



Landslide Hazard Mapping Using a Poisson Distribution: A Case Study in Penang Island, Malaysia

Lea Tien Tay, Habibah Lateh, Md Kamrul Hossain,
and Anton Abdulbasah Kamil

Abstract

Landsliding is one of the most destructive natural geohazards in Malaysia. Landslide hazard maps are very useful for urban development planning. This paper presents landslide hazard mapping using a new approach, i.e. a Poisson distribution, and compares the result with previous probabilistic approaches, i.e. frequency ratio (FR), statistical index (SI) and landslide nominal susceptibility factor (LNSF). These approaches were implemented in Penang Island to produce landslide hazard maps. The landslide causative factors considered are elevation, slope gradient, slope aspect, curvature, land cover, vegetation cover, distance from nearest road, distance from nearest stream, distance from nearest fault line, geology, soil texture and precipitation. Landslide hazard maps were assessed using the Receiver Operating Characteristics (ROC) method. Accuracy obtained for FR, SI and LNSF are 78.52 %, 78.12 % and 72.93 % respectively. Poisson distribution approach gives high accuracy of 78.51 % as FR.

Keywords

Landslide • Frequency ratio • Susceptibility • Hazard map • Poisson distribution

Introduction

Landslides are a geological phenomenon that involves downward and outward movement of rock, earth or debris. Landslides are hazardous, as they destroy properties and even claim human lives. In Malaysia, landslides happen rather frequently due to heavy rainfall during the annual monsoons: the Southwest Monsoon from late May to September and the Northeast Monsoon from November to March. In recent years from 2000 to 2009, damage due

to landslides has been particularly high (Pradhan and Lee 2010). Although it is difficult to predict landslide events, areas may be divided into different ranks according to their degree of potential hazard due to mass movement (Varnes 1984). The identification of high risk areas is important in landslide prediction and warning systems. In this paper, Penang Island is the selected area of interest for landslide hazard analysis, as it has suffered substantial damage due to landslides in recent years. In recent years, there have been many studies of landslide hazard evaluation using GIS and various evaluation techniques. Probabilistic methods such as frequency ratio have been applied many times to achieve favorable results in landslide hazard evaluation (Pradhan and Lee 2010; Lee and Talib 2005). The landslide nominal susceptibility factor has also been used to generate landslide susceptibility and hazard maps (Saha et al. 2005). To the best of our knowledge, this study is the first time the Poisson distribution has been used for landslide hazard mapping.

L.T. Tay (✉)
Universiti Sains Malaysia, Engineering Campus, 14300 Penang,
Malaysia
e-mail: tay@usm.my

H. Lateh • M.K. Hossain • A.A. Kamil
School of Distance Education, Universiti Sains Malaysia, 11800 USM
Penang, Malaysia
e-mail: habibah@usm.my

Study Area

The area selected for this study is Penang Island (Fig. 1), due to its heavy frequency of landslides over the years. Penang is one of the 13 states of Malaysia located on the North West of Peninsular Malaysia. It is bounded to the north and east by the state of Kedah, to the south by the state of Perak and to the west by the Strait of Malacca. Penang consists of both the island of Penang and a coastal strip on the mainland known as Province Wellesley. In this paper, the island of Penang is considered for landslide evaluation mapping. The island, which is separated from the mainland by a channel, covers an area of 285 km². It is located about 5° 15'N to 5° 30'N latitude and 100° 10'E to 100° 20'E longitude. Elevation of the terrain ranges from 0 to 820 m above sea level and slope gradient ranges from 0° to 87°. Penang Island is affected by fault lines that run from north to south mainly in the centre of the island. The land cover of Penang Island consists of mainly swamp, plantation, forest, grassland and urban area. Vegetation cover consists mainly of forest and fruit plantations. The temperature of Penang Island ranges from between 29 to 32 °C and the average amount of rainfall varies from 2,254 to 2,903 mm annually.

Data Sets

Data collection on a geographical database of Penang Island is done through application of remote sensing methods and GIS. Topographical, geological and various images of Penang Island were obtained from various departments in Malaysia including Meteorological department, Jabatan Pengairandan Saliran (JPS), Department of Agriculture (DOA), Minerals and Geoscience Department (JMG), Jabatan Ukur dan Pemetaan Malaysia (JUPEM) and Pusat PeGIS. Landslide occurrence points in Penang Island are also collected and transformed into a spatial database for probabilistic landslide susceptibility analysis. Most landslides happened in the mountainous terrain of the island, which is situated in the middle of the island. These landslides consist of mainly shallow rotational debris slides, debris flows and rock falls. Identification and mapping of a suitable set of instability and causative factors having a relationship with slope failures requires prior knowledge of the main causes of landslides (Guzzetti et al. 1999). To apply the probabilistic methods, data sets of landslide-causative factors were collected and constructed as a spatial database. Twelve landslide-causative factors were considered in calculating the probability of landslides.

From the topographic database, digital elevation models (DEM) were constructed with a resolution of 5-m. From the DEM, elevation was extracted. Elevation was then used to compute slope angle, slope aspect and slope curvature.



Fig. 1 Penang Island (Source. Google map)

Distance from drainage and distance from road were calculated from the drainage map and the road map, each available as digital maps respectively. Van Westen et al. (2003) suggested that buffer zones for line features, such as, rivers and roads should be set to 50 m. Similarly, distances from fault lines of Penang Island were calculated and segregated into 100 m intervals. A soil texture database of the island, which included sand, clay and urban land, was obtained as well. The land-use map consists of 17 classes of land usage. The vegetation cover map consists of 14 classes. The geologic map, showing the types of rocks and granites of Penang Island is used as one of the landslide-causative factors. Rainfall or precipitation is one of the most efficient triggering factors of landslides, as it dampens soils and washes away debris and rocks, creating landslides. There are only a few rain gauge stations in Penang Island, therefore, interpolation had to be used to prepare the precipitation map. Using 29 years of historical rainfall data (1980–2008), statistical distribution of the accumulated average precipitation was prepared using ArcGIS software's inverse weight distance interpolation method.

Methods

The probabilistic approach to landslide hazard analysis has been proven to be very useful in landslide prediction. The advantages of this approach lie in its high efficiency, low

cost, easy implementation and a better understanding of the relationships between landslides and landslide-causative factors. In probabilistic methods, each factor's data weights are calculated based on class distribution and its landslide density (Suzen and Doyuran 2004). These weights represent the importance of each factor to landslide occurrence and are used to produce a landslide hazard index.

Frequency Ratio

The frequency ratio (FR) approach is a variant of the probabilistic method that is based on the observed relationships between the distribution of landslides and landslide-causative factors. Spatial relationships and correlations between landslide occurrence and its factors are computed and these correlations show how closely related the factors are to landslide occurrence. Frequency ratio or weight of each class is computed by dividing the landslide density of a class with the total landslide density as follows

$$FR_{ab} = \frac{Aab^*}{Aab} \times \frac{A}{A^*}$$

FR = Frequency ratio of each factor.

Aab* = Area of observed landslides of class a of factor b.

Aab = Area of class a of factor b.

A = Total area of the map.

A* = Total area of observed landslides of the map.

A value of 1 is the neutral value and values higher than 1 show a higher positive correlation between the class of a factor with the landslide. To calculate the landslide hazard index (LHI), each pixel's frequency ratio on the map is summed as in the equation below. It is then inferred that areas with higher values of frequency ratio are areas that are more prone to landslides.

$$LHI(\text{Landslide hazard index}) = FR_1 + FR_2 + \dots + FR_N$$

where FR = Frequency ratio of each factor.

Statistical Index

Statistical index (SI) is another probabilistic approach that is very similar to frequency ratio. It was introduced by Van Westen and Terlien (1996). In recent years, researchers have applied this method for landslide susceptibility analysis (Long 2008) and termed the method statistical index. In this method, the weight for every class of each factor is defined as the natural logarithm of the landslide density in the class divided by the total landslide density.

Mathematically, it is almost the same as frequency ratio, as the method simply takes the natural logarithm of the frequency ratios, as shown in equation below

$$SI_{ab} = \ln\left(\frac{Aab^*}{Aab} \times \frac{A}{A^*}\right)$$

Summation of the statistical index of each factor in every pixel produces the landslide hazard index (LHI) as in equation below.

$$LHI = SI_1 + SI_2 + \dots + SI_N$$

SI = statistical index weight of each factor.

For ratios less than 1, the weight will be negative and vice versa. Negative values show decreasing correlation between the class with landslide and vice versa. In the case of no landslide occurrence, the weight of that particular class is assigned as 0. Thus, a class that has a weight of 0 shows no correlation with landslide occurrence and does not contribute to the landslide hazard index.

Landslide Nominal Susceptibility Factor

The landslide nominal susceptibility factor (LNSF) approach (Saha et al. 2005) is a modification of a method originally proposed by Gupta and Joshi (1990). The nominal risk factor is defined as follows

$$LNSF_i = Npix(S_i) * n / \left(\sum Npix(S_i) \right)$$

where Npix (Si) = number of pixels of landslide in the thematic class of i

n = number of thematic classes of respective factor

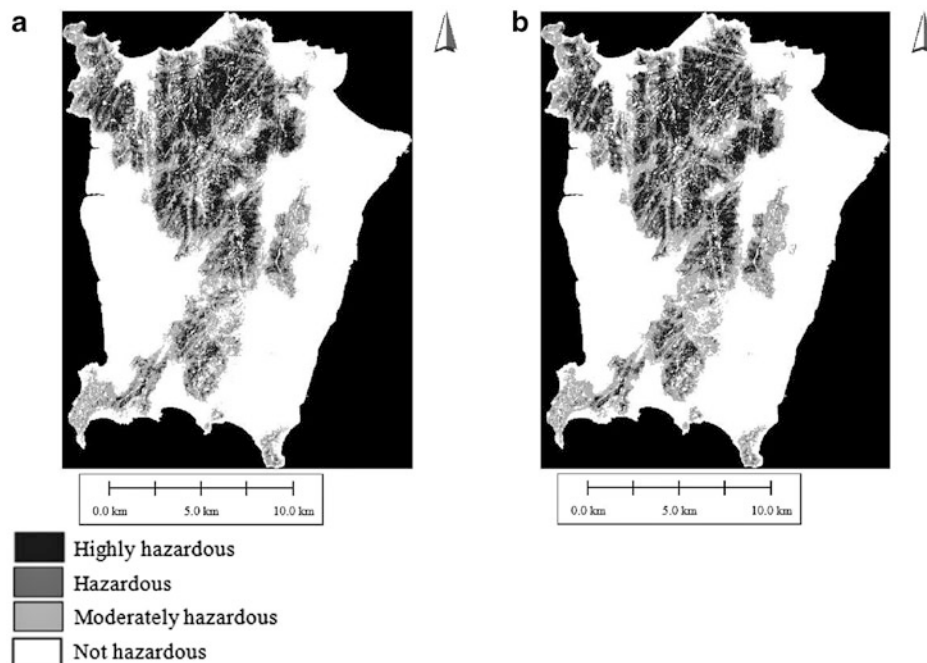
A value greater than one for LNSF means more susceptible to landsliding, whereas a LNSF value less than one means less susceptible. The nominal risk factors of each factor are summed to produce the Landslide Hazard Index (LHI) for the LNSF method.

Poisson Distribution

The Poisson distribution is a discrete distribution (Haight 1967), and it is useful in ecological studies. In this study, the Poisson distribution is used for "landslides" which is a geological event. The distribution function for the Poisson distribution is defined as

$$P[X] = \frac{\exp(-\theta) * \theta^X}{X!}$$

Fig. 2 Landslide hazard maps for Penang Island generated using (a) Frequency ratio and (b) statistical index



To estimate the probability of one landslide under a characteristic, the Poisson distribution is simplified as follows:

$$P[x = 1] = \exp(-\theta) * \theta$$

where

$$\theta = \frac{\text{Number of Landslide}}{\text{Total Number of Observations}}$$

Higher values of the probability for a characteristic shows a high risk of landslide occurrence.

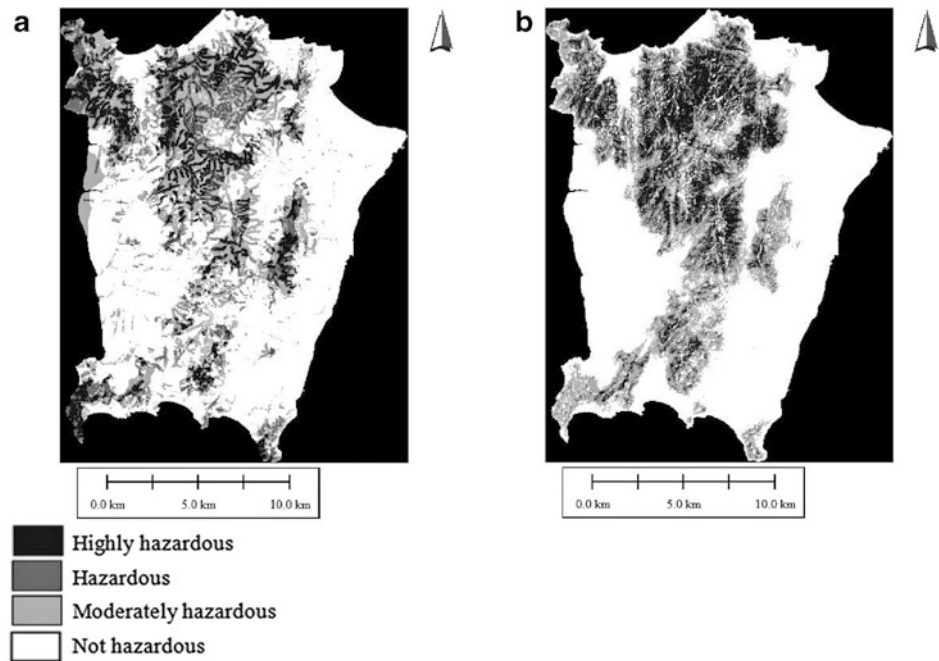
Verification and Discussion

To verify the efficiency and accuracy of each method, the ROC (receiver operating characteristics) method was employed. Two assumptions are needed in order to verify the maps. One is that landslides are related to spatial information such as slope and aspect, and the other is that future landslides will be affected by triggering factors such as rainfall (Chung and Fabbri 1999). Both assumptions are satisfied in this study. ROC curves explain how well the prediction models fit the data of landslide-causative factors with landslide occurrences to predict future landslides. Verification by ROC is done by first sorting the landslide hazard indexes (LHI) in a descending order. The ordered indexes are then divided into 100 classes and set on the y-axis, with

accumulated 1 % intervals on the x-axis (Pradhan and Lee 2010). The resulting graph shows a curve that explains how well the model and factors predict future landslides (Chung and Fabbri 1999). The area under the curves reflects the accuracies of the prediction models and is one of the commonly used accuracy statistics in natural hazard assessments (Beguria 2006). Landslide hazard maps produced using frequency ratio, statistical index, LNSF and Poisson distribution achieve accuracies of 78.52 %, 78.12 %, 72.93 % and 78.51 %, respectively. Frequency ratios and Poisson distributions are capable of producing a landslide hazard map with the highest accuracy among these four approaches in this study area. Both approaches work equally well, with most of the landslide locations found to be within the highest 20 % of the landslide hazard index. The LNSF approach is found to be less effective in producing landslide hazard maps as compared to the other three approaches.

Landslide hazard mapping is done by segregating the LHI (landslide hazard index) into several categories of risks. In this study, the indexes are categorized into four groups: Highly hazardous for the highest 10 % of the indexes (90–100 %), Hazardous for the next 10 % (80–90 %), Moderately hazardous for the next 20 % (60–80 %) and Not hazardous for the remaining 60 % (0–60 %). Figure 2 shows the landslide hazard maps produced using the frequency ratio method and statistical index. Landslide hazard maps produced using the landslide nominal susceptibility factor and Poisson distribution are shown in Fig. 3.

Fig. 3 Landslide hazard maps for Penang Island generated using (a) landslide nominal susceptibility factor and (b) poisson distribution



Conclusion

Due to the limitation of flat land in Penang Island, more hill areas are being exploited for housing and development. This creates a scenario in which a landslide assessment is needed to warn people of potential landslides in risky areas. Prediction models such as statistical or probabilistic methods are simple methods which have been proven to yield good results in landslide prediction. Landslide causative factors depend on the availability of GIS images of the research area. In this paper, landslide hazard maps of Penang Island were produced with frequency ratios, statistical indexes, nominal susceptibility factors and Poisson distributions using twelve landslide-causative factors. The frequency ratio and Poisson distribution methods are proven to produce landslide hazard maps with the highest accuracy, followed by statistical index and nominal susceptibility factor. The accuracies show that the Poisson distribution also can be used as a tool for hazard mapping and these maps can be useful in the planning of future urban infrastructure development.

Acknowledgments The authors would like to thank Jabatan Pengairan dan Saliran (JPS) and Minerals and Geoscience Department (JMG) for the data provided. The authors would also thank Universiti Sains Malaysia for the facilities provided.

References

- Beguria S (2006) Validation and evaluation of predictive models in hazard assessment and risk management. *Nat Hazards* 37:315–329
- Chung CF, Fabbri AG (1999) Probabilistic prediction models for landslide hazard mapping. *Photogramm Eng Remote Sensing* 65 (12):1389–1399
- Gupta RP, Joshi BC (1990) Landslide hazard zonation using the GIS approach—a case study from the Ramganga catchment, Himalayas. *Eng Geol* 28:119–131
- Guzzetti F, Carrara A, Cardinali M, Reichenbach P (1999) Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy. *Geomorphology* 31:181–216
- Haight FA (1967) *Handbook of the poisson distribution*. Wiley, New York, 168 p
- Lee S, Talib JA (2005) Probabilistic landslide susceptibility and factor effect analysis. *Environ Geol* 47:982–990
- Long NT (2008) Landslide susceptibility mapping of the mountainous area in a Luoi district, ThuaThien Hue province. *Vrije Universiteit Brussel, Vietnam*
- Pradhan B, Lee S (2010) Delineation of landslide hazard areas on Penang Island, Malaysia, by using frequency ratio, logistic regression, and artificial neural network models. *Environ Earth Sci* 60 (5):1037–1054
- Saha AK, Gupta PR, Sarkar I, Arora MK, Csaplovics E (2005) An approach for GIS-based statistical landslide susceptibility zoning—with a case study in the Himalayas. *Landslides* 2(1):61–69
- Suzen ML, Doyuran V (2004) A comparison of the GIS based landslide susceptibility assessment methods: multivariate versus bivariate. *Environ Geol* 45:665–679
- Van Westen CJ, Terlien MJT (1996) An approach towards deterministic landslide hazard analysis in GIS: a case study from Manizales, Colombia. *Earth Surf Process Landf* 21(9):853–868
- Van Westen CJ, Rengers N, Soeters R (2003) Use of geomorphological information in indirect landslide susceptibility assessment. *Nat Hazards* 30:399–419
- Varnes DJ (1984) *Landslide hazard zonation: a review of principles and practice*. UNESCO, Paris, 60 pp