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 Metropolitan Area, El Salvador

Combinando redes neuronales y geoestadística para evaluación de deslizamientos de tierra de el área Metropolitana de San Salvador, El Salvador

Ricardo Ríos *[1], Alexandre Ribó **[2], Roberto Mejía ***[3], and Giovanni Molina ****[4]

[1]Department of Mathematics, Science and Mathematics Faculty, University of El Salvador, El Salvador.
[2]National Institute of Health, Ministry of Health of El Salvador, El Salvador.
[3]National Institute of Health, Ministry of Health of El Salvador, El Salvador.
[4]Ministry of Environment and Natural Resources of El Salvador.

3 de junio de 2015

Resumen

Esta contribución describe la creación de 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 MARN (Ministerio de Medio Ambiente y Recursos Naturales) con un total de 939407 puntos georeferenciados con el 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 kriging 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, El Salvador, RNA, geoestadística.

Abstract

This contribution describes the creation of a landslide hazard assessment model for San Salvador, department in El Salvador. The analysis started with an aerial photointerpretation from MARN (Ministry of Environment and Natural Resources of El Salvador) 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 (ANN) were applied for the assessment of susceptibility to landslides, achieving more than 80 % of landslide were properly classified using insample and out of sample criteria. Logistic regression was used as base of comparison, obtaining this model a performance lower than ANNs

*ricardo.sv@gmail.com
**alexandre4rt@gmail.com
***robertomejia1685@gmail.com
****giova.molina@gmail.com

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, El Salvador, ANN, geostatistics.

Mathematics Subject Classification: 62P12.

1. Introduction

El Salvador, one of the smallest and most crowded nations in Central America, extends in Pacific coast about 240 kilometers westward from the Gulf of Fonseca to the border with Guatemala see figure 1. El Salvador borders an active subduction boundary between Cocos and Caribbean plates located 30 km offshore. Therefore El Salvador is affected with high seismicity and volcanic activity related to this active boundary. There is two main sources of seismic activity, the upper interface thrust that coincides with the location of the recent volcanoes and the Benioff-Wadaty zones of subducted Cocos plate where deeper intraplate earthquakes occurs ([1]). Due this tectonic and volcanic activity El Salvador has rugged relief. The main ranges cross the country with a rough west-east trend, parallel to the coast, these are separated from each other by faults and grabens. These ranges present several highly active volcanoes. The surface geology in El Salvador is almost entirely volcanic, dominated by upper Tertiary to Holocene volcanic rocks, only sparse outcrops of sedimentary and plutonic crops are located in the northern ranges, in the border with Honduras ([2]). Some of the most recent volcanic layers are formed by poor consolidated ashes and tuffs highly erodible ([3]).

Throughout the year the country experiences a tropical climate with two seasons, a dry season (November to April) and wet season (May - October). The climate of El Salvador is generally warm. In the dry season there is very less rainfall but during rainy seasons heavy showers take place. El Salvador's interior regions remain dry throughout the year. Periodically El Salvador is affected by tropical storms and occasionally by hurricanes.

Like in other parts of Central America, landslides in El Salvador constitute an important natural hazard due to prevailing steep terrain covered with poor consolidated volcanic materials and the frequent occurrence of extreme precipitation events and intense earthquakes. This problem is exacerbated by the extreme deforestation and the consequent high level of rates of erosion. Poverty, overpopulation and uncontrolled urbanization that characterizes the Salvadoran human settlements converts El Salvador in a country with high landslide risks. One example of this high risk is the devastating effect of the Las Colinas landslide triggered by a major earthquake (Mw 7.6) occurred on January of the 13, 2001 in Santa Tecla, a major city located close to San Salvador the Salvadoran capital ([4]). A huge amount of soil mass (about 200,000 m3) was thrown off the rim of El Balsamo range and flushed many houses and produced more than 500 deaths. Together with this event this earthquake triggered several landslides along the country, especially in the Metropolitan Area of San Salvador (MASS) ([5]). Other example is the large number of heavy-rainfall induced landslides occurred during Hurricane Mitch on October and November 1998 ([6]).

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. It was proposed an Artificial Neural Network model (ANN) which are a family of statistical learning models inspired by biological neural networks. The model proposed is used to estimate the susceptibility to landslide. From the results obtained by the model, a map was derived using kriging which is a method of interpolation spatial from the geostatistical approach.

2. Brief State of Art

Since the pioneering work ([7]), several mathematics and statistics models have been proposed to model landslide susceptibility: deterministic models ([8], [9] and [10]), probabilistic models ([11], [12] and [13]).

It has been used popular classification models such as logistic regression ([22] and [23]) , neural networks ([14],[15], [16]) and [17] and support vector machine ([20] and [21]).

According to ([24]) the magnitude of a possible slide is difficult to foresee as it depends on the magnitude of the triggering event and the environmental conditions (e.g., height of water table) at the moment of the event. Because of these complex relationships between the dependent variable and causal factors, and since that neural networks are particularly useful for detect complex non-linear relationships in large datasets, that is why this kind of model was chosen, despite the disadvantages such as greater computational burden and proneness to overfitting.

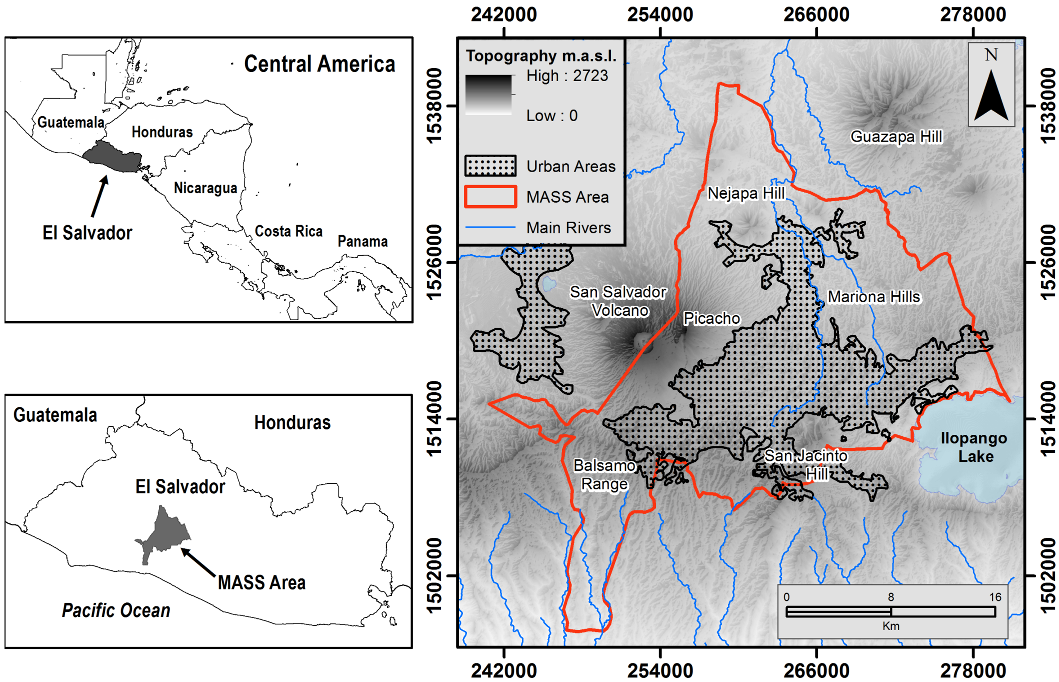

Figure 1: Maps from El Salvador

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 MARN which is presented in the Figure **??**, where white areas represent regions of landslide occurrence which were processed using the software ILWIS.

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 MARN:

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

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

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.

Before estimating the parameters of the neural network model, it is necessary standardize the input variables. In particular for classification problems is more suitable to scale inputs to $[-1, 1]$ rather than $[0, 1]$ ([19]). The following scaling was applied to each input variable:

$$x*_{k,i} = \frac{x_{k,i} - \mu}{\sigma} \tag{5}$$

Where μ and σ is respectively the mean and the standard deviation for the ith input variable applied to the kth case.

The method used for estimating the parameters of the model was an Hybrid Method ([18]): Firstly heuristic genetic algorithm using a package developed in the R statistical software specifically developed for this purpose, was used to obtain a good estimation of the parameters of the model.

The R package can be accessed from the following web address:

`https://goo.gl/PHaaG2`

Once a good estimate was obtained, this was occupied as the initial values for the conjugate gradient method implemented in the optim function of the R statistical software, to obtain a better estimation of the parameters of the model.

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:

$$Z(s) = \mu + \delta(s) \tag{6}$$

with a constant mean

$$E\left(Z(s)\right) = \mu \tag{7}$$

and a variogram defined as

$$\lambda(h) = \frac{1}{2}E\left(Z(s) - Z(s+h)\right)^2 \tag{8}$$

Ordinary kriging in terms of the covariance function

The predictor assumption is

$$\hat{Z(s_0)} = \sum_{i=1}^{n} w_i Z(s_i) \tag{9}$$

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

$$min\ \sigma_e^2 = E\left[Z(s_0) - \hat{Z(s_0)}\right]^2 \tag{10}$$

In order to make spatial predictions using ordinary kriging, an R script was developed which can be accessed in the following web address:

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.

In order to achieve the above steps a R scrit was developed which can be downloaded in the following address:

First of all one thousand fifty georeferenced point with their probability of ocurrence which was calculated using neural network, were selected randomly 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 showed a $R^2 = 0.42$ against $R^2 = 0.40$ and $R^2 = 0.39$ of the models spherical and gaussian respectively. With the exponential model were made spatial predictions on a grid of 1500×1500 and after that, the spatial predictions were imported to Geotiff raster format.

The map of landslide hazards is showed in the figure 2.

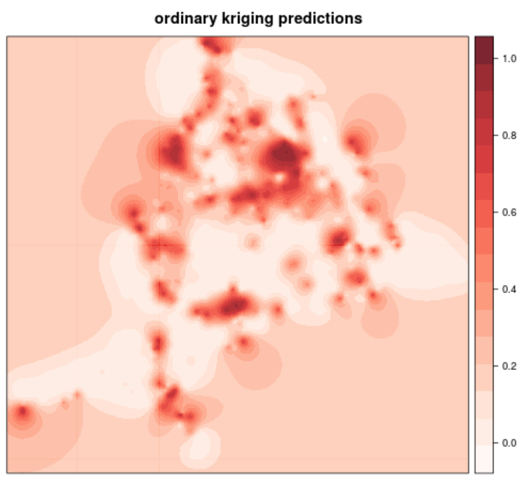

Figure 2: Map of landslide hazards

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 kriging method. The map of landslide hazard 2 was generated by kriging 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 MARN who provided the data and useful information about landslide modelling techniques.

References

[1] Dewey J.W. White R.A., Hernández D.A. Seismicity and tectonics of El Salvador *Chapter 27 in Geological Society of America Special paper 375 in Natural Hazards in El Salvador of Rose W.I., Bommer J.J., López D.L., Carr M.J. and Major J.J. 367-378. (2004)*

[2] Weber H.S., Wiesemann G., Lorenz W., Scmith-Thorné M. Mapa geológico de Republica de El Salvador escala 1 : 100 000 (6 sheets). *Bundesanstalt für Geowissenschaften und Rohstoffe, Hannover, Germany.*

[3] Bommer J.J., Rodriguez C.E. Earthquake-induced landslides in Central América *Engineering Geology.* (63):1889-220, 2002

[4] Evans S.G. Bent A.L. Las Colinas landslide, Santa Tecla: A higly destructive flowslide triggered by January 13, 2001, El Salvador earthquake. *Chapter 3 in Geological Society of America Special paper 375 in Natural Hazards in El Salvador of Rose W.I., Bommer J.J., López D.L., Carr M.J. and Major J.J. 25-37. (2004)*

[5] Jibson R.W., Crone A.J., Harp E.L., Baum R.L., Major J.J., Pullinguer C.R., Escobar D., Martinez M., Smith M.E. Landslides triggered by the 13 January and 13 February 2001, earthquakes. *Chapter 6 in Geological Society of America Special paper 375 in Natural Hazards in El Salvador of Rose W.I., Bommer J.J., López D.L., Carr M.J. and Major J.J. 69-88. (2004)*

[6] Crone A.J., Baum R.L., Lidke D.J., Sather D.N.D., Bradley L.-A., Tarr A.C. Landslides Induced by Hurricane Mitch in El Salvador-An Inventory and Description of Selected Features *USGS Open File Repord 01-444. (2002)*

[7] 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`.

[8] 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/`.

[9] 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`.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] Artificial neural networks applied to landslide susceptibility assessment. *Geomorphology*, 66(1–4):327 – 343, 2005. ISSN 0169-555X URL http://www.sciencedirect.com/science/article/pii/S0169555X04002272.

[17] 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.

[18] Paul D. McNelis Estimation of a Network with Evolutionary Computation *Academic Press*, (11):59-84, 2005. ISBN 9780124859678, URL http://www.sciencedirect.com/science/article/pii/B9780124859678500038.

[19] Warren S. Sarle URL `ftp://ftp.sas.com/pub/neural/FAQ.html#questions`. Accessed 31 May, 2015.

[20] Ballabio, Cristiano and Sterlacchini, Simone Support vector machines for landslide susceptibility mapping: the Staffora River Basin case study, Italy *Mathematical geosciences*, (11):1, 47–70, 2012. Springer.

[21] Tien Bui, Dieu and Pradhan, Biswajeet and Lofman, Owe and Revhaug, Inge Landslide susceptibility assessment in vietnam using support vector machines, decision tree, and Naive Bayes Models *Mathematical Problems in Engineering*, (11):2012, 2012. Hindawi Publishing Corporation.

[22] Akgun, Aykut A comparison of landslide susceptibility maps produced by logistic regression, multi criteria decision, and likelihood ratio methods: a case study at Izmir, Turkey *Landslides*, (11):9, 1, 93–106, 2012 Springer.

[23] Gaskill, Jacob and Zuber, Brian and Nordman, Erik Analyzing Landslide Susceptibility in St. Vincent and the Grenadines Using Co-Kriging and Logistic Regression *The 2015 IMAGIN Award, Michigan United States*. URL `http://www.imagin.org/awards/sppc/2015/papers/jacob_gaskill_paper.pdf`.

[24] Van Westen, CJ and Van Asch, Th WJ and Soeters, R Landslide hazard and risk zonation why is it still so difficult? *Bulletin of Engineering geology and the Environment*. (65):2, 167–184, 2006 Springer.