PFG 2010 / 1, 017 – 032
Stuttgart, März 2010

Article

Remote Sensing and GIS-based Landslide Susceptibility Analysis and its Cross-validation in Three Test Areas Using a Frequency Ratio Model

Biswajeet Pradhan, Dresden; Saro Lee, Daejon, Korea & Manfred F. Buchroithner, Dresden

Keywords: Landslide Susceptibility, Cross Validation, Remote Sensing, GIS, Frequency Ratio, Malaysia

Summary: The paper presents the results of the cross-validation of a frequency ratio model using remote sensing data and GIS for landslide susceptibility analysis in the Penang, Cameron, and Selangor areas in Malaysia. Landslide locations in the study areas were identified by interpreting aerial photographs and satellite images, supported by field surveys. SPOT 5 and Landsat TM satellite imagery were used to map landcover and vegetation index respectively. Maps of topography, soil type, lineaments and land cover were constructed from the spatial datasets. Nine factors which influence landslide occurrence, i. e. slope, aspect, curvature, distance from drainage, lithology, distance from lineaments, soil type, landcover, and NDVI, were extracted from the spatial database and the frequency ratio of each factor was computed. For all three areas the landslide susceptibility was analysed using the frequency ratios derived not only from the data for the respective area but also using the frequency ratios calculated from each of the other two areas (nine susceptibility maps in all) as a cross-validation of the model. For verification, the results of the analyses were then compared with the field-verified landslide locations. Among the nine cases, the case of Cameron based on the Cameron frequency ratio showed the highest accuracy (83%), and the case of Selangor based on the Penang frequency ratio showed the lowest accuracy (70%). Qualitatively, the model yields reasonable results which can be used for preliminary landslide hazard mapping.

Zusammenfassung: *Räumlich basierte Analyse von Erdrutschgefährdung und ihre Kreuzvalidierung in drei Untersuchungsgebieten auf Grundlage eines Frequenzquotienten-Modells.* Der Artikel präsentiert die Ergebnisse der Anwendung eines probabilistischen Modells unter Verwendung von Fernerkundungsdaten und GIS für Erdrutschgefährdungsanalysen in den Gebieten Penang, Cameron und Selangor in Malaysien. Die Erdrutsche wurden durch die Interpretation von Luft- und Satellitenbildern, gestützt auf Geländekartierungen, identifiziert. Zur Kartierung von Landbedeckung und Vegetationsindex wurden LANDSAT TM-Satellitenbilder verwendet. Von diesen raumbezogenen Datensätzen wurden digitale Karten von Topographie, Bodentypen, Lineamenten und Landbedeckung hergestellt. Neun Erdrutsch-bestimmende Faktoren, und zwar Hangneigung, Exposition, Hangkrümmung, Abstand zu Entwässerungslinien, Lithologie, Abstand zu Lineamenten, Bodentyp, Landbedeckung und NDVI, wurden aus der raumbezogenen Datenbank extrahiert und die Häufigkeitsquotienten für jeden Faktor berechnet. In Summe neun Karten der Erdrutschgefährdung wurden nicht nur auf der Basis der Daten der betreffenden Gebiete, sondern als Gegenprobe für die Validität der Methode auch mittels der Frequenzquotienten der jeweils anderen beiden Gebiete, erstellt. Zur Verifikation wurden die Analyseergebisse jedes Untersuchungsgebietes mit den tatsächlichen Erdrutschlokalitäten verglichen. Von den neun Fällen zeigte das Cameron-Gebiet auf der Basis des Cameron-Häufigkeitsquotienten die höchste Genauigkeit (83%), wohingegen das Selangor-Gebiet auf der Basis des Penang-Frequenzquotienten die geringste (70%) aufwies. Das Modell zeitigt sinnvolle qualitative Ergebnisse, die für Hangrutschungsrisiko-Kartierung verwendet werden können.

www.schweizerbart.de
DOI: 10.1127/1432-8364/2010/0037
1432-8364/10/0037 $ 4.00

1 Introduction

Landslides, are a recurrent problem throughout most of Malaysia. According to local newspaper reports (The Star 2008) in the years 2006 to 2008 but also in 2009 heavy rainfalls triggered landslides and mud flows along east coast highways in Peninsular Malaysia, in Sabah (East Malaysia) as well as in the island state of Penang. The areas hit hardest are along the Cameron Highland. in the mountainous state of Pahang on Peninsular Malaysia. These landslides cost millions of dollars of property loss and even lives. The extent of the damages could be minimized if a long-term early warning system predicting the mass movements in the landslide-prone areas would have been in place.

The landslides that occurred along the New Klang Valley Express Highways (NKVE) Region in the year 2003 have alerted the highway authorities and other governmental organizations towards the seriousness of landslide management and prevention. The October 2002 landslide in Kuala Lumpur which completely destroyed few houses and killed six members of a family is still in the public's memory. Landslides in Malaysia are mainly triggered by tropical rainfall and flash floods causing failure of the rock surface along fracture, joint and cleavage planes. The geology of the country is quite stable but continuous development and urbanization lead to deforestation and erosion of the covering soil layers, thus causing serious threats to the slopes.

In the past Penang Island, Cameron Highland and the area of Selangor faced numerous landslide and mudflow events, and much damage was caused in these areas. Most of these landslides have been triggered by heavy rainfall. However, only little effort has been made to assess or predict these events which caused serious damages. Through scientific analyses of these landslides, one can assess and predict landslide-susceptible areas and even the events as such, and thus reduce landslide damages through proper preparation and/or mitigation. Therefore, understanding the landslides and preventing them is one of the serious challenges, not only for Malaysia. To achieve this objective, techniques of landslide susceptibility analysis were validated and subsequently cross-validated in the three study areas using a frequency ratio model.

2 Previous Work

Many studies have been carried out on landslide hazard evaluation using GIS and Geoinformation-related techniques. Guzzetti et al. (1999) conducted GIS-based studies in the Umbria and Marches regions of Central Italy and also summarized many case studies of landslide hazard evaluation along the Apennines Mountains. Reports of landslide analyses using GIS and probabilistic models were also published (Temesgen et al. 2001; Dai et al. 2001; Pistochi et al. 2002; Akgun et al. 2008; Clerici et al. 2006; Pradhan et al. 2006; Lee et al. 2004a; Lee 2005; Lee & Lee 2006; Pradhan & Lee 2009). Most of the above studies have been conducted using the regional landslide inventories derived from aerial photographs. Guzzetti et al. (1999) developed statistical models using logistic regression for landslide hazard mapping (Tunuslouglu et al. 2008; Lamelas et al. 2008; Wang & Sasa 2005; Suzen & Doyuran 2004; Lee 2005; Lee & Pradhan 2006, 2007; Pradhan et al. 2008, Pradhan 2010). The geotechnical and the safety factor models are also good tools to spatialize landslide hazard analysis, and they have the potential to develop scenarios by changing the input parameters (Shou & Wang 2003). All these models provide solutions for integrating information levels and mapping the outputs. Recently, other new methods have been applied for landslide hazard evaluation using data mining, fuzzy logic, safety factor and artificial neural network models (Ercanoglu & Gokceoglu 2002; Tangestani 2004; Lee et al. 2003, 2004b; Pradhan & Lee 2008a; Pradhan et al. 2009a, 2009b, 2009c 2009d, 2010). The spatial results of these approaches are generally appealing, and they give rise to qualitatively and quantitatively map the landslide-susceptible and hazard areas.

The main difference between this study and the approaches described in the aforementioned publications is that the frequency ratio model was validated and also cross-validated in three study areas.

The landslide occurrences in the study areas were detected by interpretation of aerial

photographs and by field survey. A map showing the most recent landslide inventory was developed based on aerial photographs in combination with GIS for Penang Island, Cameron Highland and Selangor. Remote sensing methods, using aerial photographs and satellite images were employed to obtain significant and cost-effective information on landslides. In this study aerial photographs at scales of 1:10 000 – 1:50 000, taken between 1981 and 2005, were used to map the landslide locations. The inventory maps were prepared by a structural geologist with a profound knowledge in airphoto interpretation. These landslides can be seen in aerial photographs by interpreting breaks in the forest canopy, bare soil, and other typical geomorphic characteristics of landslide scars. Nine landslide-related factors, namely slope, aspect, curvature, distance from drainage, lithology, distance from lineaments, landcover, soil types and normalised difference vegetation index (NDVI) were either directly extracted or calculated from the spatial database or the Landsat TM and SPOT 5 satellite images. Using the detected landslide locations and factors, a landslide analysis method based on a frequency ratio model was applied and validated. To achieve this, the calculated and extracted factors were put into a 10×10m grid (Arc/Info GRID type), and then converted into ASCII data for use with the frequency ratio model. Subsequently, frequency ratio values of each factor were determined and landslides susceptibility maps constructed. Then the ratio values were applied to the other two study areas. Finally, the landslide susceptibility analysis results were validated and cross-validated using the landslide locations of all three study areas. The validation was performed by comparing all existing landslides and landslide susceptibility analysis results for the study areas.

3 Study Areas and Spatial Datasets

Three study areas, which have been badly affected in recent years, Penang Island, Cameron Highland and Selangor, were selected as suitable study areas for the present research. Penang Island lies between 35º15' N and 5º30' N, and 100º10'E and 100º20'E, and covers an area of 285 km^2 (Fig. 1). The bedrock geology of the study area consists mainly of granite. Cameron Highland lies between 4° 32‘ N and 4° 23‘ N, and 101° 22‘ E and 101° 31‘ E, and covers an area of 660 km^2. The geology of the Cameron Highland consists of mostly two types of litho types: igneous and metamorphic rocks. The third study area, Selangor, lies between 3º 23' 53.6" E and 3º 45' 18.05" E, and 101º 30' 55.33" N and 101º 3' 36.3" N, and covers an area of 8,179.28 km^2. The bedrock geology of the study area consisits of granite and gneiss. In all the three study areas landslides occurred when the maximum daily rainfall was 208 mm.

Maps relevant to landslide occurrence in the study areas were constructed in spatial vector datasets using the ARCInfo version 9.0 GIS software. These included topographic maps at a scale of 1:25,000, soil maps at 1:25,000 and geology maps at 1:63,300. A land-use map was extracted from Landsat TM satellite images with a resolution of 30 m. Data layers and overall methdology used in the analysis are shown in Fig. 2. Contour lines and spot heights were extracted from the topographic map and subsequently Digital Elevation Models (DEMs) were constructed for all study areas. Using the DEMs, slope, aspect and curvature were calculated. Soil types, litho types and distance from drainage were acquired from soil, geology and topographic maps respectively. The location of lineaments was extracted from structural map and further refined using up-to-date SPOT 5 satellite images. Then the distance from the lineaments was calculated in ArcGIS with 100 m increments based on the Euclidean distance method. The lithology map was prepared from the geological map. The lineament buffer was calculated in 100 m intervals and classified into 10 equal area classes. SPOT 5 scenes of 5 January 2005 (Penang Island and Cameron Highland), and 19 April 2005 (Selangor) were classified to map the different landcover classes. The landcover maps were prepared using SPOT 5 images (2.5 m spatial resolution) applying a supervised classification supplemented with field survey. An overall classification accuracy of up to 89% was achieved. Finally,

the Normalized Difference Vegetation Index (NDVI) maps were generated from LANDSAT TM satellite images acquired 25 January 2005, 7 March 2005, and 15 September 2005. The NDVI value was calculated using the formula NDVI = (IR – R) / (IR + R), where IR is the energy reflected in the infrared portion of the electromagnetic spectrum, and R is the energy reflected in the red portion of the electromagnetic spectrum.

For all three study areas the datasets were divided into grids with 10 × 10 m cells. The Penang Island dataset resulted in 2493 rows and 1887 columns, and the cell number being 4704,291. In 463 of them landslides had occurred. The Cameron dataset was composed of 2418 rows and 1490 columns with 3602,820 cells, landslides occurring in 324 of them. The Selangor dataset was composed of 1088 rows and 992 columns (total 1079,296 cells). Landslides occurred in 327 cells.

4 Frequency Ratio Model

Frequency ratio approaches are based on the observed relationship between the distribution of landslides and each landslide-related factor in order to reval the correlation between landslide locations and the geo-factors determining the study areas (Lee & Pradhan 2007b). Using the frequency ratio model, the spatial

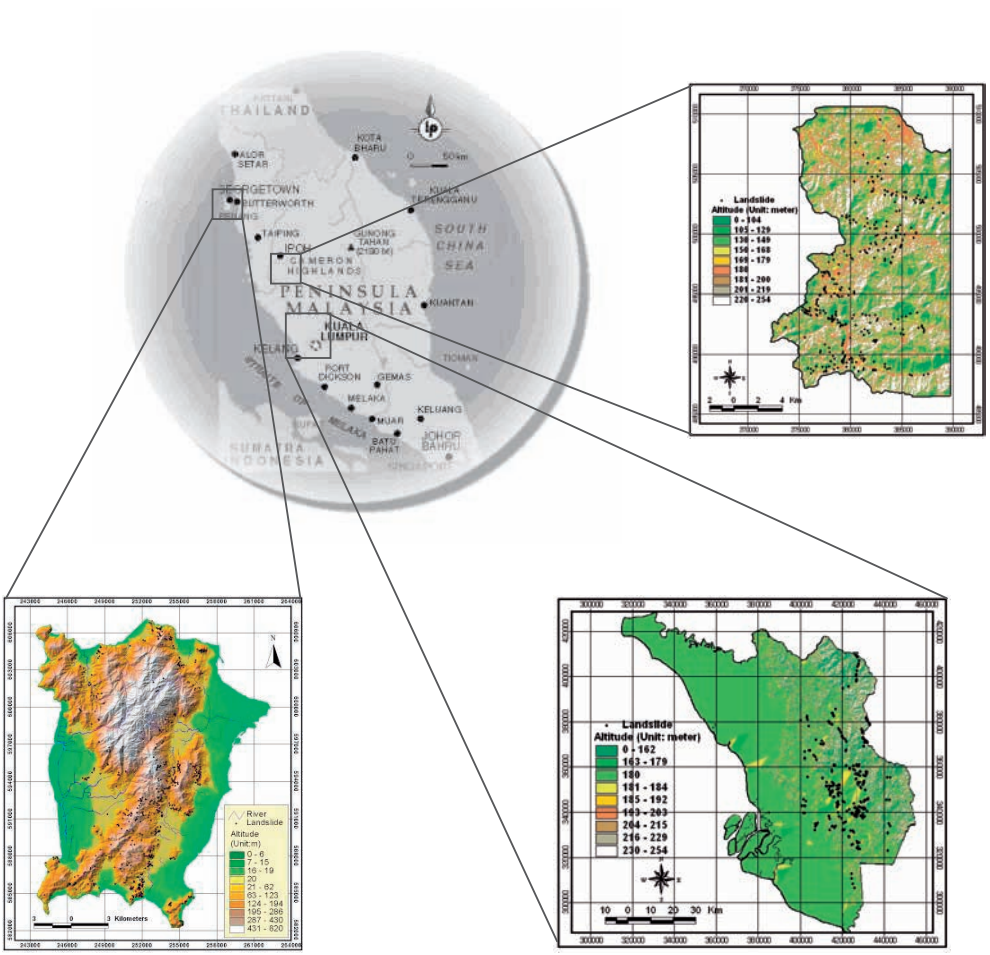


Fig. 1: Study areas Penang Island, Cameron Highland and Selangor on the Malaysian Peninsula. Black dots indicate field-verified landslide locations.

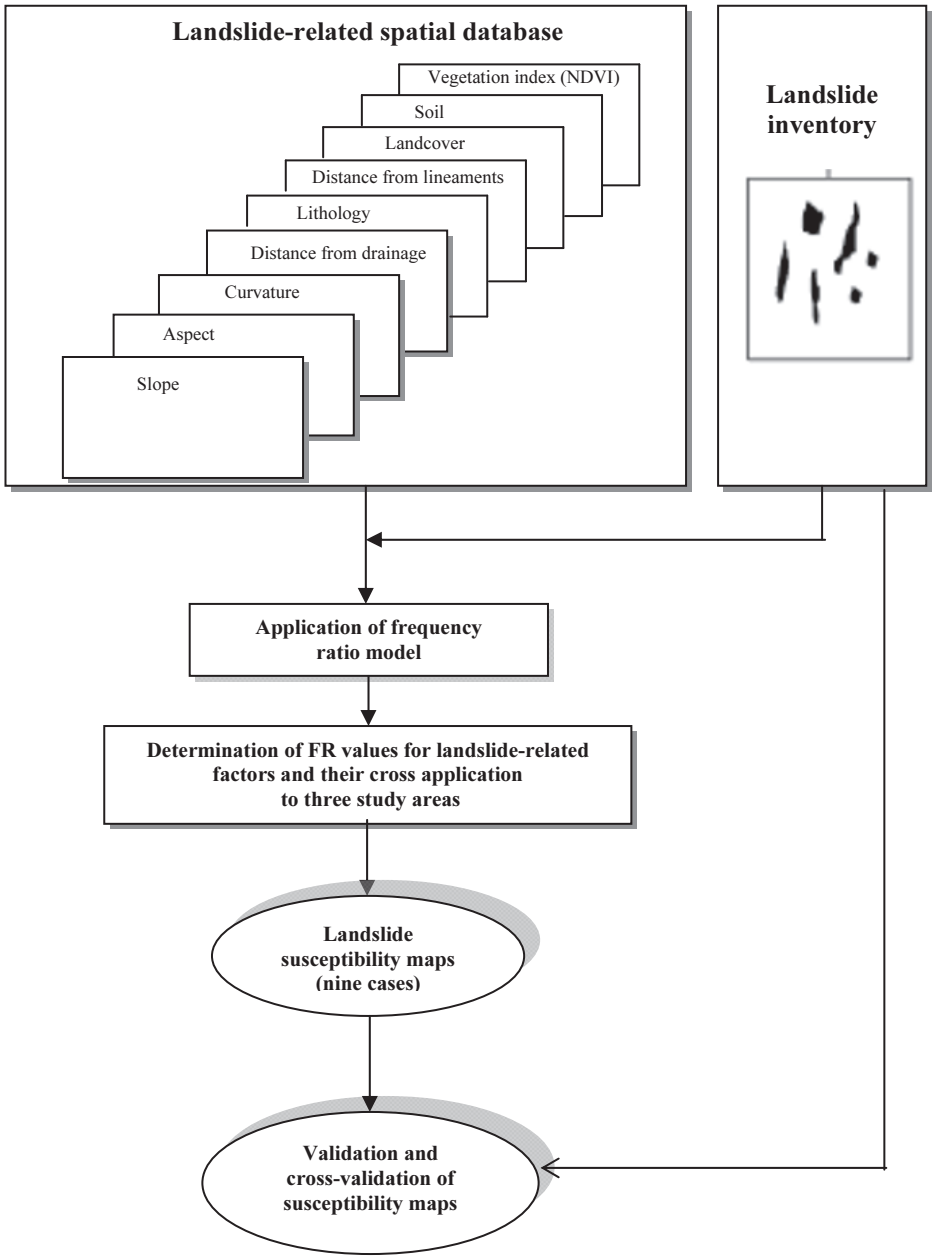


Fig. 2: Data layers and flow diagram showing the overall methodology.

relationships between landslide occurrence location and each of the factors contributing to the occurrence of landslides (slope, aspect, curvature, distance from drainage, lithology, distance from lineaments, soil, landcover, and NDVI) were derived. The frequency ratios of each factor's type or range were calculated from their relationship with landslide events for three study areas. They are listed in Tab. 1. The frequency ratio denotes a ratio between

Tab. 1: Frequency ratio of landslide-related factors for Penang Island, Cameron Highland and Selangor.

Factors/class		Penang area					Cameron area					Selangor area				
		No. of total pixels	% of total pixels ⓐ	No. of land-slide pixels	% of lands-lide pixels ⓑ	Ratio ⓑ/ⓐ	No. of total pixels	% of total pixels ⓐ	No. of land-slide pixels	% of lands-lide pixels ⓑ	Ratio ⓑ/ⓐ	No. of total pixels	% of total pixels ⓐ	No. of land-slide pixels	% of lands-lide pixels ⓑ	Ratio ⓑ/ⓐ
Slope	0 ~ 15°	1709800	57.87	53	11.45	0.20	113257	38	44	13.58	0.35	67777334	82.86	115	35.17	0.42
	16 ~ 25°	765189	25.90	152	32.83	1.27	114813	39	100	30.86	0.79	8426979	10.30	87	26.61	2.58
	26 ~ 35°	360229	12.19	157	33.91	2.78	55949	19	125	38.58	2.03	4648328	5.68	68	20.80	3.66
	> 35°	119564	4.05	101	21.81	5.39	10826	4	55	16.97	4.62	940212	1.15	57	17.43	15.16
Aspect	Flat	1199400	40.59	13	2.80	0.07	0.00	0	0	0	0	32746440	40.04	0	0.00	0.00
	North	206629	6.99	41	8.85	1.27	487967	18.38	43	13.27	0.72	5541254	6.77	55	16.82	2.48
	Northeast	207860	7.03	51	11.01	1.57	317441	11.96	70	21.60	1.81	6678670	8.17	45	13.76	1.69
	East	228674	7.74	60	12.95	1.67	304579	11.47	50	15.43	1.34	6077711	7.43	27	8.26	1.11
	Southeast	236988	8.02	82	17.71	2.21	307210	11.57	62	19.14	1.65	6181783	7.56	28	8.56	1.13
	South	205108	6.94	58	12.53	1.80	305365	11.50	28	8.64	0.75	5564596	6.80	29	8.87	1.30
	Southwest	206970	7.01	52	11.23	1.60	318819	12.01	22	6.79	0.57	6721149	8.22	35	10.70	1.30
	West	228117	7.72	54	11.66	1.51	305257	11.50	15	4.63	0.40	6105258	7.46	46	14.07	1.88
	Northwest	235036	7.95	52	11.23	1.41	307938	11.60	34	10.49	0.90	6175992	7.55	62	18.96	2.51
Curvature	Concave	770757	26.09	50	10.80	0.41	1234845	46.52	50	15.43	0.33	13288765	16.25	66	20.18	1.24
	Flat	1419529	48.04	45	9.72	0.20	242494	9.13	0	0.00	0.00	55283859	67.59	15	4.59	0.07
	Convex	764496	25.87	368	79.48	3.07	1177237	44.35	274	84.57	1.91	13220229	16.16	246	75.23	4.65
Distance from drainage	0-50m	919481	31.11	117	25.26	0.81	121747	45.86	148	45.67	1.00	25436680	31.09	122	37.30	1.20
	51-100m	648322	21.94	114	24.62	1.12	787380	29.65	95	29.32	0.99	17379195	21.24	101	30.88	1.45
	101-150m	453610	15.35	80	17.27	1.13	409687	15.43	48	14.81	0.96	11097910	13.57	43	13.14	0.97
	151-200m	299500	10.13	53	11.44	1.13	160484	6.04	24	7.40	1.23	6568267	8.03	35	10.70	1.33
	201-250m	189645	6.41	41	8.85	1.38	50232	1.89	6	1.85	0.98	3790284	4.63	14	4.28	0.92
	251-300m	120824	4.08	14	3.02	0.74	15235	0.57	1	0.30	0.54	2455628	3.01	8	2.44	0.81
	> 301m	323400	10.94	44	9.50	0.87	14206	0.53	2	0.61	1.15	15064893	18.42	4	1.22	0.07
Lithology	Igneous rock	2195706	76.65	461	99.56	1.30	1900435	71.58	216	66.66	0.93	52626677	64.88	203	62.07	0.96
	Alluvium	668834	23.34	2	0.43	0.02	-	-	-	-	-	-	-	-	-	-
	Metamorphic rock	-	-	-	-	-	754263	28.41	108	33.33	1.17	28127205	34.67	122	37.30	1.08
	Sedimentary rock	-	-	-	-	-	-	-	-	-	-	52626677	64.88	203	62.07	0.96

Distance from lineament	0-200m	341528	11.55	62	13.39	1.16	73338	24.86	122	37.65	1.51	2766801	3.38	17	5.19	1.54
	201-500m	490760	16.60	117	25.26	1.52	85885	29.11	98	30.24	1.04	4066326	4.97	41	12.53	2.52
	501-1000m	667520	22.59	118	25.48	1.13	73054	24.76	52	16.04	0.65	6532091	7.99	43	13.14	1.65
	1001-2000m	670438	22.68	90	19.43	0.86	46175	15.65	38	11.72	0.75	11288943	13.80	54	16.51	1.20
	2001-4000m	641690	21.71	68	14.68	0.68	16487	5.58	14	4.32	0.77	16185017	19.78	91	27.82	1.41
	>4001m	142846	4.83	8	1.72	0.36	8	0.00	0	0	0.00	40953679	50.07	81	24.77	0.49
Soil	KNJ	363486	12.64	15	3.23	0.26	-	-	-	-	-	4913270	6.032	0	0	0
	SLR-KGG	5691	0.19	0	0	0	-	-	-	-	-	14262446	17.51	0	0	0
	TMG-AKB-LAA	257476	8.95	26	5.61	0.63	-	-	-	-	-	2852442	3.50	18	5.50	1.58
	RGM-BTG	202745	7.05	65	14.03	2.01	-	-	-	-	-	-	-	-	-	-
	STP	1486944	51.71	313	67.60	1.31	2604027	98.09	297	91.66	0.93	15366929	18.86	82	25.07	1.33
	ULD	558937	19.43	42	9.07	0.47	50671	1.90	27	8.33	4.37	5957934	7.31	123	37.61	5.14
	BRH-OCM	-	-	-	-	-	-	-	-	-	-	823554	1.01	0	0	0
	PET	-	-	-	-	-	-	-	-	-	-	17371280	21.33	0	0	0
	MCA-TVY-GMI	-	-	-	-	-	-	-	-	-	-	46951	0.06	0	0	0
	MUN-SBN	-	-	-	-	-	-	-	-	-	-	1816066	2.23	20	6.11	2.74
	SDG-BGR-MUN	-	-	-	-	-	-	-	-	-	-	8738739	10.73	24	7.33	0.68
	RGM-JRA	-	-	-	-	-	-	-	-	-	-	3146337	3.86	34	10.39	2.71
	PRG	-	-	-	-	-	-	-	-	-	-	236370	0.29	0	0	0
	SDG-KDH	-	-	-	-	-	-	-	-	-	-	2999777	3.68	16	4.89	1.33
	MLD	-	-	-	-	-	-	-	-	-	-	2927283	3.59	10	3.05	0.86
Land cover	CLEAR LAND	75815	2.47	4	0.86	0.35	-	-	-	-	-	-	-	-	-	-
	MANGROVE	144557	4.72	13	2.80	0.59	-	-	-	-	-	35366176	43.24	127	38.83	0.9
	ROCK	10256	0.33	0	0	0	-	-	-	-	-	-	-	-	-	-
	PRI_FOREST	1587492	51.87	364	78.61	1.52	1972384	74.66	125	38.58	0.52	-	-	-	-	-
	GRASS	269083	8.79	15	3.23	0.37	110259	4.17	26	8.02	1.92	-	-	-	-	-
	COCONUT	88328	2.88	2	0.43	0.15	-	-	-	-	-	-	-	-	-	-
	RUBBER	674704	22.04	53	11.44	0.52	137986	5.22	30	9.25	1.78	6981927	8.54	4	1.22	0.14
	TIN MINE	2168	0.07	5	1.07	15.25	-	-	-	-	-	4465198	5.45	47	14.37	2.63
	MUD	83739	2.73	0	0	0	-	-	-	-	-	-	-	-	-	-
	SEC_FOREST	27618	0.90	4	0.86	0.96	358942	13.58	121	37.34	2.75	2956866	3.61	1	0.30	0.08
	CULTIVATED	93632	3.05	1	0.21	0.08	-	-	-	-	-	-	-	-	-	-
	PALM OIL	2960	0.09	0	0	0	-	-	-	-	-	6394776	7.81	17	5.19	0.66
	WATER BODY	63	0.00	0	0	0	2174	0.08	0	0	0	39539	0.04	0	0	0
	CUTTING	-	-	-	-	-	30434	1.15	13	4.01	3.49	-	-	-	-	-
	SETTLEMENT	-	-	-	-	-	29310	1.10	9	2.77	2.50	25588370	31.28	131	40.06	1.28
NDVI	100~ -50	12873	0.44	1	0.21	0.49	234	0.07	0	0	0.00	31960	0.039	0	0	0.00
	-50~ -0	563909	19.47	62	13.39	0.69	2505	0.84	3	0.92	1.09	3738742	4.57	32	9.78	2.14
	0~ 50	224249	77.42	380	82.07	1.06	59506	20.17	116	35.80	1.77	42440711	51.88	212	64.83	1.25
	500~ -100	76911	2.65	20	4.31	1.63	232702	78.89	205	63.27	0.80	35581439	43.50	83	25.38	0.58

occurrence and absence of landslides in each cell. In the relation analysis, the ratio is that of the area where landslides occurred in the entire area, so that a value of 1 represents an average value. If the value is bigger than 1, this implies a higher correlation, and values lower than 1 stand for lower correlations.

For the sake of simplicity only the frequency ratio of the Penang dataset is discussed here (Tab. 1, column 2). In the case of slope, the relationship between landslide occurrence and slope gradient shows that steeper slopes have a higher landslide probability. For slopes of 15° and less the frequency ratio was 0.2 which indicates a very low probability of landslide occurrence. For slopes above 26° the ratio was > 2, thus indicating a high probability of landslide occurrence. As the slope angle increases, the shear stress in the soil or other unconsolidated material generally increases, too. 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 aspect (Tab. 1), landslides were most abundant on south- and southwest-facing slopes. The frequency of landslides was lowest on west-, northwest-, and northwest-facing slopes, except in flat areas. The curvature values represent the morphology of the relief. A positive curvature indicates that the surface is upwardly convex at that pixel. A negative curvature indicates that the surface is upwardly concave at that pixel. A value of zero indicates that the surface is flat. As shown in Tab. 1, the higher a positive or negative curvature value, the higher is the probability of landslide occurrences. Flat areas had a frequency ratio of 0. Concave areas had a frequency ratio of 0.41. The reason for this is that subsequent to heavy rainfall, a concave slope contains more water and retains this water for a longer period which could lead to a slope failure triggering a landslide.

Convex areas had a frequency ratio of 3.07. The reason for this is that a convex rounded hilltop slope could expose to heavy rainfall causing repeated dilation and contraction of loose debris on an inclined surface that might induce a creeping or mudslide. Analyses were carried out to assess the influence of drainage lines on landslide occurrence. For this purpose, the proximity of landslides to drainage lines was identified by buffering (Tab. 1). It can be seen that as the distance from a drainage line increases, the landslide frequency generally decreases. At a distance of < 250 m, the ratio was > 1, indicating a high probability of landslide occurrence, and at distances > 251 m, the ratio was < 1, indicating very less probability. This can be attributed to the fact that terrain modification caused by gully erosion may influence the initiation of landslides. However, at a distance of < 50 m, the frequency ratio is 0.81 which is due to the lower number of previously triggered landslides. In the case of lithology it was found that the frequency ratio was lower (0.97) in alluvium types of rocks, and higher (1.30) in igneous areas. In the case of the distance from lineamens, the closer the distance to a lineament was, the greater was the landslide-occurrence probability. For distances of < 100 m, the ratio was > 1, indicating a high probability of landslide occurrence, and for distances of > 1000 m, the ratio was < 1, thus indicating a low probability. As the distance from lineament decreases, the fracturing of the rock increases, and the degree of weathering increases, thus resulting in greater chances of landslides.

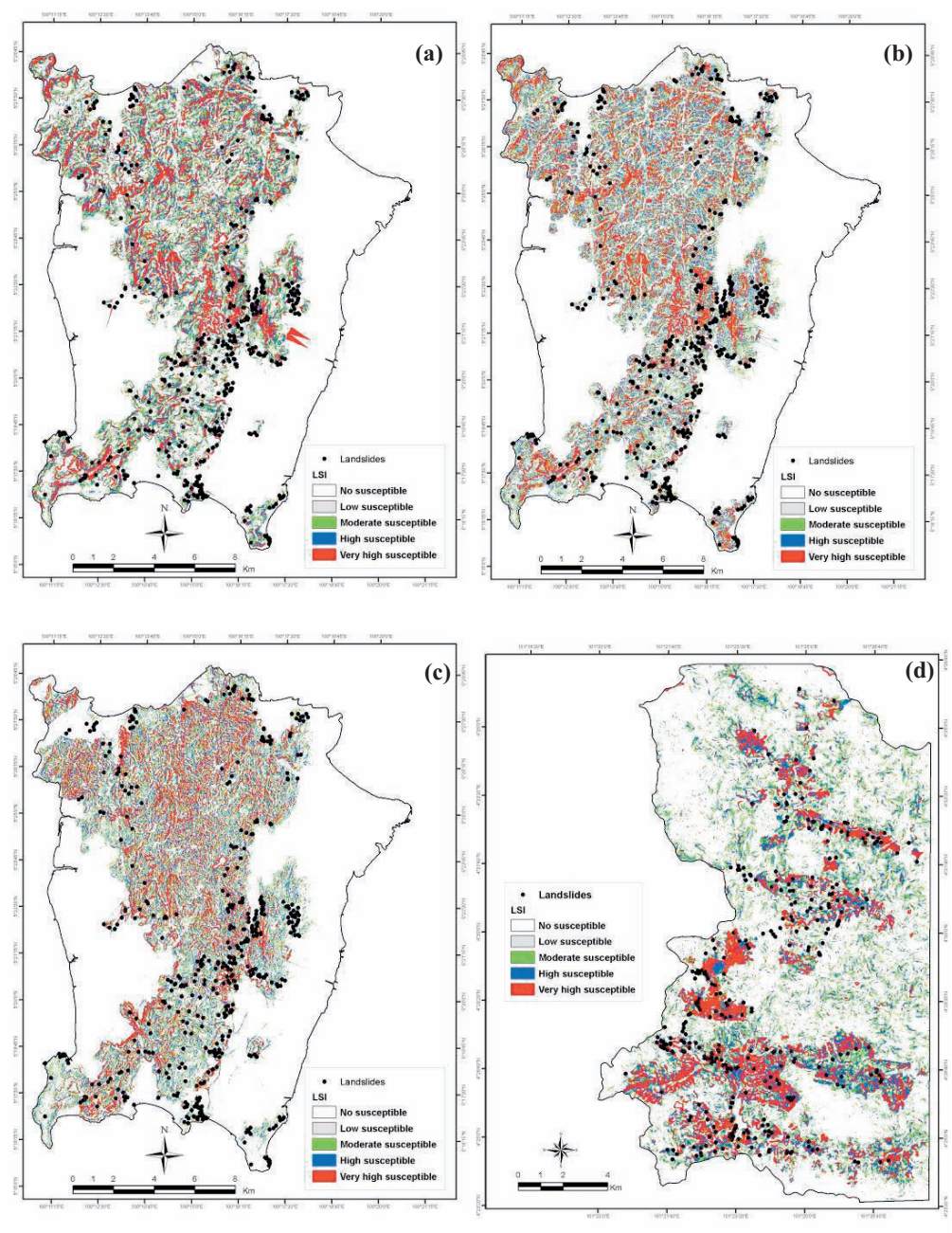
As for the soil type (Tab. 1), the frequency ratio was comparatively higher for RGM-BTG series (2.01) and STP (1.31). This indicates that the landslide probability increases with steeper land. In the case of landuse (Tab. 1), the landslide-occurrence values were higher in tin mine areas (15.25) and primary woods (1.52) but lower in hard rock areas and dense forest. Regarding the vegetation index, for NDVI values above 0 the frequency ratio was > 1, which indicates a high landslide-occurrence probability, and for NDVI values below 0 the frequency ratio was < 1, indicating a low landslide-occurrence probability. This result implies that the landslide probability decreases with the increase of the vegetation index value. This appears in the first instance unusual but can be explained by the fact that more vegetation develops along tectonic zones of weakness. Similar findings and explanations can be given for the Cameron and Selangor areas (Tab. 2, column 3 and 4).

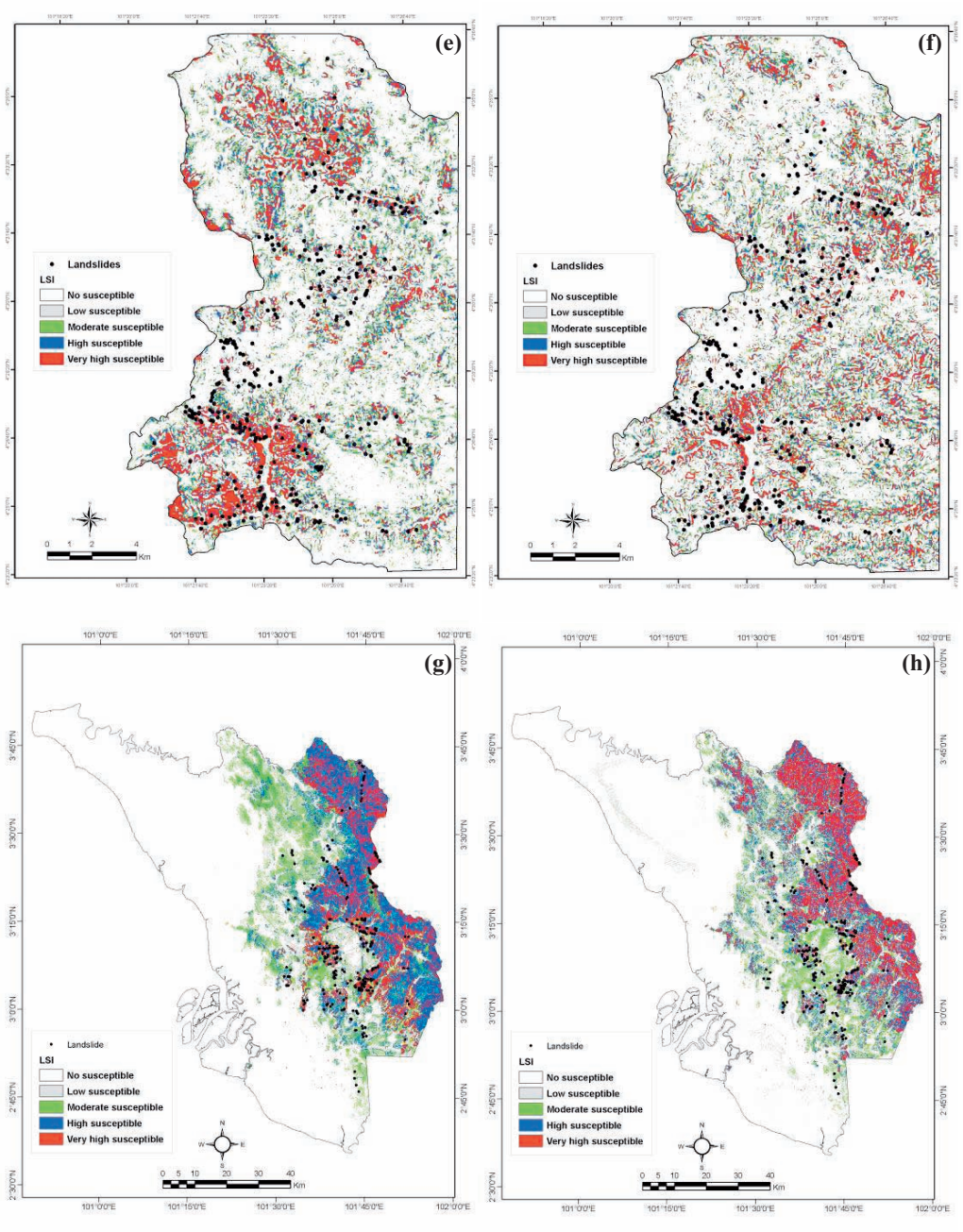
5 Application of Frequency Ratio Model for Landslide Susceptibility Analysis

The frequency ratio valuess were used for calculating the landslide susceptibility index and mapping. The ratios of each factor type or range were summed to calculate the landslide susceptibility index, as shown in Eq. (1).

$$LSI = \sum Fr \tag{1}$$

Where *Fr* is the frequency ratio of each factor type or range.





For landslide susceptibiliy mapping, the frequency ratios were applied to the study area from which they were derived, as well as to the other two areas. That is, the calculated frequency ratios from each dataset (for Penang, Cameron, and Selangor) were applied to all datasets (Penang, Cameron and Selangor). Overall, there were nine cases for mapping. Thus, the calculated ratings from the Penang datasets were applied to Penang, Cameron and Selangor. Similarly, the calculated ratings from the Cameron datasets were applied to

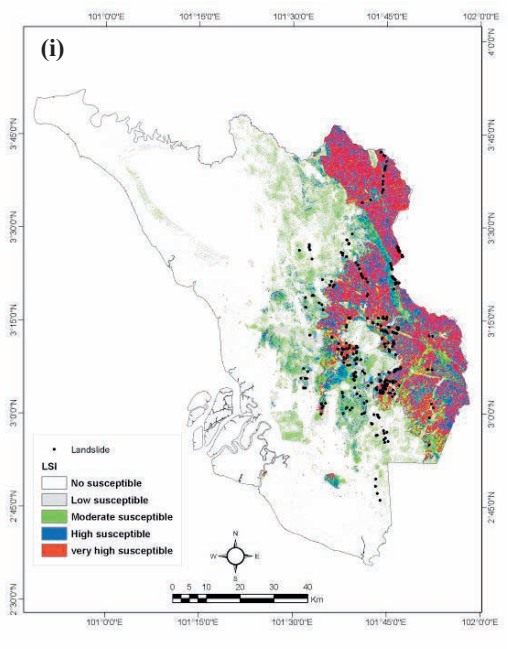


Fig. 3: The nine calculated landslide susceptibility maps of the study areas using the frequency ratio model. (a) Landslide susceptibility map of Penang based on frequency ratio of the Penang datasets; (b) Landslide susceptibility map of Penang based on frequency ratio of the Cameron datasets; (c) Landslide susceptibility map of Penang based on frequency ratio of the Selangor datasets; (d) Landslide susceptibility map of Cameron based on frequency ratio of the Cameron datasets; (e) Landslide susceptibility map of Cameron based on frequency ratio of the Penang datasets; (f) Landslide susceptibility map of Cameron based on frequency ratio of the Selangor datasets; (g) Landslide susceptibility map of Selangor based on frequency ratio of the Selangor datasets; (h) Landslide susceptibility map of Selangor based on frequency ratio of the Penang datasets; and (i) Landslide susceptibility map of Selangor based on frequency ratio of Cameron datasets.

Penang, Cameron and Selangor, and those from the Selangor datasets were applied to Penang, Cameron and Selangor, giving nine sets to be mapped. Using the frequency ratio (Tab. 1) and Eq. (1), the *LSI* values were computed for the nine cases. If no ratio was available for a certain class, the average value (i. e., 1) was used.

Hence, as presented in Fig. 3, nine landslide susceptibility maps were calculated. Fig. 3a presents Penang based on the Penang ratings, Fig. 3b the same area based on the Cameron ratings, and Fig. 3c, on the Selangor ratings. Then the calculated landslide susceptibility indices (*LSI*) were grouped into four groups by equal area classfication (highest 10%, second-highest 10%, third-highest 20% and remaining 60%) for easy visual interpretation of the landslide susceptibility. The landslide suscepti-bility increases with the height of the *LSI* value. The patterns of the identical study areas proofed to be very similar, but there were some differences in the distribution of the index values. Fig. 2 (a), for example, represents a landslide susceptibility map of Penang calculated by using frequency ratios from the Penang datasets. Here the minimum, mean and maximum values of each *LSI* are 1.6, 9.15 and 20.38, respectively. Similarly, Fig. 2 (b) shows a landslide susceptibility map of Penang calculated on the basis of frequency ratios from the Cameron datasets. In this case, the minimum, mean and maximum *LSI* are 4.96, 11.77 and 19.33.

6 Validation of the Susceptibility Maps

To validate the applied landslide susceptibility calculation method, two basic assumptions are required. One is that landslides can be related to spatial information (such as topography, soil, lithology, lineaments, drainage, land cover and NDVI) and the other one is that future landslides will be triggered by a specific impact factor, such as rainfall. In this study, these two assumptions are reasonably fulfilled because the landslides are related to spatial information and all the landslides were caused by heavy rainfall in Penang, Cameron and Selangor (Pradhan & Lee 2008).

The results of the landslide susceptibility analysis were displayed in the maps of Penang, Cameron, and Selangor which were separately computed on the basis of each of the

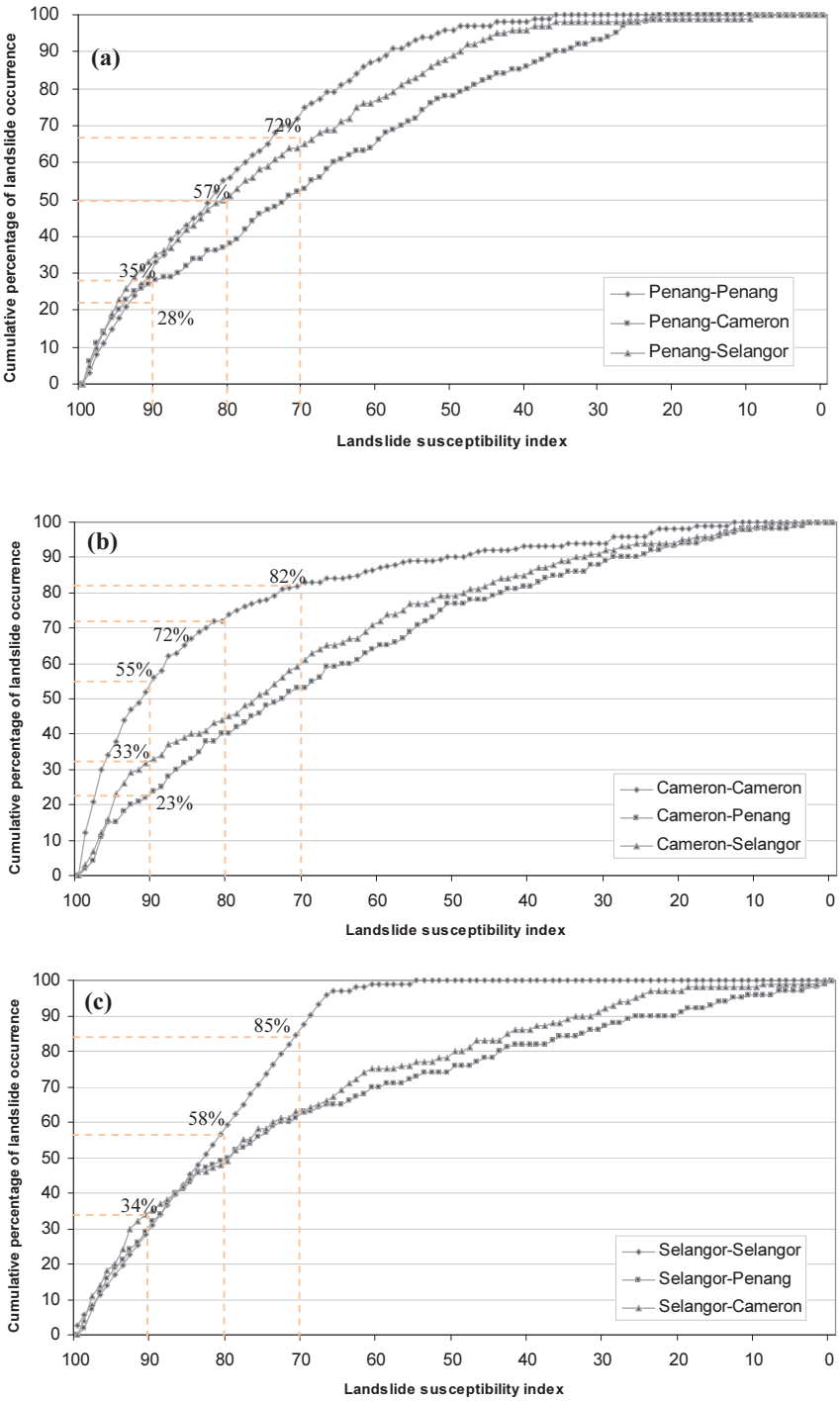


Fig. 4: Illustration of cumulative frequency diagrams showing the landslide susceptibility index ranking (x-axis) in cumulative percents of landslide occurrence (y-axis). (a) Validation result of Penang based on three areas; (b) Validation result of Cameron based on three areas; and (c) Validation result of Selangor based on three areas.

Tab. 2: Values of "areas under the curve" approach depicted in Fig. 4.

Casea	Penang-Penang	Penang-Cameron	Penang-Selangor	Cameron-Cameron	Cameron-Penang	Cameron-Selangor	Selangor-Selangor	Selangor-Penang	Selangor-Cameron
Area	0.0835	0.7058	0.7708	0.8399	0.7109	0.7215	0.8361	0.7030	0.7369
Area ratio	80.35%	70.58%	77.08%	83.99%	71.09%	72.15%	83.61%	70.30%	73.69%

[a] The frequency ratio of the second area was applied to the first area

Penang, Cameron and Selangor factors and subsequently validated and cross-validated using all landslide locations in these areas. The maps of Penang, calculated by means of the Penang, Cameron and Selangor ratings, were validated using the entire landslide locations in Penang, Cameron and Selangor. Also, the maps of Cameron, calculated on the basis of the Penang, Cameron and Selangor parameters, were validated using landslide locations in Penang, Cameron and Selangor. Likewise, for the study area of Selangor the corresponding procdure was applied. Therefore, overall validations were performed in nine cases.

A comparative depiction of the results like the one given in Fig. 4 illustrates how well the nine landslide susceptibility maps match relity. To obtain Fig. 4, the relative ranks of landslide susceptibility maps and landslide occurrence were calculated for each case, and the validation results were divided into classes of accumulated area ratios according to the percentage of the landslide susceptibility indices.

The above procedure was applied to each of the study areas. In the case of the application of the Penang frequency ratio to the study area of Penang (Fig. 4a), the 90 – 100% class with the highest 10% of probability of a landslide contains 35% of the landslides in that area. The 0 – 20% class contains 57%, and the 0 – 30% class contains 72% of all landslides in Penang. As for the application of the Cameron frequency ratio to Cameron (Fig. 4b), the 90 – 100% class with the highest possibility (10%) of a landslide contains 55% of the landslides in Cameron. The 0 – 20% class contains 72% and the 0 – 30% class contains 82% of the landslides in Cameron. For Selangor (Fig. 4c), the corresponding figures read 34%, 55% and 85% of all landslides occurring in Selangor.

The above procedure was also adapted for the other two study areas. When applying the Cameron and Selangor frequency ratios to Penang (Fig. 4a), the 90 – 100% class with the highest 10% possibility of landslides contains 28% of the landslides occurring in Cameron and 31% of the landslides of Selangor. In the case of the application of the Penang and Selangor frequency ratios to Cameron (Fig. 4b), the 90 – 100% class with the highest possibility of landslides contains 23% of the landslides of Penang area and 33% of the landslides of Selangor. When applying the Penang and Cameron frequency ratio to Selangor (Fig. 4c), the 90 – 100% class with the highest possibility of landslides contains 31% of the landslides occurring in Penang and 31% of those of Cameron.

To compare the result quantitatively, the areas under the curves were recalculated. If the total area is found to be 1 which means a perfect prediction accuracy. Hence, the areas under curves can be used to assess the prediction accuracy qualitatively. They are shown in Tab. 2. In the case of Penang based on Penang frequency ratio value, the area ratio was 0.8035, thus implying a prediction accuracy of 80.35%. In the case of Penang based on Cameron frequency ratio, the area ratio was 0.7058 and the prediction accuracy 70.58%. When applying the Selangor frequency ratio to Penang, the area ratio was 0.7708, and the prediction accuracy 77.08%. In the case of Cameron based on Cameron frequency ratio, the area ratio was 0.8399, and the prediction accuracy is 83.99%. In the case of Cameron based on Penang frequency ratio, the area ratio was 0.7109 and the prediction accuracy is 71.09%. Further "under the curve" values and the corresponding prediction accuracies can be retrieved from Tab. 2.

7 Conclusions and Discussion

The frequency ratio-based cross application approach was successfully used for the three study regions Penang Island, Cameron Highland, and Selangor. The frequency ratio model permitted to determine the ratings for the input layers and produce nine sets of landslide susceptibility maps after the cross application of the ratings to the three study areas. This allows drawing the following conclusions from the experience gained in these study areas with different geological and geomorphological environment.

For the landslide susceptibility analysis and the establishemnt of a landslide-related GIS database of all three study areas landslide locations were mapped using aerial photographs. For the landslide susceptibility analysis, the frequency ratio model was applied, validated, and cross-validated for the three study areas using the landslide database. Then, the results were validated by calculating the correlation between actual landslide locations and probable occurrences.

The calculated ratings based on the frequency ratio showed a similar trend for each study area. Among the nine generated landslide susceptibility maps, case of Cameron based on the "Cameron weight" showed the highest accuracy (83.99%), and Selangor based on the "Penang Island weight" the lowest (70.30%). Generally, however, the validation results showed a satisfying agreement between the susceptibility map and the landslide locations verified in the field.

In the present study only a susceptibility analysis based on the described nine parameters was performed. The spatial distribution of precipitations, especially of rainfall intensities, is very difficult to map and to model accurately. Intensive tropical monsoon rainfalls in the form of very local storms or torrential precipitation are the most frequent triggering factor. They seriously hamper to determine the rainfall distribution. The rain-gauge network in study areas is not dense enough to adequately record the precipitation. Therefore the classical relationship between the topographic parameters, landslides and total rainfall could not be assessed. If, however, data on landslide-causing parameters such as rainfall, earthquake shaking, or slope cutting exist, then a probability analysis including these values could also be made. Similarly, if factors relevant to the vulnerability of buildings and other property were available, a risk analysis could also be performed.

Landslides are among the most hazardous natural disasters in Malaysia. The Government and research institutions are trying to analyze the landslide hazard and risk and to show its spatial distribution over the regions. The use of multi temporal radar data such as TerraSAR for observing the landslides and residues in the research phase could be one of the prominent future directions. In the same line, there is a lot of work to be done to investigate the landslide causative parameters and their direct relationship between the triggering of future landslides.

Landslide susceptibility maps are of great help for planners and engineers to identify suitable locations for development. These results can be used as basic data to assist slope management and land-use planning.

Acknowledgements

The first author would like to thank the Alexander von Humboldt Foundation, Germany, for awarding a visiting scientist position at the Dresden University of Technology, Germany. The authors gratefully acknowledge two anonymous reviewers for their constructive comments which significantly improved the quality of the paper.

References

Akgun, A., Dag, S. & Bulut, F., 2008: Landslide susceptibility mapping for a landslide-prone area (Findikli, NE of Turkey) by likelihood-frequency ratio and weighted linear combination models. – Environmental Geology **54** (6): 1127–1143.

Clerici, A., Perego, S., Tellini, C. & Vescovi, P., 2006: A GIS-based automated procedure for landslide susceptibility mapping by the Conditional Analysis method: the Baganza valley case study (Italian Northern Apennines). – Environmental Geology **50:** 941–961.

Dai, F.C., Lee, C.F., Li, J. & Xu, Z.W., 2001: Assessment of landslide susceptibility on the natural terrain of Lantau Island, Hong Kong. – Environmental Geology **40:** 381–391.

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

Guzzetti, F., Carrarra, 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.

Lamelas, M.T., Marinoni, O., Hoppe, A. & Riva, J., 2008: Doline probability map using logistic regression and GIS technology in the central Ebro Basin (Spain). – Environmental Geology **54:** 963–977.

Lee, S., Ryu, J.H., Lee, M.J. & Won, J.S., 2003: Landslide susceptibility analysis using artificial neural network at Boeun, Korea. Environmental Geology **44:** 820–833.

Lee, S., Choi, J. & Min, K., 2004a: Probabilistic Landslide Hazard Mapping using GIS and Remote Sensing Data at Boeun, Korea. – International Journal of Remote Sensing **25:** 2037–2052.

Lee, S., Ryu, J.H., Won, J.S. & Park, H.J., 2004b: Determination and application of the weights for landslide susceptibility mapping using an artificial neural network. – Engineering Geology **71:** 289–302.

Lee, S., 2005: Application of logistic regression model and its validation for landslide susceptibility mapping using GIS and remote sensing data. – International Journal of Remote Sensing **26:** 1477–1491.

Lee, S. & Lee, M.J., 2006: Detecting landslide location using KOMPSAT 1 and its application to landslide-susceptibility mapping at the Gangneung area, Korea. – Advances in Space Research **38:** 2261–2271.

Lee, S. & Pradhan, B., 2006: Probabilistic Landslide Risk Mapping at Penang Island, Malaysia. – Journal of Earth System Science **115** (6): 661–672.

Lee, S. & Pradhan, B., 2007: Landslide hazard mapping at Selangor, Malaysia using frequency ratio and logistic regression models. – Landslides **4:** 33–41.

Pistocchi, A., Luzi, L. & Napolitano, P., 2002: The use of predictive modeling techniques for optimal exploitation of spatial databases: a case study in landslide hazard mapping with expert system-like methods. – Environmental Geology **41:** 765–775.

Pradhan, B., 2010: 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; doi: 10.1016/j.asr.2010.01.006 (on-line first).

Pradhan, B., Lee, S. & Buchroithner, M.F., 2009: Use of geospatial data for the development of fuzzy algebraic operators to landslide hazard mapping: a case study in Malaysia. – Applied Geomatics **1:** 3–15.

Pradhan, B., Lee, S., Mansor, S., Buchroithner, M.F. & Jallaluddin, N., 2008: Utilization of optical remote sensing data and geographic information system tools for regional landslide hazard analysis by using binomial logistic regression model. – Journal of Applied Remote Sensing **2:** 1–11.

Pradhan, B. & Lee, S., 2008a: 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 Frontiers **14** (6): 143–152.

Pradhan, B. & Lee, S., 2009a: Landslide risk analysis using artificial neural network model focusing on different training sites. – International Journal of Physical Sciences **3** (11): 1–9.

Pradhan, B. & Lee, S., 2009b: Delineation of landslide hazard areas on Penang Island, Malaysia, by using frequency ratio, logistic regression, and artificial neural network models. – Environmental Earth Sciences; doi: 10.1007/S12665-009-0245-8 (on-line first).

Pradhan, B. & Lee, S., 2009c: Landslide susceptibility assessment and factor effect analysis: backpropagation artificial neural networks and their comparision with fregquency ratio and bivariate logistic regression modelling. – Environmental Modelling & Software; doi: 10.1016/j.envsoft.2009.10.016 (on-line first).

Pradhan, B. & Lee, S., 2009d: Regional landslide susceptibility analysis using backpropagation neural network model at Cameron Highland, Malaysia. – Landslides; doi: 10.1007/S10346-009-0183-2 (on-line first).

Pradhan, B., Lee, S. & Buchroithner, M., 2010: A GIS-based back-propagation neural network model and its cross-application and validation for landslide susceptibility analyses. – Computers, Environment and Urban Systems; doi: 10.1016/j.compenvurbsys.2009.12.004 (on-line first).

Pradhan, B., Singh, R.P. & Buchroithner, M.F., 2006: Estimation of Stress and Its Use in Evaluation of Landslide Prone Regions Using Remote Sensing Data. – Advances in Space Research **37:** 698–709.

Shou, K.J. & Wang, C.F., 2003: Analysis of the Chiufengershan landslide triggered by the 1999 Chi-Chi earthquake in Taiwan. – Engineering Geology **68:** 237–250.

Suzen, M.L. & Doyuran, V., 2004: A comparison of the GIS based landslide susceptibility assessment methods: multivariate versus bivariate. – Environmental Geology **45:** 665–679.

Tangestani, M.H., 2004: Landslide susceptibility mapping using the fuzzy gamma approach in a GIS, Kakan catchment area, southwest Iran. – Australian Journal of Earth Sciences **51:** 439–450.

Temesgen, B., Mohammed, M.U. & Korme, T., 2001: Natural Hazard Assessment Using GIS and Remote Sensing Methods, with Particular Reference to the Landslides in the Wondogenet Area, Ethiopia. – Phys. Chem. Earth (C) **26:** 665–615.

Tunusluoglu, M.C., Gokceoglu, C., Nefeslioglu, H.A. & Sonmez, H., 2008: Extraction of potential debris source areas by logistic regression technique: a case study from Barla, Besparmak and Kapi mountains (NW Taurids, Turkey). – Environmental Geology **54** (1): 9–22.

Wang, H.B. & Sassa, K., 2005: Comparative evaluation of landslide susceptibility in Minamata area, Japan. – Environmental Geology **47:** 956–966.

Address of the Authors:
Dr. Biswajeet Pradhan, Saro Lee & Prof. Dr. Manfred Buchroithner, Technische Universität Dresden, Institut für Kartographie, D-01062 Dresden, Tel.: +49-351-463-33099, Fax: +49-351-463-37028, e-mail: Biswajeet.Pradhan@mailbox.tu-dresden.de; biswajeet@mailcity.com

Manuskript eingereicht: April 2009
Angenommen: November 2009