

METHODOLOGY

Open Access



Performance of frequency ratio and logistic regression model in creating GIS based landslides susceptibility map at Lompobattang Mountain, Indonesia

Abdul Rachman Rasyid^{1,2*} , Netra P. Bhandary¹ and Ryuichi Yatabe¹

Abstract

The purposes of this study is to create a landslide susceptibility map (LSM) for Lompobattang Mountain area in Indonesia. The foot of the Lompobattang Mountain area suffered flash flood and landslides in 2006, which led to significant adverse impact on the nearby settlements. There were 158 identified landslides covering a total area of 3.44 km². Landslide inventory data were collected using google earth image interpretations. The landslide inventories were prepared out of the past landslide events, and future landslide occurrence was predicted by correlating landslide causal factors. In this study landslide inventories were divided into landslide data for training and landslide data for validation. The LSM was prepared by Frequency Ratio (FR) and Logistic Regression (LR) statistical methods. Lithology, distance from the road, distance from the river, distance from the fault, land use, curvature, aspect, and slope degree were used as conditioning parameters. Area under the curve (AUC) of the Receiver Operating Characteristic (ROC) was used to check the performance of the models. In the analysis, the FR model results in 85.8 % accuracy in the AUC success rate while the LR model was found to have 86.9 % accuracy. However, the accuracy of both these models in AUC predictive rate is the same at around 85.1 %. The LR model is 6.34 % higher than the FR model in comparison to its accuracy for ratio of landslide validation. The landslide susceptibility map consist of the predicted landslide area, hence it can be used to reduce the potential hazard associated with the landslides in this study area.

Keyword: GIS, Landslide susceptibility, Frequency ratio, Logistic regression, Validation, Indonesia

Background

Earthquakes, intense rainfall, and snowmelt are general triggering factors of landslides. Other factors can be geology, land cover, slope geometry, solar radiation, surface and subsurface hydrology, and human activities. In Indonesia, landslides are serious problem that cause debris flow or flash flood disasters every year during or after heavy rainfalls. During 2005 to 2014, around 1926 landslide events were reported which resulted in loss of 1035 human casualties and 853 disappearance, and in

the last one decade the trend has increased (Badan Nasional Penanggulangan Bencana Indonesia 2015). The government and research institutes have been attempting to minimize the loss through appropriate landuse planning and information dissemination about landslide susceptibility.

Landslide susceptibility, hazard and risk zoning are parts of landuse planning. As first stage of landslide hazard mitigation, landslide susceptibility mapping must provide important information to support decisions for urban development, which considerably reduces potential landslide damage. In other words, landslide susceptibility maps are produced to help humans to recognize and adapt to landslide hazard mitigation procedures (Pourghasemi et al. 2012).

* Correspondence: rachman_rasyid@yahoo.com

¹Graduate School of Science and Engineering, Ehime University, 3 Bunkyo, Matsuyama 790-8577, Japan

²Department of Architecture Engineering Faculty, Hasanuddin University, Makassar 90245, Indonesia

A number of researchers have put their efforts to increase the accuracy of landslide susceptibility mapping up until today. A variety of methods have been applied to include qualitative and quantitative modeling. Westen et al. (1997) classified the general techniques of analyzing landslide zoning using GIS techniques into heuristic, statistical and deterministic approaches. More recently, some researchers have created landslide susceptibility maps using statistical models, and some of them combine those models with other approaches such as frequency ratio (FR) and logistic regression (LR) methods (e.g., by Lee and Pradhan 2007, Oh et al. 2008, and Solaimani et al. 2013). FR was combined with analytical heuristic approach (AHP) by Demir et al. (2013) and Reis et al. (2012), and combination using FR, AHP, LR and artificial neural network (ANN) model was proposed by Park et al. (2013). Integrated techniques such as FR, weight of evidence (WoE) and deterministic methods have been applied by Cervi et al. (2010) and Yilmaz and Keskin (2009). Association models like WoE, AHP and fuzzy logic to combine multiple factor layers to create landslide susceptibility map was introduced by Suh et al. (2011).

Statistical techniques involve large amounts of data to obtain reliable results (Yilmaz 2009), and they are usually suitable for wide area studies. Statistical methods use sample data based on the relationship between landslides and causal factors. The combination of both data is evaluated in an objective way. In this study we apply two statistical methods, namely FR and LR models. The FR model consists of simple procedure and is modest, while the LR model needs complex procedure for preparing data using a statistical software and only limited data in processing need to be considered (Park et al. 2013 and Demir et al. 2015).

The main objective of this study is to create a landslide susceptibility map of Lompobattang Mountain. The susceptibility map was prepared by summing the weight parameter values from frequency ratio model and an equation established by using logistic regression model. Validation of the results is emphasized in this study in order to reduce any uncertainty that may occur during prediction and to increase the accuracy of the model. To achieve this, the landslide inventory data were divided into training data (data used to obtain weight of parameters in FR analysis used in the equation obtained from LR model) and the validating data which were used to examine the level of precision. The ROC curve and AUC were used to validate the model.

Verification is applied to get the best appropriate coefficient of landslide causal factors in the LR model. To do this, the variable of equation is established by means of using equal number of landslide and non-landslide pixels. For comparison, the analysis was also carried out

by using landslide merged with 50 and 100 % of non-landslide pixels. Next, the ratio was obtained by overlaying landslide data for validation into the landslide susceptibility map.

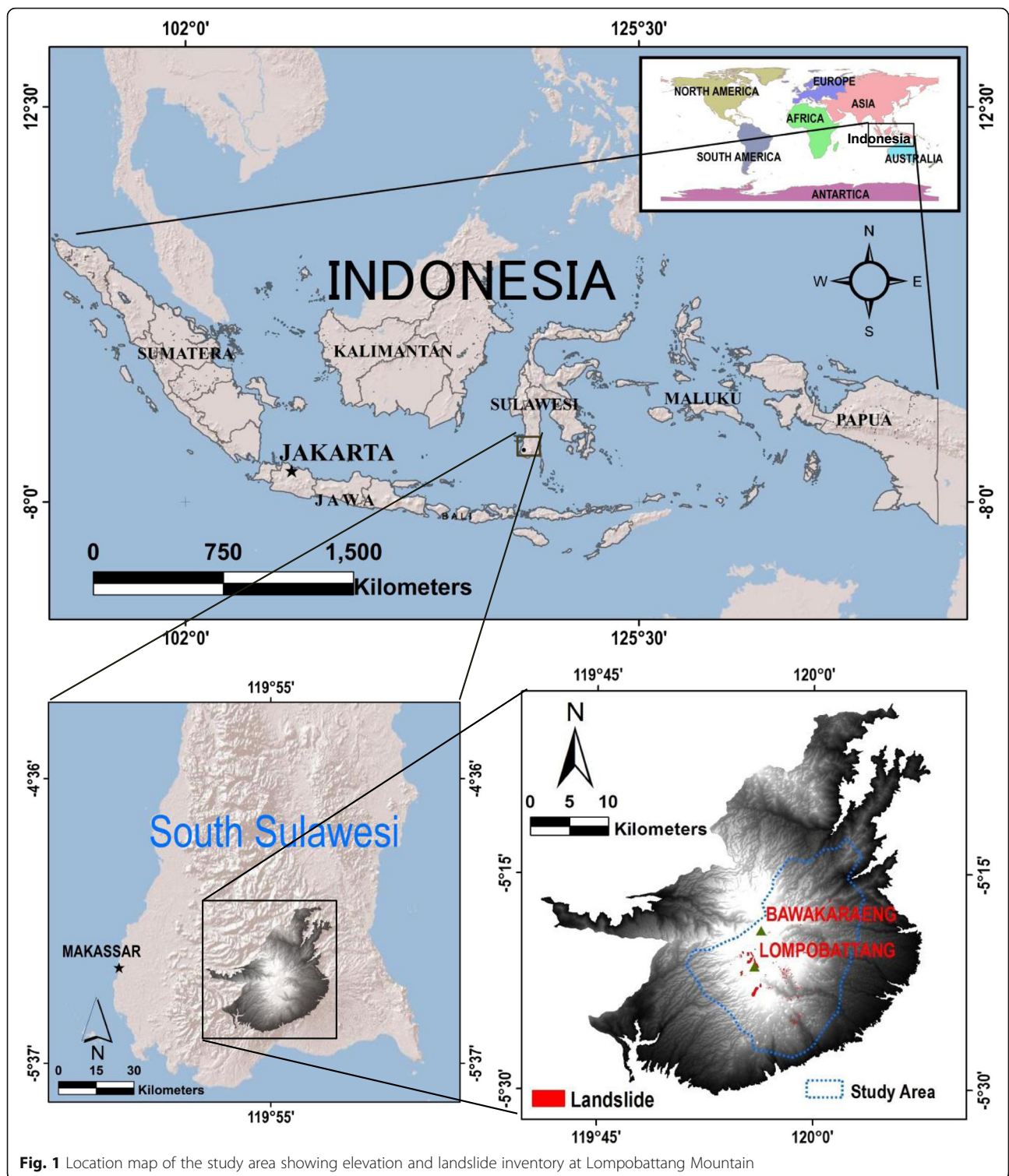
The spatial database of landslides and landslide causal factors to be used in the susceptibility analysis was prepared in the GIS environment, which has been used as a major tool of spatial analysis in landslide studies. Satisfactory results have been obtained in landslide susceptibility analysis (Shirzadi et al. 2012) and effective modeling in slope instability analysis (Dai and Lee 2002).

Study area

Bawakaraeng and Lompobattang mountains are located in Southern South Sulawesi Province and are surrounded by the districts that have high economic growth rate. Both of these mountains have important role in supporting that growth. This area provides a fertile land but frequently suffers from landslide disasters. Landslide disasters occur almost every year, especially during the rainy season, which induce flash floods and debris flows in the upstream. On March, 26 2004, a huge landslide occurred at Mt. Bawakaraeng with a volume of about 200 million m³, a width of about 1600 m and a length of about 750 m. The earth materials and debris from the landslide covered the valley along the river, causing destruction of environment and river ecosystem. Geomorphologically, such topographic features and rise of groundwater level are the main cause of the landslide (Tsuchiya et al. 2009). On June 20, 2006 heavy rainfall triggers landslides and flash floods at Mt. Lompobattang. Settlements at Sinjai, Bulukumba, Bantaeng, Jenepono and Bone regions on the foot of Lompobattang Mountain were heavily impacted. Nearly 214 fatalities, 45 missing, and around 6400 displaced were reported (Direktorat Cipta Karya Kementerian PUPERA 2006).

Lompobattang Mountain is located at 119°50'–120°04' E and 5°12'–5°28' S with altitude about 2876 m above sea level and has a total area coverage of 351.742 km² (Fig. 1). There are about 93 settlements in this area with six hydrologic watershed system; Jeneberang, Lantebong, Kelara, Apparang, Bijawang and Tangka. Based on geological maps (Sukanto and Supriatna 1982), the volcanic rocks of Lompobattang Mountain consist of agglomerates, lava, breccia, and tufa deposition, which form a broad strato-volcano and quarter lompobattang volcanic (qlv) were estimated from volcanic rock Pleistocene.

The climate of Sulawesi Island is tropical with special characteristics of two seasons within a year. The northeast monsoon gives rise to rainy season between November and May (December to January has maximum rainfall) and the southwest monsoon causes the dry season from June to October. The annual rainfall data recorded at Malino station from year 2011 to 2014 was 3643 to 5474 mm. The



average annual rainfall is 4424 mm for over 25 years (1978 to 2003). The monthly rainfall is more than 700 mm in the month of February and rises up to 900 mm in January (Tsuchiya et al. 2009). Due to

increase in rainfall intensity, the probability of landslide occurrence, particularly shallow landslides increases and is very sensitive to short-lasting high intensive rainfall (Hasnawir and Kubota 2012).

Data preparation

To create landslide susceptibility map, selection of appropriate data to be used is important, which helps to yield successful results. To create spatial database of landslide inventories and landslide causal factors in the predicted area, management and selection of data should be accurate. For the analysis of FR values Microsoft Excel was used, whereas Statistical Package for the Social Sciences (SPSS) was used to establish LR model.

Landslide inventories

Landslide inventories can be developed from field surveys, by interpretation of remotely sensed images such as based on the spectral characteristics, shape, contrast and the morphological expression (Kanungo et al. 2006), or aerial photographs (Ayalew and Yamagishi 2005) and google image interpretation (Xu et al. 2013). Landslides from 2004 to 2014 were collected from google earth

image interpretations of Lompobattang Mountain. From this, a total of 158 landslides were identified, which cover an area of 3.44 km². Most of the landslides are of shallow type with minimum and maximum landslide area of 708 m² and 512,765 m² (0.51 km²) respectively. The study area was limited to an altitude of 500 m, as no landslide data were found below this altitude (Fig. 2). Using the landslide data from Google Earth to GIS environment, we have to digitize the time series data from google earth image interpretation. Then, these files were saved as GIS compatible (kml) format and the data was again subsequently changed into shapefile and then into raster format.

Landslide causal factors

In susceptibility mapping, it is important to assume that future landslides will occur in the same condition that caused the past landslides. There are no strict guidelines

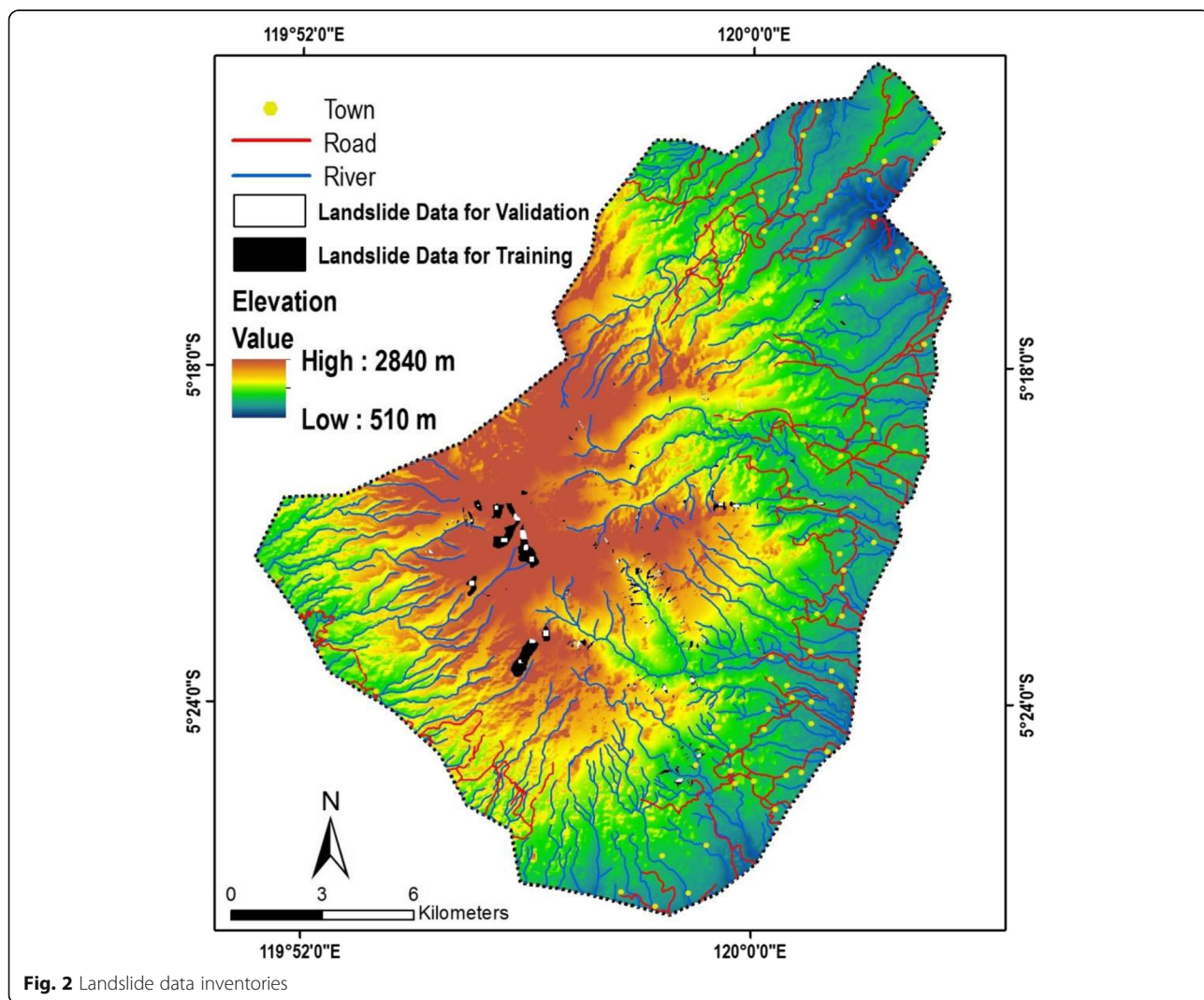


Fig. 2 Landslide data inventories

for the selection of causal factors to be used in logistic regression analysis, and as such, the covariates selected vary widely between studies (Ayalew and Yamagishi 2005). In addition, the determination of landslide causal factors were also associated with the availability of data. The entire landslide causal factors that we have used in this paper also fall in this category. Landslide data were used as dependent variable of eight causal factors including slope, curvature, aspect, distance from fault, distance from road, distance from river, lithology and landuse pattern which were selected as independent variables for the landslide susceptibility mapping (Fig. 3). All of these data are commonly used in landslide susceptibility mapping. Budimir et al. (2015) mentions that in a total of 37 parameters usually used slope, aspect, and lithology, are significantly used particularly on studies related to rainfall-induced landslides. In fact, the relevance of the spatial data combination used in the prediction became an important issue in landslide susceptibility mapping (Dewitte et al. 2010).

The geology of the area was digitized from the Geology Map of Geological Research Institute, produced by the government board at a scale of 1:250.000 (Sukanto and Supriatna 1982). This map includes the current study area. The geology includes lithology, rock type and structure (fault or lineament). Lithology is a part of basic data or parameters for landslide map analysis. In fact, Ermini et al. (2005) mentioned that lithology is a classic variable that controls landslide hazard. It is related to the material strength, because they have varied composition and structure for different type of rocks (Kanungo et al. 2006), and the resistance to driving forces depend on the rock strength, in which the strongest rocks would be more resistance. Lineaments are the structural features, which describe the zone/plane of weakness, fractures, and faults along which landslide susceptibility is higher. It has generally been observed that the probability of landslide occurrence increases at sites close to lineaments, which not only affect the surface material structures but also make contribution to terrain permeability causing slope instability. For this purpose, distance from fault was used to analyze the relationship between landslide occurrences. The proximity distance from fault was identified by buffering from lineament or fault map.

The topographic data used in the analysis include slope, aspect and curvature. These data were derived from ASTER DEM with a spatial resolution of 30 m. Using arc-toolbox raster surface in ArcGIS, the slope angle, slope aspect and curvature were derived. On a slope of uniform isotropic material, increased slope correlates with increased likelihood of failure. In this study, we have used seven slope categories, 0–5°, 5–10°, 10–20°, 20–30°, 40–50°, and above 50°, which were considered and represented in the form of slope thematic data layer. Likewise the aspect map plays a

significant role in slope stability assessment (Chauhan et al. 2010). In this study, aspect is divided into nine classes namely, flat, N, NE, E, SE, S, SW, W, and NW. To describe the variances among classes, aspect maps displayed the distribution of each direction in the topography by using different colors to each cell of the study area (Quan and Lee 2012). Profile curvature was reclassified into three classes namely concave, flat and convex. The curvature values represent the morphology of the topography. In case of profile curvature, generally related to the puddle condition after heavy rainfall. Profile curvature slope contains more water and retains water from heavy rainfall for a longer period (Lee and Thalib 2005).

Besides topographic factors and geology, landuse (cover) is a key factor responsible for landslide occurrences. The incidence of landslide is inversely related to the vegetation density. The landuse map was derived from Landsat 7 with 30 m × 30 m pixel, and its was established by BPDAS Jeneberang Walanae in 2014 a board for watershed issued at Ministry of Forestry in Indonesia (Balai Pengelolaan Daerah Aliran Sungai Jeneberang Walanae 2014). The landuse maps are usually classified into several classes, but in this study, forest (including primary and secondary), bushes, crop land (agriculture), and grass land were considered. Drainage lines and landslide occurrence in hilly area have strong association between them due to erosional activity. The distance from river was calculated by buffering and analyses of river lines that were derived from topographic map of scale 1:50.000 called Peta Rupa Bumi Indonesia (RBI) prepared by the government. The class starts from 0 to 50 m and ends with > 300 m. Similarly, distance from river and distance from road were also derived from topographic map.

Independent variables and dependent variables are used as input maps and then processed by converting them into raster maps of 30 m × 30 m pixel size. The study area includes 390,837 pixels and the landslide data used in the model include 3827 pixels.

Methods

Frequency ratio

The relationship between the landslide occurrence area and the landslide causal factors could be deduced from the relationship between areas where landslides had not occurred and the landslide causal factors. In order to identify the closeness of their relationship, a simple statistical technique has been applied to derive it with the frequency ratio approach. Furthermore, FR model became valuable in ranking the preferred causative factors based upon their ability to control a landslide incident (Kannan et al. 2013), because FR can describe clearly the difference of each score between landslide causal factors in class and landslide occurrence. Thus, the number of landslide occurrence pixels on the area must be combined between

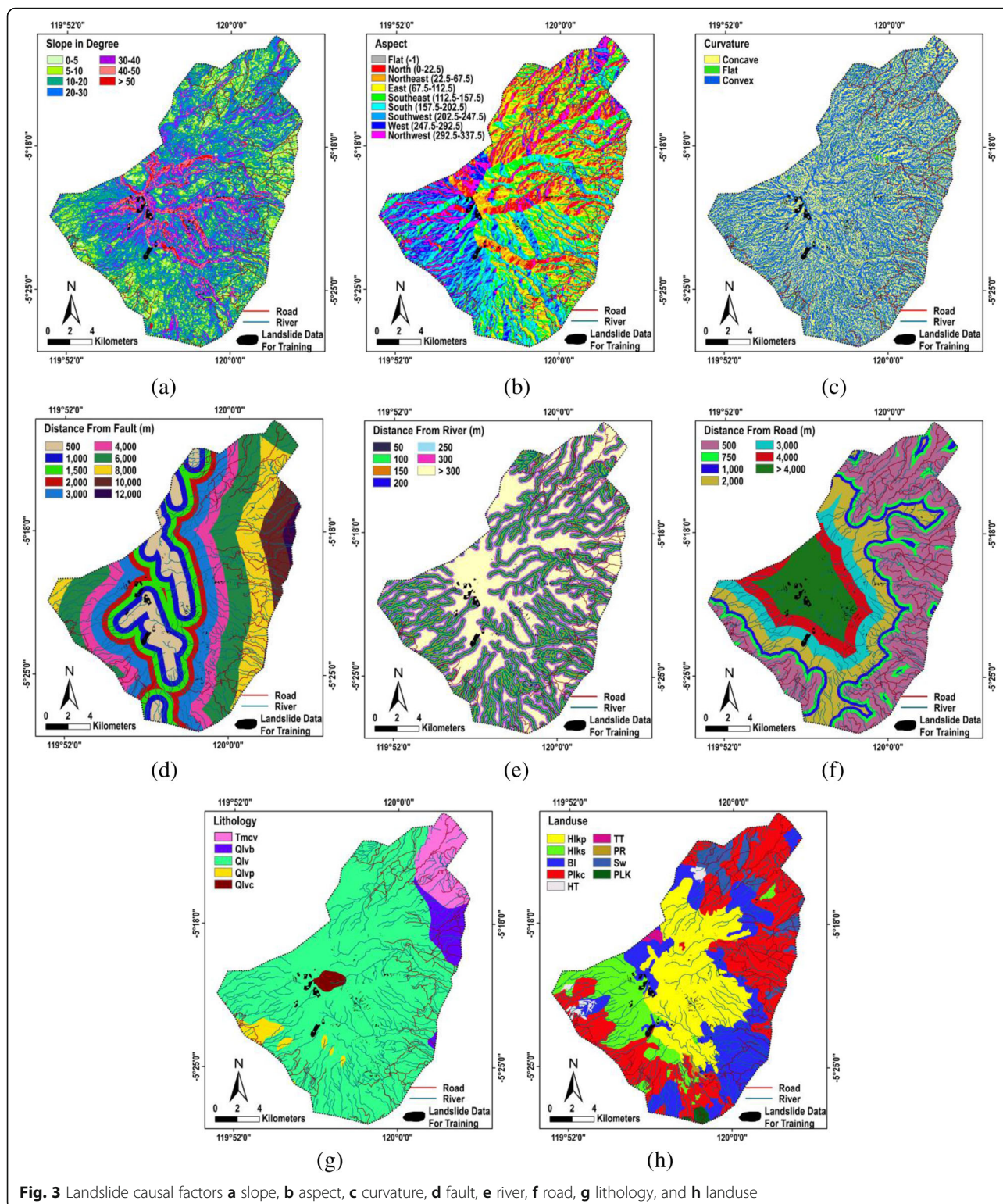


Fig. 3 Landslide causal factors **a** slope, **b** aspect, **c** curvature, **d** fault, **e** river, **f** road, **g** lithology, and **h** landuse

causal factors. Then the ratio for each factors were calculated by dividing the landslide occurrence ratio with the ratio of each class in causal factors (Lee and Thalib 2005). A ratio value in each class shows the level of relationship

the given factors attribute between landslide occurrences and when the ratio more than one means a stronger correlation then a lower ratio than one suggest a lower correlation (Lee and Pradhan 2006).

The frequency ratio value can be calculated in the following manner;

$$FR = \frac{(ij) / \sum PixL}{P(ij) / \sum Pix} \tag{1}$$

(Where, $PixL(ij)$ number of pixel with landslide within class i of j parameter, $Pixl(ij)$ Number of pixel in class i of j parameter, $\sum PixL$ total pixel of j parameter, and $\sum Pix$ total pixel of the area).

Logistic regression model

The landslide susceptibility index was obtained by logistic regression model. A simple introduction of logistic regression is available in Chau and Chan (2005) who define it as the probability of landslide occurrence divided by the probability of no landslide occurrence. It is useful for predicting the presence or absence of a characteristic or outcome based on values of a set of predictor variables. Generally, in logistic regression, the spatial prediction is modeled by a dependent variable and independent variables (Shirzadi et al. 2012) and it is useful when the dependent variable is binary or dichotomous. Furthermore, Lee (2005) has stated that advantage of logistic regression model 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. The probabilities of the regression can be understood as the probability of one state of the dependent variable as they are constrained to fall in the range of values from 0 to 1 (Xu et al. 2013) with zero indicating a 0 % probability of landslide occurrences and one indicating a 100 % probability (Dai et al. 2004).

The logistic regression is based on logistic function expressed as follows,

$$P = 1 / (1 + \exp^{-z}) \tag{2}$$

Where P : is the probability of landslide occurrence that estimated values varies from 0 to 1. Variable Z is landslide causal factors and assumed as a linear combination of the causal factors X_i ($i = 1, 2, \dots, n$) as

$$Z = B_0 + B_1X_1 + B_2X_2 + \dots + B_nX_n \tag{3}$$

Where B_i are the coefficient of landslide causal factors.

Validation and verification

In addition to decrease inaccuracy of prediction and probability, validation could raise the reliability. During prediction modeling, the most important and the absolute essential component is to carry out a validation of the predicted results (Chung and Fabbri 2003). In this study, the landslide inventories were divided into two

parts; one for training and the other for validation. This study uses 3117 (81 %) pixels of landslide inventories for generating the model and 710 (19 %) pixels for validation. The main assumption in selecting of landslide data for training and for validation is randomly on any part of landslide occurrence of the study area and also based on representation of the landslide area. To illustrate the procedure, a small part of the landslide prone area was chosen as data for validation. The size, area, depth of landslide and its distribution significantly varies from place to place.

Moreover, we used ROC curve to plot the predicted probability to comprehend issues of accuracy, criterion selection, and interpretation. In order to validate the landslide susceptibility map, AUC curve was used as a measure of overall fit and comparison of modeled prediction. The success rate was determined from the AUC of training data set, and the prediction rate was calculated from the AUC of the validation dataset. The ROC curves are significant for evaluating the predictive accuracy of a chosen model particularly in dichotomous statistical modeling such as logistic regression (Gorsevski et al. 2006), and the area under the curve obtained from the ROC (receiver operating characteristics) plot is the most preferred and applicable type of statistical assessment (Akgun et al. 2012). The predicted probabilities generated by the logistic model can be viewed as a continuous indicator to be compared with observed binary response variable.

In this study, next validation process of showing the level of accuracy of the LSM is by calculating the ratio of landslide data for validation that falls into each susceptibility class (Fig. 4). The general assumption is that most of the landslides for validation should fall on high to higher susceptibility class.

Results and discussions

The application of frequency ratio

The frequency ratio method was used to find the correlation between landslide locations in the past and each factor that affects landslides. In general, factor classes with a frequency ratio value of >1 will have higher probability landslide occurrence. The number of pixels of each class of causal factors were automatically counted by using the reclassify tool in ArcGIS software and the number pixels of landslide occurrence in each class of causal factors was found on overlaying them. By using the Eq. 1, the ratio of each class was calculated by dividing the number of pixels in each factor's class by the total number of pixels in the entire study area. Then the frequency ratio values of each factor classes were computed by dividing the landslide percentage by the area of percentage as in Table 1.

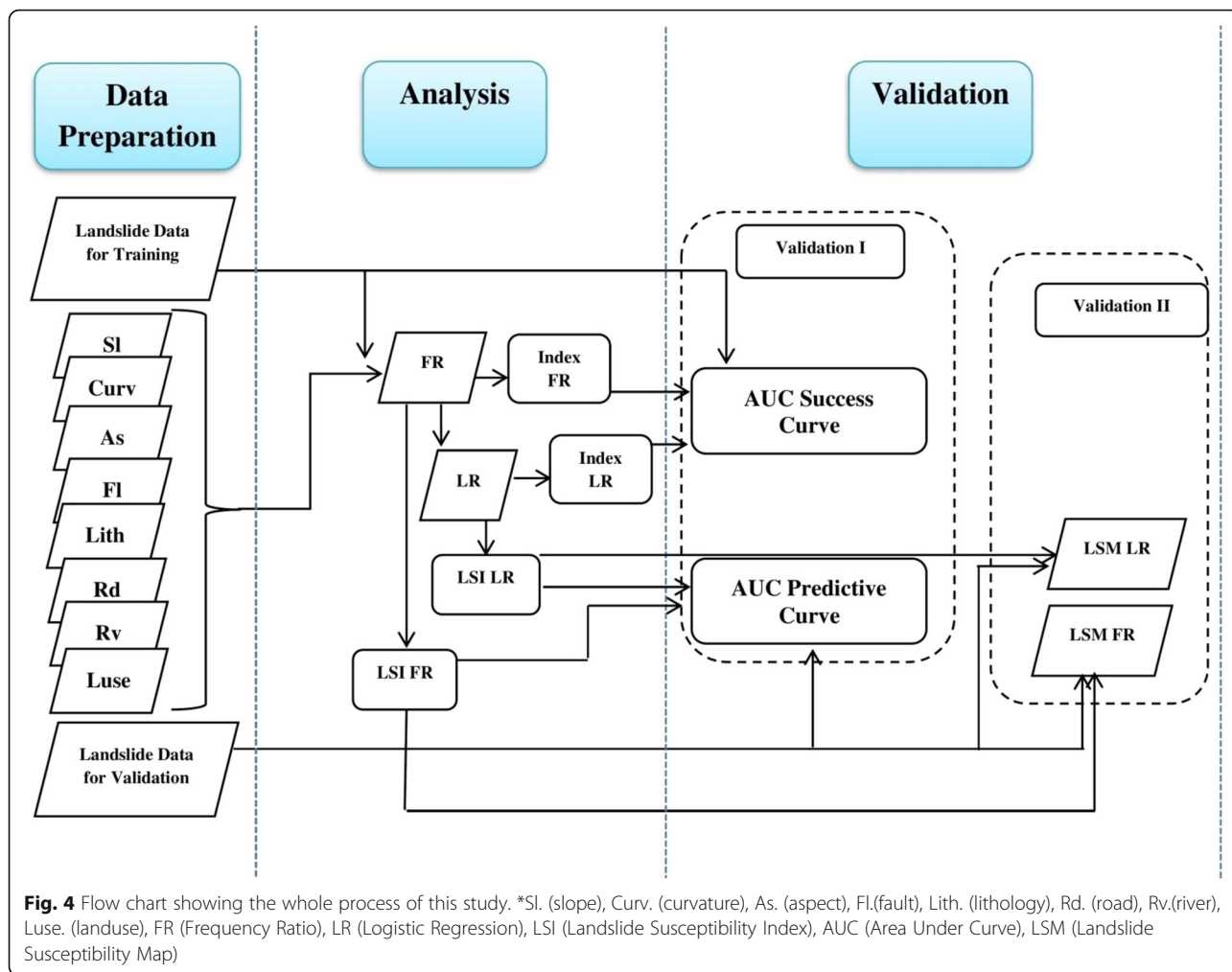


Figure 5 shows correlation between landslide occurrence and each class of presence and absence landslides inventories between each class of landslide causal factors. In the case of the relationship between landslide occurrence and slope angle, slope below 30° has a ratio of <1, which indicates a very low probability of landslide occurrence. For slope above 30°, the ratio is >1, which indicates a high probability of landslide occurrence.

In curvature class, the values represent the morphology of topography. A convex indicates a positive value, a concave indicates negative, and zero value indicates flat surface. Comparing frequency ratio values of both concave and convex, it is understood that the probability of landslide occurrence is almost similar, with slightly higher probability of landslide occurrence in case of concave curvature. This might be due to the accumulation of water in these classes. However, in the case of flat surfaces, the probability of landslide occurrence is very low. In the case of aspect class, the south, southwest and

west facing slopes, frequency ratio is >1, which indicates a high probability of landslide occurrence.

In the case of lithology classes, only Qlv has a ratio of >1 among the five lithology classes, which indicates high probability of landslide occurrence. Quarter lompobattang volcanic (Qlv) is one of the volcanic and sediment formation in South Sulawesi area. In case of distance from fault, river and road, ratio to distance/proximity is used to understand the level of influence on landslide occurrence. Distance from fault below 1000 m has a ratio of >1. This shows that as distance from the fault decrease, the probability of landslide occurrence increases. In case of distance from road, the frequency ratio value is higher at a distance class of > 3000 m. Similarly, for the distance from river above 300 m has ratio of >1. In case of distance from rivers and distance from roads, the landslide densities are higher for distance classes far away. Forests and bushes in landuse classes have a frequency ratio value of >1. Nevertheless in the case of agriculture and grass land the ratio is <1.

Table 1 Frequency ratio value for each landslide causal factors

Factor	Number of landslide	% landslide (x)	Number of pixel in class	% class (y)	Fr (x/y)
Topography					
Slope Class in Degree					
0–5	64	2.05	27,459	7.03	0.29
5–10	136	4.36	67,750	17.33	0.25
10–20	472	15.14	137,617	35.21	0.43
20–30	609	19.54	86,336	22.09	0.88
30–40	840	26.95	45,954	11.76	2.29
40–50	734	23.55	19,573	5.01	4.70
> 50	262	8.41	6141	1.57	5.35
Curvature Class					
Concave	1616	51.84	192,998	49.38	1.05
Flat	15	0.48	5424	1.39	0.35
Convex	1486	47.67	192,408	49.23	0.97
Aspect Class					
Flat (–1)	5	0.16	1933	0.49	0.32
North (0–22.5)	159	5.10	22,786	5.83	0.87
North East (22.5–67.5)	353	11.32	53,942	13.80	0.82
East (67.5–112.5)	207	6.64	57,312	14.66	0.45
South East (112.5–157.5)	242	7.76	65,237	16.69	0.47
South (157.5–202.5)	540	17.32	56,441	14.44	1.20
South West (202.5–247.5)	834	26.76	46,535	11.91	2.25
West (247.5–292.5)	372	11.93	30,493	7.80	1.53
North West (292.5–337.5)	231	7.41	35,952	9.20	0.81
North (337.5–360)	174	5.58	20,199	5.17	1.08
Geology					
Lithology Class					
Tm cv (Volcanics of camba formation)	5	0.16	28,534	7.30	0.02
Ql vb (Quarter lombpattang volcanics breccia)	13	0.42	17,403	4.45	0.09
Ql v (Quarter lombpattang volcanics)	3080	98.81	332,091	84.97	1.16
Ql vp (Quarter lombpattang volcanics parasitic)	-	-	8643	2.21	-
Ql vc (Quarter lombpattang volcanics center)	19	0.61	4159	1.06	0.57
Distance From fault (m)					
0–500	886	28.42	34,527	8.83	3.22
500–1000	952	30.54	35,187	9.00	3.39
1000–1500	296	9.50	35,483	9.08	1.05
1500–2000	177	5.68	27,949	7.15	0.79
200–3000	411	13.19	50,855	13.01	1.01
3000–4000	263	8.44	47,642	12.19	0.69
4000–6000	63	2.02	81,203	20.78	0.10
6000–8000	69	2.21	53,027	13.57	0.16
8000–10,000	-	-	21,147	5.41	-
10,000–12,000	-	-	3810	0.97	-

Table 1 Frequency ratio value for each landslide causal factors (*Continued*)

Proximity					
Distance from Road (m)					
0–500	114	3.66	155,239	39.72	0.09
500–750	84	2.69	32,697	8.37	0.32
750–1000	70	2.25	24,331	6.23	0.36
1000–2000	463	14.85	63,341	16.21	0.92
2000–3000	488	15.66	36,745	9.40	1.67
3000–4000	331	10.62	31,195	7.98	1.33
> 4000	1567	50.27	47,282	12.10	4.16
Distance from River (m)					
0–50	213	6.83	58,994	15.09	0.45
50–100	220	7.06	56,989	14.58	0.48
100–150	227	7.28	51,007	13.05	0.56
150–200	285	9.14	44,044	11.27	0.81
200–250	293	9.40	37,016	9.47	0.99
250–300	291	9.34	30,154	7.72	1.21
> 300	1588	50.95	112,626	28.82	1.77
Landuse					
Landuse Class					
Primary Dry Forest	994	31.89	99,453	25.45	1.25
Secondary Dry Forest	757	24.29	46,591	11.92	2.04
Bushes	977	31.34	83,215	21.29	1.47
Mix Dryland Agriculture	382	12.26	133,251	34.09	0.36
Forest Plant	-	-	3962	1.01	-
Open Land	-	-	1601	0.41	-
Grass Land	7	0.22	1112	0.28	0.79
Paddy Field	-	-	1533	5.00	-
Dryland Agriculture	-	-	2112	0.54	-

To create landslide susceptibility index, all the ratio of raster map landslide causal factors were summed as follows as:

$$LSI = FR_1 + FR_2 + \dots + FR_n \quad (4)$$

Where FR_1 , FR_2 , FR_3 ... FR_n are the frequency ratio raster maps of landslide causal factors. Index value using frequency ratio fall in range 1.52 to 21.1. The higher value of LSI indicates a higher susceptibility to landslide and if LSI value lower indicates lower susceptibility to landslide (Lee and Pradhan 2007).

Logistic regression model

Frequency ratio values show correlation between landslides and each class of landslide causal factors in numerical format. The frequency ratio raster maps of landslide causal factors with landslide and non-landslide points was extracted using ArcGIS tool and saved into

dbf format. Then a logistic regression equation was obtained by using SPSS software (Meten et al. 2015b).

A complete set for logistic regression analysis must contain a set of independent variables (landslide causal factors) and dichotomous dependent variables (landslide inventories). Fixing the sample size to create an equation in logistic regression analysis can be done in two ways, i.e., using all pixel landslide causal factors in study area and using equal number of dependent and independent variables to reduce bias in the sampling process (Ramani et al. 2011). In this study, the logistic regression model is developed using equal proportion of landslide and non-landslide pixels in ten iterations and using 50 % and all non-landslide data as comparison.

The constant and coefficient of independent variables were provided by logistic regression analysis using SPSS. In case of using the number of equal proportion of non-landslide pixels, they were selected randomly by SPSS. Hence, this study proposes to investigate ten

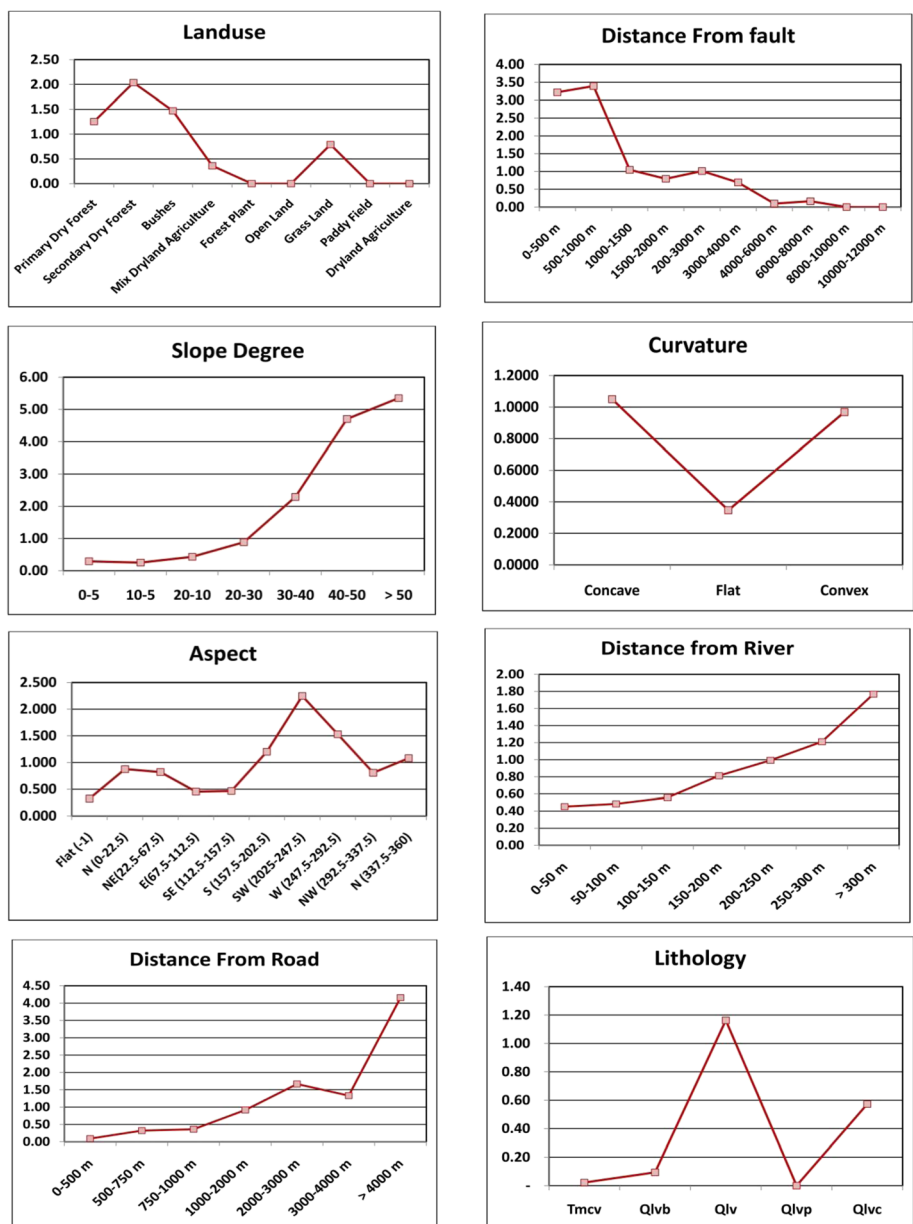


Fig. 5 Frequency ratio of the landslide causal factors

iteration to get best result and sense of fairness as shown in Table 2.

Using the logistic regression model, the landslide occurrence probability was computed, and if values are closer to one, landslides are more likely occur.

Validation

In landslide modeling, validation of predictive landslides is an important part of the procedures for landslide susceptibility mapping (Bui et al. 2012). The success rate and prediction rate can be obtained by comparing the landslide susceptibility results at known landslide locations.

In SPSS software, AUC of success rate was derived by linking the landslide index in FR model using landslide data for training. Subsequently, the AUC of predictive rate was obtained by using landslide data for validation.

There are two steps to get the value of the AUC curve as validation for fit of model using logistic regression in this study. An equal number of landslide and non-landslide data for training with landslide causal factors were combined as a merge variable in SPSS. Then binary logistic was chosen to establish the variables in equation and probability result. Subsequently, the AUC for success rate were obtained for each trial equation using

Table 2 Logistic regression coefficient of landslide causal factors using equal proportion of landslide and non-landslide pixels

No. test	Variables in the equation								
	Aspect	Curvature	Fault	Lithology	Landuse	River	Road	Slope	Constant
1	0.485	1.200	0.544	1.528	0.051	0.200	0.311	0.428	-5.867
2	0.560	1.200	0.506	1.778	0.214	0.212	0.265	0.414	-6.300
3	0.554	0.871	0.501	2.021	0.100	0.198	0.259	0.474	-6.152
4	0.605	1.231	0.484	1.814	0.130	0.236	0.288	0.382	-6.312
5	0.572	0.798	0.492	1.692	0.110	0.189	0.293	0.451	-5.737
6	0.578	1.303	0.508	1.965	0.139	0.181	0.289	0.385	-6.506
7	0.573	1.008	0.518	1.922	0.111	0.174	0.305	0.447	-6.243
8	0.482	1.112	0.517	1.709	0.058	0.212	0.302	0.439	-5.973
9	0.547	1.091	0.520	1.666	0.046	0.151	0.300	0.417	-5.875
10	0.618	1.033	0.495	1.748	0.132	0.150	0.296	0.380	-6.012
11	0.761	0.558	0.441	2.147	0.443	0.481	0.245	0.281	-10.664
12	0.763	0.562	0.438	2.143	0.468	0.485	0.245	0.279	-11.389

landslide data for training. The next step was to extract each test into regression model (Eqs. 2 and 3), and then by using ArcGIS 10.0 Software, the Landslide Susceptibility Index (LSI) maps were produced.

The AUC predictive curve was counted based on LSI map as test variable and landslide data for validation as state variable. Table 3 shows results of AUC curve for both success rate and predictive rate for each test. As comparison, the eleventh and twelfth tests were conducted by using 50 % and all of the non-landslide pixels respectively and the variables in equation were produced in SPSS. The same procedures for equal number of landslide and non-landslide pixels were used to obtain AUC success and predictive rate.

The closeness of success rate and predictive rate values show how the logistic regression helps in landslide prediction in the future (Meten et al. 2015a). The AUC curve determined by using validation dataset should be approximately equal to the AUC curve determined by using the training dataset, but it is generally lower than the success curve, because the landslide data on validating area are not used for modelling (Ngadisih et al. 2013).

In general, the AUC of ROC curves representing excellent, good, and valueless tests were plotted on the graph. To classify the accuracy of a diagnostic test, the

value ranges from 0.50 to 0.60 (fail), 0.60–0.70 (poor), 0.70–0.80 (fair), 0.80–0.90 (good), and 0.90–1.00 (excellent). The results show that the entire test falls in good category because the value ranges from 0.858 to 0.869 in success rate and 0.839 to 0.855 in predictive rate.

This study conducts one more validation to choose the best statistical model for creating landslide susceptibility map and the best equation in logistic regression approach from the 12 tests. The sum of FR value and equation of the LR models were used to create landslide susceptibility map (LSM) by reclassifying LSI of the models using natural breaks method. Overlaid landslide data validation on LSM will describe another level of accuracy beside AUC curve.

The natural breaks method or Jenks optimization method has been used widely especially by planners and it is designed to determine the best arrangement of values into different classes. This method maximizes the variance between classes and reduces the variance within classes. The five classes include very low, low, moderate, high and very high describing the level of landslide susceptibility (proneness) in the study area. The level of accuracy of the landslide susceptibility map was verified by overlaying with the landslide data for validation.

Table 3 shows the results of overlaid landslide data for validation on LSM for LR model using equal number of

Table 3 AUC to ROC curve of success and predictive rate and ratio of landslide validation on landslide susceptibility map using FR and LR model

Model	FR	LR											
		1	2	3	4	5	6	7	8	9	10	11	12
Success rate	0.858	0.867	0.864	0.866	0.861	0.866	0.864	0.869	0.865	0.862	0.858	0.859	0.839
Predictive rate	0.851	0.850	0.851	0.855	0.849	0.853	0.848	0.851	0.852	0.850	0.848	0.858	0.839
H + VH	77.88	83.66	83.80	83.94	82.11	84.22	82.82	84.23	83.94	83.8	82.11	32.39	30.28

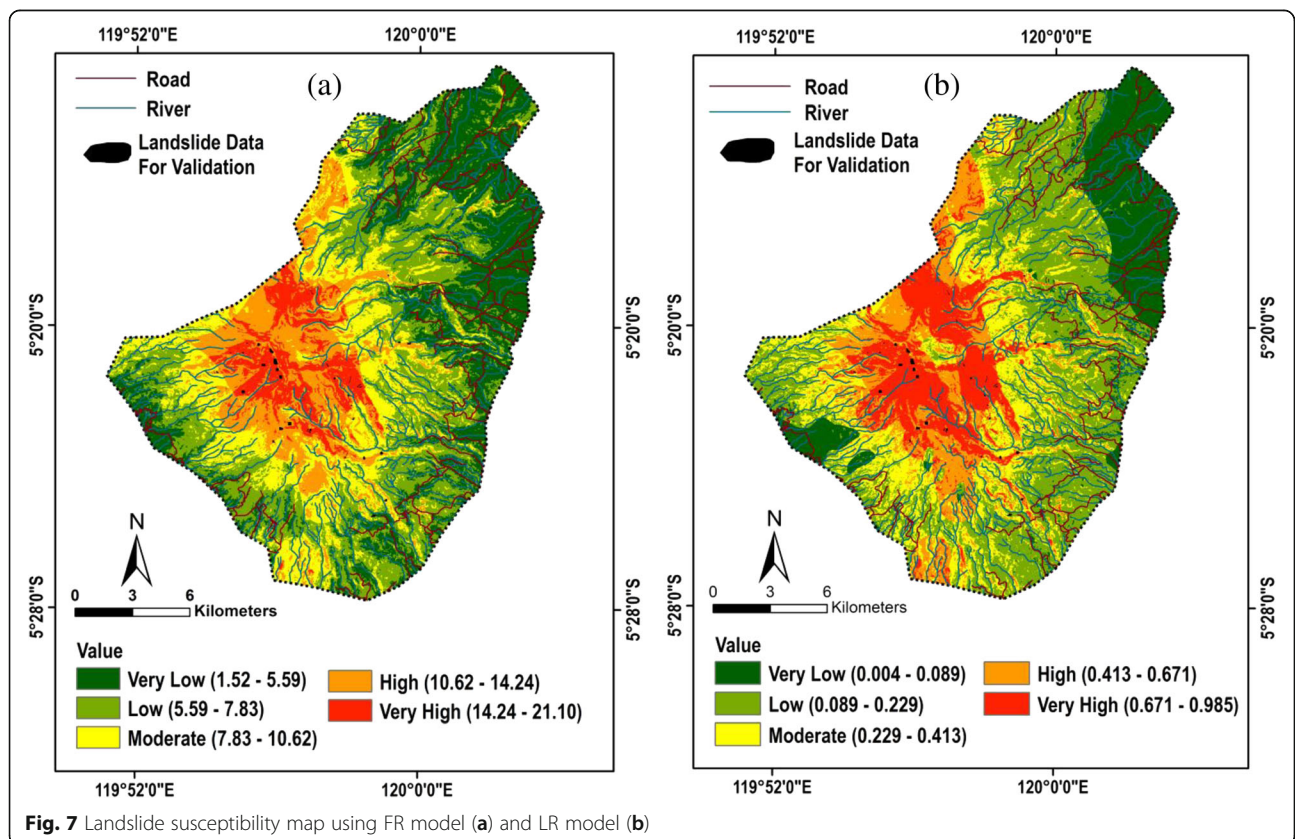
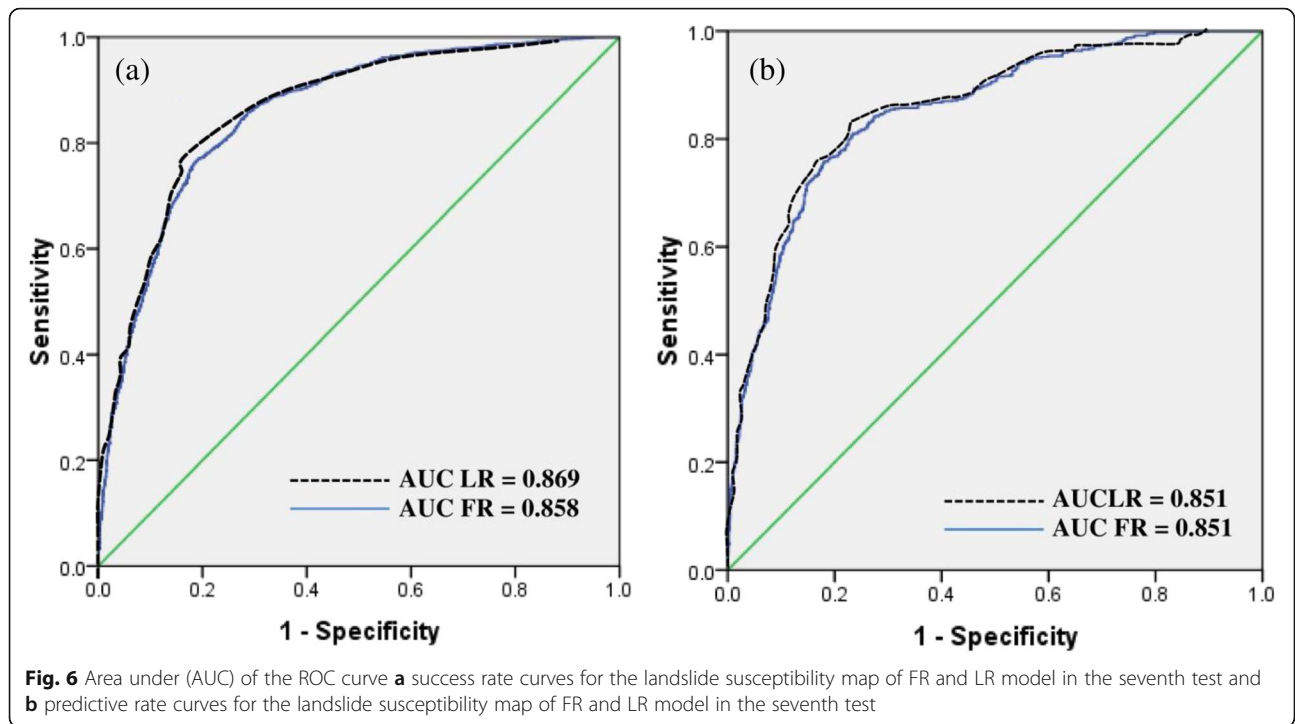


Table 4 The characteristics of susceptibility classes on LSM

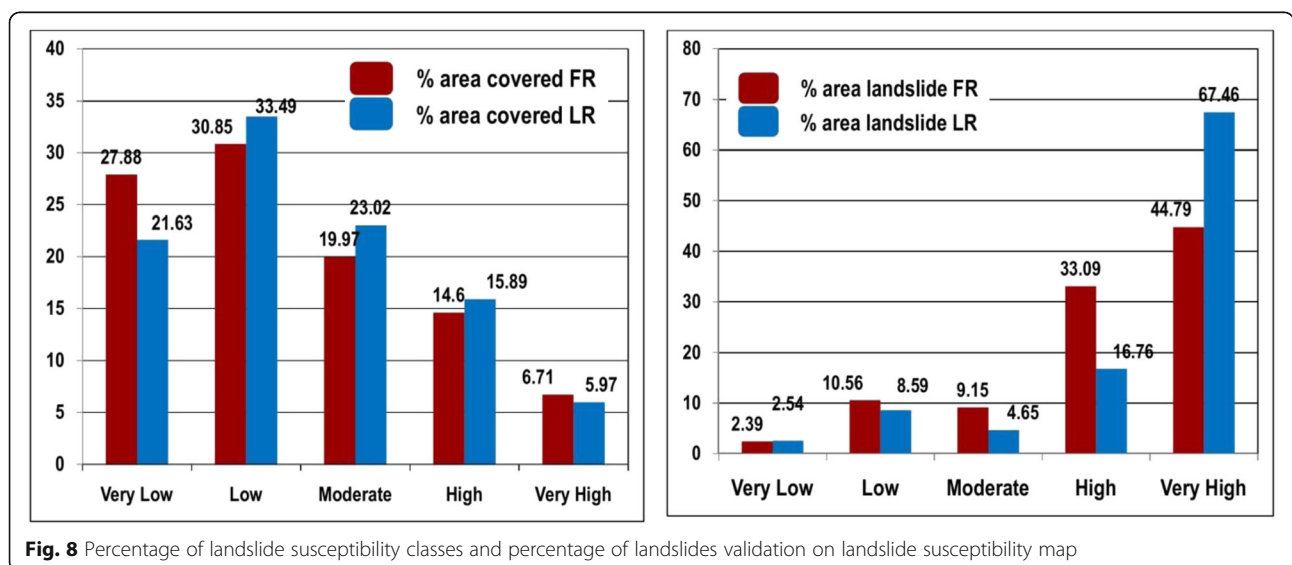
Class number	Reclassified index value	Susceptibility Class	Number of pixels	% area covered	Number of landslide validation pixels	% area landslide validation covered
Frequency Ratio Model						
1	1.52–5.59	Very Low	108,965	27.88	17	2.39
2	5.59–7.83	Low	120,552	30.85	75	10.56
3	7.83–10.62	Moderate	78,033	19.97	65	9.15
4	10.62–14.24	High	57,067	14.60	235	33.09
5	14.24–21.10	Very High	26,213	6.71	318	44.79
			390,830		710	
Logistic regression model						
1	0.004–0.089	Very Low	84,529	21.63	18	2.54
2	0.089–0.229	Low	130,898	33.49	61	8.59
3	0.229–0.413	Moderate	89,960	23.02	33	4.65
4	0.413–0.671	High	62,094	15.89	119	16.76
5	0.671–0.985	Very High	23,349	5.97	479	67.46
			390,830		710	

landslide and non landslide pixels (test 1–10) were better than for FR model, and at this point this study concludes that the seventh test in LR model was the best fit of model because the value is the highest.

The AUC of FR and LR model trial seventh was found to be 0.858 and 0.869, which shows that the model is capable of identifying landslide with 85.8 and 86.9 % accuracy respectively. In case of AUC curve for predictive rate, it was found to be 0.851 for FR and LR model, which was lower than AUC curve for success rate of the model (Fig. 6). The curve of the model and validation proves that the susceptibility model is acceptable and the model could be applied to predict the potential landslides in future.

As an interesting point to be noticed in Table 3, the eleventh and twelfth tests have a good result in AUC curve, which are 0.859 and 0.839 in success rate and 0.858 and 0.839 in predictive rate respectively. However, overlaying LSM using landslide data validation in those tests shows that the result decreases significantly to 32.39 and 30.28 % landslide covered on high to very high class. This indicates that by using equal number of landslide and non-landslide pixels with landslide causal factors to determine the variable of equation is the most reliable method to create a landslide susceptibility map.

Figure 7 shows the landslide susceptibility map using FR model and the seventh test equation of LR model. The LSM by LR model was obtained using the



coefficient values of landslide causal factors as in the equation below

$$Z = 1.922 (\text{lithology}) + 1.008 (\text{curvature}) + 0.573 (\text{Aspect}) + (\text{fault}) + 0.447 (\text{slope}) + 0.305 (\text{road}) + 0.174 (\text{river}) + 0.111 (\text{landuse}) - 0.6243 \quad (5)$$

The ranges of index value of each model in five classes were established using natural breaks method.

Can et al. (2005) and Bai et al. (2010) stated two important guidance for validating landslide susceptibility map i.e. 1) the high to very high class should cover only small areas and 2) landslide data validation should lie in high or very high classes. Table 4 shows the characteristics of susceptibility class for FR and LR models. It shows that the ratio of high to very high susceptibility class covers small area. The ratio was obtained by dividing the number of pixels in each class on LSM to the total number of pixels. Furthermore, ratio of landslide data for validation that fall on the LSM has a high value on high to very high class compared to very low to low class. The ratio was calculated by dividing the number of landslide for validation pixels, which lies on each susceptibility class to the total number of landslide for validation pixels. This method is similar to FR model or the density method.

In general, the procedure of creating landslide susceptibility map begins with use of data of landslide occurrence as dependent variable and landslide causal factors as independent variable. Logically, landslide data covers small area and occasionally in the form of scattered areas in the entire study area. The accuracy of the predicted future landslide that laid on the LSM should have lower ratio in the class of low to very low class and higher in the high to very high class. Figure 8 shows, in the case of the density ratio of area in low to very low class on the LSM both the statistical models contain more than 50 % of total area and the landslide data for validation falls in the low to very low class on the LSM map showed the ratio of below 12 %.

Conclusions

Besides creating landslide susceptibility maps, this research shows the performance of Frequency Ratio (FR) and Logistic Regression (LR) models as well. Two stages of validation were carried out in this study. First, performances of each landslide model were tested using AUC curve for success and predictive rate, which is more than 83 %. In the second stage, ratio of landslides falling on high to very high class of susceptibility was obtained, which indicates the level of accuracy of the model. In the FR model, 77.88 % landslides fall in the range of high to very high class while in LR model, it is 84.23 %. Both

the models show satisfactory results although LR model using equal number of landslide and non-landslide pixels shows slightly accurate results in total. From the logistic regression equation, it can be concluded that the landslide causal factors (i.e., lithology, curvature, aspect, distance from fault and slope) have a significant influence in causing landslides. The FR model is easy to apply, while LR model is a complex procedure. This study also shows that predicting future landslides by using logistic regression could be the best choice although the result will be more accurate on a larger scale, particularly at topographic map and geological map. Susceptibility mapping is an essential tool to delineate areas prone to landslide, and it has become important information for decision makers and government.

Acknowledgements

This research was supported by DIKTI Indonesia Scholarship Batch 1, 2013 and under a collaboration between Ehime University and Hasanuddin University of Indonesia. The authors are also thankful to Dr. Matebie Meten and Dr. Ilham Alimuddin for their comments during the preparation of this paper.

Authors' contributions

All of the authors performed the research. As first author, ARR has mostly participated in the whole process, including compiling data, analyzed data and map out the result. ARR wrote the draft of the manuscript with advice and supervision from NPB and RY. Both the co-authors have given the final approval of the version to be published. All authors read and approved the final manuscript.

Competing interests

The authors declare that they do not have any financial or non-financial competing interests with any individuals or institution.

Received: 13 May 2016 Accepted: 1 November 2016

Published online: 08 November 2016

References

- Akgun, A., E.A. Sezer, H.A. Nefeslioglu, C. Gokceoglu, and B. Pradhan. 2012. An easy-to-use MATLAB program (MamLand) for the assessment of landslide susceptibility using a Mamdani fuzzy algorithm. *Computers and Geosciences* 38: 23–34. doi:10.1016/j.cageo.2011.04.012.
- Ayalew, L., and H. Yamagishi. 2005. The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. *Geomorphology Science Direct* 65: 15–31.
- Badan Nasional Penanggulangan Bencana Indonesia. 2015. <http://dibi.bnpb.go.id/DesInventar>. Accessed 9 Dec 2015.
- Bai, S.-B., J. Wang, G.-N. Lu, P.-G. Zhou, S.-S. Hou, and S.-N. Xu. 2010. GIS-based logistic regression for landslide susceptibility mapping of the Zhongxian segment in the Three Gorges area, China. *Geomorphology* 115: 23–31. doi:10.1016/j.geomorph.2009.09.025.
- Balai Pengelolaan Daerah Aliran Sungai Jeneberang Walanae. 2014. *Laporan Penetapan Klasifikasi Daerah Aliran Sungai Wilayah Kerja*. Makassar, Indonesia: BPDAS Jeneberang Walanae. Tahun 2014.
- Budimir, M.E., P.M. Atkinson, and H.G. Lewis. 2015. A systematic review of landslide probability mapping using logistic regression. *Landslides* 12: 419–436. doi:10.1007/s10346-014-0550-5.
- Bui, D.T., B. Pradhan, O. Lofman, I. Revhaug, and O.B. Dick. 2012. Spatial prediction of landslide hazards in Hoa Binh province (Vietnam): A comparative assessment of the efficacy of evidential belief functions and fuzzy logic models. *Catena* 96: 28–40. doi:10.1016/j.catena.2012.04.001.
- Can, T., H.A. Nefeslioglu, C. Gokceoglu, H. Sonmez, and T.Y. Duman. 2005. Susceptibility assessments of shallow earthflows triggered by heavy rainfall at three catchments by logistic regression analyses. *Geomorphology* 72: 250–271. doi:10.1016/j.geomorph.2005.05.011.
- Cervi, F., M. Berti, L. Borgatti, F. Ronchetti, F. Manenti, and A. Corsini. 2010. Comparing predictive capability of statistical and deterministic methods for landslide

- susceptibility mapping: a case study in the northern Apennines (Reggio Emilia Province, Italy). *Landslides* 7: 433–444. doi:10.1007/s10346-010-0207-y.
- Chauhan, S., M. Sharma, M.K. Arora, and N. Gupta. 2010. Landslide Susceptibility Zonation through ratings derived From Artificial Neural Network. *International Journal of Applied Earth Observation and Geoinformation* 12: 340–350.
- Chau, K.T., and J.E. Chan. 2005. Regional bias of landslide data in generating susceptibility maps; Case of Hong Kong Island. *Landslides* 2: 280–290.
- Chung, C.J., and A.G. Fabbri. 2003. Validation of Spatial Prediction Models for Landslide Hazard Mapping. *Natural Hazards* 30: 451–472.
- Dai, F.C., C.F. Lee, L.G. Tham, K.C. Ng, and W.L. Shum. 2004. Logistic regression modelling of storm-induced shallow landsliding in time and space on natural terrain of Lantau Island, Hong Kong. *Bulletin of Engineering Geology and the Environment* 63: 315–327. doi:10.1007/s10064-004-0245-6.
- Dai, F., and C. Lee. 2002. Landslide characteristics and slope instability modeling using GIS, Lantau Island, Hong Kong. *Geomorphology* 42: 213–228.
- Demir, G., M. Aytakin, and A. Akgun. 2015. Landslide susceptibility mapping by frequency ratio and logistic regression methods: an example from Niksar–Resadiye. *Arabian Journal of Geosciences* 8: 1801–1812. doi:10.1007/s12517-014-1332-z.
- Demir, G., M. Aytakin, A. Akgun, S.B. Iklizler, and O. Tatar. 2013. A comparison of landslide susceptibility mapping of the eastern part of the North Anatolian Fault Zone (Turkey) by likelihood-frequency ratio and analytic hierarchy process methods. *Natural Hazards* 65: 1481–1506. doi:10.1007/s11069-012-0418-8.
- Dewitte, O., C.-J. Chung, Y. Cornet, M. Daoudi, and A. Demoulin. 2010. Combining spatial data in landslide reactivation susceptibility mapping: A likelihood ratio-based approach in W Belgium. *Geomorphology* 122: 153–166. doi:10.1016/j.geomorph.2010.06.010.
- Direktorat Cipta Karya Kementerian PUPERA. 2006. Indonesia. http://ciptakarya.pu.go.id/dok/banjir_susel/index.htm. Accessed 9 Dec 2015.
- Ermini, L., F. Catani, and N. Casagl. 2005. Artificial Neural Networks applied to landslide susceptibility assessment. *Geomorphology* 66: 327–343.
- Gorsevski, P.V., P.E. Gessler, R.B. Foltz, and W.J. Elliot. 2006. Spatial prediction of landslide hazard using logistic regression and ROC analysis. *Transactions in GIS* 10(3): 395–415.
- Hasnawir, and T. Kubota. 2012. Rainfall threshold for shallow landslide in Kelara Watershed, Indonesia. *International Journal of Japan Erosion Control Engineering Technical note* 5(No.1): 86–92.
- Kannan, M., E. Saranathan, and R. Anabalagan. 2013. Landslide vulnerability mapping using frequency ratio model: a geospatial approach in Bodi-Bodimettu Ghat section, Theni district, Tamil Nadu, India. *Arabian Journal of Geosciences* 6: 2901–2913. doi:10.1007/s12517-012-0587-5.
- Kanungo, D., M. Arora, S. Sarka, and R. Gupta. 2006. A comparative study of conventional, ANN black box, fuzzy and combined neural and fuzzy weighting procedures for landslide susceptibility zonation in Darjeeling Himalayas. *Engineering Geology* 85: 347–366.
- Lee, S. 2005. Application and cross-validation of spatial logistic multiple regression for landslide susceptibility analysis. *Geosciences Journal* 9(No.1): 63–71.
- Lee, S., and B. Pradhan. 2006. Probabilistic landslide hazards and risk mapping on Penang Island, Malaysia. *Journal of Earth System Science* 115(6): 661–672.
- Lee, S., and B. Pradhan. 2007. Landslide hazard mapping at Selangor, Malaysia using frequency ratio and logistic regression models. *Landslides* 4: 33–41. doi:10.1007/s10346-006-0047-y.
- Lee, S., and J.A. Thalib. 2005. Probabilistic landslide susceptibility and factor effect analysis. *Environmental Geology* 47: 982–990.
- Meten, M., N.P. Bhandary, and R. Yatabe. 2015a. Effect of Landslide Factor Combinations on the Prediction Accuracy of Landslide Susceptibility Maps in the Blue Nile Gorge of Central Ethiopia. *Geoenvironmental Disaster* 2: 9. doi:10.1186/s40677-015-0016-7.
- Meten, M., Bhandary, N. P., and R. Yatabe. 2015. GIS-based Frequency Ratio and Logistic Regression Modelling for Landslide Susceptibility Mapping of Debre Sina area in Central Ethiopia. *Journal of Mountain Science* 12(6). doi:10.1007/s11629-015-3464-3.
- Ngadisih, Yatabe, R., Bhandary, N. P., and R. K. Dahal. 2013. *Integration of statistical and heuristic approaches for landslide risk analysis: a case of volcanic mountains in West Java Province, Indonesia*. Georisk. doi:10.1080/17499518.2013.826030.
- Oh, H.J., Lee, S., Chotikasathien, W., Kim, C. H., and J. H. Kwon. 2008. Predictive landslide susceptibility mapping using spatial information in the Pechabun area of Thailand. *Environmental Geology* doi:10.1007/s00254-008-1342-9.
- Park, S., C. Choi, B. Kim, and J. Kim. 2013. Landslide susceptibility mapping using frequency ratio, analytic hierarchy process, logistic regression, and artificial neural network methods at the Inje area, Korea. *Environmental Earth Sciences* 68: 1443–1464. doi:10.1007/s12665-012-1842-5.
- Pourghasemi, H.R., B. Pradhan, and C. Gokceoglu. 2012. Application of fuzzy logic and analytical hierarchy process (AHP) to landslide susceptibility mapping at Haraz watershed, Iran. *Natural Hazards* 63: 965–996. doi:10.1007/s11069-012-0217-2.
- Quan, H.-C., and B.-G. Lee. 2012. GIS-Based Landslide Susceptibility Mapping Using Analytic Hierarchy Process and Artificial Neural Network in Jeju (Korea). *KSCE Journal of Civil Engineering* 16(7): 1258–1266.
- Ramani, S.E., K. Pitchaimani, and V.R. Gnanamanickam. 2011. GIS based landslide susceptibility mapping of Tevankarai Ar Sub-watershed, Kodaikkanal, India using binary logistic regression analysis. *Mountain Science* 8: 505–517.
- Reis, S., A. Yalcin, M. Atasoy, R. Nisanci, T. Bayrak, M. Erduran, C. Sancar, and S. Ekerin. 2012. Remote sensing and GIS-based landslide susceptibility mapping using frequency ratio and analytical hierarchy methods in Rize province (NE Turkey). *Environmental Earth Sciences* 66: 2063–2073. doi:10.1007/s12665-011-1432-y.
- Shirzadi, A., L. Saro, O.H. Joo, and K. Chapi. 2012. A GIS-based logistic regression model in rock-fall susceptibility mapping along a mountainous road: Salavat Abad case study, Kurdistan, Iran. *Natural Hazards* 64: 1639–1656.
- Solaimani, K., S.Z. Mousavi, and A. Kavian. 2013. Landslide susceptibility mapping based on frequency ratio and logistic regression models. *Arabian Journal of Geosciences* 6: 2557–2569. doi:10.1007/s12517-012-0526-5.
- Suh, J., Y. Choi, T.-D. Roh, H.-J. Lee, and H.-D. Park. 2011. National-scale assessment of landslide susceptibility to rank the vulnerability to failure of rock-cut slopes along expressways in Korea. *Environmental Earth Sciences* 63: 619–632. doi:10.1007/s12665-010-0729-6.
- Sukanto, R., and S. Supriatna. 1982. *Geologic Map of The Ujungpandang, Benteng, and Sinjai Quadrangles, Sulawesi*. Bandung, Indonesia: Geological Research and Development Centre.
- Tsuchiya, S., K. Sasahara, S. Shuin, and S. Ozono. 2009. The large-scale landslide on the flank of caldera in South Sulawesi, Indonesia. *Landslides* 6: 83–88. doi:10.1007/s10346-009-0143-x.
- Westen, C.J.V., N. Rengers, M.T.J. Terlien, and R. Soeters. 1997. Prediction of the occurrence of slope instability phenomena through GIS-based hazard zonation. *Geologische Rundschau* 86: 404–414.
- Xu, C., X. Xu, F. Dai, Z. Wu, H. He, F. Shi, X. Wu, and S. Xu. 2013. Application of an incomplete landslide inventory, logistic regression model and its validation for landslide susceptibility mapping related to the May 12, 2008 Wenchuan earthquake of China. *Natural Hazards* 68: 883–900. doi:10.1007/s11069-013-0661-7.
- Yilmaz, I. 2009. Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: A case study from Kat landslides (Tokat—Turkey). *Computers & Geosciences* 35: 1125–1138. doi:10.1016/j.cageo.2008.08.007.
- Yilmaz, I., and I. Keskin. 2009. GIS based statistical and physical approaches to landslide susceptibility mapping (Sebinkarahisar, Turkey). *Bulletin of Engineering Geology and the Environment* 68: 459–471. doi:10.1007/s10064-009-0188-z.

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com