate spatially distributed landslide susceptibility just landslide susceptibility is given in the georeferenced points and showed in the map of susceptibility.

3. Landslide inventory and data sources of input variables

The Landslide inventory was developed from photographic analysis on the study area, using aerial photo from SNET-MARN which is presented in the Figure 1, where white areas represent regions of landslide occurrence which were processed using the software ILWIS.

Figure 1: Aerial photo of MASS

Once that the white areas were georeferenced, a percentage of 0.5% of the total points georeferenced shows landslide ocurrence. In addition to the landslide information the following data sources of input variables was given by SNET-MARN:

1. **Geomorphology:** refers to landforms that result from lithospheric dynamics of geographic area.

2. **Slope:** derived from digital elevation model MARN-SNET

3. **Geology:** Description of the geology of the area from the map German Geological Mission (scale 1:100,000).

4. **Rainfall intensity:** Maximum rainfall recorded in the geographical area.

5. **Peak ground accelaration:** Maximum ground acceleration expressed in Gal for a return period of 500 years, this is the least that has detailed information. This information was obtained from RSIS II project.

6. **Road distance:** Distance in kilometers to the nearest road.

7. **Fault distance:** Distance in kilometers to the nearest fault. This information was obtained from German Geological Mission (scale 1:100,000).

4. Artificial Neural Network model for discrete choice

The logistic regression is a special case of neural network whose output variable is discrete, logistic regression represents a neural network with a neuron in the hidden layer. The following adaptation of a multilayeredfeed-forward artificial neural network known as MLP (Multilayer Perceptron) may be used for modeling binary classification model, where $x's$ are the observed values in the input variables, $w's$ and $\lambda's$ are the parameters of the model, p_i is the predicting probability for a network with k^* input characteristics and j^* neurons:

$$n_{j,i} = w_{j,0} + \sum_{k=1}^{k^*} w_{j,k} x_{k,i} \tag{1}$$

$$N_{j,i} = \frac{1}{1 + \exp^{-n_{j,i}}} \tag{2}$$

$$p_i = \sum_{j=1}^{j^*} \lambda_j N_{j,i} \tag{3}$$

$$\sum_{j=1}^{j^*} \lambda_j = 1, \lambda_j \geq 0 \tag{4}$$

In the context of the research problem p_i and the number k^* represent probability of landslide ocurrence and the number of input variables or causes associated with landslide ocurrence respectively.

The method used for estimating the parameters of the model was an Hybrid Method: First of all Genetic Algorithms were used to avoid getting stuck in a local rather than a global minimum of error function, after that a local search was applied again to obtain a new local minimun.

5. Spatial Prediction

In standard statistical problems, correlation can be estimated from a scatterplot, when several data pairs x, y are available. The spatial correlation between two observations of a variable $z(s)$ at locations s_1 and s_2

cannot be estimated, as only a single pair is available. To estimate spatial correlation from observational data, we therefore need to make stationarity assumptions before we can make any progress. One commonly used form of stationarity is intrinsic stationarity, which assumes that the process that generated the samples is a random function $Z(s)$ composed of a mean and residual:

$$Z(s) = \mu + \delta(s) \tag{5}$$

with a constant mean

$$E\left(Z(s)\right) = \mu \tag{6}$$

and a variogram defined as

$$\lambda(h) = \frac{1}{2} E\left(Z(s) - Z(s+h)\right)^2 \tag{7}$$

Ordinary kriging in terms of the covariance function

The predictor assumption is

$$Z(\hat{s_0}) = \sum_{i=1}^{n} w_i Z(s_i) \tag{8}$$

It is a weighted average of the sample values, and $\sum_{i=1}^{n} w_i = 1$ to ensure unbiasedness. The w_i's are the weights that will be estimated.

Kriging minimizes the mean squared error of prediction

$$min\ \sigma_e^2 = E\left[Z(s_0) - Z(\hat{s_0})\right]^2 \tag{9}$$

6. Results

The data were randomly divided into three sub-samples, the first was used for the estimation of the parameters of the neural network (training set), the second was used for choosing the model with more generalization capability (validation set) and the last was used to evaluate how well the model generalize outside of the data set used for estimation (test set).

To determine the number of neurons in the hidden layer was used the method of trial and error, starting with few neurons and increasing progressively the number of neurons in the hidden layer. Table 1 summarizes how well the models fit on the training, validation and test sets.

Table 1: Summary of classification accuracy on the training, validation and test sets for neural network models

Percentage score			
Número de neuronas	Train set	Validation set	Test set
2	0.7011	0.7020	0.7124
3	0.7206	0.7072	0.7401
4	0.7425	0.7197	0.7458
5	0.7601	0.7604	0.7740
6	0.7823	0.7704	0.7604
7	0.7853	0.7704	0.7763
8	0.7942	0.7865	0.7878
9	0.8009	0.7871	0.7901
10	0.8321	0.8132	0.8009
11	0.8194	0.7792	0.7542
12	0.8230	0.8006	0.8159
13	0.8408	0.8006	0.8031
14	0.8373	0.8017	0.8127
15	0.8517	0.8210	0.8247
16	0.8246	0.8017	0.8028
17	0.8543	0.8283	0.8246
18	0.8467	0.8022	0.8418
19	0.8341	0.7991	0.8274

The model with 17 neurons in the hidden layer was chosen, after that a logistic regression model was fitted as a basis of comparison. Table 2 presents the logistic regression fits, clearly the logistic regression's performance was lower than the worst neural network model.

Table 2: Summary of classification accuracy on the training, validation and test sets for logistic regression

Percentage score			
Model	Train set	Validation set	Test set
logistic regression	0.6710	0.6748	0.663

To generate the map of landslide hazards, the following steps were followed based on geostatistics methodology: Exploratory analysis, Variogram modelling and Spatial prediction using ordinary kriging and validation of this results. First of all one thousand data were used for the map generation process leaving the others for validation of the model. The Exploratory analysis showed the presence of spatial correlation in the data, this was done making scatter plot of pairs $Z(s_i)$ and $Z(s_j)$, grouped according to their separation distance, also the data did not show the presence of anisotropic effect. As regards the estimation of the variogram, the spherical, gaussian and exponential models were fitted but the validation of these models on data that were not taken into account in the process of estimating the variogram showed that the best model was the exponential, since that it presented a $R^2 = 0.42$ against $R^2 = 0.40$ and $R^2 = 0.39$ of the models spherical and gaussian respectively. The map of landslide hazards is showed in the figure 2.

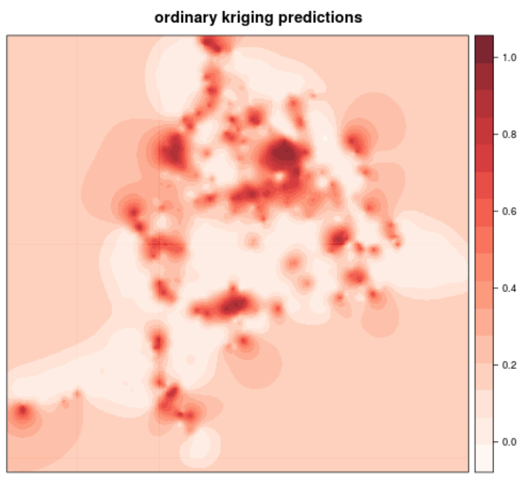

Figure 2: Map of landslide hazards

7. Discussion and conclusions

A landslide hazard assessment study was carried out in Metropolitan Area of San Salvador (MASS). The study started with the construction of a landslide inventory and analysis of the causal factors related to the occurrences of landslide. The problem of modelling landslide generated by different causes is very complex and for this reason the study proved the efficacy of the neural network model with a percentage of correct classification around 80% against other models such as logistic regression with a percentage of correct classification under 70%.

In the process of estimating the weights of the model an heuristic technique was used to obtain a better solution after that a local search was used. It is better than use only backpropagation algorithm or another local search method to estimate weights since that when these are used there is a strong danger of getting stuck in a local minimum rather than a global minimum for a vector of weights.

The problem with the neural network approach is that is difficult to estimate the weights of the model due to intensive computation involved in the present study the estimation of a neural network model takes between 4 and 10 hours for this reason we could not assess the statistical significance of the input variables in the neural network processes using boostrapping. For all of the above parallel computing must be used rather than serial computing in estimating weights of neural networks models.

The results obtained in the geostatistics methodology showed that the spatial the data satisfies the conditions for applying krigin method. The map of landslide hazard 2 was generated by krigin method and can be used by non-experts in the landslide phenomenon for many purposes such as territorial planning, prevention and mitigation of natural dissasters and so on.

Other important variables could not be obtained such as landslide types (Rockfalls, Topples, Slides, etc) and temporal information which can improve the predictive ability of the model.

8. Acknowledgment

The authors are indebted to M.SC. Giovanni Molina from SNET who provided the data and useful information about landslide modelling techniques.

References

[1] Alberto Carrara. Multivariate models for landslide hazard evaluation. *Journal of the International Association for Mathematical Geology*, 15 (3):403–426, 1983. ISSN 0020-5958. URL `http://dx.doi.org/10.1007/BF01031290`.

[2] G. Capparelli and P. P. Versace. Landslide susceptibility from mathematical model in sarno area. *Hydrology and Earth System Sciences Discussions*, 10(10):12643–12662, 2013. URL `http://www.hydrol-earth-syst-sci-discuss.net/10/12643/2013/`.

[3] Biswajeet Pradhan, Hyun-Joo Oh, and Manfred Buchroithner. Weights-of-evidence model applied to landslide susceptibility mapping in a tropical hilly area. *Geomatics, Natural Hazards and Risk*,

1(3):199–223, 2010. URL http://www.tandfonline.com/doi/abs/10.1080/19475705.2010.498151.

[4] Bettina Neuhäuser, Bodo Damm, and Birgit Terhorst. Gis-based assessment of landslide susceptibility on the base of the weights-of-evidence model. *Landslides*, 9(4):511–528, 2012. ISSN 1612-510X. URL http://dx.doi.org/10.1007/s10346-011-0305-5.

[5] Campbell R.H. Brookshire D.S. Bernknopf, R.L. and C.D Shapiro. A probabilistic approach to landslide hazard mapping in cincinnati, ohio, with application for economic evaluation. *Bulletin of the Association of Engineering Geologists*, XXV(1):39–56, 1988.

[6] Chang-JoF. Chung and AndreaG. Fabbri. Validation of spatial prediction models for landslide hazard mapping. *Natural Hazards*, 30 (3):451–472, 2003. ISSN 0921-030X. URL http://dx.doi.org/10.1023/B\%3ANHAZ.0000007172.62651.2b.

[7] S. Lee, J. Choi, and K. Min. Probabilistic landslide hazard mapping using gis and remote sensing data at boun, korea. *International Journal of Remote Sensing*, 25(11):2037–2052, 2004. URL http://www.tandfonline.com/doi/abs/10.1080/01431160310001618734.

[8] C. Melchiorre, E.A. Castellanos Abella, C.J. van Westen, and M. Matteucci. Evaluation of prediction capability, robustness, and sensitivity in non-linear landslide susceptibility models, guantánamo, cuba. *Computers & Geosciences*, 37(4):410 – 425, 2011. ISSN 0098-3004. URL http://www.sciencedirect.com/science/article/pii/S0098300410003286.

[9] Xu Zeng-wang. Gis and ann model for landslide susceptibility mapping. *Journal of Geographical Sciences*, 11(3):374–381, 2001. ISSN 1009-637X. URL http://dx.doi.org/10.1007/BF02892323.

[10] Artificial neural networks applied to landslide susceptibility assessment. *Geomorphology*, 66(1–4):327 – 343, 2005.

[11] E. Yesilnacar and T. Topal. Landslide susceptibility mapping: A comparison of logistic regression and neural networks methods in a medium scale study, Hendek region (Turkey). *Engineering Geology*, 79(11):251-266, 2005. ISSN 0013-7952, URL http://www.sciencedirect.com/science/article/pii/S0013795205000384.