

Analysis of landslide hazard mapping of penang island Malaysia using bivariate statistical methods

Ilyas A Huqqani, Lea Tien Tay, Junita Mohamad Saleh

School of Electrical and Electronics Engineering, USM Engineering Campus, Malaysia

Article Info

Article history:

Received Jan 7, 2019

Revised Mar 18, 2019

Accepted Apr 11, 2019

Keywords:

Bivariate statistical

Frequency ratio

Information value

Landslide hazard mapping

Modified information value

ABSTRACT

Landslide is one of the disasters which cause property damages, infrastructure destruction, injury and death. This paper presents the analysis of landslide hazard mapping of Penang Island Malaysia using bivariate statistical methods. Bivariate statistical methods are simple approach which are capable to produce good results in short computational time. In this study, three bivariate statistical methods, i.e. Frequency Ratio (FR), Information Value (IV) and Modified Information Value (MIV) are used to generate the landslide hazard maps of Penang Island. These bivariate statistical methods are computed using MATLAB tool. Landslide hazard map is categorized into 4 levels of hazard. The accuracy of each method and effectiveness in predicating landslides are validated and determined by using Receiver of Characteristics curve. The accuracies of FR, IV and MIV methods are 79.58%, 79.14% and 79.37% respectively.

*Copyright © 2019 Institute of Advanced Engineering and Science.
All rights reserved.*

Corresponding Author:

Lea Tien Tay,

School of Electrical and Electronics Engineering,

USM Engineering Campus,

14300 Nibong Tebal, Penang, Malaysia.

Email: tay@usm.my

1. INTRODUCTION

Landslide is a geological phenomenon in which involves any movement of debris or soil down a slope or a mass of rock considered [1]. The probability of movement of the soil downhill under the influence of the gravity, precipitation and slope incline is also known as landslide [2]. It basically involves the shift of ground and rock due to soil erosion. Landslides have been declared as significant geological hazards and a natural source of devastation around the world. Many studies have been carried out by several researchers in many landslides hazardous area around the world. Landslides damage the infrastructures, properties and claim human lives. The main focuses of studies are inventory of landslides, impact of geological, topographical, hydrological and anthropogenic factors on the occurrences of landslides and the role of several triggering factors like rainfall and other calamities. There are several causative factors of landslide occurrences. These factors are divided into various categories like geomorphology, geology, metrological soil, land cover and hydrologic conditions [2-3].

Landslides occur frequently in Malaysia due to heavy rainfall throughout the year. Most of the landslides happen in annual monsoon i.e. Southwest monsoon (May – September) and Northeast monsoon (November – March). It has been noted that many damages have been done due to landslides in Malaysia from 2000 to 2009 [4-5]. A lot of research and methods have been proposed to reduce and minimize the damages caused by landslides [6-7]. This is accomplished by predicting the risky areas and act as a warning so that proper measures can be taken [4], [8-9]. It is not possible to predict a landslide occurrence in time and space, therefore a region of landslide is considered and is classified into different categories of potential hazards [2].

GIS and various evaluation techniques have been used for evaluating landslide hazards. The most attractive methods for computing the landslide hazards maps are bivariate statistical methods such as frequency ratio, statistical index and information value [6], [10-14]. Although advanced computational tools such as artificial neural network and decision tree have been applied successfully in generating landslide hazard maps, these techniques usually consume long processing time as compared to bivariate statistical methods [15-18]. Besides that, long computation time is also a challenging problem in the real-time online system. Up-to-date landslide map provides very crucial information for warning system which needs to be updated in short time. To overcome the computation time issue, bivariate statistical methods such as Frequency Ratio (FR) and Information Value (IV) are proposed in this paper for their easy implementation and fast in producing hazard maps with good accuracy [19]. This paper presents the analysis of landslide hazard mapping of Penang Island using bivariate methods, i.e. Frequency Ratio (FR), Information Value (IV) and Modified Information Value (MIV). These bivariate statistical approaches are computed using MATLAB software.

2. STUDY AREA

Due to frequent occurrences of landslide events in Penang, Malaysia, Penang Island is selected as the study area in this paper. Penang is situated on the North West of Malaysia Peninsula. It is bounded with the state of Kedah (North and East), the state of Perak (South) and the Strait of Malacca and Sumatra, Indonesia (West). Penang consists of Island as well as coastal strip on the mainland which is also named as Province Wellesley. In this research, the landslide hazard analysis study area is the island of Penang. The area of island of Penang is 285km². It is located at 5°15'N to 5°30'N latitudes and 100°10'E to 100°20'E longitudes. The land of Penang Island basically consists plants, forest, greenland, swamp and urban areas. The ranges of slope and elevation of the terrain above the sea level are 0° to 87° and 0 to 820m respectively. The temperature of Penang Island ranges from 29°C to 32°C. The average amount of precipitation lies between 2254mm to 2903mm yearly. The major types of vegetation in Penang Island are forest and fruit plantation. This island is also shaped by the fault lines. These fault lines are basically present from North to South in the center of the island. Figure 1 shows the map of Penang Island.



Figure 1. Penang Island (Google map)

3. DATA SET

Remote sensing method and Geographical Information System (GIS) are used to collect data on geological database of Penang Island. Numerous images of Penang Island are acquired from several governments sectors in Malaysia like Department of Meteorological Malaysia, Department of Irrigation and Drainage (DID), Department of Agriculture (DOA), Department of Minerals and Geo-Science (JMG) etc. The occurrence points of landslide in Penang Island are also accumulated and converted into spatial datasets. These dataset with resolution of 5 m is used in this study. The landslide occurrence points of the entire Island are collected. The hilly topography of the island is in the central of the island and considered as the most occurred landslide area. The landslides mainly involve shallow rotational debris slides, debris flows and rock falls.

The relationship of causative factors with slope failure and identification and mapping of instability should be done to know the main cause of landsliding [20]. Before applying the probabilistic approaches, the datasets of landslide causative parameters are gathered and transformed into spatial dataset. The twelve

(12) landslide causative parameter are taken in account in computing of Frequency Ratio (FR), Information Value (IV) and Modified Information Value (MIV) of landslides. By using topographic database, a digital elevation model (DEM) is developed with a resolution of 5 m. Elevation is obtained from DEM and then further used for calculating the slope gradient angle, slope aspect and slope curvature. The road map and drainage map are used to find the distance from road and drainage respectively. The buffer zones for line features should be prepared as 50 m [21-22]. The distances from fault lines are divided into 100m segregations. The soil texture of the island is also acquired, and it consists of six types of soil texture. Geology map covers six types of different granites of Penang Island. The natural triggering factors of landslide in island is precipitation. Due to rainfall, the soil dampens, debris and rocks wash away. The inverse weight distance interpolation method is used to produce the precipitation map because of limited rain measurement stations in the island. The landslide causative parameters maps are shown in the Figure 2. These maps are obtained by using the ARCGIS software.

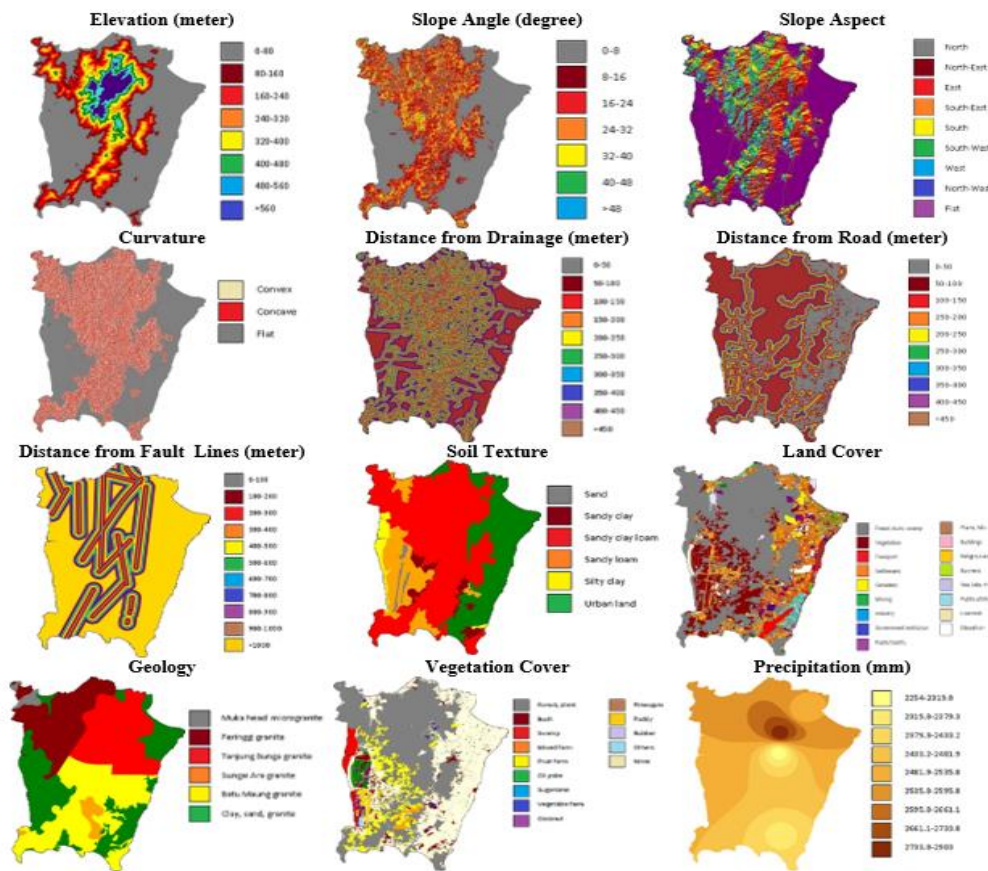


Figure 2. The maps of landslide causative parameters

4. METHODOLOGY

The bivariate statistical methods are used to produce the landslide hazard maps. These methods have several advantages such as the results obtained are highly efficient and accurate, low cost, straightforward execution and well perception of landslides and landslide causing factors relations. Each factors data weight is computed by using these methods. It depends on class allocation and landslide intensity [23]. These weights are very important in representing each factor to landslide occurrence. The landslide hazard index (LHI) is also produced by using factor’s data weights.

The main and important type of bivariate statistical is Frequency Ratio (FR) approach. It is based on the examined relationships between the landslide distribution and landslide contributing factors. The weightage of all class is obtained by dividing the class landslide density with the overall landslide density. If FR is greater than 1 then higher correlation of landslide occurs and vice versa. The FR of every pixel of the map is summed up to calculate LHI. It is concluded that the areas that have higher FRs are more inclined to landslide. FR and LHI is calculated by (1) and (2) respectively.

$$FR_i = \frac{N_{pix(S_i)}/N_{pix(N_i)}}{\sum N_{pix(S_i)}/\sum N_{pix(N_i)}} \quad (1)$$

$$LHI_{FR} = FR_{total} = \sum_{i=1}^n FR_i \quad (2)$$

where, $N_{pix(S_i)}$ Total number of pixels with landslide in class i ,
 $N_{pix(N_i)}$ Total number of pixels with class i in the study area,
 $\sum N_{pix(S_i)}$ Total number of pixels containing landslide,
 $\sum N_{pix(N_i)}$ Total number of pixels in the study area,
 n Number of classes of the causative factors.

Another model of bivariate statistical that is used in this study is an Information Value (IV) model. It is used to compute the information value of each class for all landslide causative factors. The (3) is utilized to form the information value for all classes of the factors [24-25].

$$IV_i = \ln \left(\frac{N_{pix(S_i)}/N_{pix(N_i)}}{\sum N_{pix(S_i)}/\sum N_{pix(N_i)}} \right) \quad (3)$$

The total information value for all pixels is computed using (4) as:

$$LHI_{IV} = IV_{total} = \sum_{i=1}^n IV_i \quad (4)$$

From (3), it is observed that if IV_i has a higher positive value then more landslides occurrence and vice versa. It is also noticed that a class has no landslide, IV_i is replaced by the minima of the minimum of each class. To overcome this issue, a new method is proposed. The model is known as Modified Information Value (MIV) model and can be computed in (5).

$$MIV_i = \log_2 \left(\frac{N_{pix(S_i)}/N_{pix(N_i)}}{\sum N_{pix(S_i)}/\sum N_{pix(N_i)}} + 1 \right) \quad (5)$$

$$LHI_{MIV} = MIV_{total} = \sum_{i=1}^n MIV_i \quad (6)$$

The main three conditions are observed from (5). First, if $MIV_i = 0$, then no landslide exists in a particular class. Second, if $MIV_i = 1$, it means that the proportion of landslide pixels in class i is equal to the proportion of the whole study area. Thirdly $MIV_i > 1$, it means that the more landslides pixels present in the class i than those in the whole study area. The higher the MIV, the greater landslides proportion in the class and vice versa.

5. RESULTS AND ANALYSIS

In this study, the accuracy (AUC) of landslide hazard mapping are verified by using Receiver Operating Characteristic (ROC) method. There are two assumptions taken to validate the maps. First, the landslides are linked to spatial data like topography height and slope gradient. Second, the triggering factors such as rainfall affects the future landslide [26].

The first step in verification by ROC method is to arrange the LHI in a descending sequence. These indexes are then segregated into one hundred (100) subclasses on y-axis with 1% separations on x-axis [4]. The curve illustrates the accuracy of the approaches and factors predict the upcoming landslides. The accuracy of the approach is obtained by area under the curve [26-27] The accuracies of each bivariate statistical method are obtained as 79.58% for FR, 79.14% for IV and 79.37% for MIV from the ROC curves shown in Figure 3. It is clearly shown that FR method is more accurate than the others. However, MIV model gives much closer accuracy with FR method than IV model. Landslide hazard maps are computed by LHI obtained from three bivariate statistical methods. LHI is classified into four categories.

- a) Highly hazardous for the maximum 10% of LHI (90-100%)
- b) Hazardous for the following 10% of LHI (80-90%)
- c) Moderately hazardous the next 20% of LHI (60-80%)
- d) Non- hazardous for the remaining 60% of LHI (0-60%)

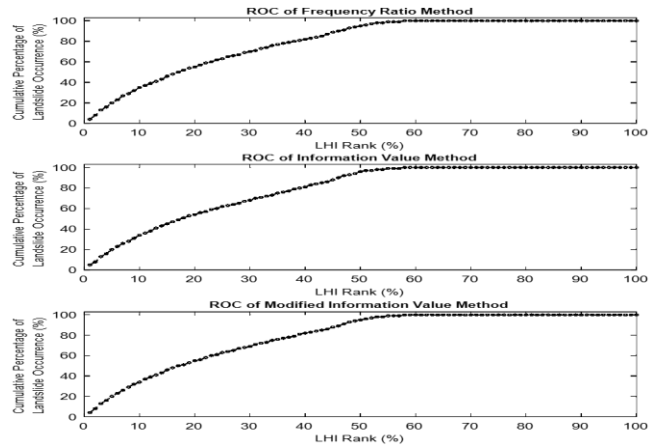


Figure 3. ROC curves of FR, IV and MIV methods

Landslide hazard maps of Penang Island generated using FR, IV and MIV are shown in Figure 4 (a), (b) and (c) respectively. In these figures, red color indicates for highly/seriously hazardous area, green color for hazardous areas, blue color for moderately hazardous areas and white color for non-hazardous areas.

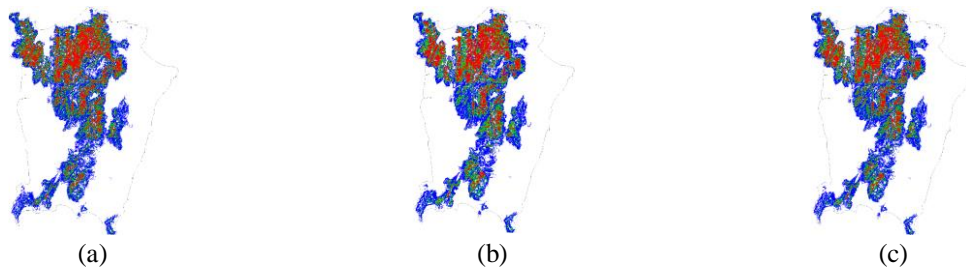


Figure 4. Landslide hazard maps (a) FR (b) IV (c) MIV

6. CONCLUSION

In Malaysia, landslide has caused many casualties and damages. It is one of the severe threatening disaster occurs in nature and is generally originated under complex environmental situations. However, a lot of construction and development has carried out in the hilly regions due to the limited flat terrain in Penang Island. Therefore, a landslide alerting scheme is regularized to warn people about the occurrence of landslides. In this paper, landslide hazard mapping of Penang Island is analyzed using bivariate statistical approaches. Bivariate statistical approaches are simple methods which take short computation time and produce results with good accuracy. FR, IV and MIV models are used to compute the landslide hazard maps of Penang Island. The accuracies of FR (79.58%), IV (79.14%) and MIV (79.37%) models are calculated by applying the ROC curve. The bivariate methods applied in this paper, i.e. frequency ratio, information value and modified information value provide good results in predicting the landslide events.

ACKNOWLEDGEMENTS

The author would like to thank Kementerian Pendidikan Malaysia (KPM) for providing the technical support and financial support (FRGS – Geran Penyelidikan Fundamental 203/PELECT/6071390) in this project.

REFERENCES

[1] D. M. Cruden, "A simple definition of a landslide," *Bull. Int. Assoc. Eng. Geol. - Bull. l'Association Int. Géologie l'Ingénieur*, vol. 43, no. 1, pp. 27–29, 1991.
 [2] D. J. Varnes, "Landslide hazard zonation: a review of principles and practice," *Natural Hazards*, vol. 3(63), 1984.

- [3] J. N. Hutchinson, "Landslide Hazard Assessment. Keynote Paper. In: Bell, D.H., Ed., Landslides," in *6th International Symposium on Landslides*, pp. 1805-1841, 1995.
- [4] B. Pradhan and S. Lee, "Delineation of landslide hazard areas on Penang Island, Malaysia, by using frequency ratio, logistic regression, and artificial neural network models," *Environ. Earth Sci.*, vol. 60(5), pp. 1037-1054, 2010.
- [5] Satoshi Murakami, Tay Lea Tien, Rohayu Bte Che Omar, Tomomi Nishigaya, Nazirah Aziza, R. Roslan, I. N. Z. Baharuddin, Habibah Hj Lateh, Naoki Sakai, "Landslides Hazard Map in Malay Peninsula by Using Historical Landslide Database and Related Information," *Journal of Civil Engineering Research*, vol. 4(3A), pp. 54-58, 2014. DOI: 10.5923/c.jce.201402.08
- [6] G. Q. Anh, D. T. Tran, D.C. Nguyen, T. B. Dieu, "Flexible Configurations of Wireless Sensor Network for Monitoring of Rainfall-Induced Landslide", *Indonesian Journal of Electrical Engineering and Computer Science (IJECS)*, vol 12, no. 3, pp 1030-1036, Dec 2018. DOI: 10.11591/ijeecs.v12.i3.pp1030-1036
- [7] N. Ya'acob, N. Tajudin, A.M Azize, "Rainfall-landslide early warning system (RLEWS) using TRMM precipitation estimates", *Indonesian Journal of Electrical Engineering and Computer Science (IJECS)*, vol 13, no. 3, pp 1259-1266, Mar 2019. DOI: 10.11591/ijeecs.v13.i3.pp1259-1266
- [8] S. Lee and J. Ab Talib, "Probabilistic landslide susceptibility and factor effect analysis," *Environ. Geol.*, vol. 47, pp. 982-990, May 2005.
- [9] Lea Tien Tay, Mutasem Sh. Alkhasawneh, Habibah Lateh, Md Kamrul Hossain, "Landslide Hazard Mapping of Penang Island Using Poisson Distribution with Dominant Factors," *J. Civ. Eng. Res.*, 2014.
- [10] S. Lee and B. Pradhan, "Landslide hazard mapping at Selangor, Malaysia using frequency ratio and logistic regression models," *Landslides*, vol. 4, no. 1, pp. 33-41, 2007.
- [11] K. L. K.- Wern, T. L. Tien, and H. Lateh, "Landslide hazard mapping of Penang island using probabilistic methods and logistic regression," in 2011 IEEE International Conference on Imaging Systems and Techniques, pp. 273-278, 2011.
- [12] Q. Wang et al., "Landslide Susceptibility Mapping Based on Selected Optimal Combination of Landslide Predisposing Factors in a Large Catchment," *Sustainability*, vol. 7, no. 12, pp. 16653-16669, 2015.
- [13] Lea Tien Tay, Mutasem Sh. Alkhasawneh, Umi Kalthum Ngaha and Habibah Lateh, "Landslide Hazard Mapping with Selected Dominant Factors: A Study Case of Penang Island, Malaysia", *American Institute of Physics (AIP) Conference Proceedings (ICOMEIA 2014)* ISBN : 978-0-7354-1304-7, DOI: 10.1063/1.4915868, Volume number: 1660, <http://dx.doi.org/10.1063/1.4915868>, 2015.
- [14] Lea Tien Tay, Habibah Lateh, Md Kamrul Hossain, Anton Abdulbasah Kamil, "Landslide Hazard Mapping Using Poisson Distribution: A Case Study in Penang Island, Malaysia", *Landslide Science for a Safer Geoenvironment*, vol. 2, Part IV, pp 521-525, 2014. doi 10.1007/978-3-319-05050-8_80.
- [15] Mutasem Sh. Alkhasawneh, Umi Kalthum Ngah, Lea Tien Tay, Nor Ashidi Mat Isa, Mohammad Subhi Al-batah, "Determination of Important Topographic Factors for Landslide Mapping Analysis Using MLP Network". *The Scientific World Journal*, Vol. 2013, Article IS 415023, 12 pages, <http://dx.doi.org/10.1155/2013/415023>
- [16] Mutasem Sh. Alkhasawneh, Umi Kalthum Ngah, Lea Tien Tay, Nor Ashidi Mat Isa, "Determination of importance for comprehensive topographic factors on landslide hazard mapping using artificial neural network", *Environmental Earth Sciences*, DOI 10.1007/s12665-013-3003-x, 2013
- [17] Mutasem Sh. Alkhasawneh, Umi Kalthum Ngah, Lea Tien Tay, Nor Ashidi Mat Isa, Mohammad Subhi Al-Batah, "Modeling and Testing Landslide Hazard Using Decision Tree", *Journal of Applied Mathematics*, vol. 2014, Article ID 929768, 9 pages, <http://dx.doi.org/10.1155/2014/929768>
- [18] Mutasem Sh. Alkhasawneh, Lea Tien Tay, Umi Kalthum Ngah, Mohammad Subhi Al-batah, Nor Ashidi Mat Isa, "Intelligent landslide system based on discriminant analysis and cascade-forward back-propagation network", *Arabian Journal for Science and Engineering (AJSE)*, April 2014, DOI 10.1007/s13369-014-1105-8
- [19] Biswajeet Pradhan, and Saro Lee, "Probabilistic landslide hazards and risk mapping on Penang Island, Malaysia," *J. Earth Sys. Sci.*, vol. 115, no. 6, pp 661-672, December 2006.
- [20] F. Guzzetti, A. Carrara, M. Cardinali, and P. Reichenbach, "Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy," *Geomorphology*, vol. 31, no. 1, pp. 181-216, 1999.
- [21] C. J. van Westen, N. Rengers, and R. Soeters, "Use of Geomorphological Information in Indirect Landslide Susceptibility Assessment," *Nat. Hazards*, vol. 30, no. 3, pp. 399-419, Nov. 2003.
- [22] Van Westen, C.J., "Application of Geographic Information System to landslide hazard zonation," *ITC-Publication*, no. 15, pp. 245, 1993.
- [23] M. L. Süzen and V. Doyuran, "A comparison of the GIS based landslide susceptibility assessment methods: multivariate versus bivariate," *Environ. Geol.*, vol. 45, no. 5, pp. 665-679, Mar. 2004.
- [24] W. Chen et al., "Landslide susceptibility mapping based on GIS and information value model for the Chencang District of Baoji, China," *Arab. J. Geosci.*, vol. 7, pp. 4499-4511, 2014.
- [25] K. L Yin and T. Z Yan, "Statistical Prediction models for slope instability of metamorphosed rocks," in In: *Proceedings of 5th Int Symp on Landslides, Lausanne, Switzerland*, vol. 2, 1988.
- [26] C.-J. F. Chung and A. Fabbri, "Probabilistic prediction models for landslide hazard mapping," *Photogramm. Eng. Remote Sensing*, vol. 65, pp. 1389-1399, 1999.
- [27] Beguria S, "Validation and evaluation of predictive models in hazard assessment and risk management," *Nat Hazards* 37, pp 315-329, 2006.