

Use of Remote Sensing Data and GIS to Produce a Landslide Susceptibility Map of a Landslide Prone Area Using a Weight of Evidence Model

Biswajeet PRADHAN^{a*}, Hyon-Joo OH^b, and Manfred BUCHROITHNER^c

^{a,c}*Institute for Cartography, Faculty of Forestry, Geo and Hydro-Science,
Dresden University of Technology, 01062 Dresden, Germany;
Email. Biswajeet24@gmail.com*

^b*Korea Institute of Geoscience & Mineral Resources (KIGAM),
30, Gajeong-Dong, Yuseong-Gu, Daejeon, 305-350, Korea*

Abstract. Preparation of landslide susceptibility maps is important for engineering geologists and geomorphologists. However, due to complex nature of landslides, producing a reliable susceptibility map is not easy. In this paper, the weights-of-evidence model (a Bayesian probability model) was applied to the task of evaluating landslide susceptibility using GIS. Using landslide location and a spatial database containing information such as topography, soil, landcover, geology, and lineament, the weights-of-evidence model was applied to calculate each relevant factor's rating for the studied area. In the topographic database the factors were slope, aspect, distance to road, and curvature; in the soil database they were soil texture, soil material, and topographic type; lithology was derived from the geological database; land-cover information extracted from Landsat TM satellite imagery; and lineament data derived from SPOT 5 satellite imagery. Tests of conditional independence were performed for the selection of factors, allowing 33 combinations of factors to be analyzed. For the analysis of mapping landslide susceptibility, the contrast values, W+ and W-, of each factor's rating were overlaid spatially. The results of the analysis were validated using the previous actual landslides locations in the study area. The combination of slope, curvature, topography, distance to road and distance to drainage, showed the best results. The results can be used for hazard prevention and land-use planning.

Keywords. Landslide, Weights-of-Evidence, Susceptibility, GIS, Remote Sensing, Malaysia

Introduction

In recent times a series of landslides have been occurring in Malaysia [1-7]. Most of them have happened on cut slopes or on embankments alongside roads and highways in mountainous areas. Some of these landslides occurred in residential areas and near high-rise apartments, few major and catastrophic landslides within the last decade [7]. Landslide hazard and risk analyses, like many other forms of risk management of either natural or civil engineering hazards [8-18], is a relatively new discipline. In tropical countries like Malaysia, most landslides are triggered by heavy rainfall. In the study area, the Cameron Highlands in Malaysia, much damage was caused on these occasions. Therefore, in order to mitigate any damage arising from landslides, it is necessary to scientifically assess susceptible areas. Thus, this study aims to carry out a landslide susceptibility analysis by applying the weights-of-evidence approach with the validation of the results. As a basic analysis tool, a Geographic Information System (GIS) was used for spatial and data management and manipulation.

Many attempts regarding landslide susceptibility mapping using various deterministic and heuristic approaches have been described in literature. Recently, there have been studies for landslide hazard evaluation using GIS [15-18]. In this study, landslide susceptibility was evaluated using GIS

based on the weights-of-evidence model. The model was originally developed for mineral potential assessment [11-13].

The main difference between this study and the references is that the GIS based weight of evidence model was applied to the landslide susceptibility mapping and validated with landslide test location data. Following heavy rainfall in 2008, 2009 and 2010 the Cameron Highlands study area experienced substantial landslide damage and were selected as a suitable application site for regional landslide susceptibility mapping (Figure 1). A detailed description of the study area can be found in Pradhan and Lee [6].



Figure 1. Location map of the study area.

1. Materials and methodology used

For the landslide-hazard analysis, the main steps were data collection and construction of a spatial database from which the relevant factors were extracted, followed by assessment of the landslide hazard using the relationship between landslide and landslide-related factors, and validation of the results [6]. A key assumption of this approach is that the past is the key to the future. In other words, the potential (occurrence possibility) of landslides can be comparable to the actual frequency of landslides. A spatial database of the Cameron Highlands related to landslide susceptibility modelling (Table 1) was compiled. This database includes maps, images, tabular data, vectors and rasters from different custodians. They existed in different formats, scales, accuracies and geometries and were almost unusable for direct implementation to this research. For this reason, these data were converted into appropriate data input for analysis and weight of evidence modelling. The GIS and remote sensing data used in the present study are shown in Table 1. Nine landslide conditioning factors were identified based on the previous paper published by Pradhan and Lee [6]. In the studied area, it is also observed that most of the landslides have taken place immediately after heavy rainfall [6]. Hence, rainfall is considered as the triggering factor. In the first stage, a landslide inventory map was prepared by mapping the landslide areas in the Cameron Highlands area by interpretation of aerial photographs, high resolution satellite imagery, previous reports and extensive field surveys. The landslide inventory showing the geographical distribution of past and recent landslides was developed from 1:10,000 – 1:50,000 scale aerial photographs, and this was used to evaluate the fre-

quency and distribution of shallow landslides in the area. A total of 389 landslide location points were compiled from various data sources. These landslides were then classified and sorted out based on their modes of occurrence using the classification scheme proposed by [19]. Topography, soil, lithology, structure, land cover, and lineament databases were constructed for the analysis (Table 1). Maps relevant to landslide occurrence were constructed from a vector type spatial database using ARC/INFO GIS software. These included 1:25,000 scale topographic maps, 1:100,000 scale soil maps and 1:63,300 scale geology maps. Using the detected landslide locations and the constructed spatial database, landslide susceptibility mapping was performed by employing weight-of-evidence model. For this, the calculated and extracted factors were mapped into a 10 m × 10 m grid in ARC/INFO GRID format. Next, using the weights-of-evidence method, spatial relationships between the landslide location and each of the landslide-related factors, such as topography, soil, geology, land cover and distance from lineament were analyzed. The spatial relationships were used as each factor's rating in the overlay analysis. Subsequently, tests of conditional independence were performed for the selection of the factors to be used in landslide susceptibility mapping. The factors' ratings were summed to calculate a landslide susceptibility index, and landslide susceptibility was mapped for 9 combinations of the factors. Finally, the results of a comparison of the 9 different combinations were validated using landslide test location data.

Table 1. GIS data used in this study

Data layers	Data type	Source of data
Landslide inventory	Polygon coverage	Landslide inventory database, aerial photos, field surveys
Topographic map	Line and Point coverage	National Mapping Agency, Malaysia
Geological map	Polygon and Line coverage	Minerals and Geoscience Department, Malaysia
Soil map	Polygon coverage	Department of Irrigation and Drainage, Malaysia
Drainage	Line coverage	Department of Agriculture, Malaysia
Normalized Differentiated Vegetation Index (NDVI)	GRID	LANDSAT TM image
Landcover	GRID	LANDSAT TM image

1.1. Weight-of-evidence- Theory

The weights-of-evidence model has many advantages compared to the other statistical methods. Weights-of-evidence is a data-driven method which is basically the Bayesian approach in a log-linear form using the prior and posterior probability and is applied where sufficient data are available to estimate the relative importance of evidential themes by statistical means [20]. In the present study, the following formulation of the Bayesian probability model, known as the weights-of-evidence model, was applied to landslide susceptibility mapping as synthesized from [20,21]. For a given number of units cells, $N(D)$, containing an occurrence, D and the total number of unit cell in the study area, $N(T)$, the prior probability of an occurrence is expressed by

$$P(D) = \frac{N(D)}{N(T)} \quad (1)$$

For details about the weights-of-evidence please see [20,21].

1.2. Evidences and calculation of weights-of-evidence

Local knowledge of the landslide susceptibility in the Cameron Highlands suggested 9 binary predictor patterns of topographic, soil, geologic, land cover, and lineament features that were likely to be useful evidence for predicting landslide susceptibility. For each of the landslide-related factors, the weights and contrast were calculated using the weights-of-evidence method (Tables 2). The total

number of pixels in the study area was 3,602,820, and the total number of landslide occurrences was 324. The magnitude of the contrast, C , was determined from the difference, W^+ and W^- , and provided a measure of the spatial association between a set of points and the patterns. C is positive for a positive spatial association, and negative for a negative spatial association.

1.3. Pair-wise test of conditional independence and interpretation

The conditional independence was tested prior to the integration of the predictor patterns to map the landslide susceptibility. All the pairs of binary predictor patterns were tested. The χ^2 table for testing conditional independence is shown in Table 2. The χ^2 values for testing the conditional independence between all pairs of binary patterns for each factor were calculated at the 99% significance level and 1 degree of freedom. If the χ^2 value in the contingency table is below 6.64, the pair of binary predictor patterns is independent. An example, using the contingency table to test the conditional independence between the binary patterns of slope aspect (Factor 1) and binary patterns of the others (Factor 2), is shown in Table 3. It can be seen that the pairs of slope aspect and slope angle show a conditional independence, because the χ^2 is 0.03 (Slope angle present 0.02 + Slope angle absent 0.01) relative to the criterion that the χ^2 value is 4.81 at the 99% significance level. This implies that these binary predictor patterns could be used together to map the landslide susceptibility. However, the pairs of slope aspect and landcover show conditional dependence, because the χ^2 value is 6.40 (Slope aspect present, 5.13 + Slope aspect, 1.27) relative to the criterion of χ^2 -value of 4.81 at the 99% significance level. This implies that these binary predictor patterns could not be used together to map the landslide susceptibility.

Table 2. Calculated χ^2 values for testing conditional independence between all pairs of binary patterns with each factor in the 99% significance level (The grey area is dependent).

	Slope aspect	Plan curvature	Distance from drainage	Distance from lineament	Lithology	Landcover	Slope angle	Soil
Plan curvature	0.05							
Distance from drainage	0.05	2.42						
Distance from lineament	0.71	0.21	0.12					
Lithology	2.10	0.30	0.14	0.43				
Landcover	6.40	0.20	0.16	1.85	0.96			
Slope angle	0.03	9.85	0.91	4.69	0.53	0.22		
Soil	1.15	0.00	3.41	7.88	0.38	10.44	0.57	
ndvi	3.36	4.81	6.19	0.73	2.71	0.14	0.07	1.15

2. Results and discussion

2.1. Application of weight-of-evidence to landslide susceptibility mapping

By means of the weights-of-evidence method the spatial relationship and the contrast value between landslide-occurrence location and each landslide-related conditioning factor was derived. The contrast was set to the rating of each landslide conditioning factor, because the contrast is related to the landslide occurrence probability. To generate the binary predictor pattern of the 9 factors, the spatial database was reclassified into a binary pattern as “favourable” and the other formations as “non-favourable”. To generate the binary predictor factor, we determine the ratings or weight for which the spatial association between the landslide occurrences and each factor is optimal. The optimum cut-off for the binary predictor factor was determined by calculating the $C/S(C)$, standardized value of contrast. The W^+ and W^-

values were used as ratings for each factor. It is observed that, for slope class $> 33^\circ$ the value of $C/S(C)$ is 16.35, which is the highest among the slope classes. Therefore, the slope class $> 33^\circ$ was selected as the optimum cut-off value. As a result, the slope class $> 33^\circ$ is considered as binary favourable and the other classes were used as binary unfavourable classes in the case of slope. Similarly, for the slope class $> 33^\circ$, the value of W^+ and W^- equals to 1.56 and -0.25, respectively.

Table 3. Contingency table to test conditional independence between binary pattern of slope aspect and other factors.

	Slope aspect present			Slope aspect absent			Totals
	Observed	Expected	χ^2	Observed	Expected	χ^2	
Distance from drainage present	9	9.53	0.03	39	38.47	0.01	48
Distance from drainage absent	46	45.47	0.01	183	183.53	0	229
Totals	55	55	0.04	222	222	0.01	277
Curvature present	47	46.46	0.01	187	187.54	0	234
Curvature absent	8	8.54	0.03	35	34.46	0.01	43
Totals	55	55	0.04	222	222	0.01	277
NDVI present	16	11.12	2.14	40	44.88	0.53	56
NDVI absent	39	43.88	0.54	182	177.12	0.13	221
Totals	55	55	2.69	222	222	0.67	277
Soil present	7	4.96	0.84	18	20.04	0.21	25
Soil absent	48	50.04	0.08	204	201.96	0.02	252
Totals	55	55	0.92	222	222	0.23	277
Slope angle present	14	14.49	0.02	59	58.51	0	73
Slope angle absent	41	40.51	0.01	163	163.49	0	204
Totals	55	55	0.02	222	222	0.01	277
Landcover present	29	20.85	3.19	76	84.15	0.79	105
Landcover absent	26	34.15	1.95	146	137.85	0.48	172
Totals	55	55	5.13	222	222	1.27	277
Lithology present	23	18.47	1.11	70	74.53	0.28	93
Lithology absent	32	36.53	0.56	152	147.47	0.14	184
Totals	55	55	1.68	222	222	0.42	277
Distance from lineament present	11	8.94	0.48	34	36.06	0.12	45
Distance from lineament absent	44	46.06	0.09	188	185.94	0.02	232
Totals	55	55	0.57	222	222	0.14	277

If the LSI_c and LSI_w value are high, this means a higher susceptibility to landslide; a lower value means a lower susceptibility to landslides. Before the integration of the landslide conditioning parameters the statistical validity of the resulting susceptibility maps was examined by applying an overall test of conditional independence. The 9 combinations of the factors were determined to be conditionally independent. First, for the combination of all factors, the landslide susceptibility map was calculated using all the factors. The LSI_c value index is used for the interpretation. The index was classified into equal areas and grouped into 10 classes for visual interpretation. After the test of conditional independence was performed, the 9 cases of combinations (Table 3) of the factors were analyzed, and the LSI_w values were calculated etc. These indexes were also classified into equal areas, and for the purpose of visual interpretation grouped into 10 classes (Figure 2).

2.2. Validation of the susceptibility maps

Validation was performed by comparison with existing landslide data, using a success rate curve. The success rate illustrates how well the estimators perform with respect to the landslides used in

constructing those estimators [13]. The range was classified by area, which was equal to about 1%. For example, in all cases, an index rank above 10%, could explain 51% of the landslides, and in case S3, an index rank above 10% could explained 45% of the landslides. Also, in all cases, an index rank above 20%, could explain 78% of the landslides, and in case S3, an index rank above 20% could explain 68% of the landslides. Among the 9 cases considered, case C1 shows the best result, 48% and 62% of landslides for rank of above 10% and 20% in the index. In the case C1, we used plan curvature, distance from drainage, distance from lineament, lithology and landcover after the test of conditional independence.

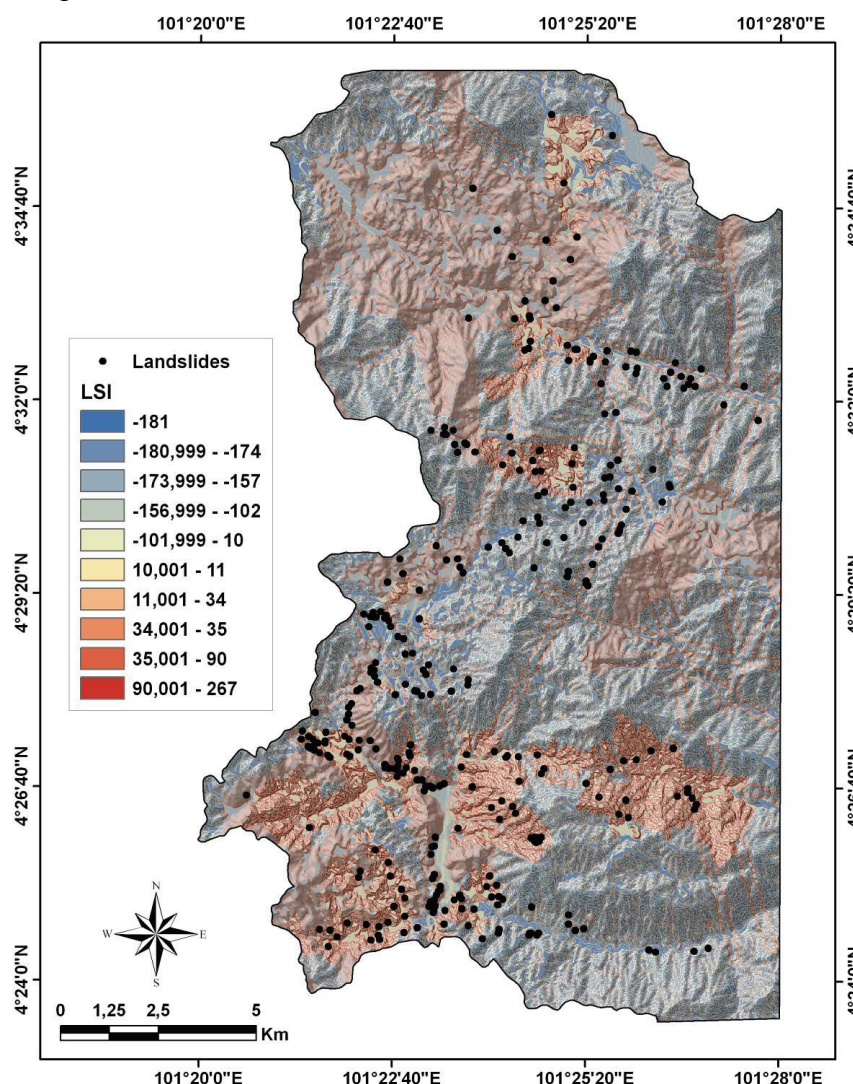


Figure 2. Predictive landslide susceptibility map based on weights-of-evidence analysis; using all contrasts.

3. Conclusion

In this study, the relationships between landslide and landslide conditioning factors were investigated for the Cameron Highlands using the weights-of-evidence method. Subsequently, a test of independence between each factor was performed, and 9 combinations of the factors were derived. For the analysis, the calculated contrast, and values were determined as rating for each factor, and were overlaid to calculate a landslide susceptibility index (LSI). Finally, landslide susceptibility map was constructed using these LSI values and the resultant maps were validated employing success rate curve method.

References

- [1] B. Pradhan, S. Lee, Utilization of optical remote sensing data and GIS tools for regional landslide hazard analysis by using an artificial neural network model at Selangor, Malaysia. *Earth Science Frontier*, **14(6)**, (2008), 143–152
- [2] B. Pradhan, S. Lee, Landslide risk analysis using artificial neural network model focusing on different training sites. *International Journal of Physical Science* **3(11)**, (2009), 1–15
- [3] B. Pradhan, Remote sensing and GIS-based landslide hazard analysis and cross-validation using multivariate logistic regression model on three test areas in Malaysia. *Advances in Space Research*, **45(10)**, (2010), 1244-1256
- [4] B. Pradhan, S. Lee, Delineation of landslide hazard areas on Penang Island, Malaysia, by using frequency ratio, logistic regression, and artificial neural network models. *Environmental Earth Science*, **60**, (2010), 1037-1054
- [5] B. Pradhan, S. Lee, Landslide susceptibility assessment and factor effect analysis: backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modeling. *Environmental Modelling and Software*, **25(6)**, (2010), 747-759
- [6] B. Pradhan, S. Lee, Regional landslide susceptibility analysis using backpropagation neural network model at Cameron Highland, Malaysia. *Landslides*, **7(1)**, (2010), 13-30
- [7] B. Pradhan, S. Lee, M.F. Buchroithner, A GIS-based back-propagation neural network model and its cross-application and validation for landslide susceptibility analyses. *Computers Environment and Urban Systems*, **34**, (2010), 216-235
- [8] S. Lee, J. Choi, Landslide susceptibility mapping using GIS and the weight-of evidence model. *International Journal of Geographical Information Science*, **18(8)**, (2004), 789-814
- [9] S. Lee, B. Pradhan, Landslide hazard mapping at Selangor, Malaysia using frequency ratio and logistic regression models. *Landslides*, **4**, (2007), 33-41
- [10] S. Lee, B. Pradhan, Probabilistic landslide risk mapping at Penang Island, Malaysia. *Journal of Earth System Sciences*, **115(6)**, (2006), 661-672
- [11] J. Mathew, V.K. Jha, G.S. Rawat, Weights of evidence modeling for landslide hazard zonation mapping in part of Bhagirathi valley, Uttarakhand. *Current Science*, **92**, (2007), 628–638
- [12] B. Neuhauser, B. Terhorst, Landslide susceptibility assessment using “weights of evidence” applied to a study area at the Jurassic escarpment (SW—Germany). *Geomorphology*, **86**, (2007), 12–24
- [13] B. Pradhan, Manifestation of an advanced fuzzy logic model coupled with geoinformation techniques for landslide susceptibility analysis. *Environmental & Ecological Statistics*, (in press), DOI: 10.1007/s10651-010-0147-7 (2010)
- [14] I. Das, S. Sahoo, C. Van Westen, A. Stein, R. Hack, Landslide susceptibility assessment using logistic regression and its comparison with a rock mass classification system, along a road section in the northern Himalayas (India). *Geomorphology*, **114(4)**, (2010), 627-637
- [15] M. Ercanoglu, C. Gokceoglu, Assessment of landslide susceptibility for a landslide-prone area (north of Yenice, NW Turkey) by fuzzy approach. *Environmental Geology*, **41**, (2002), 720–730
- [16] S. Lee, U. Chwae, K. Min, K. Landslide susceptibility analysis and verification using the Bayesian probability model. *Environmental Geology*, **43**, (2002), 120-131
- [17] C.J. Van Westen, E. Castellanos, S.L. Kuriakose, Spatial data for landslide susceptibility, hazard, and vulnerability assessment: An overview. *Engineering Geology*, **102**, (2008), 112-131
- [18] A. M. Youssef, B. Pradhan, A.F.D. Gaber, M.F. Buchroithner, Geomorphological Hazard Analysis along the Egyptian Red Sea Coast between Safaga and Quseir. *Natural Hazards and Earth System Science*, **9**, (2009), 751-766
- [19] D.J. Varnes, with IAEG Commission on landslides and other mass movements: landslide hazard zonations: a review of principles and practices. UNESCO, Paris, (1984) p 63
- [20] F. P. Agterberg, G.F. Bonham-Carter, Q. Cheng, D.F. Wright, Weights of evidence modeling and weighted logistic regression for mineral potential mapping. In *Computers in Geology, 25 Years of Progress*, Edited by J. C. Davis, and U. C. Herzfeld (Oxford: Oxford University Press), (1993), 13 – 32
- [21] G.F. Bonham-Carter, F.P. Agterberg, D.F. Wright, Weights of evidence modeling: a new approach to mapping mineral potential In *Statistical Applications in the Earth Sciences* edited by F. P. Agterberg, and G. F. Bonham-Carter. Geological Survey Canada Paper **89-9**, (1989), 171–183.

