Environ Geol (2006) 50: 847–855
DOI 10.1007/s00254-006-0256-7

ORIGINAL ARTICLE

Saro Lee
Touch Sambath

Landslide susceptibility mapping in the Damrei Romel area, Cambodia using frequency ratio and logistic regression models

Received: 17 October 2005
Accepted: 28 February 2006
Published online: 4 April 2006

S. Lee (✉)
Geoscience Information Center,
Korea Institute of Geoscience &
Mineral Resources (KIGAM),
30, Gajeong-dong, Yuseong-gu,
Daejeon 305-350, Korea
E-mail: leesaro@kigam.re.kr
Tel.: +82-42-8683057
Fax: +82-42-8619714

T. Sambath
Department of Geology,
General Department of Mineral Resources,
Ministry of Industry, Mines and Energy,
Phnom Penh, Kingdom of Cambodia

Abstract This study applied, tested and compared a probability model, a frequency ratio and statistical model, a logistic regression to Damre Romel area, Cambodia, using a geographic information system. For landslide susceptibility mapping, landslide locations were identified in the study area from interpretation of aerial photographs and field surveys, and a spatial database was constructed from topographic maps, geology and land cover. The factors that influence landslide occurrence, such as slope, aspect, curvature and distance from drainage were calculated from the topographic database. Lithology and distance from lineament were extracted and calculated from the geology database. Land cover was classified from Landsat TM satellite imagery. The relationship between the factors and the landslides was calculated using frequency ratio and logistic regression models. The relationships, frequency ratio and logistic regression coefficient were overlaid to make landslide susceptibility map. Then the landslide susceptibility map was compared with known landslide locations and tested. As the result, the frequency ratio model (86.97%) and the logistic regression (86.37%) had high and similar prediction accuracy. The landslide susceptibility map can be used to reduce hazards associated with landslides and to land cover planning.

Keywords Landslide · Frequency ratio · Logistic regression · GIS · Cambodia

Introduction

Cambodia is a part of the Indochina craton, which has been stable since Late Triassic time. Geological hazards such as earthquakes or volcanic eruptions are rare. But, landslides have been observed in mountainous areas and occur mainly in areas of steep slopes and only in the rainy season. Because mountainous areas are not highly low populated or in most cases, un-inhabitanted, there are few effects of landslides on the property. However, landslide can cause problems to road networks in the highlands and mountain valleys. It is therefore necessary to assess and manage areas that are susceptible to landslides in order to mitigate any damage associated with them. Among the many causes, landslides triggered by heavy rainfall are the most common throughout Cambodia. The resultant need to predict landslide occurrences has led to the development of numerous stochastic and process-based models, with increasing emphasis on the use of a GIS. In this study, this is carried out by applying the frequency ratio and logistic regression models, with testing of the results in the Damre Romel area, Cambodia using GIS.

Using GIS as the basic analysis tool for landslide hazard, mapping can be effective for spatial data management and manipulation for the analysis. In this regard, there have been many studies of landslide hazard mapping using GIS. For example, Guzzetti et al. (1999)

summarized many landslide hazard evaluation studies. Many studies have applied probabilistic models (Jibson et al. 2000; Luzi et al. 2000; Parise and Jibson 2000; Rautelal and Lakhera 2000; Baeza and Corominas 2001; Lee and Min 2001; Clerici et al. 2002; Donati and Turrini 2002; Lee et al. 2002a, b; Zhou et al. 2002; Lee and Choi 2003, Lee et al. 2004b, 2004c, Lee and Choi 2004). One of the statistical models available, the logistic regression models, has also been applied to landslide hazard mapping (Atkinson and Massari 1998; Dai et al. 2001, Dai and Lee 2002; Ohlmacher and Davis 2003; Lee 2004e), as has the geotechnical model and the safety factor model (Gokceoglu et al. 2000; Romeo 2000; Carro et al. 2003; Shou and Wang 2003; Zhou et al. 2003). As a new approach to landslide hazard evaluation using GIS, data mining using fuzzy logic and artificial neural network models have also been applied (Ercanoglu and Gokceoglu 2002; Pistocchi et al. 2002; Lee et al. 2003a, b, 2004a). The difference in this study is the application and comparison of GIS-based methods to landslide susceptibility mapping in Cambodia.

The study area (Fig. 1) is at high altitude and represents an outcrop of continental sandstone. The latter contains siltstone, shale, breccia and other lithological horizons of variable thickness, which easily weather and erode. The area was originally forested by large trees, but these have been removed by the local population, and only bushes remain. One formation is composed of deeply weathered, Jurassic–Cretaceous dacite and tuff in which the Peang Lovea Landslide has occurred. The upper vegetated layer of dacite is weathered and broken into blocks that unstably overlie the tuff. The landslide happened after a heavy storm, when groundwater destabilized the weathered tuff layer and caused the tuff and overlying unconsolidated dacite to slide downslope. The study area has large recent alluvium which is flat and has no landslide. The area occupies 47% of the study area. So, the area was eliminated for more precisious landslide susceptibility analysis (Fig. 1).

Theory: frequency ratio and logistic regression

The relationship between the landslide occurrence area and the landslide-related factors could be deduced from the relationship between areas where landslides had not occurred and the landslide-related factors. To represent this distinction quantitatively, one of the probability models, the frequency ratio, were used. The frequency ratio is the ratio of the probability of an occurrence to the probability of a non-occurrence for given attributes (Bonham-Carter 1994). In the case of a landslide, if we set the landslide occurrence event to be represented by a factor, "B‘", and this factor's attributes are denoted by "D", then the frequency ratio of D is the conditional probability ratio. If this ratio is greater than 1, then the relationship between a landslide and the factor's range or type is strong. If the ratio is less than 1, then the relationship between a landslide and the factor's range or type is weak.

The spatial relationship between a landslide occurrence location and each landslide-related factor was derived using the frequency ratio model. Therefore, the rating of each factor's type or range was assigned as the relationship between a landslide and the value of each factor's type or range, i.e., the ratio of the number of cells where landslides had not occurred to the number of cells where landslides had occurred. The landslide susceptibility index (LSI) was calculated by summation of each factor's ratio value using Eq. (1).

$$\mathrm{LSI} = \sum \mathrm{Fr}(\mathrm{Fr}\text{: Rating of each factor's type or range}) \quad (1)$$

Logistic regression, which is a multivariate analysis model, is useful for predicting the presence or absence of a characteristic or outcome based on values of a set of predictor variables. The advantage of logistic regression is that, through the addition of an appropriate link function to the usual linear regression model, the variables may be either continuous or discrete, or any combination of both types, and they do not necessarily have normal distributions. In the present situation, the dependent variable is a binary variable representing the presence or absence of landslides. Quantitatively, the relationship between the occurrence and its dependency on several variables can be expressed as:

$$p = 1/(1 + e^{-z}) \quad (2)$$

where p is the probability of an event occurring. In the present situation, the value p is the estimated probability of landslide occurrence. The probability varies from 0 to 1 on an S-shaped curve and z is the linear combination. It follows that logistic regression involves fitting an equation of the following form to the data:

$$z = b_0 + b_1x_1 + b_2x_2 + \ldots + b_nx_n \quad (3)$$

where b_0 is the intercept of the model, the b_i (i=0, 1, 2, ..., n) is the slope coefficients of the logistic regression model, and x_i (i=0, 1, 2, ..., n) Are the independent variables. The linear model formed is then a logistic regression of presence or absence of landslides (present conditions) on the independent variables (pre-failure conditions).

Data and methodology

Landslides may occur as a consequence of a number of determining and triggering factors. In order to assess susceptibility from landslide it is therefore necessary to

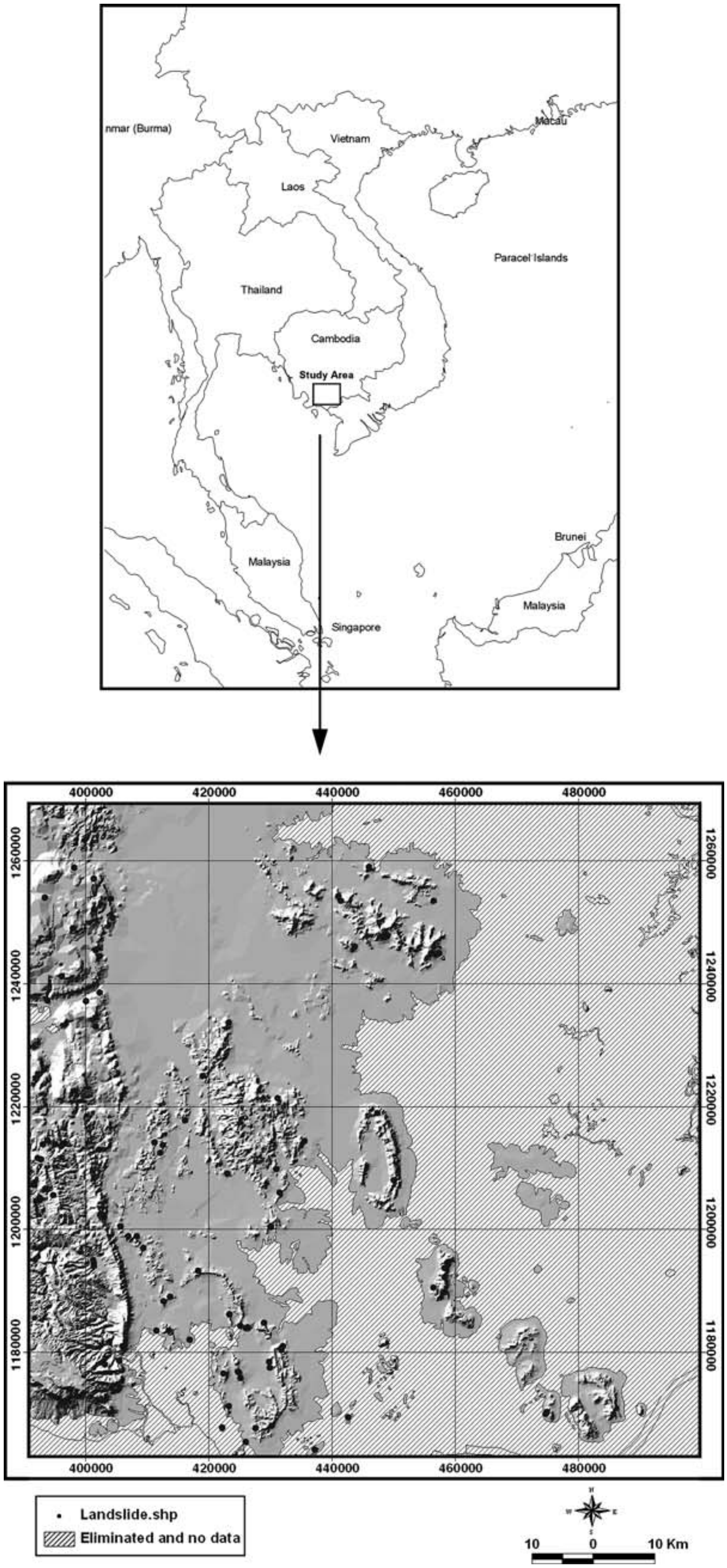

Fig. 1 Study area and landslide location with hillshade map

identify and analyze the factors leading to a landslide. Data preparation involved the digitization or creation of GIS database, which include the topographical, geomorphological, geological and land cover data. A digitized map of landslide location, which detected from satellite imagery and field surveys was produced, and these digital data were input into the GIS. A vector-to-raster conversion was undertaken to provide raster data of landslide areas. Factor maps related to landslide occurrence were constructed in a vector-type spatial database. These included topographic and geological maps. A land cover map was classified from Landsat TM satellite imagery with 30 m resolution. The resultant GIS database included slope, aspect, curvature, distance to drainage, lithology, distance to lineament and land cover data. The study area was divided into a grid with 30×30 m cells, occupying 3,653 rows and 3,653 columns: totaling 6,560,383 grid-cells (only study area) and landslides fell into 89 of these.

Contour (20-m interval) and survey base points that had an elevation value read from the topographic map were extracted, and a digital elevation model (DEM) was constructed. Using the DEM, the slope gradient, slope aspect and curvature were calculated. The slope gradient of a surface refers to the maximum rate of change in elevation across a region of the surface and the slope aspect of a surface is the compass direction maximum rate of change in z in the downward direction. The curvature is a morphological measure of the topography. A positive curvature indicates that the surface is upwardly convex at that cell, and a negative curvature indicates that the surface is upwardly concave at that cell. A value of zero indicates that the surface is flat. The distance from drainage was calculated in 1 m intervals. The type of geology of an area plays an important factor in the development of landslide. So, the lithology from the geological map was used and the distance from lineament was calculated in 1 m intervals. Land cover data was classified from a LANDSAT TM image using unsupervised classification method and field survey. The seven classes such as urban, water, forest, agricultural area, grass, shrubland and barren area were extracted for land cover mapping.

Using the detected landslide locations and the constructed spatial database, landslide analysis models were applied and tested. To represent the distinction quantitatively, frequency ratio and logistic regression models were used. For this analysis, the calculated and extracted factors were mapped to a 30 m resolution grid. The raster data were converted for the statistical program used. Then, using a frequency ratio and logistic regression model, the spatial relationships between the landslide location and each landslide-related factor, such as topography, geology and land cover, were analyzed. Then, in the frequency ratio model, the relationship was used as each factor's rating. Using the rating, the factors have been overlaid and landslide susceptibility index was calculated. In the logistic regression model, a formula of landslide occurrence possibility was extracted using the relationships. This formula was used to calculate the landslide susceptibility index. The indexes were mapped to represent landslide susceptibility. Finally, the susceptibility maps were tested using known landslide locations and success rates were calculated for quantitative testing. In this study, GIS software, ArcView 3.3 and ARC/INFO 9.0 NT version and statistical software, SPSS 12.0, were used as the basic analysis tools for spatial management and data manipulation.

Relationship between landslides and factors

The relationship between areas where a landslide has occurred and landslide-related factors can be distinguished from the relationship between areas without past landslides and landslide-related factors. To represent this distinction quantitatively, the frequency ratio was used. The factors chosen, such as the slope, aspect, curvature, distance from drainage, lithology, distance from lineament and land cover were evaluated using the frequency ratio method to determine the level of correlation between the location of the landslides in the study area and these factors. The approaches are based on the observed relationships between each factor and the distribution of landslides.

Table 1 shows the relationship between landslide occurrence and each factor. Topographic factors, such as slope, aspect, curvature and distance from drainage were used. In the case of the relationship between landslide occurrence and slope, as the slope increases, the landslide frequency generally increases. For example, below a slope of 10°, the ratio was <1, which indicates a low probability of landslide occurrence. For slopes above 11°, the ratio was >1, which indicates a high probability of landslide occurrence. This means that the landslide probability increases according to slope angle. As the slope angle increases, then the shear stress in the soil or other unconsolidated material generally increases. Gentle slopes are expected to have a low frequency of landslides because of the generally lower shear stresses associated with low gradients. Steep natural slopes resulting from outcropping bedrock, however, may not be susceptible to shallow landslides. In the case of the relationship between landslide occurrence and aspect, landslides were most abundant on west-facing and northeast-facing slopes. The frequency of landslides was highest on west-facing and northeast-facing slopes, except in flat areas and lowest on east-facing and southeast-facing slopes. In the case of the relationship between landslide occurrence and curvature, the concave area has the higher probability of a landslide occurrence than convex area. Flat areas had a low curvature value

Table 1 Frequency ratio and logistic regression coefficient

Factor	Range or type	No. of pixels in domain	Percent of domain	No. of landslide	Percent of landslide	Ratio[a]	Logistic regression coefficients
Topographic slope (degree)	0	343,847	52.41	8	8.99	0.17	0.0330
	1	693,982	10.58	5	5.62	0.53	
	2–3	392,790	5.99	0	0.00	0.00	
	4–6	359,004	5.47	9	10.11	1.85	
	7–9	310,386	4.73	8	8.99	1.90	
	10–12	288,298	4.39	11	12.36	2.81	
	13–16	341,098	5.20	13	14.61	2.81	
	17–20	289,975	4.42	7	7.87	1.78	
	21–25	257,371	3.92	19	21.35	5.44	
	25 <	189,062	2.88	9	10.11	3.51	
Topographic aspect	Flat	2,739,287	41.75	4	4.49	0.11	−1.5888
	N	358,915	5.47	6	6.74	1.23	−0.1311
	NE	660,304	10.07	20	22.47	2.23	0.5998
	E	640,678	9.77	12	13.48	1.38	0.0009
	SE	564,204	8.60	8	8.99	1.05	−0.1305
	S	417,839	6.37	11	12.36	1.94	0.3293
	SW	460,592	7.02	9	10.11	1.44	−0.0279
	W	363,967	5.55	12	13.48	2.43	0.2974
	NW	354,597	5.41	7	7.87	1.46	0
Topographic curvature	Concave (−)	1,509,847	23.01	39	43.82	1.90	0.2365
	Flat (0)	3,538,785	53.94	18	20.22	0.37	0.1474
	Convex (+)	1,511,751	23.04	32	35.96	1.56	0
Distance from drainage (river) (m)	0–120	697,633	10.63	1	1.12	0.11	0.0002
	123–241	653,039	9.95	2	2.25	0.23	
	247–381	657,598	10.02	10	11.24	1.12	
	384–524	652,275	9.94	4	4.49	0.45	
	530–690	654,737	9.98	9	10.11	1.01	
	692–882	658,428	10.04	11	12.36	1.23	
	885–1,116	650,262	9.91	11	12.36	1.25	
	1,120–1,431	645,745	9.84	11	12.36	1.26	
	1,432–1,910	645,524	9.84	10	11.24	1.14	
	1,911–5,512	645,142	9.83	20	22.47	2.29	
Distance from lineament (m)	0–216	660,025	10.06	14	15.73	1.56	−0.0001
	218–450	656,577	10.01	14	15.73	1.57	
	453–700	656,029	10.00	11	12.36	1.24	
	702–989	656,722	10.01	12	13.48	1.35	
	990–1,320	657,891	10.03	10	11.24	1.12	
	1,321–1,741	655,189	9.99	13	14.61	1.46	
	1,742–2,312	655,517	9.99	3	3.37	0.34	
	2,313–3,156	654,214	9.97	8	8.99	0.90	
	3,157–4,857	654,223	9.97	1	1.12	0.11	
	4,858–42,738	653,996	9.97	3	3.37	0.34	
Geology (rock type)	Andesite	2,075	0.03	0	0.00	0.00	−0.2675
	Black shale, sandstone and Jaspers	51,954	0.79	2	2.25	2.84	10.8132
	Claystone	263,108	4.01	8	8.99	2.24	11.3244
	Conglomerate	37,558	0.57	2	2.25	3.93	11.5496
	Diorite and gabbros	9,190	0.14	0	0.00	0.00	3.0323
	Eluvial sand	310,545	4.73	1	1.12	0.24	9.5934
	Fine grained granite	83,389	1.27	1	1.12	0.88	9.9905
	Granite	57,495	0.88	0	0.00	0.00	0.2236
	High alumina granite	785	0.01	0	0.00	0.00	0.6726
	Hornfelse	5,975	0.09	0	0.00	0.00	0.9703
	Jaspers	695	0.01	0	0.00	0.00	0.6641
	Lateritic pisolith and crust	7,737	0.12	0	0.00	0.00	1.2249
	Limestone	22,789	0.35	3	3.37	9.70	12.4811
	Metaconglomerate	1,300	0.02	0	0.00	0.00	−1.1451
	Old alluvium	3,733,958	56.92	13	14.61	0.26	9.662
	Quartzite	381,075	5.81	28	31.46	5.42	11.5771
	Rhyolite	256,009	3.90	5	5.62	1.44	10.6383
	Sandstone and microbreccias	70,194	1.07	1	1.12	1.05	9.8804
	Sandstone, shale and marl	116,964	1.78	2	2.25	1.26	9.9791

Table 1 (Contd.)

Factor	Range or type	No. of pixels in domain	Percent of domain	No. of landslide	Percent of landslide	Ratio[a]	Logistic regression coefficients
	Sandstone	1,089,722	16.61	21	23.60	1.42	10.3545
	Shale	53,169	0.81	2	2.25	2.77	11.1684
	Shale and sandstone	1,768	0.03	0	0.00	0.00	−0.5815
	Tuff and volcanic breccias	219	0.00	0	0.00	0.00	0.9929
	Undiscriminated plutonic rock	2,710	0.04	0	0.00	0.00	0
Land use	Urban	1,062	0.02	0	0.00	0.00	−10.6018
	Agriculture	1,320,770	20.13	4	4.49	0.22	−3.0640
	Grasslands	393,791	6.00	0	0.00	0.00	−11.7395
	Shrublands	1,089,383	16.61	15	16.85	1.01	−2.7208
	Forest	3,739,591	57.00	69	77.53	1.36	−3.0570
	Water	4,730	0.07	0	0.00	0.00	−11.7062
	Soils and rocks	11,056	0.17	1	1.12	6.67	0

[a] % landslide/% domain

of 0.29. The reason for this is that following heavy rainfall, a concave slope contains more water and retains this water for a longer period.

Analysis was carried out to assess the influence of drainage on landslide occurrence. For this purpose, the distance to a drainage was identified by buffering. In the case of the relationship between landslide occurrence and distance from drainage, as the distance from a drainage increases, the landslide frequency generally increases. At a distance of >about 500 m, the ratio was >1, indicating a high probability of landslide occurrence, and at distances <about 500 m, the ratio was <1, indicating a low probability. This can be attributed to the fact that terrain modification caused by gully erosion and undercutting may influence the initiation of landslides.

In the case of the relationship between landslide occurrence and lithology, the frequency ratio was higher in shale, sandstone, claystone, conglomerate, limestone, rhyolite and quartzite, at >2.00. In the case of the relationship between landslide occurrence and time stratigraphic unit, the frequency ratio was higher in Ante-Permian, Cambrian-Silurian, Cambrian-Upper Silurian Devonian-Carboniferous Jurassic-Cretaceous, Lower-Middle Triassic, Permian, and Triassic Upper Jurassic-Cretaceous, at >2.00.

In the case of the relationship between landslide occurrence and distance from a lineament, the closer the distance was to a lineament, then the greater was the landslide-occurrence probability. For distances to a lineament of about <2,500 m, the ratio was >1, indicating a high probability of landslide occurrence, and for distances to a lineament of about >2,500 m, the ratio was <1, indicating a low probability landslide occurrence. This means that the landslide probability decreases with increasing distance from a lineament. As the distance from a lineament decreases, the fracture of the rock increases, and in addition, the degree of weathering increases.

In the case of the relationship between landslide occurrence and land cover, landslide occurrence values were higher in forest and shrublands areas, and lower in urban, agriculture and grasslands areas. The reason for this is that landslides occurred mainly in mountainous areas simply correlate to slope.

Application of the frequency ratio and logistic regression models

For calculation of the frequency ratio, the area ratio for landslide occurrence and non-occurrence was calculated for the class or type of each factor, and an area ratio for the class or type of each factor to total area was calculated. So, frequency ratios for the class or type of each factor were calculated by dividing the landslide occurrence ratio by the area ratio. The frequency ratios are shown in Table 1. The frequency ratios of each factor's type or class were summed to calculate the landslide susceptibility index (LSI), as shown in Eq. (4)

$$\begin{aligned} \mathrm{LSI} = {} & \mathrm{SLOPE_r} + \mathrm{ASPECT_r} + \mathrm{CURVATURE_r} \\ & + \mathrm{DRAINAGE_r} + \mathrm{GEOLOLOGY_r} \\ & + \mathrm{LINEAMENT_r} + \mathrm{LANDCOVER_r} \end{aligned} \tag{4}$$

(where $\mathrm{SLOPE_r}$ is frequency ratio of slope; $\mathrm{ASPECT_r}$, is frequency ration of aspect; $\mathrm{CURVATURE_r}$ is frequency ratio of curvature; $\mathrm{DRAINAGE_r}$ is frequency ratio of distance from drainge; $\mathrm{GEOLOLOGY_r}$ is frequency ratio of geology; $\mathrm{LINEAMENT_r}$ is frequency ratio of distance from lineament; $\mathrm{LANDCOVER_r}$ is frequency ratio of landcover. The frequency ratio values are listed in Table 1; z is a parameter).

If the LSI value is high, it means a higher susceptibility to landslide; a lower value means a lower susceptibility to landslides. The LSI value index is shown in Fig. 2 for interpretation. The index was classified into

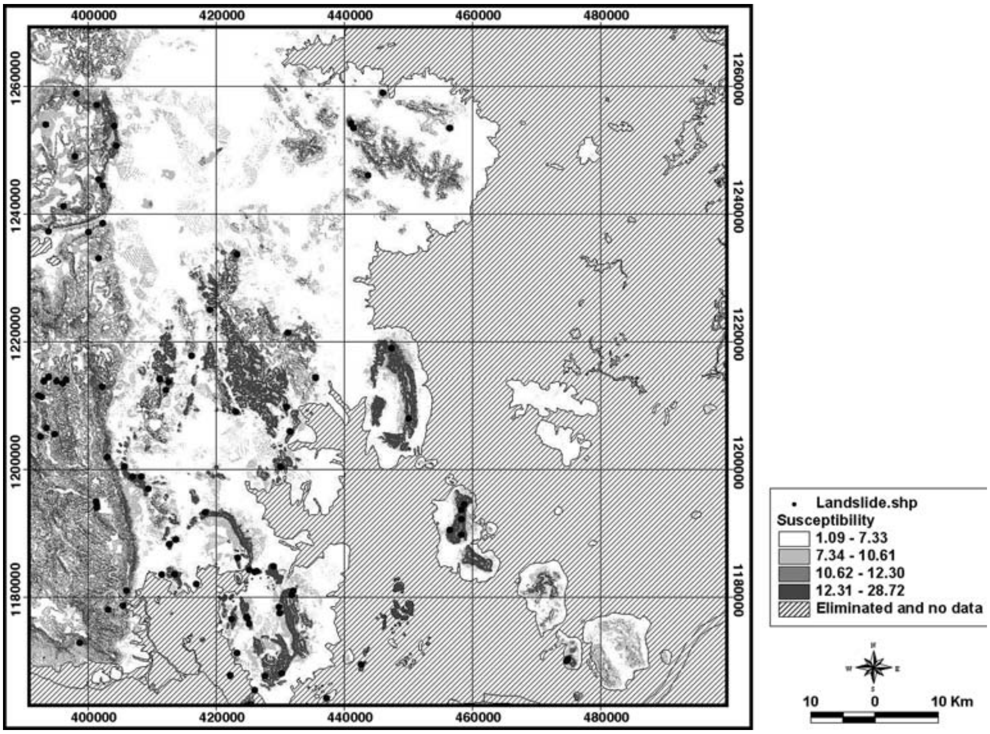

Fig. 2 Landslide susceptibility map based on frequency ratio

equal areas and grouped into five classes for visual and easy interpretation. The minimum value is 0.56 and maximum value is 61.50, the mean value is 8.69 and the standard deviation value is 9.58.

A key concept for understanding the tests used in logistic regression is that of log likelihood. Usually, however, the overall significance is tested using the chi-squared test, which is derived from the likelihood of observing the actual data under the assumption that the model that has been fitted is accurate. It is convenient to use −2 times the log (base e) of this likelihood (−2LL). The log likelihood value (−2LL) here is 1996.337. Several criteria can be used to guide entry: these include the greatest reduction in the –2LL values.

A statistical program was used to calculate the correlation of a landslide event to each factor. Firstly, all factors were constructed in the database and then logistic regression coefficients of the factors were calculated (such as those in Table 1). The coefficients of the logistic regression model were estimated using the maximum-likelihood method. In other words, coefficients that make the observed results most likely are selected. Since the relationship between the independent variables and the probability is nonlinear in the logistic regression model, an iterative algorithm is necessary for parameter estimation (Dai and Lee 2002). In Table 1, there are positive associations, such as slope and negative associations, and distance from lineament. After interpretation, Eqs. (2) and (3), which predict the landslide-occurrence possibility, were created.

$$\begin{aligned} z = {} & (0.0330 \times \mathrm{SLOPE}) + (0.0002 \times \mathrm{DRAINAGE}) \\ & + (-0.0001 \times \mathrm{LINEAMENT_b}) + \mathrm{ASPECT_b} \\ & + \mathrm{CURVATURE_b} + \mathrm{GEOLOLOGY_b} \\ & + \mathrm{LANDCOVER_b} - 18.8654 \end{aligned} \tag{5}$$

(where SLOPE is slope value; CURVATURE is curvature value; DRAINAGE is distance from drainge value; LINEAMENT is distance from lineament value; $\mathrm{ASPECT_b}$, $\mathrm{CURVATURE_b}$, $\mathrm{GEOL_b}$, $\mathrm{LANDCOVER_b}$ are logistic regression coefficient values listed in Table 1; z is a parameter).

Finally, the probability that predicts the possibility of landslide occurrence, for the study area, was calculated using the spatial database, data from Table 1 and Eqs. (2) and (5). The distribution of calculated possibility is shown in Fig. 3. The values were classified by equal areas and grouped into five classes for visual interpretation. The minimum value is 0.00 and maximum value is 0.01112. The mean value is 0.00000543 and the standard deviation value is 0.00002506.

Testing of the landslide susceptibility maps

The landslide susceptibility analysis result was tested using known landslide locations. Testing was performed by comparing the known landslide location data with the landslide susceptibility map. The comparison results are shown in Fig. 4 as a line graph. The success rates in Fig. 4

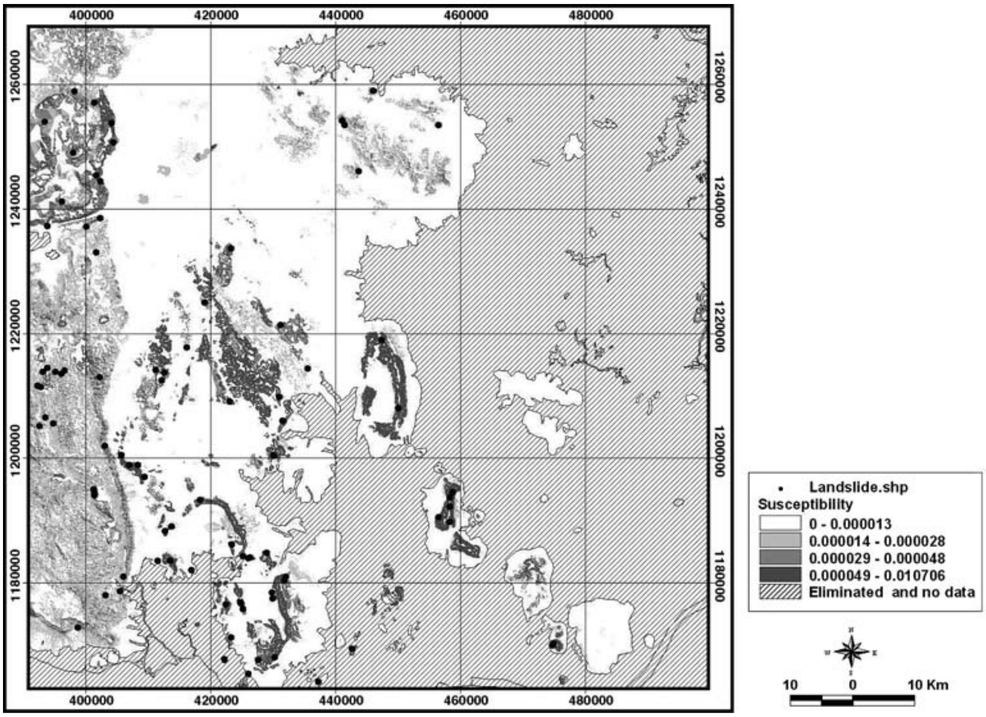

Fig. 3 Landslide susceptibility map based on logistic regression

illustrate, explains how well the model and factor predict the landslide. To obtain the relative ranks for each prediction pattern, the calculated index values of all cells in the study area were sorted in descending order. Then the ordered cell values were divided into 100 classes, with accumulated 1% intervals. The above procedure was also adapted for the landslide occurred cells by comparing the 100 classes obtained with the distribution on the study area. Then, the graph was made by connecting the two classified value. For example, in the case of frequency ratio model, 90–100% (10%) class of the study area where the landslide susceptibility index had a higher rank could explain 49% of all the landslides. In addition, the 80–100% (20%) class of the study area where the landslide susceptibility index had a higher rank could explain 78% of the landslides. In the case of logistic regression model, 90–100% (10%) class of the study area where the landslide susceptibility index had a higher rank could explain 51% of all the landslides. In addition, the 80–100% (20%) class of the study area where the landslide susceptibility index had a higher rank could explain 78% of the landslides.

To compare the results quantitatively, the areas under the curve were re-calculated as the total area is one which means perfect prediction accuracy. So, the area under a curve can be used to assess the prediction accuracy qualitatively. The area under the curve is shown in Fig. 4. In the case of frequency ratio model used, the area ratio was 0.8697 and we could say the prediction accuracy is 86.97%. In the case of logistic regression model used, the area ratio was 0.8637 and we could say the prediction accuracy is 86.37%.

Discussion and conclusions

In this study, a probabilistic and statistical approach for estimating the susceptible areas of study area of Cambodia using a GIS was applied and tested. As the result, the frequency ratio model (86.97%) and logistic regression model (86.37%) showed high accuracy.

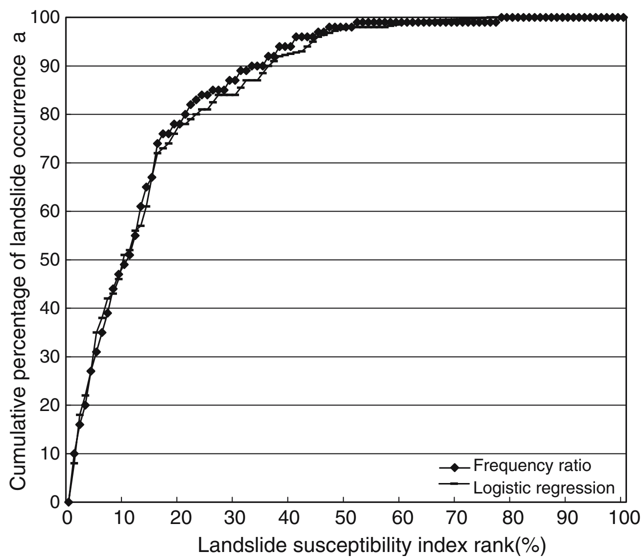

Fig. 4 Illustration of cumulative frequency diagram showing landslide susceptibility index rank (x-axis) occurring in cumulative percent of landslide occurrence (y-axis)

Moverover, frequency ratio model and the logistic regression model showed a similar accuracy.

Although up until now, no study has been made on the landslide disasters of Cambodia, such research is indicated and must pay attention to three prioritized areas. The first area is the highlands of southwestern Cambodia, where landslides have very often damaged the road systems linking Phnom Penh to the Krong Preah Sihanouk port town and to Koh Kong Province. The second prioritized area is Pailin City, lying in an intermontane plain. This urban region is rapidly developing and the surrounding highland area will be populated in the near future. The third area indicated for landslide study is that surrounding the Angkor monument. Some of the ancient temples were built on the tops of unstable hills that may easily collapse through landsliding.

References

Atkinson PM, Massari R (1998) Generalized linear modeling of susceptibility to landsliding in the central Apennines, Italy. Comput Geosci 24:373–385

Baeza C, Corominas J (2001) Assessment of shallow landslide susceptibility by means of multivariate statistical techniques. Earth surf process landforms 26:251–1263

Bonham-Carter, GF (1994) Geographic information systems for geoscientists, modeling with GIS. Pergamon Press, Oxford, pp 398

Carro M, De Amicis M, Luzi L, Marzorati S (2003) The application of predictive modeling techniques to landslides induced by earthquakes: the case study of the 26 September 1997 Umbria-Marche earthquake (Italy). Eng Geol 69:139–159

Clerici A, Perego S, Tellini C, Vescovi P (2002) A procedure for landslide susceptibility zonation by the conditional analysis method. Geomorphology 48:349–364

Dai FC, Lee CF (2002) Landslide characteristics and slope instability modeling using GIS, Lantau Island, Hong Kong. Geomorphology 42:213–228

Dai FC, Lee CF, Li J, Xu ZW (2001) Assessment of landslide susceptibility on the natural terrain of Lantau Island, Hong Kong. Environ Geol 40:381–391

Donati L, Turrini MC (2002) An objective method to rank the importance of the factors predisposing to landslides with the GIS methodology: application to an area of the Apennines (Valnerina; Perugia, Italy). Eng Geol 63:277–289

Ercanoglu M, Gokceoglu C (2002) Assessment of landslide susceptibility for a landslide-prone area (north of Yenice, NW Turkye) by fuzzy approach. Environ Geol 41:720–730

Gokceoglu C, Sonmez H, Ercanoglu M (2000) Discontinuity controlled probabilistic slope failure risk maps of the Altindag (settlement) region in Turkey. Eng Geol 55:277–296

Guzzetti F, Carrarra A, Cardinali M, Reichenbach P (1999) Landslide hazard evaluation: a review of current techniques and their application in a multiscale study. Central Italy. Geomorphology 31:181–216

Jibson WR, Edwin LH, John AM (2000) A method for producing digital probabilistic seismic landslide hazard maps. Eng Geol 58:271–289

Lee S (2004e) Application of likelihood ratio and logistic regression model for landslide susceptibility mapping using GIS. Environ Manage 34:223–232

Lee S, Choi U (2003) Development of GIS-based geological hazard information system and its application for landslide analysis in Korea. Geosci J 7:243–252

Lee S, Choi J (2004) Application of a weight-of-evidence model to landslide susceptibility analysis. Int J Geogr Inf Sci 18:789–814

Lee S, Min K (2001) Statistical analysis of landslide susceptibility at Yongin, Korea. Environ Geol 40:1095–1113

Lee S, Chwae U, Min K (2002a) Landslide susceptibility mapping by correlation between topography and geological structure: the Janghung area, Korea. Geomorphology 46:49–162

Lee S, Choi J, Min K (2002b) Landslide susceptibility analysis and verification using the Bayesian probability model. Environ Geol 43:120–131

Lee S, Ryu JH, Min KD, Won JS (2003a) Landslide susceptibility analysis using GIS and artificial neural network. Earth Surf Process Landforms 27:1361–1376

Lee S, Ryu JH, Lee MJ, Won JS (2003b) Landslide susceptibility analysis using artificial neural network at Boun, Korea. Environ Geol 44:820–833

Lee S, Ryu JH, Won JS, Park HJ (2004a) Determination and application of the weights for landslide susceptibility mapping using an artificial neural network. Eng Geol 71:289–302

Lee S, Choi J, Woo I (2004b) The effect of spatial resolution on the accuracy of landslide susceptibility mapping: a case study in Boun, Korea. Geosci J 8:51–60

Lee S, Choi J, Min K (2004c) Landslide Hazard Mapping using GIS and Remote Sensing Data at Boun, Korea. Int J Remote Sens 25:2037–2052

Luzi L, Pergalani F, Terlien MTJ (2000) Slope vulnerability to earthquakes at subregional scale, using probabilistic techniques and geographic information systems. Eng Geol 58:313–336

Ohlmacher GC, Davis JC (2003) Using multiple logistic regression and GIS technology to predict landslide hazard in northeast Kansas, USA. Eng Geol 69:331–343

Parise M, Jibson WR (2000) A seismic landslide susceptibility rating of geologic units based on analysis of characteristics of landslides triggered by the 17 January, 1994 Northridge, California earthquake. Eng Geol 58:251–270

Pistocchi A, Luzi L, Napolitano P (2002) The use of predictive modeling techniques for optimal exploitation of s patial databases: a case study in landslide hazard mapping with expert system-like methods. Environ Geol 41:765–775

Romeo R (2000) Seismically induced landslide displacements: a predictive model. Eng Geol 58:337–351

Shou KJ, Wang CF (2003) Analysis of the Chiufengershan landslide triggered by the 1999 Chi-Chi earthquake in Taiwan. Eng Geol 68:237–250

Zhou CH, Lee CF, LI J, Xu ZW (2002) On the spatial relationship between landslides and causative factors on Lantau Island, Hong Kong. Geomorphology 43:197–207

Zhou G, Esaki T, Mitani Y, Xie M, Mori J (2003) Spatial probabilistic modeling of slope failure using an integrated GIS Monte Carlo simulation approach. Eng Geol 68:373–386